Federated
European
Genome-phenome
Archive
Norway

System description Federated EGA Norway

The general overview of Federated EGA, its specific purpose, and how the nodes in the Federated service are collaborating is documented in the FEGA Norway fact sheet available from the main FEGA Norway website https://ega.elixir.no.

The intended purpose of this document is to describe how the FEGA Norway service to archive Norwegian sensitive human genome and phenome data in Norway has been implemented. This system description will provide a more complete overview of the implementation as well as relevant pointers to additional documents. For an even more detailed analysis of security and risk aspects, we refer to the ROS and DPIA documents of FEGA Norway.

Table of contents

Solution outline - fully based on TSD security

The technical foundation of FEGA Norway is TSD, the national service for sensitive data hosted by USIT at UIO, that is part of the Sigma2 portfolio of supported national e-infrastructure services.

The FEGA Norway service archives all submitted data in a dedicated TSD project, only available to the FEGA Norway operations team and TSD operators. All data transfers of sensitive data to and from this TSD project, are all executed over the standard TSD API interfaces, secured and maintained by TSD staff, and are the same APIs that TSD's own upload and download services like the TSD data portal (https://data.tsd.usit.no/) are using.

The full system description of TSD can be found here.

FEGA Norway end users are only able to perform data uploads to the FEGA Norway TSD project, and if approved by a Data Controllers DAC, access a dataset for download through the TSD File API services. No other access to the TSD project is facilitated for the end users.

Additional information about the FEGA Norway service and Federated EGA network in general is provided on an additional web server, also hosted by USIT, at the https://ega.elixir.no address. Here the end users will find all necessary information needed to get started, and find software needed to encrypt their data, a login page to generate authentication tokens, software to upload and download encrypted data through the TSD file API.

Data transfer to and from the FEGA Norway TSD project

All data transferred to and from FEGA Norway is encrypted using strong encryption. We use the Global Alliance for Genomics and Health (GA4GH) encryption standard - crypt4gh v1 (web, spec doc), designed to securely and efficiently encrypt large genome scale datasets, such that only the intended recipient can decrypt the data. This is achieved in the crypt4gh standard, by using asymmetric encryption with public-private key pairs, so that no secrets or password information has to be exchanged between the parties sharing the sensitive data between each other.

In addition to the data itself being encrypted, all transfers to and from FEGA Norway (through the TSD API for file transfers) are done in an encrypted communication channel over the https protocol.

Data encryption at rest in the FEGA Norway archive

When a data set is fully uploaded and quality controlled in FEGA Norway, the data is archived as crypt4gh encrypted files inside the FEGA Norway TSD project. These can only be decrypted with the private key of the FEGA Norway service. This private key is safeguarded in digital and analogue forms inside the TSD infrastructure, according to FEGA Norway and TSD joint operational procedures.

Storing FEGA Norway archived datasets encrypted at rest, provides an additional layer of security relative to a standard TSD project processing of sensitive data.

User identification

As mentioned in the introduction/service outline section, FEGA Norway end users do not have access as TSD users to the FEGA Norway TSD project. They authenticate themselves using the international Life Science login service operated by ELIXIR Europe and several other Life Science ESFRI infrastructures. Upon successful identification, a token representing the user is used to access the TSD File API services to execute file transfer operations.

Submission of data files to FEGA Norway

The user first encrypts the data using FEGA Norway's public key, easily available from the ega.elixir.no website, as well as their own private key - kept only accessible to themselves. This encryption must be done in a trusted and secure environment of the Submitter, potentially fully offline.

Then the encrypted data must be moved to an online system, from where they can be uploaded over the internet to FEGA Norway services. To upload an encrypted file to FEGA Norway, the submitter has to first authenticate themselves towards the Life Science login service, and then use a command-line tool to connect to FEGA Norway and the TSD File API. The tool communicates over https with the standard TSD File API interface to transfer the encrypted data file, where it gets stored in a user specific inbox folder inside the FEGA Norway TSD project. The data is not temporarily stored to disk in any way between the client transferring the data and the destination inside TSD.

After encrypted files have been uploaded to FEGA Norway, the non-sensitive meta-data describing the dataset and the research it relates to, is entered into the Federated EGA Submission Portal. Upon completing the meta-data submission, this trigger a further quality control and verification process of the uploaded encrypted data before it's stored in it's final archival destination and the data is removed from the user's inbox area.

More complete information and guidance on the submission process can be found here.

Retrieval of data files from FEGA Norway

Access to download a FEGA Norway hosted data set is only granted upon direct instruction from the data controller of the data set. It happens using the same 3 mechanisms as described for uploading a data set:

- The data set is first re-encrypted with the public key of the requester that has been granted access by the data controller
- The encrypted dataset is moved to user specific outbox area, where only the authenticated user will have access through the standard TSD File API services.

- The requester has to first authenticate towards the Life Science Login and use the FEGA Norway command line tool to connect to TSD File API and download the encrypted files.

After receiving the data, the requester need to follow the instructions from the data controller in the agreed Data Access Agreement on how to use and process the data further. In Norway, this will often include requirements of all processing of un-encrypted data to happen inside pre-agreed secure e-infrastructures such as TSD, Hunt Cloud and SAFE.

FEGA Norway restricted access levels for operations team

The FEGA Norway operational model has currently 6 levels of access inside the FEGA Norway TSD project, ranging from helpdesk personnel only needing access to pre-programmed routines for standard operations, with no direct access to the encrypted data, via the main FEGA Norway engineers deploying the services themselves, service responsible coordinators with authorization power and all the way to the TSD engineers operating the TSD infrastructure.

This model effectuates that FEGA Norway staff jointly operating the service only is granted access on a need-to-know basis, limiting the the staff with access to the archived data and encryption keys to a minimum.

Security assessments

The security measures of TSD and the FEGA Norway specific implementations described above on top of these are assessed separately in two different Risk and Vulnerability Analyses (ROS - Risiko Og Sårbarhetsanalyse):

1. TSD ROS
2. FEGA Norway ROS (available upon request)