2nd Joint Personalized Health Day Switzerland

The future of data-driven medicine in Switzerland

Abstract Booklet

24 August 2022

Strategic Focus Area
Personalized Health and Related Technologies

Swiss Personalized Health Network
SPHN

Poster Titles

SPHN | Swiss Personalized Health Network
Haus der Akademien | Laupenstrasse 7
CH-3001 Bern
nds@sphn.ch | www.sphn.ch

PHRT | Personalized Health and Related Technologies
CLP | Clausiusstrasse 45
CH-8092 Zürich
phrt-office@ethz.ch | www.sfa-phrt.ch

1. CLINICAL DATA MANAGEMENT AT THE SWISS UNIVERSITY HOSPITALS

Horki P[1], Kalt KT, Hirschi P, Zoergiebel S, Mueller-Breckenridge M, Crameri K

[1]*SIB Swiss Institute of Bioinformatics | Personalized Health Informatics Group*

Keywords: Data sharing, clinical data integration, metadata catalogue, SPHN Interoperability Framework

During healthcare processes, the patient story is spread over multiple applications, harbouring a wealth of data. However, the data is stored in various disparate systems or databases. To leverage the potential of healthcare data for both patient treatment and medical research, it is necessary to obtain an overall view of a patient's data across the various systems, and to ensure the availability of data for repeated and continuous use.

Through a complex and technically - but also from a governance point of view - challenging process of data integration from heterogeneous primary source systems into central data management structures, the university hospitals enable the efficient secondary use of routine care data and thus allow an easy access to an enormous breadth and depth of patient data. This means significantly less effort required to obtain data for researchers, and ensures data availability for a wide variety of purposes also outside of research.

A further, laborious work-step in the hospitals concerns the implementation of standardized encoding or mapping to terminologies/ontologies of the data according to the SPHN Interoperability Framework. Thanks to this process, it is possible to exchange interoperable data between different sites. In a related work, SPHN is currently developing an innovative and sustainable pipeline making these steps even easier and more cost-effective. Finally yet importantly, a first proxy of an SPHN metadata catalogue provides researchers with a good overview of what the university hospitals can deliver in a timely and cost-effective manner. This catalogue will continue to grow during the life of the new SPHN Data Streams and ongoing work at the university hospitals.

2. BIOMEDIT: SWITZERLAND'S SECURE INFRASTRUCTURE FOR ANALYSING BIOMEDICAL DATA

Kapoor S, Crameri K, Appleton O, Gabbretti R, Fernandez Pinilla P, Fox M, Guzzi S, Österle S, Rinn B, Sengstag T, Stockinger H

[1]SIB Swiss Institute of Bioinformatics | Personalized Health Informatics Group

Keywords: BioMedIT, data sharing, technology development, data processing, data security

The advent of digital transformation in health care is producing an exponential increase in the amount of information available for each patient. The use of this information in data-driven biomedical research is critical and can lead to important changes in medicine and quality of life. Given the sensitive nature of this health-related information, data-driven and personalised health research requires special IT infrastructure and services, blending the concepts of security, scalability and performance without compromising flexibility of use.

To address the above needs, the BioMedIT network project was funded by the Swiss federal government within the framework of the Swiss Personalized Health Network Initiative (SPHN). The goal of the BioMedIT network project is the creation and operation of a national, secure IT infrastructure to support computational biomedical research and clinical bioinformatics using sensitive data. The BioMedIT network can be used by all Swiss universities, research institutions, hospitals and other interested partners, wherever there is a need to process sensitive data

In this poster, we present the BioMedIT network and its setup in three legally independent scientific IT competence platforms at ETH Zurich, SIB/UNIL in Lausanne and University of Basel. Furthermore, we provide an overview of scientific and high throughput computing services designed especially for sensitive/confidential data for personalized health and data-driven research in Switzerland.

3. A LEGAL AND ETHICAL FRAMEWORK FOR A RESPONSIBLE SHARING AND USE OF HEALTH DATA IN MULTI-CENTER RE-SEARCH PROJECTS

Maurer J[1], Erard F, Crameri K

[1]*SIB Swiss Institute of Bioinformatics | Personalized Health Informatics Group*

Keywords: Data sharing, ELSI, health data, privacy

To ensure patient privacy and to promote responsible data usage, data-sharing policies must be agreed and observed in the Swiss research community. The Swiss Personalized Health Network (SPHN) initiative has identified three major areas for responsible data sharing:

1. Protecting the privacy and integrity of patients and their personal health information,

2. Establishing an overarching framework to regulate general principles of collaboration and conditions of data sharing, and

3. Provisioning a secure IT platform for data processing and storage.

The importance of individual rights, privacy, data fairness and accountability has been illustrated in the published SPHN Ethical Framework and implemented as part of the SPHN-funded projects. An exemplary essential approach to protect patient privacy prior to data sharing is the de-identification of personal health data.

Research collaboration partners are in need of a contractual framework and governance rules of data providing institutions to regulate data sharing. An overarching framework to regulate general principles of collaboration and conditions under which data are disclosed to other parties have been set up through the release of several agreement templates.

BioMedIT, a secure IT network, provides essential security measures to protect sensitive personal data used in researchers' projects in compliance with Swiss law requirements.

With the implementation of these three pillars, SPHN became the primary driver for im-proving the nationwide ecosystem for a responsible data sharing in biomedical research while complying with regulatory requirements and addressing the conditions of data providing institutions.

4. FROM CLINICAL (ROUTINE) DATA TO FAIR RESEARCH DATA

Touré V[1], Gnodkte K, Armida J, Crameri K, Österle S

[1]*SIB Swiss Institute of Bioinformatics | Personalized Health Informatics Group*

The Swiss Personalized Health Network (SPHN) builds a nationwide infrastructure for sharing clinical and other health-related data in a secure and FAIR (Findable, Accessible, Interoperable, Reusable) manner, ensuring that data coming from different sources is interoperable and exchangeable between stakeholders. Rather than relying on existing data models, the strategy is to develop a purpose-independent description of existing data and knowledge.

A comprehensive framework was built encompassing the definition of semantics for data unambiguity and standardization, and data format specifications for data exchange and exploration. Well-defined semantic concepts connected to external standards (e.g., SNOMED CT and LOINC) function as reusable building blocks to represent various information and data types of different medical disciplines.

For a formal representation of the semantics, a graph-based approach is used to encode and connect the concepts - the Resource Description Framework (RDF) - which is part of the Semantic Web technologies. This resulted in the development of the SPHN RDF schema. To support the use of this schema and facilitate the work of stakeholders, training and documentation are provided, as well as tools such as:

1) the DCC Terminology Service, to integrate external semantic standards into the SPHN RDF schema and data;

2) the SHACLer, to automatically build SHACL rules for data validation and;

3) the SPARQLer to automatically build SPARQL queries for facilitating data exploration in triplestores.

This framework is implemented in all Swiss university hospitals and forms the basis for future data-driven research with clinical and health-related data, facilitating knowledge sharing among all stakeholders.

5. SWISS BIOREF: A NATIONAL INFRASTRUCTURE FOR GENERATING PERSONALIZED REFERENCE LIMITS FOR DIAGNOSTIC MEDICINE

Blatter T[1], Witte H, Nath U, Leichtle AB

[1]*Inselspital Bern, Department of Clinical Chemistry*

Keywords: Personalized health, Reference Limits, Research Infrastructure, Sensitive Data, Data Security

Reference limits for patient test results are in standard use across many medical disciplines, allowing physicians to identify potentially pathological test results with relative ease. The process of inferring cohort-specific reference limits is however often ignored due to the high costs and cumbersome effort associated with such a task. Determining reference limits based on data collected during daily clinical routine using fully automated computational resources may help to lower the affiliated costs and personalize the reference limits to the respective cohort population and enhance patient care.

Cohort profile: With the Swiss BioRef project, we have developed a multi-center computational framework, where specialized web applications estimate and assess patient group-specific reference limits based on clinical routine data from four major Swiss Hospitals. We have established a common legal governance and interoperability framework for our clinical partners to share their data either to a central database via the secure BioMedIT network or providing their data in a decentralized way via MedCo, the secure and encrypted data-sharing system, allowing each data provider to abide to the restrictions laid out by their cantonal ethics waivers.

Web applications: The deployed web applications, which allow intuitive and interactive data stratification by patient factors (such as age, administrative sex, and personal medical history) and laboratory analysis factors (such as device, analyzer and test kit identifier) are just made accessible for Swiss physicians and researcher. We believe that establishing an opportunity for clinical physicians and researchers to define personalized reference limits in a convenient and reproducible way is a vital part of practiced precision medicine today.

6. FROM ROUTINE TO RESEARCH DATA: AUTOMATIC FLOW AND PREPROCESSING OF ROUTINE DATA IN THE NCCR ANTIRESIST PROJECT

Grieder S[1], Büchel D, Franzeck F, Saupper C, Sluka C

[1]*University Hospital of Basel, Department of Clinical Research*

Keywords: Clinical data integration, technology development, clinical routine data, data mart, automatic data transfer

The Clinical Studies work package of NCCR AntiResist collects samples and routine data of infected patients to expand our understanding of antibiotic resistance. At the University Hospital of Basel (USB) we implement an automated data pipeline integrating hospital routine data provided by the USB Clinical Data Warehouse (CDWH) into an electronic data capture (EDC) system, where data are curated, and additional data are manually entered by the study team. The data pipeline contains a large set of quality checks and data preprocessing steps. The process involves (1) the identification of infected patients with signed / non-refused general consent for research in the CDWH, biobanks, and the EDC, (2) regular diffs between already imported and new data, (3) the transformation of routine data from free text into the EDC data model, (4) data quality checks and feedback to the study team of any inconsistency that cannot be resolved automatically, and (5) secure data transfer. A particular challenge is the expected size of the project and its amount of data, with daily executions anticipated for a period of multiple years. Our pipeline is built out of a series of R, PL/SQL, and Bash scripts and a MariaDB data mart, where intermediate data are stored and preprocessed. The tools and routines developed for this project have been set up such that they can be used as a blueprint for future projects with similar requirements.

7. GENERATION AND EVALUATION OF SYNTHETIC DATA IN A UNIVERSITY HOSPITAL SETTING

Kaabachi B[1], Despraz J, Meurers T, Prasser F, Raisaro JL

[1]CHUV, Biomedical Data Science Center

Keywords: Medical Research; Synthetic Data; Privacy; Utility; Generative Adversarial Networks

Thanks to recent advances in machine learning techniques and improved processing capacity, generative models have illustrated innovation in a wide range of disciplines. One such discipline that would benefit highly from the ability to create virtual cohorts is medical research. Yet, to fully utilize the potential of generative models, it is required that the created synthetic data retains the statistical properties of the original cohorts while protecting the privacy of the patients present in the original dataset.

In this study, we propose a method to adapt generative models to work with longitudinal data from electronic health records (EHR) (both discrete and numeric) and develop a framework for systematically evaluating the utility-privacy trade-off for different data sharing scenarios and attacker models.

8. BUILDING AN INNOVATIVE, SUSTAINABLE AND EXPANDABLE DATA PROVIDER ARCHITECTURE THAT GENERATES ADDITIONAL VALUE FOR BIOMEDICAL RESEARCH

[1]Kalt K, Österle S, Crameri K, Krauss P, Obreiter M, Teixeira D, Raisaro JL, Unterkircher A, Dietrich RA

[1]*University Hospital Zurich, Directorate for Research and Education*

Keywords: Health Data Ecosystem, SPHN Connector, Service orientation, Semantics, IT Architecture

In the past, many research initiatives relied on proprietary or common ad-hoc data models such as OMOP or i2b2, and the process of acquiring data from source systems was limited to data that met the requirements of the specific research project. In SPHN, we follow a more generic approach and develop a broad platform that follows the SPHN semantic framework to describe different structured data types using international standards. The SPHN IT Architecture Working Group – with representation from the five university hospitals – is currently developing common IT infrastructure components to transform, validate, store and deliver data to research projects or shared services. Core components such as the SPHN connector or the research data repository will not only enable faster provision of consistently high quality data, but will also allow the development of a data catalogue with information on the availability and provenance of the data. Finally, applied technologies allow federated analysis and learning by bringing the algorithm to the data, rather than delivering gigabytes of sensitive data (repeatedly) to research projects outside the healthcare institutions. Other healthcare facilities can also implement the developed infrastructure components to provide high value health data to research projects on a national level with limited effort.

Although the infrastructure is still under development, we are working with other initiatives to leverage synergies and establish the architecture as a cross-institution and cross-initiative platform that could form the basis for the future research-data ecosystem and consequently can be sustainably managed beyond the life of SPHN.

9. RESOLVING DIAGNOSTIC DEADLOCKS IN SYSTEMIC AUTOINFLAMMATORY DISEASES

Linder A[1], Portmann K, Giurgea I, Eyer K

[1]ETH Zurich, Institute of Pharmaceutical Sciences (IPW)

Keywords: Autoinflammation, microfluidics, single-cell analysis

Systemic autoinflammatory diseases (SAIDs) are a set of rare heterogenous conditions characterized by episodes of systemic inflammation. These flares are caused by the abnormal activation of the innate immune system in absence of infection or autoimmunity. Since its discovery, the genetic nature of these diseases has been rapidly unraveled with the continuous identification of mutations affecting innate immunity. Indeed, to this day, approximately 60 genes have been shown to be involved in SAIDs alongside a wide range of clinical manifestations. Therefore, diagnosis relies mainly on suspicion and genetic testing. The current diagnostic strategy, however, induces a median delay of 7.3 years from symptoms onset to diagnosis. Indeed, disease-causing mutations cannot to be identified for 40% to 60% of patients with clinical manifestation of SAIDs. Thus, there is a need to identify and define objective diagnostic criteria for patients with unknown mutations to shorten diagnosis delays and provide adequate treatments.

Here, we propose to directly measure the dysregulation and activation of innate immune cells in response to stimulation on a single-cell level. For this, we use our platform for functional single-cell measurements. First, we integrated a novel bioassay to measure quantitatively and dynamically cellular dysregulation of innate immune cells. Next, we will employ the developed assays in a proof-of-concept study of healthy donors and SAIDs patients with known or unknown mutations. This study will enable us to measure the innate immune response with an unprecedented resolution and precision, and we aim to identify objective criteria in the cellular responses to enable an accelerated and precise diagnosis. Beyond these markers, we will test the ability of the measures to improve and personalize treatments of SAIDs. Lastly, the use of this novel system might help to identify involved disease pathways and advance our understanding of the pathophysiology of SAIDs.

10. SEMI-AUTOMATED METHOD TO MAP ICD-10 CONCEPTS TO SNOMED CT BY USING A FHIR TERMINOLOGY SERVER

Mateen A[1], Triep K, Endrich O

[1]*Inselspital, Medizincontrolling*

Keywords: FHIR Terminology Server, ICD-10, SNOMED CT, Snowstorm, Webservices

The International Classification of Diseases (ICD) is the most widely used classification system globally [1] as well as in Swiss hospitals for health management and capturing clinical disease terms. Despite the fact that it is widely used, it is not designed for the documentation of clinical care due to aggregation of similar concepts. On the other hand, SNOMED CT was developed to capture clinical data at a granular level. SNOMED CT is the most comprehensive clinical terminology that provides a uniform way of describing health conditions and their associated information. The usage of SNOMED CT is considered as an important step to digitalize the health care system. In current work, we did a pilot project to show the prospects of mapping ICD-10 codes to SNOMED CT.

Methods: The University Hospital of Bern, Inselspital, provided the list of ICD-10 codes, stored in Clinical Data Warehouse. The publically available Snowstorm FHIR Terminology Server endpoint was used to retrieve the SNOMED CT concepts by using reverse look up [2].

Results: There were twenty-three ICD-10 codes, out of which fourteen codes retrieved three hundred thirty-three SNOMED CT terms. One Code T81.8 had eighty-three equivalent SNOMED concepts. This is due to the fact that SNOMED CT provides much more specific clinical terms compared to ICD-10.

References:

[1] International Classification of Diseases. Wikipedia, https://en.wikipedia.org/w/index.php?title=International_Classification_of_Diseases&oldid=1098369212 (2022, accessed 29 July 2022).

[2] Swagger UI, https://browser.ihtsdotools.org/snowstorm/snomed-ct/swagger-ui.html (accessed 27 January 2022).

11. AUTOMATED CLINICAL ROUTINE DATA EXTRACTION AND INTEGRATION FOR DATA SAFETY MONITORING BOARD (DSMB) REPORTS IN A PERSONALIZED PHASE I STUDY

[1]Neuschwander P, Schulz A, Sluka C, Franzeck F, Williner A, König D, Läubli H

[1]*University Hospital Basel,* Department of Clinical Research

Keywords: Personalized therapy, phase I study, clinical data warehouse, data safety monitoring board, automated routine data integration

Adoptive cell therapy with tumour-infiltrating lymphocytes (TILs) is a personalized cell-based immunotherapy. TILs are isolated from a patient's tumour lesion, expanded in vitro and then transfused back into the same patient after lymphodepleting chemotherapy. The ongoing phase I Base TIL trial (NCT04165967) evaluates feasibility and safety of TIL therapy with subsequent PD-1 blockade in nine patients with advanced melanoma. The independent DSMB is responsible for safeguarding the interests of the patients and met after the first and third patient. For their evaluation, the DSMB requested continuous charts of vital parameters and laboratory values. Trial data are entered into the electronic Case Report Form (eCRF) for each study visit, however not continuously. In routine care of the stationary patients, these parameters are regularly documented in the patient chart.

To obtain a holistic overview for the DSMB, these continuously recorded routine data needed to be integrated with the eCRF study data. We used a "power-user" access to the Clinical Data Warehouse (CDWH) of the University Hospital Basel to extract these data. The challenges encountered include the pseudonymisation of patient IDs, and data being distributed across different CDWH views with different structure and granularity, due to some patients being admitted to the intensive care unit for some period. The data extraction and post-processing therefore required close collaboration between the CDWH team, the PI, and study team, as well as the data scientists and project managers. Our poster describes the procedures used and the challenges they presented, as well as the conclusions and lessons learned.

12. WHAT ARE THE BOTTLENECKS TO DATA SHARING IN SWITZERLAND?

Ormond KE[1], Bavamian S, Becherer C, Currat C, Joerger F, Hiendlmeyer E, Maurer J, Staub T, Vayena E.

[1]HEST/Health Etrhics and Policy Lab ETH-Zurich

Keywords: Data sharing, bottlenecks, ethical, legal, logistics

While data sharing is strongly espoused in principle, in practice it can be challenging to implement. Little is known about the actual bottlenecks to biomedical data sharing in Switzerland. The goal of this study was to assess the obstacles to data sharing, including areas of legal, ethical and logistical bottlenecks. 11 semi-structured interviews were conducted (primarily in English) by a single interviewer with 17 individuals who represented lawyers, data protection officers, ethics officers, scientists and project managers, bioinformaticians, clinical trials services and biobank overseers. Interviews were transcribed and de-identified. Transcripts and notes taken during each interview were then thematically analyzed and discussed by the entire research team. Most respondents felt that it was not the actual process of data sharing that was the bottleneck, but rather the processes and systems around it, which were time intensive and create confusion, that served as a barrier to data sharing. The templates that have been developed by SPHN and the general consent process were overall felt to have streamlined processes significantly. Preliminary themes include (1) It's difficult to get high quality data from clinical records, (2) Areas of confusion include 'who owns the data' and the inconsistencies that are created through general consent that vary by institution, (3) Anonymization is defined differently between the EU, UK and US, and even when clearly defined it is hard to operationalize in practice (4) We need to create a culture of data sharing and recognize that data sharing is a service, not an event, which requires sustainability efforts. In order to best facilitate a culture of data sharing in Switzerland, it may require legal clarifications, further education about the process and resources to support data sharing, and an investment in sustainability on a funder and institutional level.

13. PRIVACY-PRESERVING FEDERATED BIOMEDICAL ANALYTICS

[1]Pyrgelis A, Sav S, Hubeaux JP

[1]*EPFL, School of Computer and Communication Sciences*

Keywords: Federated analytics, security, privacy, data protection, data sharing

Accurate and robust biomedical analytics requires a large amount of diverse and heterogeneous data that is usually scattered across multiple healthcare institutions or hospitals. Consequently, data sharing among these entities is an inevitable need, yet is often not feasible due to the sensitive nature of the data and strict privacy regulations such as HIPAA or GDPR. A recently proposed approach, federated analytics, allows entities to keep their data in their local premises and collaboratively obtain insights from their joint data. However, recent research has shown that federated analytics approaches suffer from several privacy vulnerabilities. To this end, we design and build novel systems that enable privacy-preserving federated analytics by leveraging multiparty lattice-based homomorphic encryption (MHE) [1]. Our systems enable decentralized trust and protect the confidentiality of the institutions' data by keeping under encryption any information exchanged throughout the whole computation process.

We demonstrate the applicability of our systems to biomedical analytics by reproducing several centralized studies in a privacy-preserving and federated setting. In particular, we provide solutions that enable Kaplan-Meier survival analysis for oncology, genome-wide association studies, and single-cell analysis for disease-associated cell classification [2, 3]. Our experimental results show that our solutions successfully preserve the utility of the data while providing end-to-end confidentiality for federated biomedical analytics. Part of this research was carried out in the framework of the PHRT/DPPH [4] and the SPHN/PHRT MedCo projects. The tools presented here are now supported by Tune Insight SA [5]; more details are provided in a companion poster submission.

[1] https://eprint.iacr.org/2020/304.pdf

[2] https://www.nature.com/articles/s41467-021-25972-y

[3] https://pubmed.ncbi.nlm.nih.gov/35607628/

[4] https://dpph.ch

[5] https://tuneinsight.com","Sinem Sav, Apostolos Pyrgelis, and Jean-Pierre Hubaux"

14. THE SWISS MOLECULAR PATHOLOGY BREAKTHROUGH PLATFORM (SOCIBP) PROJECT

Schnidrig D, Rubin MA, Rätsch G, Ng KYC, Kahles A, Benjak A, Selimovic-Hamza S

University of Bern, Department for BioMedical Research (DBMR)

Keywords: Oncology, Bioinformatics, Genomics analysis, Data sharing, Data visualization

The clinical cancer genomics landscape in Switzerland is highly diverse - a broad range of sequencing assays and data processing strategies are used across Swiss hospitals. SOCIBP has filled a major gap in oncology research in Switzerland by building the computational infrastructure necessary to streamline this diverse data into one easy-to-use resource for researchers and hospitals in Switzerland. The main achievements of this infrastructure development project are 1) the establishment of portable and reproducible bioinformatics tools to analyse heterogeneous genomic data generated for clinical care in a standardised and scalable manner and 2) the establishment of a national cBioportal instance which allows clinicians and researchers to scan genomic data from all across Switzerland to help populate or design clinical trials, or develop new hypotheses. The user interface of cBioportal enables data queries without the need of computational expertise, allowing researchers of all backgrounds to interactively visualise and analyse the data. Access to the portal is available to all qualified researchers in Switzerland via a controlled data access protocol requiring an agreement for data use and approval from the SOCIBP board. An overview of the available cohorts as well as a link to the data access request form can be found under https://socibp-cbioportal.leomed.ethz.ch/.

15. CHRONOPHENOTYPING ATRIAL FIBRILLATION USING MACHINE LEARNING

Sobel J[1], Biton S, Marcusohn E, Suleiman M, Tsutsui K, Behar JA

[1]*University of Geneva, Global health institute*

Keywords: ECG, Atrial fibrillation, machine learning, Chronophenotypes

Background: Atrial fibrillation (AF) is the most prominent type of cardiac arrhythmia, which affects an estimated 4% of the adult population over 60 and this percentage correlates with age. Circadian variation in the disease presentation has previously been reported at the population level. However, it is unknown whether there exist chronophenotypes and what their clinical outcomes are.

Methods: A collection of 24h Holter recordings and patients clinical information were collected at Rambam hospital in Haifa (Israel, N=2000) and Saitama Medical University (Japan, N=150). A deep learning method called ArNet2 was developed to classify 60 beats windows into normal sinus rhythms or AF. We used ArNet2 on both databases to detect episodes of AF. Unsupervised learning was used to identify AF patient subgroups according to AF episodes temporal distribution. Finally, we describe the difference in 5 years survival, and comorbidities, in each identified AF chronophenotype.

Results: Our work confirmed on two independent databases (geographically and ethnically distinct) diurnal variation of AF at the population level and suggested the existence of several chronophenotypes.

Conclusions: We developed a novel unsupervised learning method for chronophentyping through the analysis of the temporal distribution of AF episodes during 24 hours long ECG recordings. A total of 6 phenotypes was discovered. These sub-groups of AF patients depict distinct prognoses and characteristics which may suggest alternative treatment plans . Our work paves the way to personalized medicine based on the automated analysis of Human long term continuous physiological recordings.

16. DEEP MEDICAL IMAGING BIOBANKS

ter Voert E[1], Caprara S, Kozerke S, KonukogluE

[1]ETH Zurich, Computer Vision Laboratory, Department of Information Technology and Electrical Engineering

Keywords: Imaging Biobanks, Raw imaging data, FAIR imaging metadata, DICOM RDF Ontology, PET / CT / MRI

Hospitals generate large amounts of data which are commonly stored in diverse databases and formats. This makes it difficult to find for researchers. Moreover, to save storage space, often only the processed and clinical relevant data is stored. However, the original raw medical imaging data is of utmost importance in advanced research projects.

Our PHRT project was started to overcome these hurdles. Its goal is to build the necessary infrastructure that allows storing and sharing large-scale (raw) medical imaging datasets to serve algorithm development and clinical data-science research.

For this, we set up databases within the secure and controlled environment of the hospitals where raw imaging data can be stored. Next, we make the pseudonymized metadata of these imaging datasets available, so researchers can find them. As not every medical center has the required computing power to analyze or process the imaging data, we work together with BioMedIT who offer high-performance computing clusters and data storage solutions in secure environments. After a data request by a researcher and approval by the hospital, the imaging data of interest can be safely transferred to a project space using their secure encryption and transfer tool. A test system was recently set-up at the Balgrist University Hospital and Campus.

For the metadata we use the FAIR data ecosystem developed by SPHN. It allows to semantically annotate and link existing data. Unfortunately, it has limited possibilities for imaging datasets. Therefore, we are closely working together to further develop this.

For our project we developed an SPHN project-specific ontology which describes the DICOM header of an MR, CT and PET image in RDF. The developed Python script can then read these DICOM headers and generate RDF graphs according to the ontology. These RDF graphs can then be stored in a metadata database (GraphDB) installed at the ETH Zurich. The metadata can then also be queried using this database.

17. KAIKO'S DATA AND COMPUTE PLATFORM: AN INTEGRALIST APPROACH TO ADVANCING PERSONALIZED HEALTH RESEARCH

Crisan VC[1], Berke R, Yaman A

[1]*Kaiko*

Keywords: Multimodal data integration, digital pathology, big data processing and sharing

Two of the major challenges faced by institutions seeking to advance medical research are: data silos, and the limitations of on-premise compute. These issues perpetuate as bottlenecks across the whole spectrum of healthcare and data science. There is often incomplete oversight of the patient journey for data-driven decisions, inability to identify or mitigate biases in Machine Learning models, or inability to train multimodal models, to name just a few. To address these challenges and their implications, Kaiko is building a Data and Compute Platform that accommodates an open-ended series of "use case enabling – use case generating" cycles. In this presentation, we highlight how such an open-ended platform can scale with new use cases and their increasing complexities. We show immediate advancements from this initiative, including: productization of compute-intensive Machine Learning models in radiotherapy workflows, customized tooling for ingesting and unifying multimodal data, as well as an enterprise data model that can scale at intra- and inter-hospital levels. As a detailed case study, we present our tooling for centralizing multimodal data, which comprises open-source connectors for whole slide images (WSIs) for digital pathology, data processing pipelines, and modules for fast and secure big data sharing. Finally, we outline why we believe such data and compute platforms are key to generating a desired snowballing effect in healthcare innovation.

18. THE IARAT: A DIGITAL TOOL FOR HARMONIZING MOVEMENT DATA COLLECTION ACROSS CLINICAL CENTERS AND HOME ENVIRONMENTS

Awai C[1], Schönhammer J, Menard T, Luft AR

[1]*Cereneo foundation – Center for Interdisciplinary Research (cefir)*

Keywords: Technology development, Data collection, Harmonization, Clinical implementation

Data platforms that combine data from multiple sources and locations, are dependent upon sufficient data quality from the data provider level. Edge nodes employ standardized clinical tests, however there is significant variability between nodes in how data is collected, pre-processed, and its storage and availability. Harnessing this variability at the data aggregation layer is extremely challenging. An elegant solution would be an open digital tool for typical clinical tests that is used at the edge layer and shared between nodes. Ideally, a shared digital data collection tool provides standardized instructions to patient and data collector, is easy to use and maintain, is universally applicable for clinic and home environments, and is easily ingested into a transfer layer.

Objective. We here present an example of the development of a digital tool for the harmonized collection of movement data during a frequently used clinical test procedure in neurorehabilitation at an edge node.

Method. A digital tool was developed in a co-creation design process to collect hybrid performance scores from therapists (traditional) and wearable sensors (innovative). Data is pre-processed in an edge compute solution and provided for ingestion by a transport layer in raw and typical health formats (e.g. HL7). The tool can be used by trained personnel and has an optional educational component that enables home use.

Impact. Connected, standardized tools for data collection at the edge layer are an elegant solution to boost data quality and availability for impactful aggregation of health data.

19. SWARM LEARNING

Fontana R[1], Schultze H, Volk A

[1]Hewlett Packard Enterprise

Keywords: Privacy-preserving collaboration, share without sharing

Swarm Learning is a decentralized, privacy-preserving Machine Learning framework. This framework utilizes the computing power at, or near, the distributed data sources to run the Machine Learning algorithms that train the models. It uses the security of a blockchain platform to share learnings with peers in a safe and secure manner. In Swarm Learning, training of the model occurs at the edge, where data is most recent, and where prompt, data-driven decisions are mostly necessary. In this completely decentralized architecture, only the insights learned are shared with the collaborating ML peers, not the raw data. This tremendously enhances data security and privacy.

To illustrate [1] the feasibility of using Swarm Learning to develop disease classifiers using distributed data, we chose four use cases of heterogeneous diseases (COVID-19, tuberculosis, leukaemia and lung pathologies). With more than 16,400 blood transcriptomes derived from 127 clinical studies with non-uniform distributeons of cases and controls and substantial study biases, as well as more than 95,000 chest X-ray images, we show that Swarm Learning classifiers outperform those developed at individual sites.

[1] Nature volume 594, pages265–270 (2021), https://www.nature.com/articles/s41586-021-03583-3

20. COMMON DATA MODEL HEALTHCARE (CDMH)

Mathis R[1], IT-Logix AG

[1]*IT-Logix AG*

Keywords: Clinical Datawarehouse, Clinical data integration, Data Analytics, Data Science

IT-Logix has been working with leading university and cantonal hospitals in the field of data warehousing, data analytics and data science for many years and has been able to build up a lot of industry-specific knowhow. This industry-specific knowledge forms the foundation for our generic analytical blueprint data model for hospitals – The Common Data Model Healthcare (CDMH). The aim is to help our customers to find working solutions very quickly. The fully integrated data model is therefore technology-agnostic and can be used on all recognized database technologies.

With our blueprint model and agile project approach, we can create a high-quality clinical data warehouse for you in the shortest possible time. The creation of a transparent, data and fact-based decision-making culture and ensuring interoperability is always the central objective.

21. SHARING WITHOUT SHARING: REVOLUTIONIZING SECURE DATA COLLABORATIONS IN HEALTHCARE

Ramon J[1], Pastoriza T

[1]Tune Insight SA

Keywords: Secure data collaborations, secure federated learning, regulatory compliance, multiparty homomorphic encryption",

We will present cryptographic techniques that streamline data protection compliance and ethics approval for multi-site analyses and clinical trials. These solutions, powered by Tune Insight's technology and already deployed at SPHN Hospitals, enable the extraction of aggregate analytics from federated data from different clinical sites, with end-to-end protection guarantees and no transfer of the original data.

Tune Insight's solutions aim at resolving the information-sharing tradeoff by enabling more accurate insights on larger amounts of more relevant collective Real-World Data (RWD). This enables health institutions (pharma and hospitals) to streamline compliance and ethics approval, to facilitate the extraction of richer statistics and to build better models by securely collaborating on valuable sensitive data. This will increase the immediacy and expand the range of available information, leading to more personalized and effective treatments.

This is possible by means provable technological guarantees that regulatory compliance (GDPR, FADP, HRA, HRO) [1] to enable authorized users access only to the global insights (aggregate statistics, ML algorithms and models) built on all network data, whereas no access or transfer is granted on the local contributed patient data that remains under the control of its source.

We will also present our two SPHN-funded pilot deployments and our recent collaborations with the EPFL Laboratory of Data Security involving the MIT and Broad Institute [2] and the University of Tübingen [3].

[1] Scheibner et al. "Revolutionizing Medical Data Sharing Using Advanced Privacy Enhancing Technologies: Technical, Legal and Ethical Synthesis." JMIR, 2021

[2] Froelicher et al. "Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption." Nature Comms 12, 5910 (2021)

[3] Sav et al. "Privacy-preserving federated neural network learning for disease-associated cell classification." Cell Patterns 2022","Juan Ramon Troncoso, Romain Bouyé, Francesco Marino, Jean-Pierre Hubaux"