

[Federated] EGA

Providing global discovery and
access for sensitive human data

Salvador Capella-Gutierrez, INB/ELIXIR-ES @ BSC



EUROPEAN
GENOME-PHENOME
ARCHIVE

EMBL-EBI



Sharing sensitive data key to enable science & health



Exposure of personal or sensitive data is **potentially harmful**

Individuals have the **right to protection** of their personal data

Governments pass legislation that **codifies data protection rights**

Support aims of research & health care initiatives

EGA manages sensitive human data



Mission: Permanent archiving and sharing of **personally identifiable genetic, phenotypic, and clinical data** generated for the purposes of biomedical research projects or in the context of research-focused healthcare systems

Drive development of and implement community standards for supporting FAIR data sharing



Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

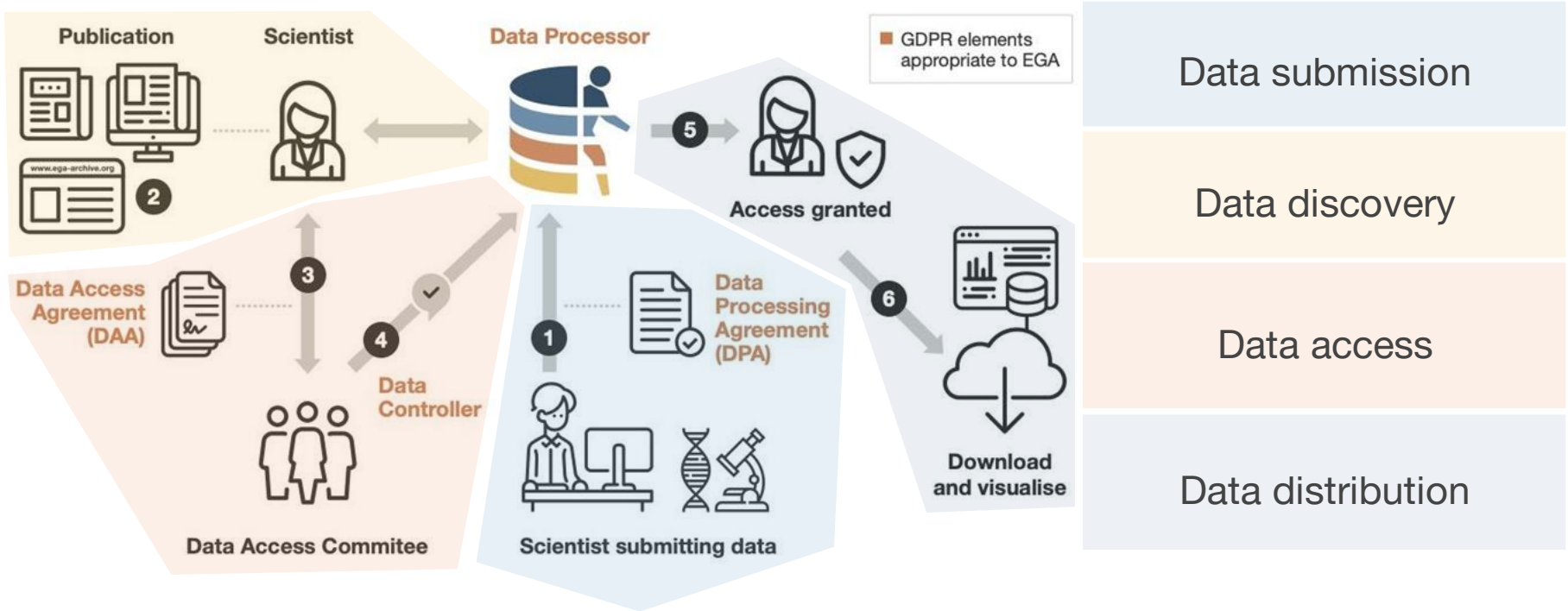


2008: EGA launched at EMBL-EBI

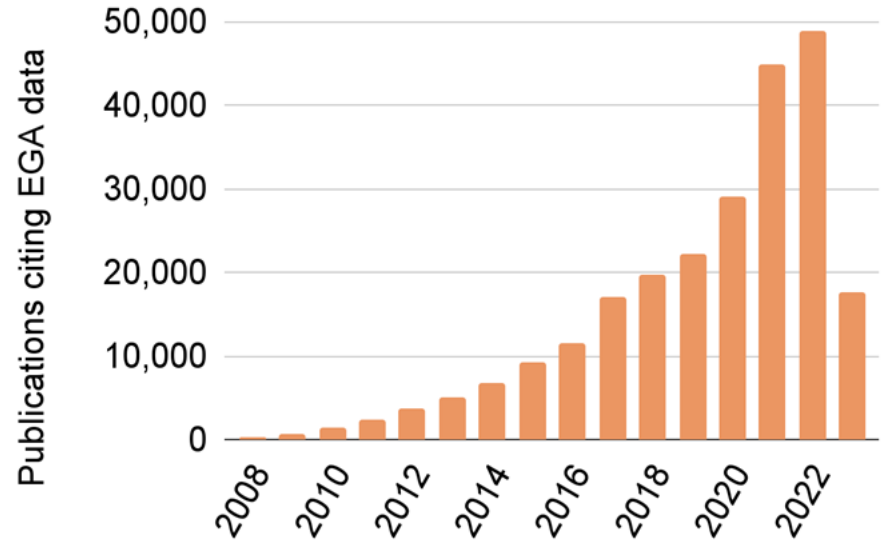
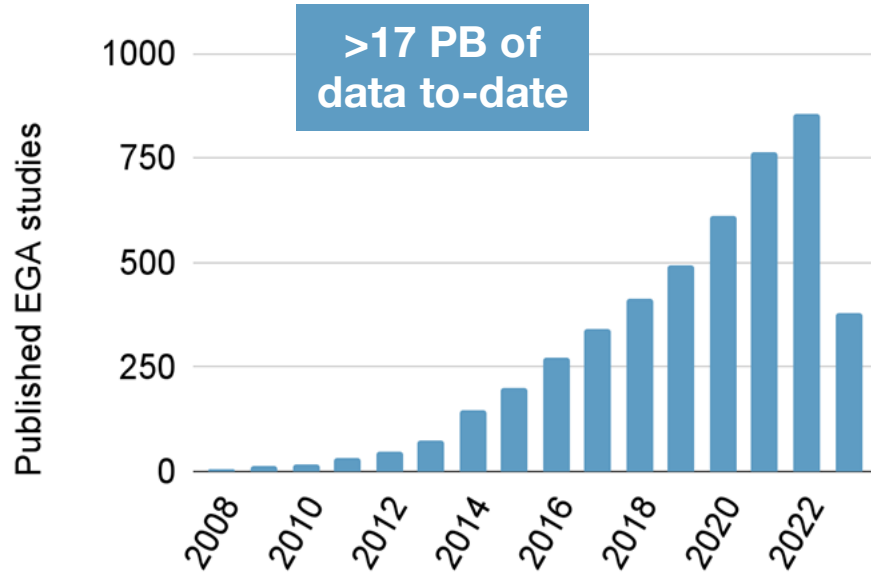


2013: CRG & EMBL-EBI enter formal collaboration to co-manage EGA

EGA data access model



EGA grows and remains key to life science research



Impact of EGA in personalised medicine

– Common diseases

EGA Studies:

- UK10K ([EGAS00001000713](#))
- WGS GoT2D ([EGAS00001001459](#))
- WES GoT2D ([EGAS00001001460](#))

Revision and reanalysis of existing genetic data by applying more powerful analytic techniques and resources generated.

A total of 70,127 subjects (70KforT2D) were retained for downstream analysis

Up to 7 loci were identified that were not previously reported as associated with T2D. Providing more candidate causal variants.

refs: [Nature Comms \(2015\)](#), [Nature \(2016\)](#), [Nature Comms \(2018\)](#)

nature
International journal of science

Article Published: 11 July 2016

The genetic architecture of type 2 diabetes

ARTICLE
Received 17 Mar 2014 | Accepted 17 Jul 2015 | Published 14 Sep 2015

Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel

nature communications

Explore content ▾ About the journal ▾ Publish with us ▾

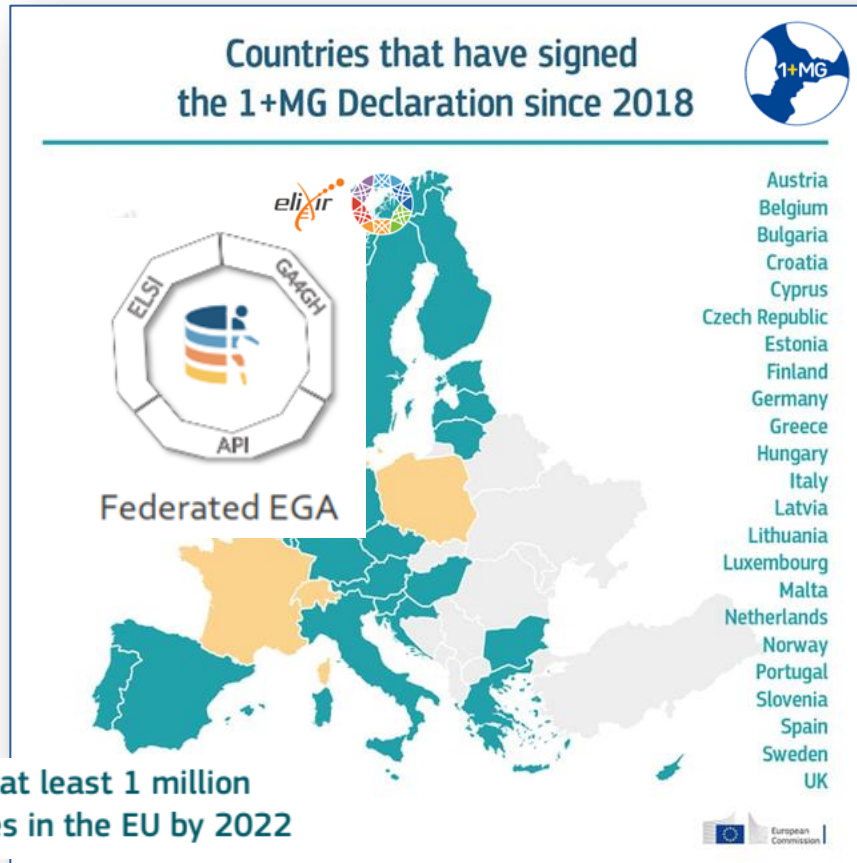
nature > nature communications > articles > article

Article | [Open Access](#) | Published: 22 January 2018

Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes

Manhattan plot showing $-\log_{10}(P)$ values across chromosomes 1-22, X, and Y. Significant associations are highlighted with colored bars and labels, including *PCSK9*, *SLC39A8*, *SLC6A14*, *SLC6A15*, *SLC6A16*, *SLC6A17*, *SLC6A18*, *SLC6A19*, *SLC6A20*, *SLC6A21*, *SLC6A22*, *SLC6A23*, *SLC6A24*, *SLC6A25*, *SLC6A26*, *SLC6A27*, *SLC6A28*, *SLC6A29*, *SLC6A30*, *SLC6A31*, *SLC6A32*, *SLC6A33*, *SLC6A34*, *SLC6A35*, *SLC6A36*, *SLC6A37*, *SLC6A38*, *SLC6A39*, *SLC6A40*, *SLC6A41*, *SLC6A42*, *SLC6A43*, *SLC6A44*, *SLC6A45*, *SLC6A46*, *SLC6A47*, *SLC6A48*, *SLC6A49*, *SLC6A50*, *SLC6A51*, *SLC6A52*, *SLC6A53*, *SLC6A54*, *SLC6A55*, *SLC6A56*, *SLC6A57*, *SLC6A58*, *SLC6A59*, *SLC6A60*, *SLC6A61*, *SLC6A62*, *SLC6A63*, *SLC6A64*, *SLC6A65*, *SLC6A66*, *SLC6A67*, *SLC6A68*, *SLC6A69*, *SLC6A70*, *SLC6A71*, *SLC6A72*, *SLC6A73*, *SLC6A74*, *SLC6A75*, *SLC6A76*, *SLC6A77*, *SLC6A78*, *SLC6A79*, *SLC6A80*, *SLC6A81*, *SLC6A82*, *SLC6A83*, *SLC6A84*, *SLC6A85*, *SLC6A86*, *SLC6A87*, *SLC6A88*, *SLC6A89*, *SLC6A90*, *SLC6A91*, *SLC6A92*, *SLC6A93*, *SLC6A94*, *SLC6A95*, *SLC6A96*, *SLC6A97*, *SLC6A98*, *SLC6A99*, *SLC6A100*.

European 1+MG initiative sets vision for data sharing



Towards access to at least 1 million sequenced genomes in the EU by 2022



Beyond One Million Genomes Project



European Genomic Data Infrastructure



Federated EGA

More genomics data generated in healthcare context

Areas of clinical uptake: infectious disease, cancer, rare disease, common/chronic



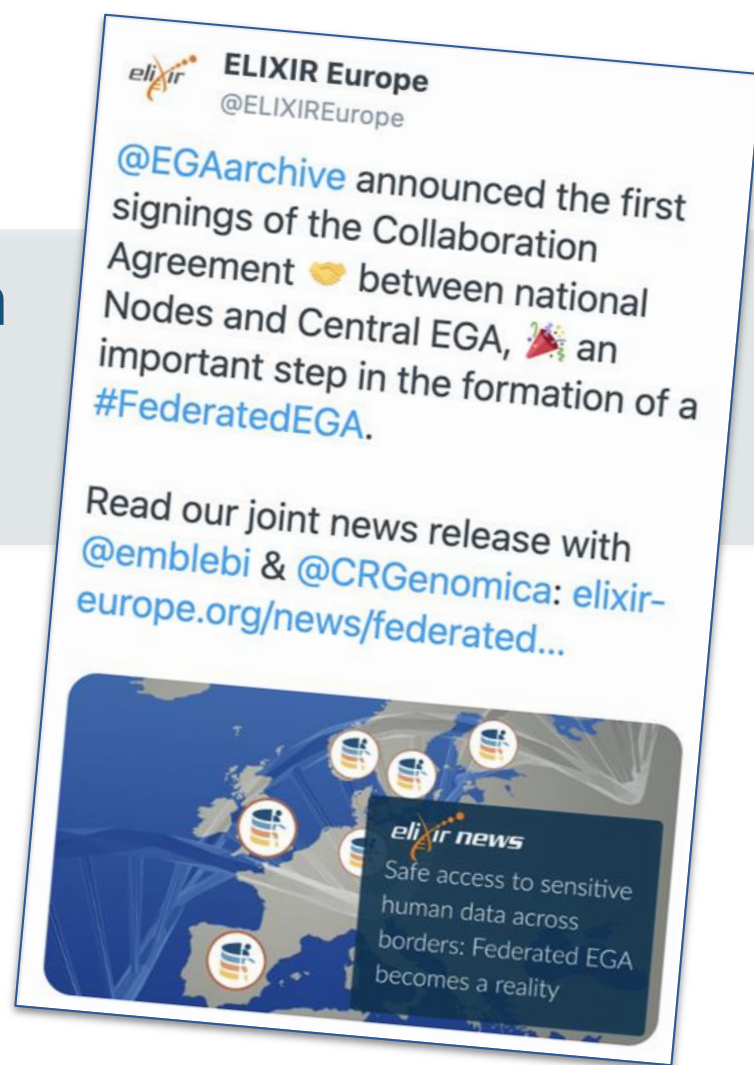
Medical Genomes

- Countries with active national medical genome projects
- Countries with some activity of medical genomics
- Countries planning medical genome projects

Cohorts

- National cohorts > 100k genotyped or sequenced at least 25k
- National cohorts > 100k people active collection now
- Planning national cohorts > 100k

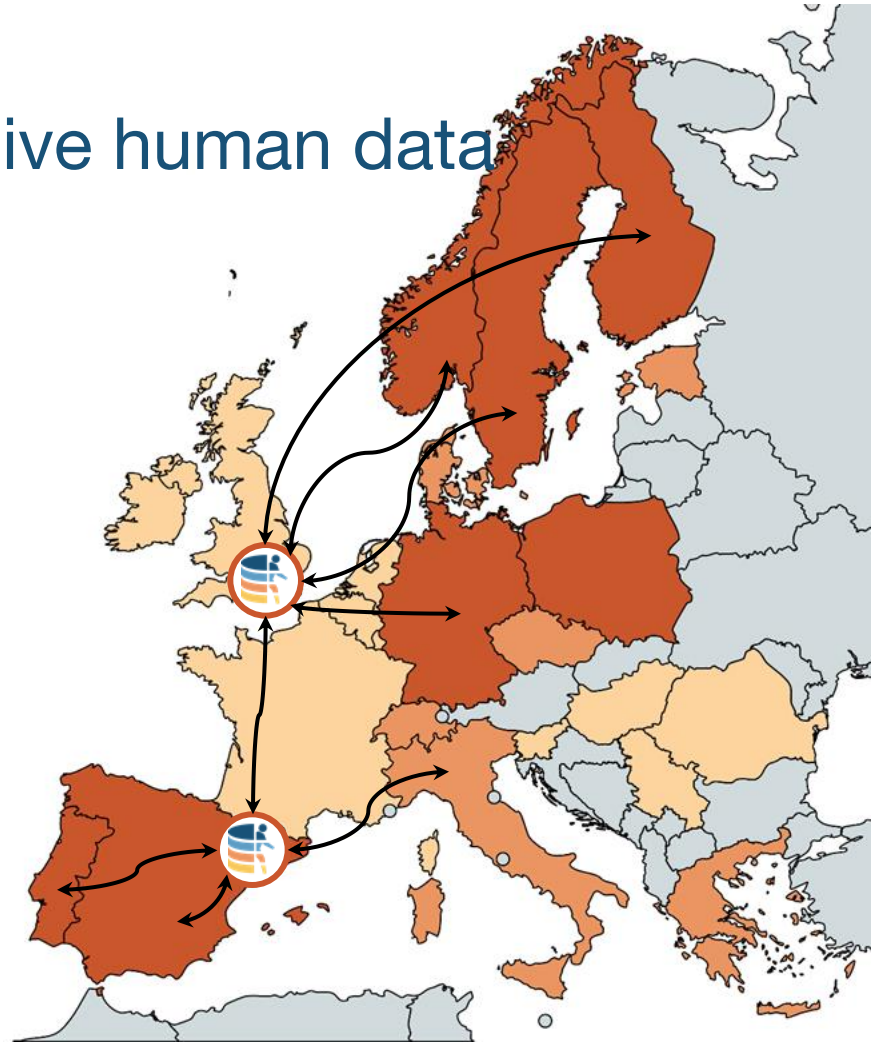
Federated EGA collaboration officially launched in September 2022








Federated EGA:

discovery & access for sensitive human data

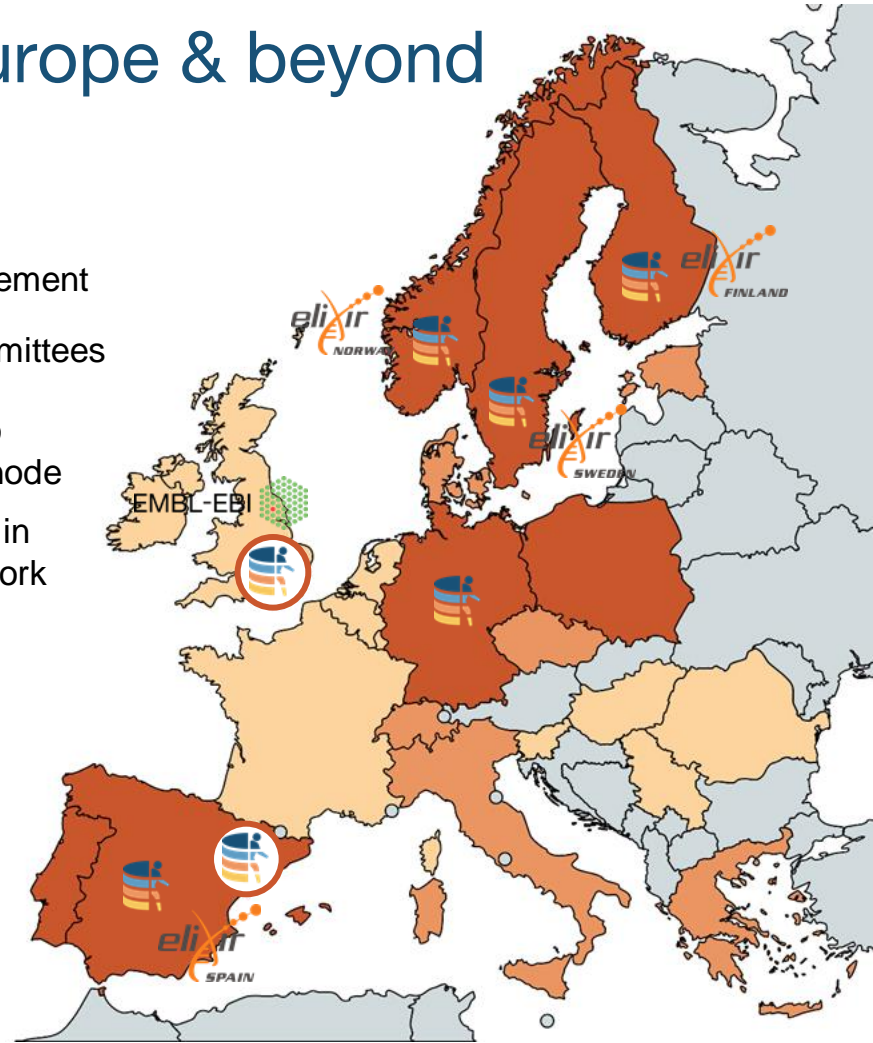
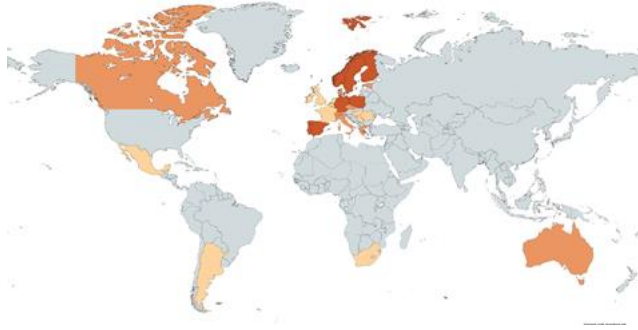
Federated EGA strives to support the discovery of and secure access to human data globally, while respecting national data protection regulations, with the goal of accelerating disease research and understanding and improving human health.



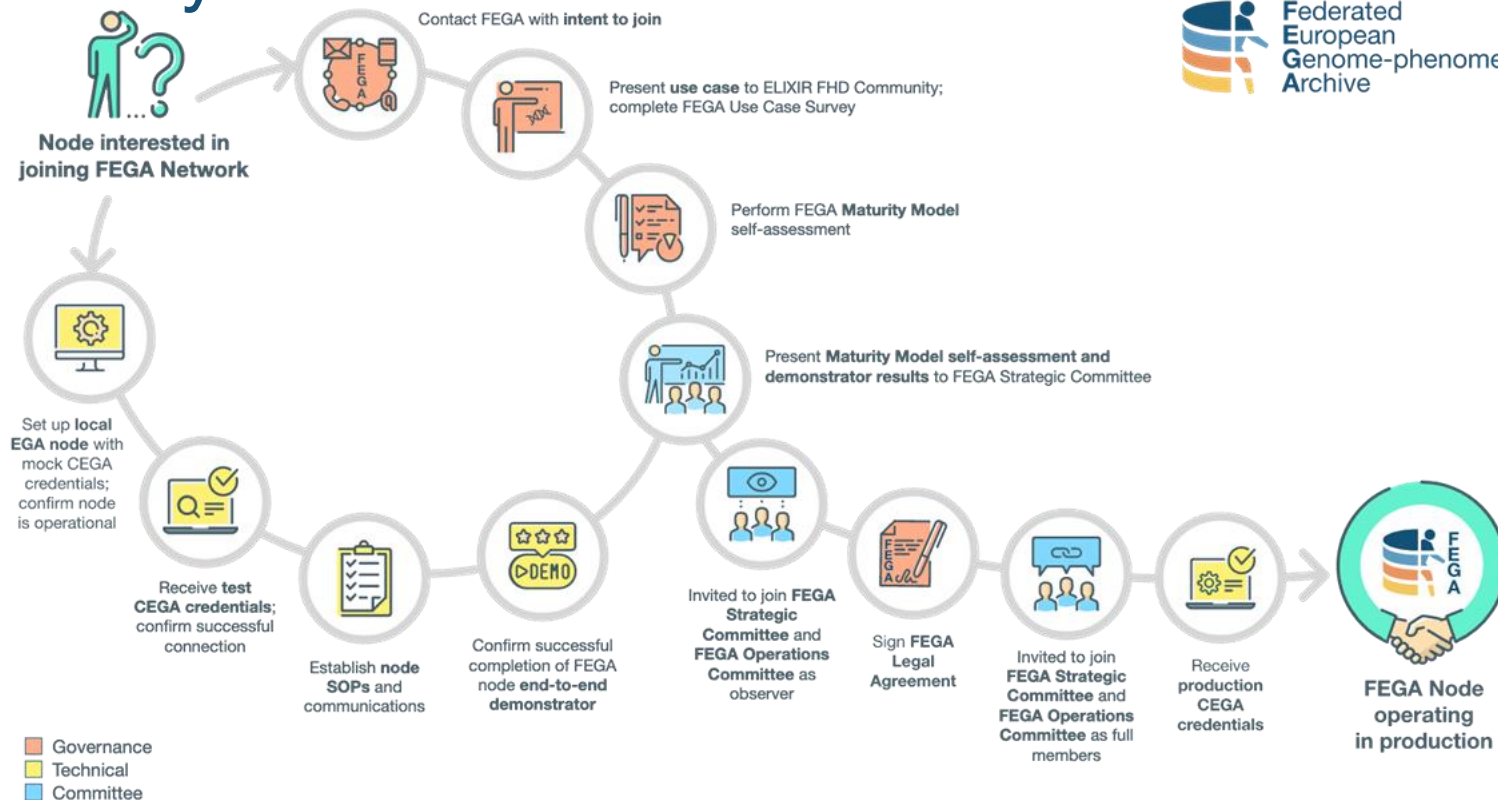
FEGA engagement across Europe & beyond

-  Central EGA
-  Signed FEGA Collaboration Agreement
-  Joined FEGA Committees
-  Engaged in work to establish a FEGA node
-  Expressed interest in joining FEGA Network

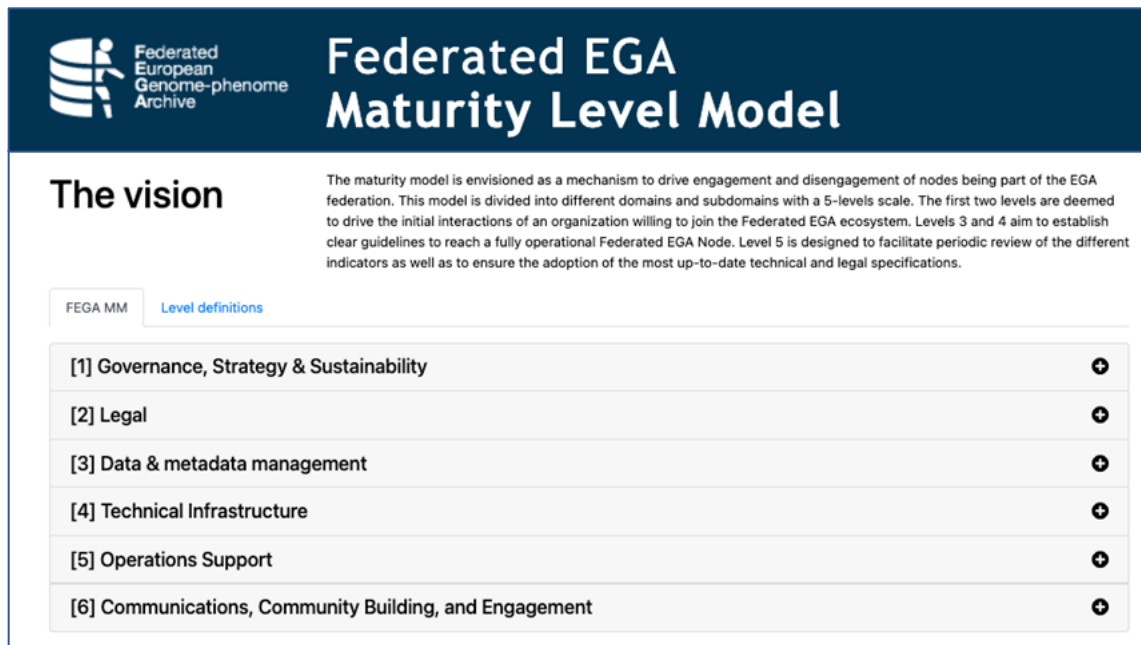
>24 nodes engaged and counting!



Joining FEAGA network follows established pathway



FEGA Maturity Model enables self-assessment



Federated EGA Maturity Level Model

The maturity model is envisioned as a mechanism to drive engagement and disengagement of nodes being part of the EGA federation. This model is divided into different domains and subdomains with a 5-levels scale. The first two levels are deemed to drive the initial interactions of an organization willing to join the Federated EGA ecosystem. Levels 3 and 4 aim to establish clear guidelines to reach a fully operational Federated EGA Node. Level 5 is designed to facilitate periodic review of the different indicators as well as to ensure the adoption of the most up-to-date technical and legal specifications.

The vision

FEGA MM Level definitions

- [1] Governance, Strategy & Sustainability
- [2] Legal
- [3] Data & metadata management
- [4] Technical Infrastructure
- [5] Operations Support
- [6] Communications, Community Building, and Engagement

Align with other Maturity Models



HEALTHYCLOUD
Health Research & Innovation Cloud



FAIRplus



Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.



<https://inab.github.io/fega-mm/>

Guidance for nodes considering joining FEGA

The screenshot displays the FEGA documentation website. On the left, there is a navigation sidebar with a search bar and a list of topics. The main content area shows the breadcrumb 'Docs > Establishing a Federated EGA Node' and a sub-section '2. Explore implemented standards'. Below this, a table lists three standards: Beacon, Crypt4GH, and Data Use Ontology (DUO). Each row in the table includes columns for Standard, Purpose, Specification Version, Supported Version, Implementation, and Publication/Preprint.

Federated European Genome-phenome Archive

Search docs

- Establishing a Federated EGA Node
 - What am I doing here?
 - What is Federated EGA?
 - How do I start?
 - What does the journey look like?
 - Acknowledgements
 - License
- TOPICS
 - Maturity Model
 - Data and Metadata Management
 - 1. Learn about data security best practices
 - 2. Explore implemented standards
 - 3. Understand data definitions and flow
 - 4. What's next?
 - Outreach and Training
 - Technical and Operational
 - Governance and Legal

Docs > Establishing a Federated EGA Node

2. Explore implemented standards

Central EGA largely adhere to **GA4GH standards**. Specific standards already implemented are summarised below:

Standard	Purpose	Specification Version	Supported Version	Implementation	Publication/Preprint
Beacon	Supports discovery of genomic variants and individuals.	v1.0.1	v0.3	Specification, Documentation, Endpoint	N/A
Crypt4GH	Enables direct byte-level compatible random access to encrypted genetic data stored in community standards (e.g. CRAM, VCF).	v1.0	v1.0	Specification, Documentation, Endpoint	DOI
Data Use Ontology (DUO)	Allows users to semantically tag datasets with usage restrictions so datasets can be automatically discoverable based on a researcher's authorization level or intended use.	2021-02-23	2021-02-23	Specification, Documentation, Endpoint	DOI

Federated European Genome-phenome Archive

Search docs

Establishing a Federated EGA Node

TOPICS

- Maturity Model
- Data and Metadata Management
- Outreach and Training
- Technical and Operational
- Governance and Legal

<https://ega-archive.github.io/FEGA-onboarding/>

What's next for EGA and Federated EGA?

- Build on **human data sharing knowledge & experiences** of experts across Europe
- Accelerate **additional national nodes** joining Federated EGA
- Continue advocating for **community standards** via Driver Projects and ELIXIR::GA4GH Strategic Partnership
- Support **transnational access to human data** across Europe & beyond via ELIXIR-coordinated initiatives/projects e.g. 1+MG, B1MG, GDI

Acknowledgements: Federated EGA



Thank you!

Questions?



EUROPEAN
GENOME-PHENOME
ARCHIVE

EMBL-EBI



CRG
Centre
for Genomic
Regulation



Barcelona
Supercomputing
Center

Centro Nacional de Supercomputación

Additional resources

- EGA: <https://ega-archive.org/>
- Joining FEGA: <https://ega-archive.github.io/FEGA-onboarding/>
- GA4GH: <https://www.ga4gh.org/>
- ELIXIR: <https://elixir-europe.org/>
- GDI: <https://gdi.onemilliongenomes.eu/>