Federated EGA: Structure and Organisation

Authors: Amy Curwin, Gabriele Rinck, Ana Teresa Freitas, Grant McNair, Thomas Keane, Jordi Rambla, Mallory Freeberg, Teresa D'Altri, EGA and FEGA Strategic Committee
January 2026, Version 2.0

1 Scope of this document

This document defines the structure and organisation of the Federated EGA (FEGA) network and forms part of the "common kernel" referenced in and aligned with a Collaboration Agreement (CA) signed by each entity entering the federation. The FEGA network is organised in three tiers: Central EGA (CEGA), FEGA Nodes and FEGA Affiliates. Herein we outline the goals, governance, commitments and service expectations within the FEGA network. Appendix 1 provides definitions of service terminology (e.g. submission, data distribution, helpdesk, etc.).

2 Federated EGA Network Structure

Central EGA Committed globally	Federated EGA Node Committed nationally	Federated EGA Affiliate Committed to a consortium/organisation
• **Hosts data.** • **Facilitates** global discovery of **shared metadata**. • **Accepts** data from all around the world. • **Has** Helpdesk. • **Chairs and participates actively** in the FEGA committees. • **Unbounded** time commitment for FEGA.	• **Hosts and/or owns data**. • **Shares metadata** to allow global discovery. • **Accepts** 3rd party data from jurisdiction. • **Has** Helpdesk responsible for jurisdiction. • **Participates actively in** the FEGA Committees • Expected to **stay** in FEGA at least **4 years**.	• **Hosts/owns data**. • **Shares metadata** to allow global discovery. • **Does not** accept 3rd party data. • **Minimal** Helpdesk (distribution only). • **Does not participate in** the FEGA Committees. • Expected to **stay** in FEGA at least **4 years**.

The Federated EGA network structure at a glance

The key characteristic distinguishing the three tiers (CEGA, FEGA Nodes and FEGA Affiliates) is the jurisdiction from which the tier accepts data for archiving. CEGA accepts data from the global scientific community, while a FEGA Node accepts data from a single country's jurisdiction, following the legal framework within that country. A FEGA Affiliate archives data originating from their own organisation or selected partner institutions (e.g. international consortium partners). Furthermore, CEGA and FEGA Nodes are considered to be "service providing tiers"—providing an archiving service to external users, either globally or within their jurisdiction—whereas FEGA Affiliates are not providing an archiving service to external users and only archive data within their legal entity.

2.1 Central EGA

CEGA, coordinated by EMBL-EBI and CRG, offers submission, long-term data archiving, and data distribution to the international scientific community. In the federation, CEGA will:

1. Govern entry and exit conditions to the federated network according to the FEGA CAs and with input from the FEGA Strategic Committee;
2. Provide rotating chairs of the FEGA Strategic Committee and FEGA Operations Committee;
3. Govern the use of the EGA brand (e.g., grant applications, websites, public presentations);
4. Convene the EGA Scientific Advisory Board;
5. Coordinate the admission of new FEGA members and the assignment to the appropriate tier;
6. Review Node and Affiliate performance in reference to the CA and report to the relevant boards and committees;
7. Support international submissions, ETL (extract, transform, and load) pipeline, and data distribution of data held at CEGA;
8. Offer international helpdesk services to all users of central services;
9. Provide helpdesk support and training to FEGA nodes;
10. Coordinate the development and maintenance of EGA APIs, tools, and resources;
11. Offer coordination and support to FEGA Affiliates to operationalise.

CEGA will provide the following production-level services to the FEGA network:
1. EGA project website capable of listing descriptions of all studies hosted in the FEGA Network;
2. Accessioning service for issuing permanent EGA identifiers for the relevant metadata objects submitted to the Node/Affiliate;
3. Service for receipt of non-personal metadata from the Node/Affiliate.

2.2 Federated EGA Node

FEGA Nodes provide EGA services (external helpdesk, submissions, archiving, and distribution) for a particular jurisdiction. Publicly shareable metadata about studies/datasets archived at

these nodes will be shared with CEGA to enable discovery. The minimum membership term will be four years. The following summarises FEGA Node commitments and service expectations, in alignment with the FEGA Node CA and FEGA Node Operations guidelines document[1].

FEGA Nodes will provide necessary services that are required for the instantiation and operation of a FEGA Node, including:
1. Data submission and validation service to the node;
2. Validation of metadata and submission of non-personal metadata to CEGA to obtain EGA accessions and promote data discovery;
3. An ETL pipeline (see Appendix 1);
4. Geo-dispersed data storage system where all submitted data is replicated in at least two locations;
5. Data distribution service providing data access to authorised users for datasets held at their Node;
6. Permissions management service for datasets archived by the Node;
7. Helpdesk service with sufficient capacity to respond to user queries, including from submitters, Data Access Committees within the jurisdiction, and authorised users for accessing data held at the Node;
8. Operational capacity to deliver node services in terms of personnel and/or infrastructure;
9. Authorise CEGA to monitor the availability of Federated services and provide summary level operational reporting to FEGA Operations Committee;
10. A system capable of supporting local data security and EGA best practice requirements;
11. A consistent user experience for users interacting with the FEGA Node, based on shared EGA SOPs.

Federated EGA Nodes will also:
1. Contribute, within their capabilities, to the development of shared EGA APIs, tools, and resources;
2. Contribute with their capabilities to helpdesk support and training to onboarding of less experienced Federated EGA Nodes;
3. Nominate members for the FEGA Strategic Committee, FEGA Operations Committee, and relevant working groups. The members will contribute meaningfully and proactively to the tasks of those committees;
4. Nominate representatives to attend EGA SAB meetings as observers.

2.3 Federated EGA Affiliate

A FEGA Affiliate may be an individual institution, a national or regional healthcare genomics initiative, consortium of institutions, and/or a project-specific consortium, either national or international. A FEGA Affiliate manages user access to data hosted by the Affiliate and provides data distribution to approved users. It is recommended that they offer the same APIs and

---

[1] https://ega-archive.org/assets/files/Federated-EGA-Node-Operations-v2.1.pdf

EMBL-EBI

CRG
Centre
for Genomic
Regulation

interfaces as CEGA but they can also develop their own APIs and interfaces. Publicly shareable metadata about studies/datasets archived at FEGA Affiliates will be shared with CEGA to enable discovery. The minimum membership term for a FEGA Affiliate will be four years. The following summarises the commitments and service expectations of a FEGA Affiliate, in alignment with the FEGA Affiliate CA.

FEGA Affiliates will provide necessary services that are required for the instantiation and operation of a FEGA Affiliate, including:

1. Submission of non-personal metadata to CEGA to obtain EGA accessions and promote data discovery (validation of metadata will be performed by CEGA Helpdesk);
2. A data storage system and a defined approach to data backup or recovery in case of data loss;
3. Data distribution service providing data access to authorised users for datasets held at the Affiliate;
4. Permissions management service for datasets archived by the Affiliate;
5. Helpdesk service with sufficient capacity to respond to user queries, including from relevant Data Access Committees, and authorised users for accessing data held at the Affiliate;
6. Operational capacity to deliver FEGA Affiliate services in terms of personnel and/or infrastructure;
7. Federated network uptime monitoring including providing summary level operational reporting to FEGA Operations Committee;
8. A system capable of supporting local data security and EGA best practice requirements.

3 Services, Infrastructure, and Governance

	Central EGA (CRG & EMBL-EBI)	Federated EGA Node	Federated EGA Affiliate
External services			
Data submission (see Appendix 1)	Offers international submission service	Offers submission service for a particular jurisdiction	Has no public submission commitments
Helpdesk support (see Appendix 1)	Provides external international helpdesk support	Provides helpdesk support for submitters in its jurisdiction and for approved users of data managed at its facilities	Provides helpdesk support for approved and prospective users of data managed at its facilities

Data Access Committees	Provides support and tooling for DAC activities	Conforms to Central EGA DAC SOPs	Manages the DACs linked to their own data/studies.
Data Distribution (see Appendix 1)	Manages worldwide distribution for data hosted at CEGA	Manages worldwide distribution for data hosted at FEGA Node	Manages worldwide distribution for data hosted at Affiliate
Data Discovery	Provides worldwide data discovery services	Integrates with EGA data discovery services	Integrates with EGA data discovery services
(F)EGA logos and branding	Full use of EGA/FEGA branding in funding requests and external communications	Full use of FEGA branding in funding requests and external communications	Full use of FEGA Affiliate branding in funding requests and external communications
Infrastructure			
Compute and storage	Sufficient compute and storage to support submission, QC, ETL, and distribution	Sufficient compute and storage to support local submission, QC, ETL, and distribution	Sufficient compute and storage to support hosting and distribution of data
Authentication and authorisation (see Appendix 1)	Hosts or proxies the AAI services to other nodes	Conforms to EGA APIs, is integrated and compatible with Central EGA	Is integrated and compatible with Central EGA
APIs	Develops and maintains core EGA APIs	Conforms to core EGA APIs	Recommended use of core EGA APIs to allow integration of services
Data replication (see Appendix 1)	At least two geographically replicated copies of all data	At least two geographically replicated copies of all data	Defined approach to data backup or recovery in case of data loss
Helpdesk service (see Appendix 1)	Hosts and coordinates Central EGA helpdesk	Helpdesk system to coordinate with Central EGA. Must	Use their own user

	system. Provides operational contacts for all EGA nodes.	have a designated operational contact person available for Central EGA requests	request/issue management mechanism. Must have at least a designated contact person available for Central EGA requests
Software development licensing	Permissive licensing for all EGA code, e.g. Apache 2.0[2]	Permissive licensing for all FEGA-related code, e.g. Apache 2.0	Recommend permissive licensing for EGA-related code
Governance			
EGA Strategic Committee (see Appendix 1)	Membership	Not member	Not member
FEGA Strategic Committee (see Appendix 1)	Co-chairs	Membership	Not member
FEGA Operations Committee (see Appendix 1)	Co-chairs	Membership	Not member
(Central) EGA Scientific Advisory Board	Required attendance, host of annual SAB meeting	Optional attendance as observer	Not required
Basis for membership	Legal agreement for entry/exit	Legal agreement for entry/exit depending on approval by Central EGA and the FEGA Strategic Committee. Minimum term four years.	Legal agreement for entry/exit depending on approval by Central EGA. Minimum term four years.

---

[2] Apache 2.0 license: https://www.apache.org/licenses/LICENSE-2.0

Appendix 1 - Glossary of Terms

Activity or Organisation	Definition of Activity
Authentication and authorisation	Authentication - verification of a person's identity. Authorisation - verification from a data controller that a user has permission to access data.
Collaboration Agreement (CA)	The Federated EGA Collaboration Agreement, signed between Central EGA (EMBL and CRG) and the node.
CEGA	Central EGA is coordinated by EMBL-EBI and CRG and offers submission, long-term data archiving, and data distribution to the international scientific community.
Common Kernel	The Common Kernel consists of the following documents published on the EGA project website (https://ega-archive.org/federated) describe the organisational structure and committees: EGA Federation: Structure and Organisation (this document), Terms of reference for the FEGA Strategic Committee, Terms of reference for the FEGA Operations Committee.
Data distribution	A service that enables authorised users to securely access or download either entire datasets (e.g. secure FTP, Aspera) or subsets (e.g. htsget API).
Data replication	A geographically separated physical copy of the raw data and metadata.
Data submission	A node has the capability to offer: • A secure user account login system • Secure upload of raw data and metadata • Validate the raw data and metadata according to current EGA protocols Full details are provided in additional Federated EGA technical and SOP documents.
Data QC	A service that can generate and publicise high level quality metrics from submitted data, primarily data from genome sequencing data.
EGA Strategic Committee	A committee formed of senior management from Central EGA nodes that provide the strategic governance for the EGA, and by extension FEGA.
ETL pipeline (Extract, transform, and load)	The software pipeline that takes an EGA submission (data+metadata) and • Validates the input (data integrity, syntax, and semantics) • Archives the data in a replicated long term storage

	● Loads the metadata into a queryable database system
FEGA/FEGA Network	The collection of Central EGA, FEGA Nodes and FEGA Affiliates, which make up the federated, distributed service for long-term archiving and sharing of all types of human data resulting from biomedical research projects where the data is consented for sharing in a controlled access context. Also known as Federated EGA or the FEGA Network.
Federated EGA Operations Committee	A committee formed of operational supervisors and leads from the FEGA nodes to review, communicate, harmonise operational aspects of the FEGA nodes (e.g. service technical compatibility, updates in policies and SOPs, helpdesk service capacity). See ToR for committee for more details.
Federated EGA Strategic Committee	A committee formed of senior management representatives from all the Federated EGAnodes to provide strategic governance and policy setting for the federated nodes. See ToR for committee for more details.
Helpdesk support	A service where external users can send operations support queries such as: ● Completing data submissions ● Finding datasets ● Data download or access issues The service should have sufficient personnel resources to provide timely responses.
ToR	Terms of Reference

Version tracking table

	Authors	Summary of changes	Approved
v2.0	Thomas Keane, Jordi Rambla, Mallory Freeberg, Teresa D'Altri, Amy Curwin, Gabriele Rinck, Ana Teresa Freitas, EGA and FEGA Strategic Committees	Update how the third tier (now Affiliate, formerly Community) participates in FEGA. Update in scope and introductory section (new figure). Updates in services, alignment with CA and ToRs.	January 2026

CRG
Centre for Genomic Regulation

v1.1	Thomas Keane, Jordi Rambla, Dylan Spalding, Paul Flicek, Arcadi Navarro, Helen Parkinson	Minor edits, additional glossary and appendix definitions	January 2020
v1.0	Thomas Keane, Jordi Rambla, Paul Flicek, Arcadi Navarro, Helen Parkinson	Based on experience of Central EGA	March 2019