

European Genome-phenome Archive

Implementation of the EU General Data Protection Regulation (GDPR)

Authors: Giselle Kerry, Mallory Freeberg, Dylan Spalding, Thomas Keane, Jordi Rambla
July 2020

1 About the EGA

The European Genome-phenome Archive (EGA) is a resource for permanent secure archiving and sharing of all types of potentially identifiable genetic and phenotypic data resulting from biomedical research projects. Its aim is to provide access to data, to foster data re-use, to enable reproducibility, and to speed up biomedical and translational research in line with the 'FAIR' (Findable, Accessible, Interoperable, and Reusable) principles. Access to EGA data must be approved by a Data Access Committee (DAC) and data must be appropriately consented for sharing. The EGA was formally launched in 2008 at the European Bioinformatics Institute (EMBL-EBI), an outstation of the European Molecular Biology Laboratory (EMBL), to address an identified need for archiving and sharing the results of genome-wide association studies from the Wellcome Trust Case Control Consortium. In late 2012, with the signing of a memorandum of understanding (and subsequent formal agreement in 2016) between EMBL-EBI and the Centre for Genomic Regulation (CRG), the EGA formally became a joint project of the two institutes. The two institutes work together to support the EGA services, including supporting submissions, web site, strategic leadership, and data infrastructure developments.

2 EMBL-EBI & GDPR

The EGA is co-managed by EMBL-EBI and CRG. EMBL-EBI is an international treaty organisation that has certain privileges and immunities¹ (e.g. exemptions from the application of national law) and also may self-regulate its activities (e.g. establish its own institutional legal framework) within the framework of its founding act of 1973. The General Data Protection Regulation² (GDPR) is a European Union (EU) regulation that legislates how organisations can share and process personal data of EU citizens. EMBL places great value in maintaining collaboration with researchers who are subject to GDPR. For that reason, it is of utmost importance for EMBL to handle data received from those collaborators in a secure and responsible manner. Mindful of its public mandate and the sensitivity of the data it handles, EMBL has always ensured a high level of data protection in its activities. Since the introduction of GDPR in May 2018, EMBL has established an internal policy on General Data Protection (IP68)³ which is within its mandate as a self-regulating international treaty

¹ https://www.embl.de/aboutus/administration/legal-services/legal_status/index.html

² <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN>

³ <https://www.embl.de/aboutus/administration/legal-services/data-protection/IP68-Data-Protection-EN-18052018.pdf>

organisation, and which EMBL deems is 'adequate' in the sense of GDPR⁴. The CRG operates within the EU and so fully complies with the GDPR.

3 EGA & GDPR

3.1 Genetic and phenotypic data

Within GDPR, there are two main actors: data controllers and data processors. Data controllers are persons or entities which determine the purposes and means that the personal data may be processed, e.g. companies, researchers, or universities. For EGA, the data controller is ultimately the data producer and the submitter(s) who submit the data to EGA. The data controller also creates a Data Access Committee (DAC) who will decide on data access permissions at EGA. Data processors are the persons or entities which process the data on behalf of a data controller. With regard to GDPR, EGA is a data processor as it processes data as instructed by the data controller. GDPR applies to any organisation which accesses personal data from an individual within the EU. Under GDPR, personal data is defined as any data that is identifiable, including names and email addresses as well as health-related and genetic data. EGA does not accept personally identifiable data except genetic and phenotypic data, so all other data submitted to EGA, such as names and addresses, must be pseudonymized. GDPR requires that data controllers implement data protection principles, such as data minimisation, to minimise the risk of data leakage, and protect the rights of the data subjects. As a data processor, EGA has a set of security policies⁵ that are followed to minimise the risk of unauthorised data access or data loss.

In its role as a data processor, EGA requires all submitters to sign a Data Processing Agreement (DPA) when the submission account is first created. This agreement is only required to be signed once per submitter, and will remain valid for future submissions to EGA. A copy of the EGA DPA can be found on the EGA website⁶.

3.2 Other personal data

The EGA also collects personal data as part of our interactions with submitters, data access committees, and researchers accessing data distributed by EGA. We have published a set of privacy notices on our website: <https://ega-archive.org/gdpr-notice>

⁴ <https://www.embl.org/info/data-protection/>

⁵ https://ega-archive.org/files/European_Genome_phenome_Archive_Security_Overview.pdf

⁶ <https://ega-archive.org/gdpr-notice>