Position paper

Common cross-border approach for processing sensitive data based on the e-infrastructure from NeIC Tryggve and de.NBI Cloud

Introduction

This paper provides a background on the infrastructure required for the trusted transfer, storage, access and processing of sensitive data across (federated) service providers. It explains the type of the planned collaboration of German de.NBI Cloud and Nordic NeIC Tryggve and, as a proposed follow-up to Tryggve, the NeIC Heilsa consortium.

The focus area of the collaboration is technical interoperability of federated services in the context of legal, operational, and semantic requirements, as well as standards for processing biomedical sensitive data in compliance with the GDPR and according to high IT security.

Background

The health and research sector is confronted with an unprecedented flood of data resulting from new patient device technologies and recent technological breakthroughs in areas such as genome sequencing, imaging and remote sensing, which are becoming part of routine clinical care at an ever faster pace. Despite this massive increase in data sources, diversity and volume, only a surprisingly small fraction of this data is currently integrated in research or routine processes in clinical practice.

Increased cooperation between the European countries is essential to overcome isolated data silos, lack of interoperability and fragmentation of initiatives. Cross-border access to genomic and clinical information is a cornerstone in health research and clinical practice. The aim is to achieve larger cohorts offering scope for new exciting clinically significant research. Sharing more genomic data will improve understanding and disease prevention, and enable personalised treatments and targeted drug prescribing.

Promising large scale projects such as GAIA-X[1], GHGA[2], BioDaten[3] or EGA[4] require trusted environments for sharing and handling sensitive data. This has also been recognized by the EU. The existing infrastructure and know-how are to be brought together to achieve a common and concrete goal: one million genomes accessible in Europe. Genomic and other health data should

---

[1] https://www.bmwi.de/Redaktion/EN/Publikationen/Digitale-Welt/project-gaia-x.html
[2] https://ghga.dkfz.de/
[3] http://www.biodaten.info/
[4] https://www.ebi.ac.uk/ega/home

be accessible across the EU member states. This will expand the data capabilities to enable improved research and clinical care.

The exploitation of the essential large data sets is mainly hindered by the lack of modern IT structures capable of aggregating and integrating this data and thereby allowing researchers and clinicians to take advantage of a holistic view on all available data that would allow optimized diagnosis and treatment. The planned cooperation of the Nordic network Tryggve and the German de.NBI Cloud is based precisely on this – the bundling of the activities that are required for cross-border data access and the use of the extended data records.

Overview of the collaborating partners

NeIC Tryggve and the proposed NeIC Heilsa "Tryggvedottir" consortium

Tryggve[5] is a collaboration between NeIC and the Nordic nodes (Finland, Norway, Sweden, Denmark, and Estonia) of the ELIXIR infrastructure. Since its establishment in 2014, Tryggve has created a Nordic expert community and technology for cross-border secure infrastructure for sensitive biomedical data, as well as supported a number of research use cases to utilise the infrastructure in real life projects. Efforts are backed up by major operational national e-Infrastructures (NO: TSD, FI: CSC ePouta, DK: Computerome, SE: SNIC Bianca)

The facilitator and major funder of Tryggve is the Nordic e-Infrastructure Collaboration, NeIC. NeIC is a distributed organisation established as part of the NordForsk to develop e-infrastructure solutions of common Nordic value. NeIC has been crucial in bringing the national stakeholders together regionally.

The high-level target of the Tryggve collaboration is to establish a Nordic federated secure platform that can be used for health research utilizing large sensitive data collections. The distribution of analysis pipelines and access to local or distributed data is developed according to global standards as defined by the Global Alliance for Genomics and Health (GA4GH), and in close interaction with ELIXIR work on these topics.

This infrastructure platform is well suited for providing the Nordic component of the European 1+ Million genomes initiative, as well as for multiple Nordic initiatives for personalised medicine and for genomics. Furthermore, the new e-Infrastructure enables applications such as artificial intelligence algorithm development on structured and internationally interoperable data resources.

de.NBI Cloud

The de.NBI Cloud[6] was founded by the German ministry for education and research (BMBF) in 2016 and is an association of the universities of Bielefeld, Freiburg, Gießen, Heidelberg and Tübingen, the German Cancer Research Center (DKFZ) and the Charité Universitätsmedizin Berlin.

The main goal of the de.NBI Cloud is the provision of storage and computing capacities as well as relevant programs and databases for data processing in the life sciences. As such, the de.NBI Cloud offers a powerful IT infrastructure in combination with flexible bioinformatics workflows and analysis tools for the life science community in Germany.

---

[5] www.neic.no/tryggve
[6] www.denbi.de/cloud

The entire system is accessible via single sign-on (SSO) and is based on the ELIXIR authentication and authorisation infrastructure (ELIXIR-AAI). In order to offer users a uniformly designed cloud platform and to be able to use synergy effects between the locations, the technical conditions of the de.NBI cloud locations are largely harmonised. In this way, the user can switch from one de.NBI cloud location to another.

The genome platforms of de.NBI Cloud at Charité in Berlin and at German Cancer Research Center (DKFZ) in Heidelberg plan to provide extensive data sets of the International Cancer Genome Consortium for processing as GAIA-X use cases. Furthermore, the de.NBI Cloud sites in Heidelberg and Tübingen are forming the backbone of the proposed German Human Genome-Phenome Archive (GHGA.de), which is on track to become a national federated EGA node. The embedding in ELIXIR and EOSC-Life as well as the planned partnership with Tryggve enable networking and sustainability in an international context.

Common approach

In order to meet the constantly increasing requirements for a secure solution for sensitive data, measures are increasingly being planned and implemented in the context of data protection and information security.

The partners of this position paper are planning to work on the following items.
- Cross-border federation model including the documentation of the architecture.
- Cross-border workflows based on standardised GA4GH interfaces enabling
  - access to distributed data from any node
  - that workloads are distributed in software containers
  - secure interoperable repositories
- AAI and REMS – joining activities
  - Description of national datasets controlled by data custodians/owners to make them discoverable
  - Identification of (research) users at a level of assurance appropriate for sensitive data processing services
  - Streamlining the processes of data access committees
- Use cases for sensitive data handling
  - The partners plan to describe and pilot a use case to test the concept and analyse the possible implementations
    - Scoping the use case in the area of cross-border analysis of sensitive data sets, starting with a proof of concept based on available sensitive data and on infrastructure from selected de.NBI Cloud and Tryggve sites.
    - The legal approval from the participating institutions will be organized and obtained.
    - In a next step, sensitive data sets and the technical cloud infrastructure will be provided by the de.NBI cloud sites, in particular the University of Tübingen, DKFZ and Charité, as well as Tryggve sites to be selected.
    - As one of the options GA4GH interfaces will be offered on the sites. GA4GH compliant workflows will be used for data analysis. (possible tools include Galaxy and Kubernetes).
  - The necessary measures in terms of GDPR and IT security have to be realised for the use cases. The legal documentation will be produced in collaboration among the partners, for instance GDPR "Data Protection Impact Assessment" and "Records of processing activities".
- Mutual training and support

Conclusion

Secure infrastructure and established access methods are key issues to enable the strongly needed integrative European landscape. The rigorous realisation of trusted and GDPR compliant environments are necessary to overcome the current limitations of data dissemination and exploitation. The collaboration and the first cross-border use cases of Tryggve and de.NBI Cloud will show how privacy and security for sensitive data handling could be realized. This planned approach could serve as a blueprint for the further cross-border cooperation of infrastructure solutions in research and healthcare.

24 February 2020