Fresh Ideas for Cancer Care

A Blueprint to Test Data Sharing within the Data Collaboration Center of the Swiss Personalized Health Network

Coaches:
Prof. Dr. Manfred Claassen, Dr. Diana Elena Coman Schmid, Dr. Katrin Crameri, Prof. Mitchell Levesque, PhD, Dr. Daniel Vonder Mühll

Students:
Nicolas Banholzer, Florian Kast, Aline Odermatt, Marc Wohlwend

Abstract

A secure and interoperable data infrastructure between hospitals and universities in Switzerland will allow researchers to get access to a large pool of health-related data. Ultimately, the findings from these research activities will lead to improvements in the efficiency and quality of healthcare.

In a national collaboration of unparalleled scale, the national initiative Swiss Personalized Health Network (SPHN) and the strategic focus of the ETH domain Personalized Health and Related Technologies (PHRT) are currently improving data interoperability between hospitals and universities. In this work, we propose a blueprint to test data sharing in SPHN, managed centrally by the Data Coordination Center (DCC) of SPHN. Thereby, we follow the chart of "Sarah the Researcher". Our blueprint serves multiple objectives, e. g., from reviewing and verifying progress on SPHN data interoperability to identifying the issues that hinder it. We illustrate how our general blueprint may be adapted to a specific case, here as part of a case study with Swiss Personalized Oncology (SPO).

1 Introduction

For the "Fresh Ideas for Cancer Care" project, we took a closer look at the Swiss national collaborative programs for personalized health: the Swiss Personalized Health Network (SPHN) and the strategic focus of the ETH domain Personalized Health and Related Technologies (PHRT), in order to develop new ideas for better, more sustainable and also cost-efficient cancer care.

In 2016, the State Secretariat for Education, Research, and Innovation (SERI) and the Federal Office of Public Health (FOPH) mandated the Swiss Academy of Medical Sciences (SAMS) to initiate and govern the SPHN, which promotes the development of a framework for personalized health and personalized medicine in Switzerland. The ultimate goal is to promote health and well-being, to prevent, diagnose and treat unfavorable health conditions more precisely, thus reducing the risk of developing such conditions and permitting more effective treatments of disease states with less adverse effects.[1]

For that purpose, SPHN firstly creates the conditions that facilitate the exchange of health-related data promoting research. The initial four years (2017-2020) were thus dedicated to develop an infrastructure enabling data interoperability. Secondly, so-called driver projects are supporting this process by improving data collection, harmonization and analysis in specific areas of biomedical research. Till date, a total of 24 SPHN projects were funded and around CHF 58.3 million were allocated into research infrastructure projects. In these projects, 33 Swiss organizations and institutions are involved. In summary, SPHN is a national collaboration of unparalleled scale.

PHRT is a strategic focus area of the ETH domain with the goal of improving the quality of healthcare and select therapeutic strategies for individual patients.[2] In order to reach this goal, large amounts health-related data will be collected and analyzed. For that purpose, scalable IT infrastructure will be built in order to collect, store and exchange the collected data. Analysis of this data will also require developing new technologies based on cutting-edge research at ETH. Thus, PHRT will provide clinics and the health sector with access to the technological know-how of the ETH domain.

The SPHN and PHRT complement each other and coordinate their activities in order to promote personalized health and personalized medicine in Switzerland. For example, SPHN and PHRT run joint driver projects and the PHRT platform also uses the secure IT infrastructure of SPHN/BioMedIT.

[1] `https://www.sphn.ch/en.html`
[2] `https://www.sfa-phrt.ch/`

2 Big Data in Healthcare

Big Data is characterized by the features volume, velocity, variety and veracity (Thouvenin 2017, p. 28). In the healthcare sector, big data is also featured by value, because big data is expected to establish a better cost-benefit ratio of medical treatments. The size of the data volume depends on availability on the one hand and accessibility on the other hand. For this reason, the SPHN initiative accomplished the signing of a cooperation agreement with all five university hospitals in Switzerland at the end of 2017.[3]

Big data in healthcare, moreover, raises big hopes of preventing many deaths in the future through a better understanding of patient data combined with insights from comparable circumstances and treatments, and diagnostic support systems (Ohmann et al. 2017, Salathé and Driessen 2016, Sprecher 2018, p. 550). To achieve this, effective and at the same time secure means for data exchange are needed. Personal data as well as sensitive personal data are protected by the right of privacy. However, privacy is relative depending on the person who is processing the healthcare data. Therefore, the relations between researchers and data donors need to be clarified. Furthermore, secure IT infrastructure is to be established.

New technologies hold up a mirror to society and make visible what was previously protected by latency. The relating discomfort with big data is fed by these (sensitive) information is becoming visible (Nassehi 2019, p. 42). Thus, it is demanded that the goal must be to create a research environment by means of technology and regulation that enables a trustworthy, non-discriminatory and socially acceptable use of data and preserves herewith the data sovereignty of the individual (Sprecher 2018, p. 551). Furthermore, a trustworthy data management system empowers data donors to take more informed decisions in the future (Mausbach 2019, Vayena and Gasser 2016, Williams et al. 2015).

2.1 Protection of Privacy through Data Protection

At the Swiss federal level, personal health related data is primarily protected by the Federal Act on Data Protection (FADP; SR 235.1). It covers data processing carried by private people as well as federal bodies. The term "data" is thereby defined as information that refers to an individual person or an identifiable person (Art. 3 lit. a FADP).

The FADP differentiates between sensitive personal data (Art. 3 lit. c FADP) and other personal data (Art. 3 lit. a FADP). The first category includes amongst others data concerning the health of a person (Art. 3 lit. c No. 2 FADP). Thus, data sharing in healthcare occurs with sensitive personal data. When processing this kind of data, a set of special requirements must

[3] `https://www.sphn.ch/en/projects/infrastructure-implementation-projects.html`

be obeyed, for example:

- For data processing, which requires the consent of the concerned person, the consent must be given expressly, which excludes an implied consent (Baeriswyl 2015, Art. 4 N 69).
- A person who reveals data especially worthy of protection to a third party without justification breaches the privacy of the data subject (Art. 12 FADP).

When health related data is used for research purposes, particular importance must be given to the requirement of purpose limitation (Art. 4 para. 3 FADP). According to this principle, data processing has to occur with a defined purpose or objective, e. g., human research or medical treatment purposes. Data acquisition in stock is against good faith and thus illegal (Baeriswyl 2015, Art. 4 N 34). According to Art. 15 para. 1 FADP, legal violations are to be enforced by the affected person.

2.2 Protection of Privacy in Human Research

The Federal Act on Research involving Human Beings (HRA; SR 810.30) strives to protect the dignity, the personality and the health of human beings in research (Art. 1 para. 1 HRA). Therefore, it contains requirements concerning information and consent (Art. 7 HRA), general rules on quality and goals of research (Art. 5 and 10 HRA), security and safety requirements (Art. 15 HRA) as well as liability provisions (Art. 19 HRA). For our research project in particular, the hereinafter described requirements are of relevance.

2.2.1 Anonymization and Pseudonymization

The fourth chapter of the HRA addresses the further use of biological material and health related personal data for research purposes. The term "further use" indicates that the data, respectively biological material, has priorly been sampled for another use, be it in the context of a medical treatment or a prior research project, and may now be used for another purpose.

In this context, the HRA categorizes biological material and health related data into three sections: uncoded, coded and anonymized. Coded biological material and data is defined as being linked to an identifiable person via a code and can therefore be tracked back to this person. Anonymized biological material and data on the other hand can not or only with undue effort be linked to a specific person. In research practice, the terms "de-identified" instead of "coded" is more common; the SPHN uses the term "pseudonymized".

The Ordinance on Human Research with the Exception of Clinical Trials (HRO; SR 810.301) defines in a more specific way what the term "coded" means. Therefore, biological material and

health-related personal data are considered as coded correctly when, from the view of a person who has no access to the key, the data is to be considered "anonymous". A person who is not involved in the research project has to keep the key separated from the data collection. This person must be named in the ethics approval application document (see chapter 4). The key must be stored separately from the material or data collection and in accordance with the principles of Art. 5 para. 1 HRO, by a person to be designated in the application who is not involved in the research project.

Any person who stores health-related personal data for research must take appropriate operational and organizational measures to protect it and, in particular, restrict the handling of the health-related personal data to those people who require this data to fulfill their duties; prevent unauthorized or accidental disclosure, alteration, deletion and copying of the health-related personal data; document all processing operations which are essential to ensure traceability (Art. 5 para. 1 HRO).

The HRA furthermore differentiates between genetic health related data and biological material on the one hand and non-genetic health related data on the other hand. This distinction is based on the argument that genetic data and biological material has a higher potential to predict future diseases and therefore a higher misuse potential. The requirements on the consent depend on the category of the data or biological material (see Art. 32 and 33 HRA).

2.2.2 Informed Consent

The participants of a research project must give their consent to any research project after being provided with sufficient information and given an adequate amount of time for consideration. The HRA specifies in Art. 16 about what and in which form the participant must be informed. The information includes inter alias type, purpose, duration and procedure of the research project. The concerned people have the right to deny or revoke their consent at all times and without explanation (Art. 7 HRA).

The people involved in research projects have the right on being informed about research results concerning their health or to renounce this information. Furthermore they have a right on information on all their personal data available (Art. 8 HRA). This is the default, but one has to consider that anonymous data cannot be linked back to a specific person, which is why the duty to inform lapses in that case consequently (Poledna 2015, Art. 8 N 16).

According to Art. 16 para. 1 phrase 1 HRA, a person may only be integrated into a research project if he or she has given his or her consent after having been sufficiently informed. Consent must be given in writing, although the Federal Council may provide for exceptions to this rule

(Art. 16 para. 1 phrase 2 HRA). Before the data subject decides on consent, he or she must be granted an appropriate period of reflection in accordance with Art. 16 para. 3 HFG.

2.2.3 Formalities

Consent to a research project must be given in writing, i. e., by analogy with Art. 13 et seq. of the Federal Act on the Amendment of the Swiss Civil Code (Part Five: The Code of Obligations; SR 220) after personal dating and signature of the declaration by the person competent for consent (Sprecher and Van Spyk 2015, Art. 16 N 23). According to prevailing doctrine, Art. 13 Code of Obligations as a legal principle of general value is also applicable in public law, in which human research is to be classified, and applies, in addition to contracts, to all declarations of intent to the extent required by law (Federal Court decision BGE 101 III 66, E. 3).

Written form means that the content of a declaration is permanently recorded by means of characters on a declaration medium, traditionally on paper documents (Schwenzer 2015, Art. 13 N 3). Font and writing instrument are irrelevant, provided that permanent embodiment is guaranteed (Schwenzer 2015, Art. 13 N 4). Smart technologies cannot permanently represent information because data merely symbolize a digital image of reality and are permanently changeable.

Of course, consent could be given by means of an electronic signature (Art. 14 para. 2bis Code of Obligations), which exists since the Federal Act on Certification Services in the Field of Electronic Signatures and Other Applications of Digital Certificates (SR 943.03) came into force. However, the digital signature has so far only been used in a few cases (Portmann 2010, p. 38). "SuisseID"[4] and its subsequent project "SwissID"[5] are also on a par with handwritten signatures and have the potential to accelerate the consent process in human research. Yet, they also lack the necessary dissemination.

Consent may be given in another form than written and documented if it is a "Category A" research project under the HRO with adults capable of making judgments, if written clarification and consent is disproportionate on the basis of the project ordinance and if the deviation from written form is indicated in the application to the responsible ethics committee (Art. 9 para. 1 HRO).

A research project corresponds to category A if the planned measures for the extraction of biological material or collection of personal data are associated with only minimal risks and burdens (Art. 7 para. 1 HRO). In order to assess the intensity of the intervention, the specific circumstances of the individual case must be considered. If the individual case must be examined, no standard procedure can be developed, which is the aim with regard to digitization. Thus, the

[4] `https://www.postsuisseid.ch/`
[5] `https://www.swissid.ch/`

way over the exception rule does not lead to the needed result.

2.2.4 Broad Consent

In the age of big data and digitization, it would be a disproportionate effort to request a written consent from every single data subject for every single research project. In practice, hospitals nowadays work with a so called "broad consent" which is obtained for the further use of biological material and health data for research purposes if the human research projects are not yet concrete. The previously described formalities also apply to the broad consent (Rudin 2015, Art. 32 N 17).

In this document, the data subject gives his permission to use his health-related data and biological material for research purposes. This procedure is controversial, because one consents to not yet defined research projects, which means the research subjects can not be duly informed about those projects; therefore, the consent can not be considered "informed" anymore. Nevertheless, most institutions work with it to the present day (Mausbach 2019, p. 8). Broad consent is permissible for research projects that pursue a biomedical purpose and are approved by an ethics committee (Baeriswyl 2015, p. 92).

2.3 Consent Procedure in Practice

Obtaining consent is challenging for hospitals for many reasons. It is not always clear who is responsible. In principle, the physician would be responsible to provide the patients with necessary information, but the task is often delegated to the people at the counter who are not sufficiently trained in all cases. In addition, the daily work at the counter of a hospital is psychosocially stressful, because the patients are usually in an exceptional situation. Added to this can be communication problems of a linguistic and cultural nature. For this reason, only 10-20% of the broad consent forms are being returned nowadays.[6]

Prototypical processes are currently being tested at the University Hospital Basel[7], whose contact person was Julia Maurer, PhD.[8] An SMS authentication of surname, first name, birthday and mobile phone number is followed by the signature of the consent on paper. The signature is then photographed and digitally recorded in the system. In contrast to this new approach for stationary stays, a photo of the patient's written consent has been taken during an ambulatory consultation by the physicians who identify the patient at the same time.

One may doubt whether the prototypical processes mean the desired simplification of the

[6] `https://www.sphn.ch/en/projects/infrastructure-development-projects.html` → 2. E-General Consent: Development and Implementation of a Nationwide Harmonized Interactive Electronic General Consent → Lay Summary

[7] `https://www.sphn.ch/en/projects/infrastructure-development-projects.html` → funded project No. 2.

[8] `https://dkf.unibas.ch/de/departement` → Clinical Study Competence Center – Regulatorik

process of obtaining consent. Although the banking business and other sensitive fields of activity have long since been digitized, this will probably only be possible in human research through an amendment to the ordinance or through a new type of two-factor authentication connected to the referring data in such a way that a subsequent change of the data can be recognized, which can also prevail in the health sector.

2.4 Operational Privacy Protection

Written consent by patients is a prerequisite for obtaining and analyzing health-related data. It then must be followed by rigorous data protection, as otherwise, patients will give no consent to share their data in the future.

Patients may feel a moral obligation to share health information for analytical purposes in order to improve care (Faden et al. 2013). But as ever greater amounts of data are stored electronically, there is generally a growing concern about data privacy in this digitized world. For instance, more than 85% of individuals have lied or misrepresented their personal information online (Rainie et al. 2013).

Individuals are particularly concerned when it comes to electronic health data (Angst and Agarwal 2009). Such data contains very sensitive information about individuals which they want to keep confidential. As a result, electronic health data must be handled securely in order to protect confidentiality. Otherwise, it is unlikely that individuals are giving consent to share their sensitive data in the future. That is, trust is an important factor determining individual willingness to disclose personal health information (Anderson and Agarwal 2011).

Data privacy can be divided into two essential goals: (i) de-identification (the analytical analogous for the legal term "de-coded") and (ii) concealment of sensitive information (Duncan and Lambert 1989). De-identification deals with the process where personal information is obfuscated such that individuals cannot be uniquely identified. Very few information is enough to identify individuals, e. g., a study by (Sweeney 2002) finds that 87% of the U.S. population are uniquely identified by a combination of gender, date of birth and the 5-digit zip code. Various mechanisms have been proposed in order to prevent identification from such data. For instance, data could be aggregated at the micro-level by removing the last digits from zip codes in order to make individual records less distinguishable.

Individuals are often less concerned about identification (except when they committed a crime), but rather, they worry about sensitive information that is linked to their identity. That is why, various techniques have been proposed in order to perturb sensitive information (For an overview, see Aggarwal and Philip 2008). The general idea is to randomize data, e. g., by

replacing, shuffling, or adding noise to data.

Different concepts exist in order to measure the strength of data privacy. For instance, the concept of k-anonymity has been proposed to quantify the strength of anonymity (Machanavajjhala et al. 2006). It states that data is k-anonymous if an individual record cannot be distinguished from at least k-1 records. Very rigorous solutions to protect sensitive information have emerged with the model of differential privacy (Dwork 2007), which evaluates algorithms for privacy protection based on mathematical guarantees regarding their risk of disclosing sensitive information.

Differentially private algorithms are a noble goal for stakeholders extremely concerned with data privacy. Yet, this perspective is one-sided: requiring high level of data privacy can make data useless for analysis (Brickell and Shmatikov 2008). There is a trade-off between data privacy and utility. On the one hand, data privacy is important in order to protect sensitive information. On the other hand, too much data privacy will obfuscate sensitive information to the degree that no findings can be derived from it.

A balance between data privacy for patients and data utility for research is particularly important when it comes to electronic health data. Sensitive information is often important in determining optimal treatments. Yet, if sensitive information is not dealt with care, researchers may be unable to obtain it in the first place.

2.5 Interoperability

The form of written consent and privacy protection varies greatly both between and within nations. As a result, hospitals may store data in different forms, or restrict the access for analytical purposes differently. This hinders interoperability, which here is the exchange of health information between hospitals.

In the following, we describe the status quo of the interoperable data infrastructure within the SPHN (and the SPHN Data Coordination Center). We point out several issues both from a legal and operational perspective. The Data Coordination Center (DCC), which is managed by the Personalized Health Informations Group of the Swiss Institute for Bioinformatics, aims at achieving technical and semantic interoperability. The DCC also provides a secure data and IT infrastructure called BioMedIT to allow exchange between the regional data nodes of the university hospitals (see Figure 1) and finally the researcher. BioMedIT is a coordinated network of IT infrastructure (nodes) to provide authorized members with a secure high-performance computing environment where they can analyze large amounts of biomedical data from patients. Currently, nodes in Basel, Lausanne and Zurich are built; the BioMedIT node in Zurich is

already operational. In summary, the DCC is supposed to provide a harmonized database which researchers can access in order to do their analyses on data that is pooled from all university hospitals.

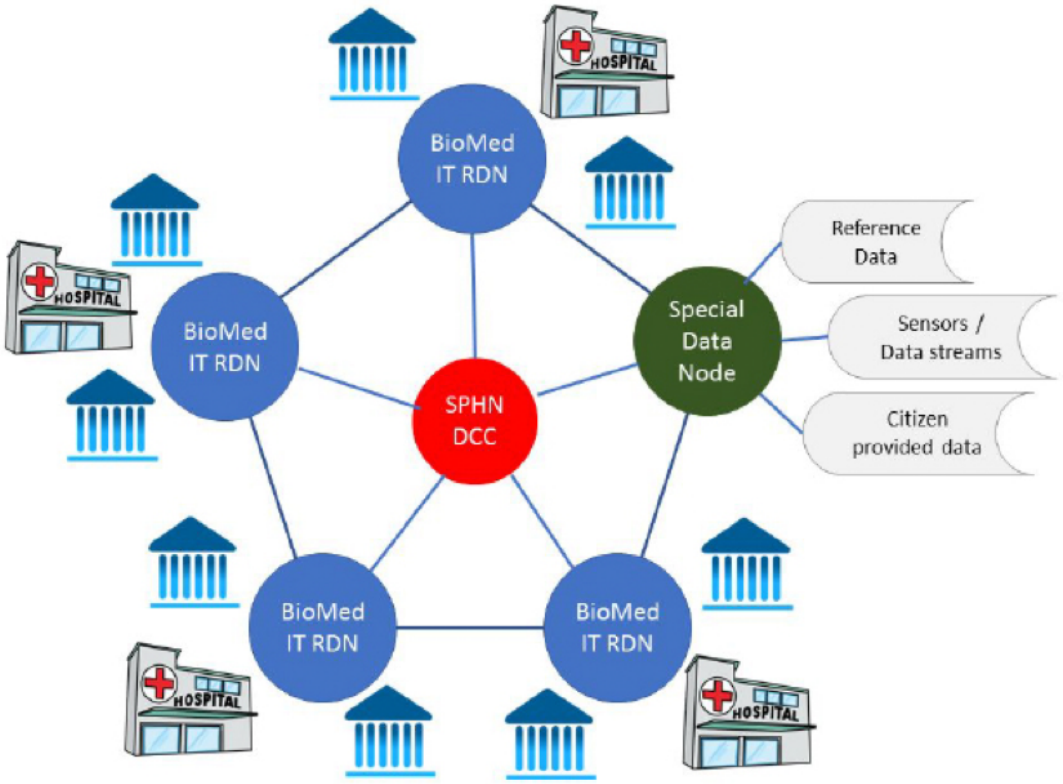

Figure 1: Infrastructure of the Data Coordination Center (DCC).

2.5.1 Legal and Ethical Perspective

The DCC has to comply with the legal framework. Herewith, the SPHN contributes to the development, the implementation and the validation of the required infrastructure to make health-related data interoperable across Switzerland. Therefore, the SPHN has developed an Information security policy[9], which applies the rules of the FADP, the HRA and their referring Ordinances as well as the Swiss Criminal Code (SR 311.0) to the BioMedIT-Infrastructure. Different EU Frameworks and already existing data protection regulations have also been taken into account.

Another example is the Data Transfer and Use agreement (DTUA) that facilitates data and material exchange in the context of academic research projects. In May 2019, a working group of the SPHN and the Swiss Biobanking Platform (SBP)[10] has developed a template for such

[9] https://dcc.sib.swiss/media/filer_public/a5/7b/a57b4ad4-c741-4be9-aefb-b0256be150c4/sphn_information_security_policy_v1.pdf

[10] https://swissbiobanking.ch/

an agreement. This template governs the transfer and use of data that is made available by a provider to a non-profit third party, which intents to use this data for own research purposes. SPHN and SBP encourages their partners as well as anyone working with data or biological material in Switzerland to adopt these templates.

Before accessing confidential data, the recipient project leader and the data provider both must sign a DTUA. To ensure that the data exchange also follows ethical guidelines, the SPHN has furthermore established a document named "Ethical Framework for responsible Data Processing in Personalized Health Research".[11] This document aims to provide ethical guidance in relation to personal data processing within the SPHN. This ethical guideline touches four general principles: respect for people, privacy, data fairness and accountability. In order to create this framework, the authors systematically analyzed multiple already existing international ethical guidelines.

2.5.2 Operational Perspective

Building an interoperable network between university hospitals is not an easy task. There are two hurdles:

- **Technical interoperability:** Requires an IT infrastructure where data can be exchanged between university hospitals or stored in a single database.
- **Semantic interoperability:** Requires harmonization of data, e. g., collecting and storing data in the same format.

Technical issues arise when hospitals or groups use different infrastructure and formats, e. g., hard- or software, to collect data. It may then be difficult to connect these separate infrastructures in order to allow for fast and easy exchange of data. If preferences about the use of hard- or software differ, it is also difficult to align infrastructures in the future. Nevertheless, while technical interoperability is not yet complete, the remaining hurdles are manageable and should be overcome over the next couple of years.

More difficult is semantic interoperability. Various standards exist about how electronic health data should be stored, e. g., the Logical Observation Identifiers Names and Codes (LOINC) is a universal standard for identifying medical laboratory observations. Stakeholders within the SPHN can agree on some common standards. However, for some electronic health data, standards are highly debated for two reasons.

[11] `https://www.sphn.ch/de/news-events-publications/publikationen.html`

Firstly, clinicians at the university hospitals have become familiar with their standards of collecting and storing data. Often they have different standards and are unwilling to align them. They are reluctant to change because either they believe their standard is the best or because they believe that having to adapt to a different standard will make them less efficient in their clinical routine.

Secondly, researchers who want to analyze data have their own preferences for standards. All researchers want harmonized data because otherwise they struggle converting data of different formats into a common one. Conversions can always be criticized and studies are less comparable when researchers choose different conversions. That is why LOINC and other universal standards are being established.

Consensus about the right standards is already very difficult to reach within clinicians and researchers; between them is even more challenging. As a remedy, SPHN creates working groups consisting of both clinicians and researchers. Each working group tries to find consensus for standards relating to their projects. Consensus is summarized via data catalogs, where data variables are defined and their format is clarified. All university hospitals are then required to provide the data as described in the catalog.

SPHN working groups create catalogs at different paces. Thereby, consensus is often more difficult to reach when universal standards do not yet exist, or when standards compete strongly. In such cases, participants of the working groups must be careful to avoid deadlocks, i. e., a point where no consensus has been reached and no further progress can be made.

3 Blueprint: Testing Data Exchange

We have described several legal and operational issues that hinder data exchange and, through various interviews with stakeholders from PHRT and SPHN, we identified data interoperability as the currently most pressing issue. Various working groups and driver projects within SPHN and PHRT are working towards solving this issue, thereby developing infrastructure, governance and concepts that ultimately lead to an interoperable data infrastructure.

Along the way, researchers are eager to get early access to the pool of data that is being generated. We believe that in some cases early access can be granted even before the infrastructure is fully developed. However, only if data security can be guaranteed. We argue that the underlying infrastructure and process for data exchange should be subject to a stress test before opening the data pool to researchers outside SPHN and PHRT.

The stress test we developed is similar to what is done in other industries. For instance, consider a software engineer: one would not want to release the software before a beta version was

thoroughly tested. Similarly, within SPHN, working groups would not want to invite researchers to access their data (e. g., via the DCC) unless the process from data request to data return has been thoroughly tested.

For an external researcher, a hampering process to obtain data could put their whole research projects on hold and lead to frustration. In order to avoid such false starts, we propose working groups to run stress tests on their projects. The conceptual idea is to test the whole process from data request to return and check whether it is comparably quick and easy for external researchers to receive and analyze the data they need for their research project. There are several ways in which a user-centered stress test could help a SPHN working group:

- **Verify progress:** Throughout the project, stress tests can be run to verify if work packages have really been accomplished as claimed.
- **Verify legal processes:** Uncover and mitigate legal hurdles, e. g., implementing an efficient consent procedure.
- **Advertise success:** SPHN is a large initiative with many stakeholders and we feel that individual stakeholders are often occupied with their daily tasks, thereby often missing the bigger picture of SPHN and how they contribute to it. Verifying that progress is made and advertising the success could encourage stakeholders to move forward with the SPHN initiative.
- **Detect issues:** Stress tests uncover what is not working and needs to be fixed.
- **Determine type of change:** Issues can be small or large and corresponding reactions need to be incremental or radical, e. g., the latter when being in a deadlock.

We have described issues with interoperability both from a legal and operational perspective. Similarly, it requires both a legal and operational framework in order to design a blueprint for testing data exchange. Legally, for instance, ethics approval and data de-identification need to be conceptualized. Operationally, for instance, researchers need to be informed about where and how to request data, which and in what form they can analyze the data, and how data needs to be returned.

3.1 Legal Framework

In Switzerland, there are no independent rights to data and due to the lack of physicality, property law does not apply to data (Hürlimann and Zech 2016, p. 19). Federal Court decision BGE 136 III 401 overcomes the central dogma of the rights of personality, according to which

consent to a violation of personality can be independently revoked at any time. The rights of personality can be the subject of contractual obligations (Karavas 2018). The Federal Court stated that an exception only applies to the personal core area of personality, in which a valid contractual agreement is to remain excluded (Federal Court decision BGE 136 III 401, E. 5.4).

The data provider delivers the data in a coded form and in a format agreed by the parties (DTUA template, p. 2).[12] He is responsible for the de-identification of the data. The data provider is furthermore responsible that he is entitled to supply the data. This means he has obtained all necessary contents and authorizations for the transfer and use of the data (DTUA template, p. 2).

On the other hand, the investigator is responsible for obtaining ethical approval for the planned research project as well as entering the DTUA with the data provider (SPHN Information Security Policy, p. 3).

3.2 Operational Framework

The operational framework of the stress test is depicted in Figure 2. It can be divided into three steps:

- **Data request:** From the perspective of an external researcher, where is my first point of contact? What formal requirements need to be satisfied upon first contact (e. g., contract, ethical approval, etc)? Is there a data catalog describing the data (e. g., variables, their type and format)?
- **Data sharing:** How do I get access to the data? Is the data de-identified and which fields are masked (e. g., name and birthdate) and how?
- **Data return:** Do I need to return the data on a specified date? Is it possible to get access to the data for a follow-up study without registering a new research project? If I encountered data issues, how can I curate the data for future researchers (e. g., marking an outlier)? Is it possible to add the results of the study to the data (e. g., a type of Machine Learning model in the form of unstructured data)?

Our concept of a stress test is very generic and needs to be adapted to the specific use case. Stress tests can be designed to test the whole process with a (pass/fail) outcome, or they can be designed to test intermediary steps if the respective use case is still under development.

[12] `https://www.sphn.ch/en/projects/elsi-activities.html`

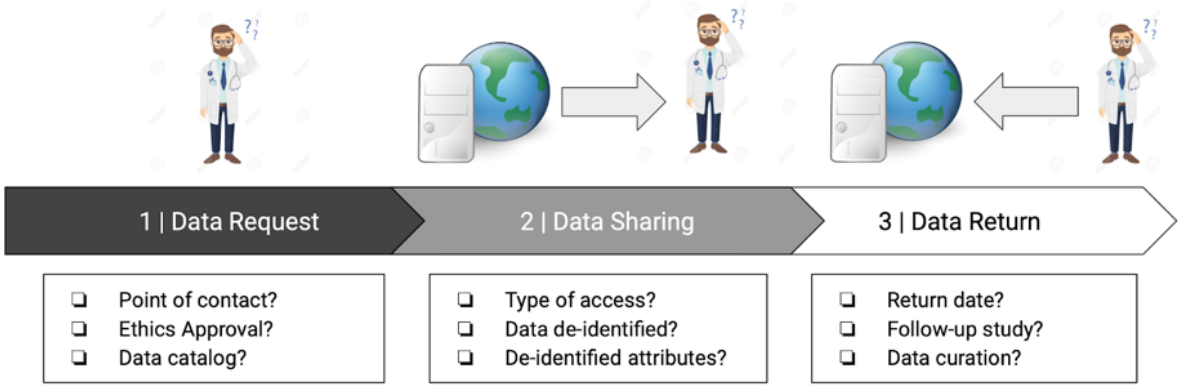

Figure 2: Operational framework for a blueprint to test data sharing.

4 Case Study: "Sarah the Researcher" Requesting Data from the Driver Project Swiss Personalized Oncology

Within SPHN, "Sarah the Researcher" represents the idea of how access is granted to outside researchers (see Figure 3). To the best of our knowledge, the described procedure was not yet tested for the majority of data. Yet, we believe this is of great importance.

The driver project Swiss Personalized Oncology (SPO) works on the swiss-wide interoperability of genomic, clinical and laboratory data from cancer patients to achieve personalized and improved treatment algorithms for cancer patients. One main deliverable is a data warehouse infrastructure capturing data from university hospitals. Work package one of the SPO contains the generation of a structured data catalog based on 20 sample cases per contributing hospital. This catalog describes the data, i. e., the definition of the variables, their type and format. It is planned to gather the data within the DCC data warehouse infrastructures. The long term goal would be a continuous collection from all consenting patients within the standard clinical flow.

We envisioned to use this catalog of the SPO to run a test of "Sarah the Researcher" and its approach to judge the applicability.

4.1 Feasibility

To start a new project, an outside researcher needs information what kind of data is available (data catalog). In SPO, they are currently creating this data catalog based on 20 samples per hospital. That is, the data catalog is not yet available to the public. Outside researchers would thus have to manually inquire what kind of data is available until the planned starting date of their research. They further need to inquire how they can access the data, i. e., retrieve the data from a database, run the analysis on a BioMedIT node, or collect the data from each hospital

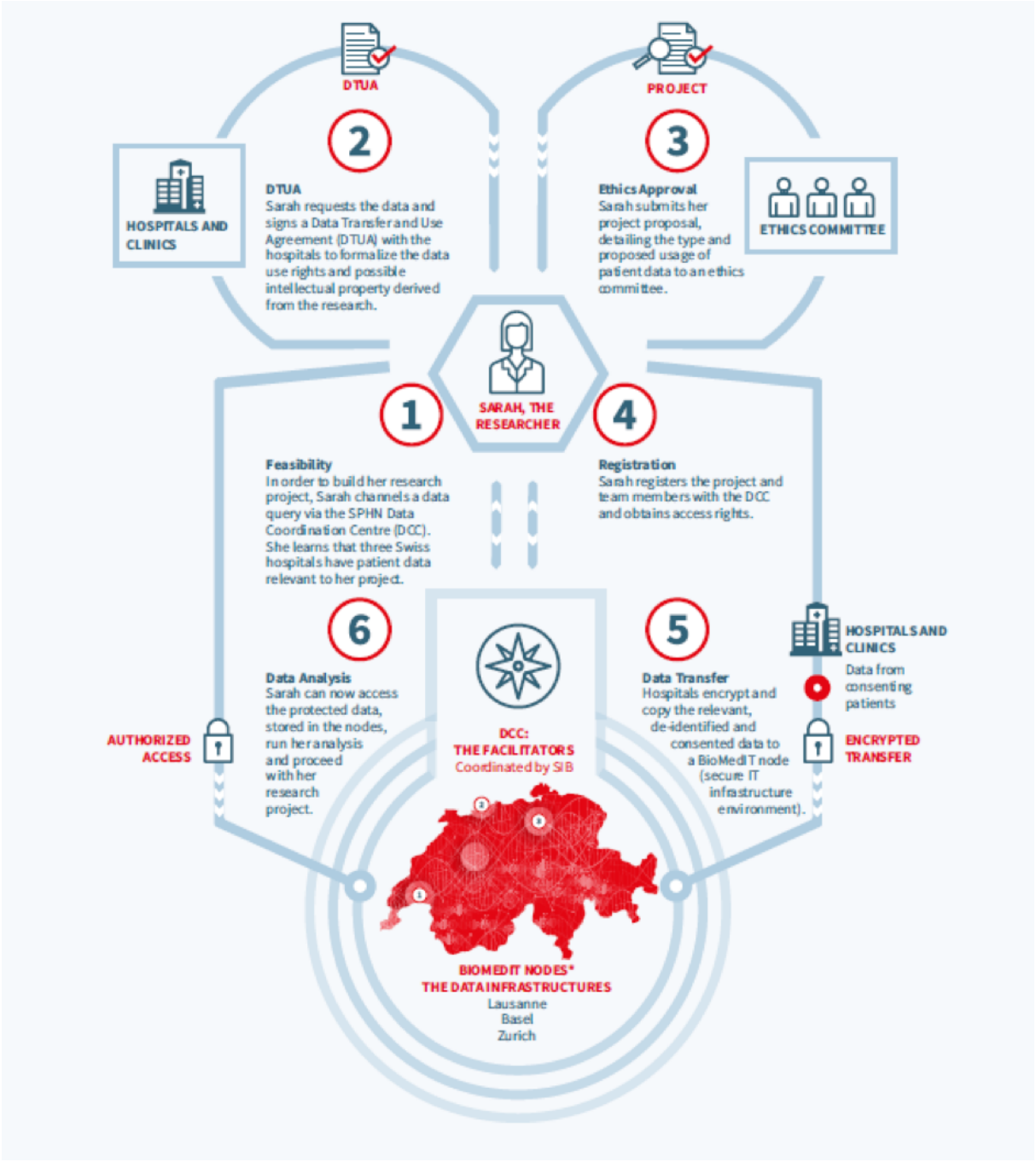

Figure 3: The concept of "Sara the Researcher".

separately. This is burdensome because each inquiry needs to be handled manually until the data catalog is established and rules of access are defined.

4.2 DTUA

The SPO's main duty is to ensure that the provided data is properly encoded and all the legal requirements concerning the sharing of biological material and genetic data as stated in the HRA and the HRO are met. Therefore, the DTUA between the SPO (providing party) and "Sarah

the Researcher" (recipient) must be concluded.

The template constitutes a key regulatory requirement which defines the rights, responsibilities and obligations of the parties involved (e. g., provider, recipient and processor) regarding permitted use, ownership, publications, intellectual property as well as liability when data is being transferred or accessed in the frame of a project (SPHN Fact Sheet: DTUA template, p. 1).

In this agreement, SPO must justify that there is an agreed reason to provide access. SPO is furthermore responsible for classifying the data according to risk and data privacy requirements, which ensure appropriate levels of security for the confidential or high-value information assets (public, internal or confidential assets).

Finally, SPO must notify the BioMedIT Node if a user in the project team does no longer need access to data or systems and ensure that access rights are removed when the project purpose ends or ceases to be valid (SPHN Information Security Police, p. 13).

If not only data but also biological material is to be transferred, a Material Transfer Agreement is to be signed. Discussing the specialty of such an agreement would go beyond the scope of this paper. The Swiss Biobanking Platform has developed a template for such an agreement.[13]

4.3 Ethics Approval

Art. 45 HRA states that research with patient and clinical data requires prior approval; annex 2 of the HRO states the necessary application documentation, which includes the following documents:

1. A base form which includes an abstract of the scientific issue in the native language of the location where the project takes place.

2. A description of the scientific issue.

3. Evidence of origin of the biological material and/or the health related personal data as well as observance of the requirements concerning informed consent respectively information concerning their right to revoke.

4. In case the research project occurs with pseudonymised biological material/health related personal data, evidence concerning their safe and correct pseudonymization must be provided. According to Art. 26 HRO, the person who holds the key to the encrypted data must also be named.

[13] `https://swissbiobanking.ch/materiel-transfer-agreement-template-2/`

5. Evidence of the safe handling with biological material/personal data, namely their preservation.

6. The project leader's Curriculum Vitae, including evidence of his/her knowledge and experience, as well as a list of further people involved in the research project including their functions and their corresponding technical competences.

7. Information concerning the infrastructure available at the place of realization of the project.

8. Possible approvals of Swiss ethic committees concerning the collection of biopsy respectively of health related personal data.

The researcher must hand in the application documents to the Ethics committee in the canton where the research project takes place. While ethical requests might pose no hurdle for some external researchers, we do assume that especially for basic and applied researchers writing a proposal for ethical approval is a frightening undertake. For the test case of SPO, this is no different. We could therefore imagine SPHN provides support and assistance for "future Sarahs". It is for example feasible that a member of the ELSI Advisory group serves as contact person if questions concerning the ethical approval arise or to control the application documents before the researcher hands them in to the cantonal ethics committee.

4.4 Registration

The project has to be registered with the SPHN and DCC. Personally, we think projects should be registered earlier. This would allow SPHN to monitor the amount of initially registered projects and, subsequently, the proportion of projects that are dropped because they were deemed unfeasible or unethical. This could also inform other researchers and prevent them from pursuing projects that were dropped in the past.

4.5 Data Transfer

Most data will be transferred from the secure hospital environment to a BioMedIT node, where it can be analyzed by running scripts on the server.[14] We compare the results of this analysis with the results obtained when performing the analysis on the original data at the hospital. Comparison will ensure that data transfer was successful and that analyses return the same results both locally and on the server. Besides consistency, further technical aspects could be

[14]Some projects, particularly those that are in their early stages, may not yet be able to combine data and send it to the BioMed IT node. Here researchers may still need to go to each hospital, collect the data and merge them manually.

evaluated such as runtime and ease of use for the researcher, e. g., some researchers may find it burdensome to run analysis on a server rather than on their local computer. Although we believe that testing data transfer for SPO will not differ greatly from other cases, there can still be specific issues with respect to the data, e. g., the amount or complexity of data that is exchanged in SPO may be larger and thus running the analysis on the server may be more complicated compared to analysis on the server for other electronic health records.

5 Conclusions

Data interoperability is crucial for biomedical research and this importance has been recognized in Switzerland with the ambitious national initiative Swiss Personalized Health Network (SPHN) and the strategic focus of the ETH domain Personalized Health and Related Technologies (PHRT). In this work, we gave an overview of legal and operational issues regarding data interoperability, and subsequently provided a legal and operational framework for testing data exchange within the data coordination center (DCC) of SPHN.

From a legal perspective, the law intents to protect the personality of persons who provide health-related data for research purposes. This legal requirements are complicating research, which is to a certain point unavoidable even if the interests of researchers and participants are mostly aligned. However, a framework too formal and restrictive may lead to a situation where research progress and innovation are thwarted. Such a scenario helps neither the researchers nor the participants, especially in a research field where many of the participants suffer from severe diseases and are in desperate need for innovative new therapies. The objective from a legal perspective must thus be to precisely examine the protective purpose of the individual legal norm and evaluate if new technical methods allow to fulfill this purpose. If so, the legal framework can eventually be interpreted in a way that suits the research while at the same time respecting the personality and dignity of the research subjects.

From an operational perspective, technical interoperability is seen as manageable, whereas semantic interoperability is more difficult because the involved stakeholders have more diverse opinions that impede consensus.

Legal and operational hurdles should not prevent progress and our advice is to test data exchange at an early stage. We developed a blueprint for this purpose along the lines of the approach of "Sarah the Researcher". Furthermore, we detailed how this may be applicable to a specific case, here the work package one of the driver project in Swiss Personalized Oncology (SPO). Our blueprint could help identifying issues that hinder data interoperability. However, since SPO is at an early stage, we could not yet validate our blueprint and thus recommend that a first test

data transfer should be executed as soon as work package one is finished. This test run will provide valuable information on how to improve and further develop the data sharing within SPO, the DCC and finally SPHN as a whole. It will also help refine the concept of "Sarah the Researcher". However, this can only be the first step and many more need to follow before SPHN can open its DCC and the contained data to outside researchers.

References

Aggarwal CC, Philip SY (2008) *Privacy-preserving data mining: Models and algorithms* (Springer Science & Business Media).

Anderson CL, Agarwal R (2011) The digitization of healthcare: Boundary risks, emotion, and consumer willingness to disclose personal health information. *Information Systems Research* 22(3):469–490.

Angst CM, Agarwal R (2009) Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. *MIS quarterly* 33(2):339–370.

Baeriswyl B (2015) Art. 4-11a. *Stämpflis Handkommentar zum DSG*.

Brickell J, Shmatikov V (2008) The cost of privacy. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, 70 (New York, New York, USA: ACM Press).

Duncan G, Lambert D (1989) The risk of disclosure for microdata. *Journal of Business & Economic Statistics* 7(2):207–217.

Dwork C (2007) An ad omnia approach to defining and achieving private data analysis. *Privacy, Security, and Trust in KDD*, 1–13 (Berlin, Heidelberg: Springer Berlin Heidelberg).

Faden RR, Kass NE, Goodman SN, Pronovost P, Tunis S, Beauchamp TL (2013) An ethics framework for a learning health care system: A departure from traditional research ethics and clinical ethics. *Hastings Center Report* 43(s1):S16–S27.

Hürlimann D, Zech H (2016) Rechte an Daten. *sui-generis* (27):85–95.

Karavas V (2018) *Körperverfassungsrecht* (Dike).

Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M (2006) L-diversity: Privacy beyond k-anonymity. *22nd International Conference on Data Engineering (ICDE'06)*, 24–24 (IEEE).

Mausbach J (2019) Dynamische Einwilligung zur Forschung am Menschen. *Jusletter* (January 28, 2019).

Nassehi A (2019) *Muster-Theorie der digitalen Gesellschaft* (C.H.Beck).

Ohmann C, Banzi R, Canham S, Battaglia S, Matei M, Ariyo C, Becnel L, Bierer B, Bowers S, Clivio L, Dias M, Druml C, Faure H, Fenner M, Galvez J, Ghersi D, Gluud C, Groves T, Houston P, Karam G, Kalra D, Knowles RL, Krleža-Jerić K, Kubiak C, Kuchinke W, Kush R, Lukkarinen A, Marques PS, Newbigging A, O'Callaghan J, Ravaud P, Schlünder I, Shanahan D, Sitter H, Spalding D, Tudur-Smith C, van Reusel P, van Veen EB, Visser GR, Wilson J, Demotes-Mainard J (2017) Sharing and reuse of individual participant data from clinical trials: Principles and recommendations. *BMJ Open* 7(12):e018647.

Poledna T (2015) Art. 8-10. *Stämplis Handkommentar zum HFG*.

Portmann R (2010) Herausforderung digitale Signatur. *digma* (1):38–39.

Rainie L, Kiesler S, Kang R, Madden M, Duggan M, Brown S, Dabbish L (2013) Anonymity, privacy, and security online. *Pew Research Center* 5.

Rudin B (2015) Art. 32-25. *Stämpflis Handkommentar zum HFG*.

Salathé M, Driessen S (2016) Generalkonsent: Eine Einheitliche Vorlage soll schweizweite Forschung erleichtern. *SAMW Bulletin* (3):1–4.

Schwenzer I (2015) *Art. 11-17* (Helbing Lichtenhahn).

Sprecher F (2018) Datenschutz und Big Data im Allgemeinen und im Gesundheitsrecht im Besonderen / V.-VII. *ZBJV* (154):519–552.

Sprecher F, Van Spyk B (2015) Art. 16-18. *Stämpflis Handkommentar zum HFG*.

Sweeney L (2002) k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10(05):557–570.

Thouvenin F (2017) Forschung im Spannungsfeld von Big Data und Datenschutzrecht: Eine Problemskizze. *SRP – Schriften zur Rechtspsychologie* (15):27–53.

Vayena E, Gasser U (2016) Between openness and privacy in genomics. *PLOS Medicine* 13(1):e1001937.

Williams H, Spencer K, Sanders C, Lund D, Whitley EA, Kaye J, Dixon WG (2015) Dynamic consent: A possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. *JMIR Medical Informatics* 3(1):e3.