Swiss Personalized Health Network

HospFAIR Program: Improvement of the implementation of the SPHN interoperability strategy and processes in the hospitals with regards to the creation of FAIR data[1]

(20.10.2021)
(10.03.2022: timelines adjusted in table 1)

---

[1] FAIR = Findable, Accessible, Interoperable and Reusable

1   Background

The Swiss Personalized Health Network (SPHN) aims at bringing Switzerland at the forefront of personalized health research by establishing nationwide interoperability of health-related information. SPHN is a nationwide research and infrastructure collaboration of hospitals and universities in Switzerland with the aim to ultimately improve the health of citizens. The network fosters the interdisciplinary exploitation of health-related information in order to develop innovative approaches in prevention, diagnosis and treatment of disease. SPHN is supporting the creation of common infrastructures and a new culture to share health data between institutions and research groups in Switzerland.

As part of the SPHN initiative, various efforts are currently underway to harmonize and define data standards in order to ensure the interoperability of health-related data in Switzerland, with a specific focus on data from the five Swiss university hospitals. In addition, national solutions with regard to the findability and accessibility of data (e.g. through the creation of [meta-]data catalogues) as well as concerning data long-term storage and data reusability by other research groups are currently being developed.

Novel technical solutions for hospitals which facilitate SPHN-related research

The overall progress of SPHN from 2016-19 has been summarized in the SPHN Report from the National Steering Board 2016-2019[2]. Notably, the following main developments have been achieved:

−   Selection and funding of 24 health data infrastructure projects;
−   Definition of a blueprint for a clinical dataset with 49 concepts and 199 component descriptions, in a compositional structure inspired by semantic and conceptual modeling strategies;
−   Development of clinical research data management systems in the five Swiss university hospitals and establishment of related infrastructures and processes;
−   Development of the BioMedIT Network at Swiss academic institutions operating under a common IT security policy;
−   Harmonized Data Transfer and Use Agreements (DTUAs) between 15 Swiss research institutions and university hospitals by creating common templates;
−   Establishment of a Federated Query System allowing feasibility queries across all five university hospitals.

---

[2] published as Swiss Academies Report in January 2021, doi.org/10.5281/zenodo.4044123,
https://sphn.ch/wp-content/uploads/2021/02/NSB_report_2016-2019_SAR_2021_single_page.pdf

Unaddressed issues and remaining hurdles

From the currently running SPHN Driver Projects and the work in relation to the SPHN semantic interoperability framework[3], SPHN has identified a series of critical and challenging issues which mainly relate to the implementation of interoperability requirements, bringing forward the explicit need for the introduction of data standards (by mapping or at the source) and professionalization of the data extraction and data delivery pipelines at the university hospitals according to the SPHN interoperability strategy requirements.

Most of the Clinical Data Warehouses (CDW) of the Swiss university hospitals are based on relational database management systems. To transform data from a relational model representation – according to the SPHN interoperability strategy – into a graph representation based on RDF, Extract Transform and Load (ETL) pipelines have to be implemented by data provider's IT. They typically include an RDF transformer step where raw data from the CDW is converted and loaded into a triple store. Then, data can be extracted and transferred into RDF files for each specific project. Before data transfers, a common quality control guideline and its technical implementation (e.g. via SHACL rules, summary statistics, etc.) shall assure that data is delivered according to the SPHN specifications. These processes are partly already in place in the Swiss university hospitals, but need support and improvements with regards to their efficiency.

In addition to the already available standards in the university hospitals (e.g. ICD-10 codes for FOPH diagnoses, CHOP codes for procedures, ATC for medication), hospitals need to code (or map) patient data in (or to) internationally recognized research data standards, such as SNOMED CT, LOINC, and UCUM. The SPHN interoperability strategy provides the framework and specifies the requirements for data delivery. Depending on the design and development status of the CDWs, different efforts are needed to achieve an automated or integrated data delivery process.

It is the ultimate goal to have streamlined processes and a cost-efficient sustainable way to convey structured data from UH to the research environment in accordance with the SPHN Interoperability Strategy, making the necessary steps for the university hospitals in their role as data providers smooth and efficient. To this aim, the **SPHN IT Architecture Working Group** (ITAC WG) started with the elaboration of an architecture and maintenance plan for a sustainable and scalable hospital data delivery framework in September 2021. Key requirement of the envisioned architecture is the sustainability aspect with regard to future data deliveries/provisions to research projects according to the SPHN Interoperability Framework and in alignment with operative routine data exchange formats at hospitals, also beyond the funding phase of SPHN. For the realization of such a harmonized architecture solution, it is of utmost importance that both the technical and human resources in the hospitals are expanded accordingly in view of the requirements for data preparation and delivery as specified by the SPHN Interoperability Framework.

---

[3] https://sphn.ch/network/data-coordination-center/the-sphn-semantic-interoperability-framework/

2 Aim of the HospFAIR program

HospFAIR is an implementation project in its nature with the aim of improving quality of shared data and systematizing data standardization and extraction. It calls to enhance and streamline the various processes in the university hospitals, which currently can represent a bottleneck for data interoperability and data sharing, through agile and iterative data production cycles. Final product of HospFAIR are large and readily deliverable sets of high-quality patient data (meeting the SPHN interoperability strategy requirements) that are reproducible at any time, because their production pipeline is systemically implemented at the CDW of the university hospitals. At the end of the first phase of the program, there should be a sustainable and scalable data extraction pipeline in place to efficiently produce a federated volume of data according to the SPHN interoperability framework.

A new Working Group, the **SPHN Data Standards and Quality Working Group** (DASAQ WG) will be set up with the mandate to elaborate a work plan for Phase I of HospFAIR and to closely accompany the entire HospFAIR program. It will be responsible for the definition of the concepts to be delivered in the realm of HospFAIR as well as for the development of a Quality Assessment plan for HospFAIR. Furthermore, the DASAQ WG representatives will elaborate UH-individual work plans for the implementation of standards at the source (LOINC/L4CHLAB, UCUM, ATC, as listed in the CA) and for the introduction of future standards (such as SNOMED CT, ICD-O) and the WG should serve as a sounding and discussion board to ensure mutual coordination and alignment of activities as well as an exchange of best practices.

Rather than aiming at generating one representative dataset for research use, the goal of HospFAIR is to support university hospitals to efficiently generate use case specific sub-datasets that are readily deliverable and that can be iteratively enriched. This will help to effectively address the biggest problems of data interoperability, data delivery and increase the ability of cost-efficient data sharing.

- In the first phase of the HospFAIR program (Phase I), the emerging HospFAIR dataset consisting of core concepts (see below) will solely be used for process-optimization and quality assurance by the DASAQ WG in close collaboration with the DCC, university hospitals, and two experts specifically assigned to perform HospFAIR data QC analyses.

- In a second phase (Phase II), the established pipelines will be used to produce use case specific sup-sets representing large numbers of patients according to specific in- and exclusion criteria and requirements in order to be able to answer specific medical questions. These datasets can be used by the HospFAIR consortium or third-party researchers for specific research or methodology-developing projects according to research protocols approved by the relevant Ethics Committee (EC). The development of the corresponding legal templates (e.g. Consortium Agreement, Data Transfer and Use Agreements) will also be part of Phase II of HospFAIR. Depending on the use case, the HospFAIR datasets can also be set up in a federated way, i.e. data stays at the individual hospitals with the possibility to explore and analyze them in a distributed, privacy preserving way, without the need of establishing Data Transfer and Use Agreements.

- Also in Phase II, preparations to introduce a genomic layer to HospFAIR shall be undertaken, including patients of whom a biosample (generally consented to be used for research purposes) is available for genomic DNA extraction. Moreover, information about additional available biospecimens in biobanks for multi-omics analysis shall be included.

– In a third phase (Phase III), genomic and multi omics analyses shall be carried out according to specific use case requirements. Costs for the data generation in the omics layer will be covered by funds outside of the HospFAIR program at a later time point and are therefore not integral part of the HospFAIR agreement with the university hospitals.

The HospFAIR program is set up as a modular program. University Hospitals have the choice to participate in the entire program, or only in individual phases. Participation in phase II and III, however, requires participation in phase I.

SPHN Interoperability Framework

Phase I
- Core dataset
- Standardization
- Quality assurance only
- Data extraction, RDF pipeline
- FAIR data management
- No research projects

Phase II
- Core data availability of a large number of patients
- Use case specific sub-sets
- Biosample availability
- Federated data exploration
- Research projects (DTUAs with UH)

Phase III
- Sample analysis
- Genomics
- Multi omics
- Data integration
- Federated analysis
- Research projects (DTUAs with UH)

BioMedIT: Secure data transfers, temp. storage and quality control (Phase I), data processing and analyses for research projects (Phase II), data integration and analyses for research projects (Phase III)
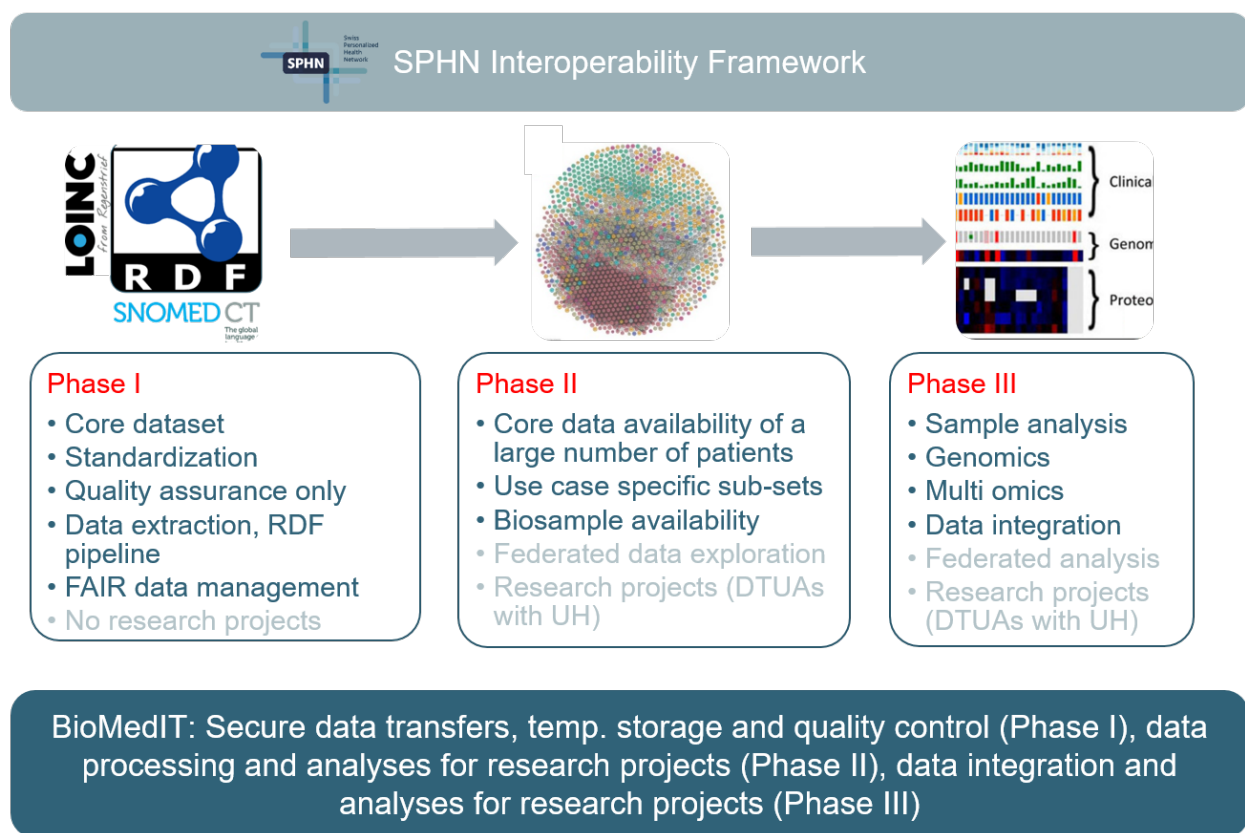
Figure 1: The HospFAIR program. Data standardization, quality assurance, and professionalization of interoperable RDF data production and delivery pipelines according to the SPHN interoperability framework. Use case specific curation of FAIR datasets in SPHN including omics data integration, which can be used for research purposes (on BioMedIT) or for data exploration purposes in a federated, privacy preserving fashion.

The HospFAIR data basis should emerge through an incremental approach by working on the different concepts in a step-wise manner, eventually (i) ensuring the compliance with the SPHN semantic framework and the implementation of controlled vocabulary, (ii) warranting that the data generation mechanisms are efficiently implemented in the hospital ICT infrastructure in alignment with the SPHN IT architecture goals, and (iii) allowing adjustments and improvements of the interoperable data generation workflow in a timely manner (e.g. with regards to the new SPHN RDF releases).

The HospFAIR program should also contribute to more clearly define roles and responsibilities on the data provision site with regard to data transformation, data preparation and quality assurance.

For the use of HospFAIR datasets in particular research projects in Phase II (beyond a federated data exploration), a Data Transfer and Use Agreement template shall be elaborated, also allowing for the use of the datasets in further research projects (with ethical approval), including the use of biosamples.

The HospFAIR program is part of a broader effort to make health-related data FAIR in Switzerland (Figure 1). In addition to the fairification of clinical data from hospitals within the framework of the HospFAIR program, SPHN and BioMedIT will provide additional funding for the fairification of omics and other molecular data as well as cohort and registry data in collaboration with the SPHN Task Forces.

3 Conditions and requirements

3.1 Conditions for the university hospitals to participate in this program

a) Investment in data standardization (according to the SPHN interoperability strategy, in particular SNOMED CT, ATC, LOINC and UCUM), implementation of the SPHN RDF schema releases, investment in the RDF data production pipeline and data quality management and assurance;

b) Dedication of 1-2 representative(s) to each of the HospFAIR working groups (ITAC and DASAQ WG) and active participation and collaboration of the WG members in order to fulfill the WG mandates;

c) Data preparation and delivery according to the SPHN interoperability strategy and creation of use case specific HospFAIR datasets for federated exploration or research purposes (by September 2022), providing the core concepts [described in 3.2 e)] of large numbers of patients per hospital (Phase II);

d) Elaboration of a templates for a Consortium Agreement and a Data Transfer and Use Agreement (DTUA) for delivery, access and use of HospFAIR datasets for specific research projects (Phase II);

e) Willingness to provide available biosamples for a subset of patients according to the project plan to be developed in Phase III;

3.2 Requirements concerning the future HospFAIR datasets:

a) Data and samples must be generally consented;

b) Usage of the datasets coming out of HospFAIR (to be generated in Phase II) by third party researchers is a mandatory requirement in accordance with the provisions of the Funding Regulations and shall be governed by respective agreements;

c) Core concepts (according to SPHN dataset[4]):
- Demographics: Data Provider Institute, Administrative Case, Healthcare Encounter, Subject Pseudo Identifier, Administrative Gender, Birth Date, Death Status, Consent

---

[4] https://sphn.ch/document/sphn-dataset/

- Clinical Measurements: Heart Rate, Systemic Arterial Blood Pressure, Body Temperature, Oxygen Saturation, Central Venous Pressure, Circumference Measure (e.g. head, hip, waist or neck), Height, and Weight.
- Diagnoses: Problem Condition (clinical condition, problem, diagnosis, or other event, situation, issue, or clinical concept that has risen to a level of concern in coded form), FOPH Diagnosis, Nursing Diagnosis, ICD-O Diagnosis, and Oncology Treatment Assessment.
- Procedures: FOPH Procedure, Radiotherapy Procedure, and Catheter.
- Medications: Drug Administration Event, Drug Prescription, and Allergy.
- Adverse events
- Lab Results: Clinical Chemistry, Hematology, Immunology, Microbiology, and Virology.
- Pathology: TNM Classification, Tumor Specimen, Tumor Stage and Tumor Grade.
- Biobanking: Sample availability information "yes/no", Biosample.

d) Complete congruence with the specifications in the SPHN dataset. In case basic pre-processing of data is needed, this lies in full responsibility of the UH (e.g. transformation of a drug product amount [e.g. IRFEN Lactab 200mg 2 pills] to amount of active ingredient [ATC code: M01AE01, 400mg]; or the introduction of the necessary level of granularity, such as "bodysite: left arm" to bodysite: arm laterality: left").

4 Purpose and use of the datasets generated in HospFAIR:

4.1 Building and benchmarking the initial dataset in HospFAIR (production, description, quality assessment, benchmarking)

The initial dataset will not be used for research, but only for quality control purposes and production improvement. All concepts mentioned above will be addressed in agile and iterative data production cycles until data is delivered according to SPHN specifications. Thorough quantitative descriptive analyses of the emerging dataset will be carried out by two experts that will be provided with funds to dedicate their time to monitoring, documentation, quality assessment and feedback to UH, in order to improve overall data quality. In addition, the processes in the university hospitals with regards to data preparation and quality control will be closely monitored and support being granted, if necessary. Subsequently, to test the usefulness of the data provided, specific subsections of patients will be identified within the HospFAIR dataset and compared to reference datasets for benchmarking. The DASAQ WG will elaborate a work plan for Phase I of HospFAIR.

4.2 Use case specific datasets generated in HospFAIR Use cases

Researchers within and outside of SPHN shall have the possibility to present use cases and related medical questions, which could be addressed with a dataset relying on the above-mentioned concepts provided by the UH through HospFAIR. Use case approaches could either be motivated by an open call or though determination by expert groups.

There can be two types of datasets generated in HospFAIR:

1. Decentralized datasets where data does not leave the UH and can be explored in a federated and privacy preserving manner (through the use of the appropriate technologies), and
2. Centralized datasets that will be merged and stored centrally on BioMedIT.

As a general rule, both forms of datasets should be made accessible to as many interested researchers as possible. With regards to the centralized datasets, a DTUA template for third party use needs to be elaborated by the university hospitals with support by the ELSI helpdesk of the DCC. Research projects that foresee to use HospFAIR datasets need to provide ethics approval and sign the respective DTUA with the university hospitals.

The detailed process of dataset generation in Phase II will be planned during Phase I and guided by the SPHN National Advisory Board.

5   Legal Framework

The contractual legal framework for HospFAIR shall be elaborated by the university hospitals in collaboration with the ELSI helpdesk of the DCC. Since the HospFAIR program is set up as a modular program, the contracts for the different phases will be set up separately.

The envisioned cornerstones of the framework are as follows:

5.1   Phase I: Initial Data Transfer and Use Agreement (DTUA) between SIB (representing SPHN) and each Data Provider (university hospital)

The basic responsibilities for Phase I will be specified in an Agreement between SPHN and the university hospitals. A standard DTUA/DTPA (with the BioMedIT network institutions as the processor) will be set up between the UH, BioMedIT and the SIB PHI Group defining the data that the Data Provider will transfer to BioMedIT in Phase I of the project, as well as the use of the data for QC and benchmarking as well as the technical aspects (including information security) of transferring and storing the data on BioMedIT.

An ethics approval for Phase I of HospFAIR will not be required, since the data will solely be used for quality control purposes and deleted after completion of Phase I. In any case, a non-competence statement will be requested from the EKNZ.

5.2   Phase II: Acceptable use policies for data exploration; Consortium Agreements and Data Transfer and Use Agreements (DTUA) for research projects

The basic responsibilities for Phase II will be specified in an addendum to the Collaboration Agreements between SPHN and the university hospitals. Since the federated data exploration (where data stays at the UH and only aggregated, anonymous results will be communicated to the researcher/user) does not require the standard agreements for data transfer and use, acceptable use policies for accessing the federated data exploration technology in each use case has to be set up. For research projects on BioMedIT, a Consortium Agreement and a DTUA allowing for further research use (with the BioMedIT network institutions as the processor) will define for every use case:

–   The data that the Data Provider will transfer to BioMedIT as part of a HospFAIR dataset in Phase II of the project.

- The use of the data in the individual use cases according to an EC-approved protocol.

- The technical aspects (including information security) of transferring and storing the data on BioMedIT

5.3 Third use Data Use Agreement (DUA) between a Data User (i.e. principal investigator) and the Data Providers (i.e. university hospitals)

A standardized DUA (with the BioMedIT network institutions as the processor) for third parties to use the HospFAIR datasets will have to be accepted and signed by the institution of the principal investigator requesting access to a HospFAIR dataset. This DUA will define:

- The responsible data user (principal investigator)

- The use of the data for research- or methods-driven projects, given those projects comply with the Data Transfer and Use Agreement (DTUA, 5.2), the approval by the relevant EC, and all other requirements set by the UH;

- A standardized MTA will define for every university hospital (or other provider of biosamples) and SPHN the rights and obligations of the provider and the recipient with respect to the materials and any derivatives, and allow for omics analysis and further use of the corresponding data for research projects in accordance with the DTUA.

- The allowed use of the data in the context of a specific (research) project, for which an ethics approval is provided.

The legal framework for Phase III will be set up separately according to the one for Phase II.

6 Funding

SPHN will allocate a total of up to CHF 4.3 million over a 36-month period for the HospFAIR program. Funding for research and methodology development projects using the HospFAIR dataset as well as biosample analyses will be supported through other funding instruments in the framework of the SPHN funding period 2021-2024 and will not be described hereunder.

The participation of the university hospitals to the setup of HospFAIR will be supported with approximately CHF 3.5 million and will be described in an addendum to the Collaboration Agreements signed between SAMW/SPHN and the university hospitals.

SPHN will allocate approximately 0.8 Mio CHF for project management, technical support, data analysis and QC assurance in the realm of the HospFAIR program.

Funding by SPHN requires matching contributions by the applicants' institutions.

6.1 Timelines and deliverables for the HospFAIR program

Table 1: Timelines, responsibilities and deliverables for the HospFAIR program

Date	Description
Q4 2021	Preparation: Finalization of the HospFAIR program description and contractual agreement drafts, approval by the NSB Deliverables (Responsibility: DCC and MO): − HospFAIR program description − Addendum to the CA between SPHN and UH for phase I
Q2 2022 – Q1 2023	Phase I: Data standardization, quality assurance, and professionalization of interoperable RDF data production and delivery pipelines at the UH according to the SPHN interoperability framework Deliverables (Responsibility: UH): − D1: UH have delegated 1-2 representatives for each SPHN WG: the SPHN Data Standards and Quality Working Group (DASAQ WG) and the SPHN IT Architecture Working Group (ITAC WG) − D2: UH have signed a DTUA for HospFAIR Phase 1 − D3: 1 FTE in each UH affiliated with the central CDW team responsible for the SPHN interoperability framework and data delivery pipeline (implementation and contact point) D4: UH has implemented a data production pipeline and iteratively transferred data according to the work plan from the DASAQ WG and the ITAC WG and agreed standards of the SPHN data set (https://sphn.ch/document/sphn-dataset/), including the following sub-deliverables: − Budget for consultancy or further education of UH personnel involved in the data preparation and delivery pipeline according to SPHN interoperability framework requirements − Implementation of the (project unspecific) SPHN RDF Schema including the above specified concepts − Full validation on the UH side of each RDF dataset to be sent out In close alignment with the IT Architecture WG developments: − Provision of sufficient compute power for the data validation and the overall data delivery pipeline (comparable to a low latency large size database; exact requirements will be determined according to the individual set-ups) − Flexibility of the pipeline to react to new RDF schema releases and additional requirements by the projects − Automation of repetitive steps − Deployment of a triple store with the capacity to load at least all data of one data delivery including incremental loads (incl. licensing costs) − Provision of the final validation reports and pipeline runtimes of each data delivery to DCC D5: UH has delivered a report describing the data production pipeline and the achievement of KPIs Deliverables of the SPHN IT Architecture WG: see separate Mandate

	Deliverables of the SPHN Data Standards and Quality Working Group: see separate Mandate
Q1/2023 – Q4/2023	Phase II: to be elaborated
Q1/2024 – Q4/2024	Phase III: to be elaborated

7   Legal basis

The SPHN HospFAIR program is supported in accordance with the Funding Regulations of SPHN and with the mandate of SERI. The initial setup of the HospFAIR dataset will be funded under the category of "Infrastructure Implementation". The participation of the university hospitals will be governed through an addendum to the Collaboration Agreements signed between SAMS/SPHN and the university hospitals.

8   Financial frame, controlling and reporting

8.1   Financial frame

The overall budget for the HospFAIR program and therewith SPHN's maximum financial engagement is capped to an amount of CHF 4.3 Mio. A maximum of **CHF 700'000 (seven hundred thousand Swiss Francs) will be allocated per university hospital.** The remaining CHF 0.8 Mio will be allocated to fund the project management, the quality assessment and the data analysis of the HospFAIR dataset. The following payment plan for the participating university hospitals is foreseen:

Table 2: Payment plan for HospFAIR (hospital collaboration part)

Payment schedule	Amount [CHF]
Prepayment after signature of the HospFAIR agreement for Phase I	150k
Payment upon completion of Deliverables D1-D3 (30.06.2022)	100k
Payment after approval of the final report for Phase I	100k
Prepayment after signature of the HospFAIR agreement for Phase II	150k
Individual payments after approval of individual use cases	Max. 200k

8.2   Controlling and Reporting

The provisions outlined in Article 8 of the Collaboration Agreement shall apply also for the HospFAIR program. The SPHN National Steering Board will regularly receive technical reports by HIT-STAG and the DCC, based on qualitative and quantitative indicators, on implementation of agreed milestones.

Unused and unmatched funds must be refunded to SPHN and may not be put to any other use.

9 Appendix

9.1 Table 3: of abbreviations

ELSI	Ethical-Legal-Social-Issues
IAB	International Advisory Board
IT	Information Technology
MO	Management Office
NSB	National Steering Board
PH	Personalized Health
PHRT	Personalized Health and Related Technologies
SAMS	Swiss Academy of Medical Sciences
SERI	State Secretariat for Education, Research and Innovation
SIB	Swiss Institute of Bioinformatics
SNSF	Swiss National Science Foundation

9.2 Glossary

Terms are defined in the Ethical Framework for Responsible Data Processing available at
http://www.sphn.ch/.

9.3 Matching funds guidelines

As required by law, the "matching funds" principle (in cash and/or in kind) is a mandatory prerequisite for receiving SPHN funding. The amount requested from SPHN must be matched with own contributions by the host institutions.

The own contributions must pertain to the scope and main goals of the SPHN initiative and can be provided as in cash or in kind. They must support the aim of SPHN and not directly relate to clinical service and other health care issues. However, building interconnections between electronic patient records and Personalized Health research data infrastructures are acceptable.

Table 2: Own contributions definitions

Contribution	Definition
cash	Only funds, which are transferred by the host institution to an account/credit line administered by the applicant are to be considered as cash contributions.
kind	Resources from the institutions' operating budget that are explicitly allocated to SPHN initiatives (projects): (1) Personnel paid from the institution's operating budget involved in SPHN initiatives (projects). A maximum of 20% of a professor's time can be claimed. (2) Earlier investments (e.g. infrastructure platforms) will be reported within the financial report of the first year. SPHN will determine the eligible amount on a case-by-case basis.