SPHN – The BioMedIT Network: A Secure IT Platform for Research with Sensitive Human Data

Author(s):
Coman Schmid, Diana; Crameri, Katrin; Oesterle, Sabine; Rinn, Bernd; Sengstag, Thierry; Stockinger, Heinz

Publication date:
2020-06-16

Permanent link:
https://doi.org/10.3929/ethz-b-000424360

Originally published in:
Studies in Health Technology and Informatics 270, https://doi.org/10.3233/SHTI200348

Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200348

SPHN – The BioMedIT Network: A Secure IT Platform for Research with Sensitive Human Data

Diana COMAN SCHMID[c,a], Katrin CRAMERI[a, 1], Sabine OESTERLE [a], Bernd RINN[c,a], Thierry SENGSTAG[b,a] and Heinz STOCKINGER[a]
on behalf of the BioMedIT network team: Leila T. ALEXANDER[a], Jonathan BARDA[a], Christian BOLLIGER[c,a], Urban BORSTNIK[c], Gerhard BRÄUNLICH[c,a], Olivier BYRDE[c], Jérôme DAUVILLIER[a], Robin ENGLER[a], Pablo ESCOBAR LOPEZ[b,a], Volker FLEGEL[a], Martin FOX[a], Sofia GEORGAKOPOULOU[b,a], Jani HEIKKINEN[b,a], Martin JACQUOT[b,a], Nicolas KOWENSKI[c], Guillermo LOSILLA[b,a], Sergio MAFFIOLETTI[c,a], Jorge MOLINA[a], Diego MORENO[c], Allen NEESER[c], Michal OKONIEWSKI[c,a], Warren PAULUS[a], Kevin SAYERS[a], Torsten SCHWEDE[b,a], Jaroslaw SURKONT[b,a], Richard WARTENBURGER[c], and Thomas WÜST[c,a] (in alphabetical order)
[a] *SIB Swiss Institute of Bioinformatics, Switzerland,*
[b] *University of Basel, Switzerland,*
[c] *ETH Zurich, Switzerland*

Abstract. The BioMedIT project is funded by the Swiss government as an integral part of the Swiss Personalized Health Network (SPHN), aiming to provide researchers with access to a secure, powerful and versatile IT infrastructure for doing data-driven research on sensitive biomedical data while ensuring data privacy protection. The BioMedIT network gives researchers the ability to securely transfer, store, manage and process sensitive research data. The underlying BioMedIT nodes provide compute and storage capacity that can be used locally or through a federated environment. The network operates under a common Information Security Policy using state-of-the-art security techniques. It utilizes cloud computing, virtualization, compute accelerators (GPUs), big data storage as well as federation technologies to lower computational boundaries for researchers and to guarantee that sensitive data can be processed in a secure and lawful way. Building on existing expertise and research infrastructure at the partnering Swiss institutions, the BioMedIT network establishes a competitive Swiss private-cloud – a secure national infrastructure resource that can be used by researchers of Swiss universities, hospitals and other research institutions.

Keywords. Personalized health, SPHN, research infrastructure, scientific IT services, federated computation, sensitive data, confidential data, data privacy, data security, health-related data, interoperability, private-cloud, service virtualization

[1] Corresponding Author: Katrin Crameri, Personalized Health Informatics Group, SIB Swiss Institute of Bioinformatics, Elisabethenstrasse 43, 4051 Basel, Switzerland; E-mail: katrin.crameri@sib.swiss

1. Introduction

The advent of digital transformation in health care has produced an exponential increase in the amount of information available for each patient. The use of this information in data-driven biomedical research has the potential to drive important changes in medicine [1]. In order to be able to leverage the potential of health-related data through biomedical research, data science and related research fields, data needs to be interoperable and available to researchers in various disciplines. In addition, strong capabilities in clinical bioinformatics and computational service infrastructure are required in order to enable the integration and interpretation of large and rich data sets. Furthermore, big data analyses and machine learning require vast amounts of data and accordingly high-performance IT infrastructures for computing and storage.

Given the sensitive nature of health-related information, data-driven personalized health research requires special IT infrastructures and services, blending the concepts of (i) security and compliance to protect the confidentiality of the data and the privacy of the research study participants; (ii) scalability and performance to be able to adapt to changing needs of users; and (iii) flexibility and ease of use to foster cutting-edge biomedical research. Security measures for Information and Communications Technology (ICT) systems are necessary to protect confidential information from unauthorized use, modification, loss or release[2]. A few years ago, these requirements were not a major concern for academic computing facilities in Switzerland because those were predominantly tailored towards the handling of (insensitive) basic research data.

Another important requirement concerning the architecture of IT infrastructures for researchers working in multidisciplinary networks on big data sets in the biomedical field, is the possibility for controlled access to data sets across research teams from different institutions, also cross borders. In addition, in the context of nation-wide collaborative research projects, technical interoperability between different IT infrastructures should be granted in order to enable reproducibility of data analysis workflows executed at distributed locations [3].

To address the needs listed above, the BioMedIT project was funded by the Swiss federal government for the period of 2017-2020 within the framework of the Swiss Personalized Health Network Initiative (SPHN) [4] and in close collaboration with the strategic focus area Personalized Health and Related Technologies (PHRT) of the Swiss Federal Institutes of Technology (ETH) domain [5]. The aim of the project is to provide all researchers in Switzerland with access to a service infrastructure for collaborative analysis of confidential data without compromising data privacy. The intention is to create and maintain a national infrastructure resource that can jointly be used by all Swiss universities, research institutions, hospitals and other interested partners.

2. The BioMedIT Network: a National Secure Infrastructure Resource for Switzerland

The BioMedIT network builds on three scientific IT competence centers – the BioMedIT nodes – in different geographical locations: one in Basel (sciCORE, operated by the University of Basel), one in Lausanne (Core-IT, operated by the SIB Swiss Institute of

[2] Key elements of an effective ICT security system include (i) Monitoring and controlling access to confidential information; (ii) Safe transmission of data, and (iii) Secure storage and disposal of data [2].

Bioinformatics) and one in Zurich (SIS, operated by ETH Zurich). Over the past two years, all three nodes established secure compute and data platforms specialized on handling confidential research data that is are subject to legal and ethical constraints. The computational service infrastructure is available to Swiss universities, research institutes, hospitals, service providers and other interested partners (*organizational clients*, see Section 2.4) that are not part of the BioMedIT network, to securely store, manage and process confidential research data. In addition, and specifically within the framework of the two Swiss national initiatives SPHN and PHRT, the BioMedIT network can be used for mono- and multi-site, individual and collaborative research projects (*project clients*, see Section 2.3). In fact, several innovative biomedical research projects [6] are already using a first version of the BioMedIT infrastructure.

2.1. Security Aspects of the BioMedIT Network

A common Information Security Policy [7] that applies to the entire BioMedIT network, defines the necessary organizational and technical measures to allow researchers to process sensitive data in a secure way[3]. This policy controls the way the confidentiality, integrity, and availability of information is handled, preventing misuse and malicious damage. In conventional high performance or high throughput systems available in scientific environments, the physical hardware and the respective software services are shared between all users. For data privacy and security reasons, BioMedIT takes a different approach: researchers involved in one particular project work completely isolated from any other project, meaning that all the data, software and resources belong exclusively to that project and there is no possibility to access it in any other way or from any other path. To this end, virtualization technologies are applied to establish private-cloud environments in which each scientific project is completely isolated from other projects with respect to data and compute resources. Access of authorized users to any project space requires two-factor authentication. Furthermore, users that are authorized to use the secure BioMedIT network can only access the infrastructure from within trusted IT environments (either from within Swiss university or university hospital networks or via VPN). Access to the Internet from the BioMedIT nodes is strictly controlled, limited to trusted and explicitly white-listed web resources.

The three BioMedIT nodes share a common security architecture to transfer, store, manage, analyze and share biomedical data while following the latest technical and legal standards required by Swiss legislation[4] as well as internal regulations of the associated institutions. For project-related data transfer, an end-to-end encryption process from the data source through the BioMedIT network to the project space is set up, based on public-key cryptography. Special consideration is given to key management which is provided as a central service to users of the network.

Furthermore, the network follows the SPHN Ethical Framework for Responsible Data Processing in Personalized Health Research [11]. Importantly, before using the secure BioMedIT network, users are requested to follow a data security awareness training. BioMedIT offers the "Data Privacy and IT Security Training" as on-line

[3] In general, sensitive data (as it is defined according to Swiss and international data protection legislation) is classified as "confidential" according to the SPHN Information Security Policy.

[4] In particular, the Federal Act on Data Protection [8], the Human Research Act [9], the EU General Data Protection Regulation [10]

training [12] or class-room course hosted regularly in different cities in Switzerland. Finally, users must pass a mandatory on-line exam.

2.2. Technical Features of BioMedIT's ICT environment

The BioMedIT nodes provide a broad range of secure services to store, manage, process and share biomedical data. For example, multi-CPU (central processing unit) and multi-GPU (graphics processing unit) compute nodes are connected by fast internal networks to large and fast parallel file systems and state-of-the-art data management services. To date, the BioMedIT infrastructure features data storage capacity approaching in total 5 PB, support for data encryption, secure backup, private-cloud and high performance computing (HPC) leveraging more than 3000 CPU cores and 324 GPUs. State-of-the-art software for data science and specialized tools for data management are provisioned alongside with long-standing expertise in scientific IT support for research data management, bioinformatics, HPC and computational analysis.

2.3. Lowering Computational Barriers for Researchers (Project Clients)

The BioMedIT network provides a flexible ICT environment, and individual project spaces can be configured based on the researcher's computational and storage needs. This includes storage and compute capacity, up-to-date data analysis and modeling software packages, as well as backup and archiving solutions. The nodes can provide secure work environments for basic exploratory data analysis but can also scale up to cover high performance and high throughput needs such as large-scale machine learning.

The BioMedIT portal, a central service of the network operated by the SPHN Data Coordination Center of SIB's Personalized Health Informatics Group, provides a single access point to the BioMedIT nodes and associated resources, simplifying access control for the researchers. Depending on the use-case, the BioMedIT network can be accessed by command-line or by web-based remote-desktop technology.

Standardized ways for secure data transfer from partner institutions, such as the five Swiss University Hospitals, the PHRT data centers and other data providers are currently being established and will be available to all projects relying on data from any of these providers.

The BioMedIT network aims to enable interoperable workflow execution, providing a way for researchers to work seamlessly across the nodes. Containerization of data workflows using Open Container Initiative (OCI) standards [13] are playing an important role for this and will at the same time improve reproducibility of results obtained by these workflows. The ultimate goal of BioMedIT is to provide a data-aware federated analysis platform where researchers can work with distributed data. This approach is especially valuable for confidential data or large data sets, which cannot easily be shared.

2.4. Outsourcing Sensitive Research Data Operation to BioMedIT (Institutional Clients)

While the focus of the BioMedIT network is to support national research projects in the SPHN and PHRT programs, the BioMedIT nodes are creating a unique data and computing environment that will be useful and potentially required for research activities outside of these programs. Clinical institutions scaling up their data processing in preparation of clinical research projects or clinical decision support strategies,

consortium projects with public and private partners, or novel ways to run clinical trials, which create large amounts of trial data and require good clinical practice (GCP) compliant data management, are examples for it. The BioMedIT nodes are offering a safe environment for these types of activities and are working on enhancing their platforms for future needs. By linking to the BioMedIT network, Swiss research institutions and hospitals can hence benefit from a trusted research IT infrastructure without the need to (re)develop costly *in-house* infrastructures and know-how.

3. Conclusion

BioMedIT provides data resources and a network of secure data nodes for doing data-driven research on sensitive (confidential) biomedical data. It lowers computational barriers by provisioning an efficient solution for cross-institutional analysis of sensitive data, and is therefore unique in Switzerland. The technical solution is comparable to related activities within ELIXIR in various European countries [14]. However, BioMedIT operates within a national network of health-care providers, infrastructure and research support groups, research projects as well as citizens, and therefore has a great potential to become a platform of reference for health data exchange and analysis. A first version of the BioMedIT network is established and already used by SPHN and PHRT projects and pilot collaborations with Swiss university hospitals. These projects help shaping new versions of the BioMedIT infrastructure to make it even more useful, efficient and secure for current and future biomedical research applications.

References

[1] C. Lovis, Unlocking the Power of Artificial Intelligence and Big Data in Medicine, *J Med Internet Res.* **21** (2019), e16607.
[2] Australian Government, A Guide for Data Integration Projects Involving Commonwealth Data for Statistical and Research Purposes https://statistical-data-integration.govspace.gov.au/topics/secure-data-management/information-and-communication-technology-security, accessed 17 January 2020
[3] Enhancing Research through IT Expertise https://www.enhancer.ch/pipeline, accessed 17 January 2020
[4] The Swiss Personalized Health Network Initiative www.sphn.ch, accessed 17 January 2020
[5] Personalized Health and Related Technologies, https://www.sfa-phrt.ch, accessed 17 January 2020
[6] SPHN Projects, Bottom-up funding: Driver Projects and Infrastructure Development Projects https://sphn.ch/network/projects, accessed 17 January 2020
[7] SPHN Information Security Policy, Version 1.0 of 23 August 2018, https://sphn.ch/wp-content/uploads/2020/01/sphn_information_security_policy_v1.pdf
[8] Federal Act on Data Protection of 19 June 1992 (Status as of 1 March 2019), SR 235.1 https://www.admin.ch/opc/en/classified-compilation/19920153/index.html
[9] Federal Act on Research involving Human Beings of 30 September 2011 (Status as of 1 January 2014), SR 810.30 https://www.admin.ch/opc/en/classified-compilation/20061313/index.html
[10] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
[11] Ethical Framework for Responsible Data Processing in Personalized Health Research, Version 2, of 7 May 2018 https://sphn.ch/wp-content/uploads/2019/11/Ethical_Framework_20180507_SPHN.pdf
[12] https://edu.sib.swiss/course/view.php?id=424, accessed 17 January 2020
[13] Open Container Initiative https://www.opencontainers.org, accessed 15 January 2020
[14] ELIXIR https://elixir-europe.org/, accessed 17 January 2020