European Genome-phenome Archive: Security Overview

The European Genome-phenome Archive (EGA) is a controlled access archive for consented human data. The EGA does not grant or deny access to data, this is done by the relevant Data Access Committee (DAC) and EGA applies these permissions to the data on behalf of the DAC (Figure 1). This document provides an overview of EGA's practices in ensuring the security of data stored at EGA. As security is a prime concern of the EGA, the EGA is a member of the Global Alliance for Genomics and Health (GA4GH - https://www.ga4gh.org/) Data Security work stream. The EGA contributes and helps develop the recommendations outlined the GA4GH Security Technology Infrastructure document[1], which defines guidelines, best practices, and standards for building and operating an infrastructure that promotes responsible data sharing in accordance with the GA4GH Privacy and Security Policy[2].

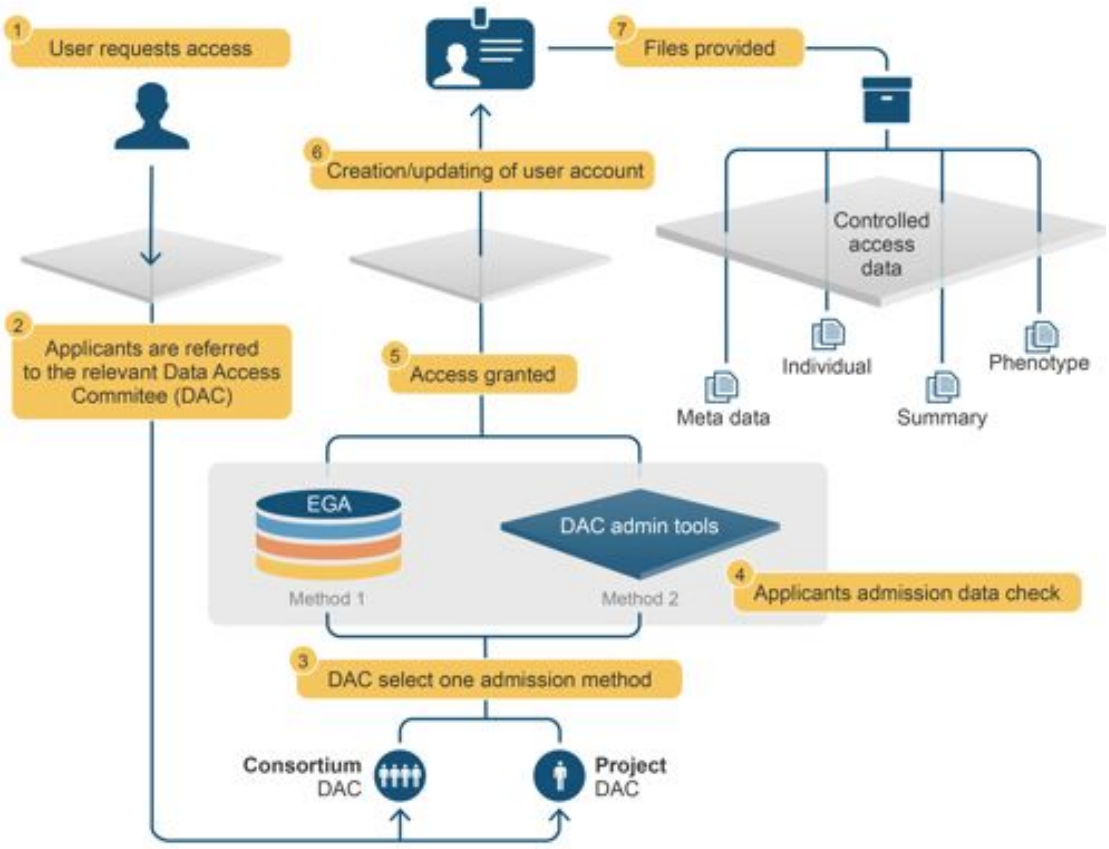

Figure 1: Process of applying for access to data held in the EGA. The user makes a request to access data controlled by a DAC. The DAC informs the EGA of the decision, and if access is granted, the EGA creates an account for the user (if the user does not already have an account) and grants permission to the data for that user.

[1] https://www.ga4gh.org/docs/ga4ghtoolkit/data-security/2016May10_REV_SecInfrastructure.pdf
[2] https://www.ga4gh.org/docs/ga4ghtoolkit/data-security/Privacy-and-Security-Policy.pdf

The key points of EGA security strategy are:

1 Regular Risk Assessment

- The EGA regularly identifies and assesses risk related to the following:
    - Breach of confidentiality,
    - Breach of privacy or autonomy,
    - Malicious or accidental corruption or destruction of data archived at EGA,
    - Disruption of services provided by the EGA.

2 Risk mitigation

- The EGA implements and maintains safeguards to minimize the risks identified above in accordance with the 5 control objectives listed in Appendix 1 and outlined in the GA4GH Security and Infrastructure document[3].
- If a breach is discovered, the EGA applies a defined protocol to minimize damage.

3 Identity and authorisation management

- The EGA authenticates the identity of individuals or software accessing controlled access data held at the EGA.
- The EGA ensures an appropriate level of assurance (LoA) is applied to the identity consistent with the risk associated with that individual, such as multi-factor authentication for DACs.
- The EGA provides the minimum access rights and privileges consistent with the user's identity, allowing access consistent with the GA4GH Privacy and Security Policy, as determined by the appropriate DAC.

4 Audit Logs

- The EGA maintains a set of logs recording:
    - Changes to user access rights,
    - Data access requests,
    - Resource usage.

5 Cryptography, communication security, and data integrity

- The EGA ensures data transmission integrity using a hash function.
- All data transmitted to or from the EGA is end-to-end encrypted.
- All data at EGA is stored using strong encryption.
- Encryption keys are not stored in the same system as the encrypted data.
- All data archived at EGA must be accompanied by a signed submission statement ensuring appropriate consent or ethical approval has been obtained, and is in accordance with all applicable laws and regulations.

The EGA has a defined protocol defining the response in the event of a security breach, and is continuing to work with the GA4GH Data Security Work Stream to help define best practice and associated standards for breach responses.

[3] https://www.ga4gh.org/docs/ga4ghtoolkit/data-security/2016May10_REV_SecInfrastructure.pdf

Appendix 1

GA4GH Control Objectives

- Control Objective 1: Implement technology safeguards to prevent unauthorized access, use, or disclosure of confidential and private data.
- Control Objective 2: Implement technology safeguards to prevent the discovery, access,and use of individuals' genomic and health-related data, and individual identities, other than as authorized by applicable jurisdictional law, institutional policy, and individual consents.
- Control Objective 3: Implement technology safeguards to prevent and detect accidental or malicious corruption or destruction of data.
- Control Objective 4: Implement technology safeguards to prevent disruption, degradation, and interruption of services enabling access to data.
- Control Objective 5: Implement technology safeguards to prevent and detect potential security attacks and misuse of authorized accesses and privileges.