

Sharing sensitive human genomic, phenotypic, & clinical data

Dr. Mallory Freeberg
Coordinator, European Genome-phenome Archive (EGA)

Tw: @MalloryFreeberg
Em: mfreeberg@ebi.ac.uk
ORCID: 0000-0003-2949-3921

Bioinformatics for Biobusiness

6 July 2022



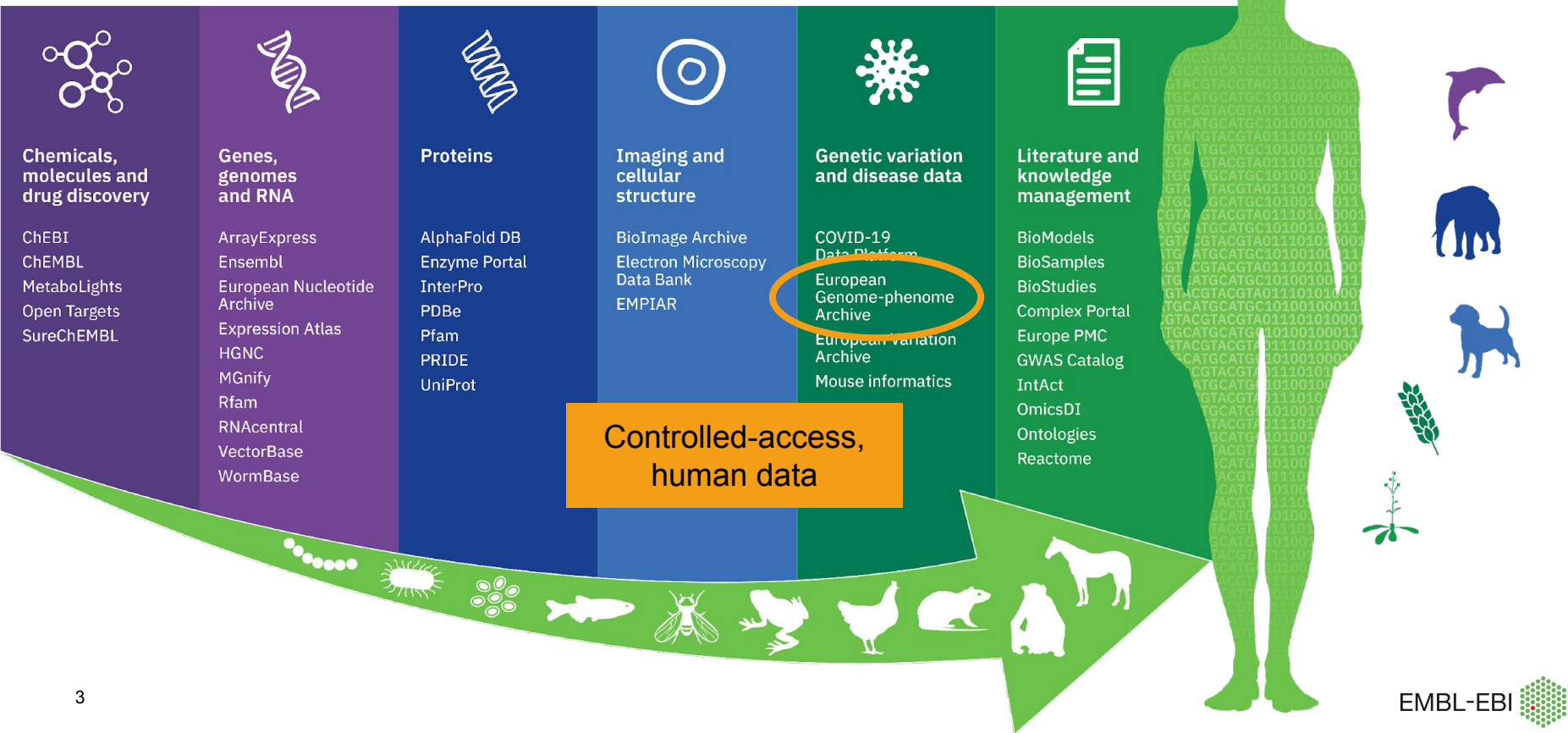
slides

What is EMBL-EBI?

- World leading source of public biomolecular data
- Our vision is to benefit humankind by advancing scientific discovery and impact through bioinformatics.
- Part of the European Molecular Biology Laboratory (EMBL), Europe's flagship laboratory for the life sciences.



Data resources at EMBL-EBI



What is special about controlled access data?



How can controlled access data be shared?



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

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

e.g. name, email, genetic characteristics, medical history, disease diagnoses

EU General Data Protection Regulations (GDPR)

UK General Data Protection Regulation and Data Protection Act 2018

Finland Data Protection Act (supplement to GDPR)

Research aims

Maximise value
Be good data stewards
Reduce waste

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 implementation of standards for supporting sharing of controlled-access human data
- FAIR principles
- Community standards



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

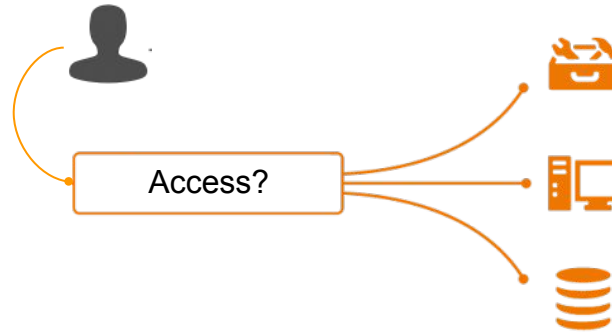
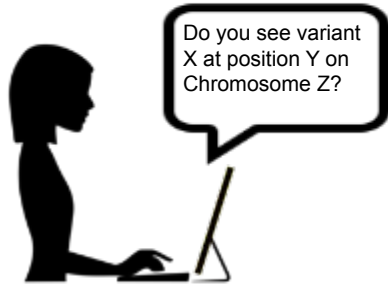


2008 EGA launched at EMBL-EBI



2012 CRG and EMBL-EBI enter formal collaboration to manage EGA

Challenges for managing controlled access data



Data discovery

How do I find data relevant to me?
How do I know these data are useful?

Data access
management

How can I access data?
How can I prevent unauthorised access?

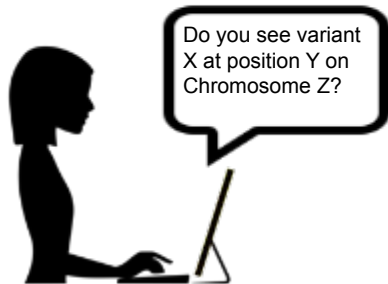
Regional data
protection regulations

What do I do if my data cannot be shared
outside regional borders?



**How can these
challenges be met?**

Data Discovery Finding relevant, useful data



Data discovery

How do I find data relevant to me?
How do I know these data are useful?

Won't talk about:

- Increase data FAIRness
 - Findable: accessions, indexed
 - Accessible: standard access protocols
 - Interoperable: ontologies, vocabularies
 - Reusable: rich metadata, data usage license
- Metadata curation and harmonisation



“As open as possible. As controlled as required”

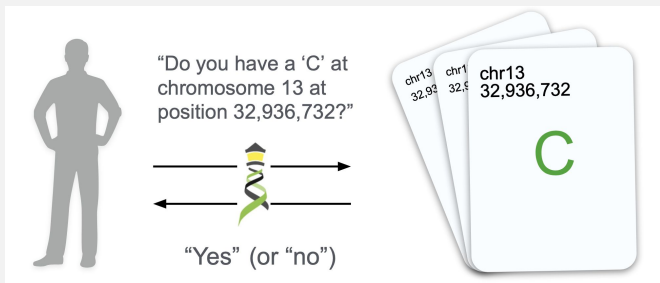
Discovery over public metadata

→ Non-identifiable, aggregated / binned

Discovery over sensitive data without direct access

→ Query tools

Data Discovery Beacon API for federated discovery



Can you provide data about focal deletions in CDKN2A in Glioblastomas from juvenile patients with unrestricted access?



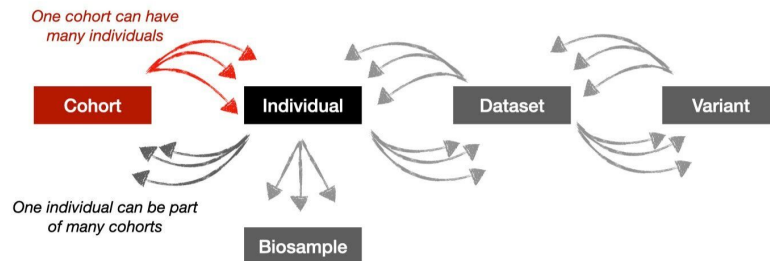
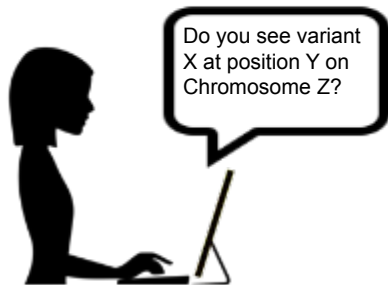
```
CDKN2A:DEL
size<1Mb
granularity:record
ncit:C3058
DUO:0000004
HP:0003621
```



Beacon defines an open standard for **federated discovery of human genomic, phenotypic, and clinical data** in biomedical research and clinical applications

Beacon v2 enables enhanced querying over **clinical and phenotypic information, curated variant annotations, and data use conditions**

Data Discovery Enhanced discovery features

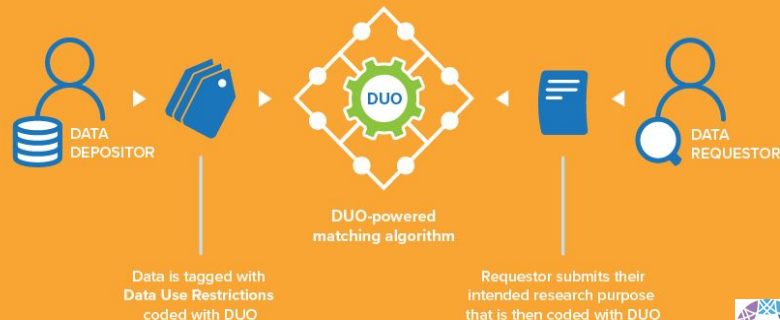


Data discovery

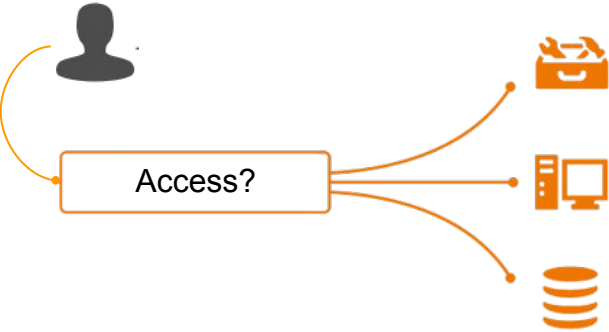
How do I find data relevant to me?
How do I know these data are useful?

Data Use Ontology (DUO) to describe data use conditions and restrictions

Automate matching with research purposes



Data Access How to manage access safely

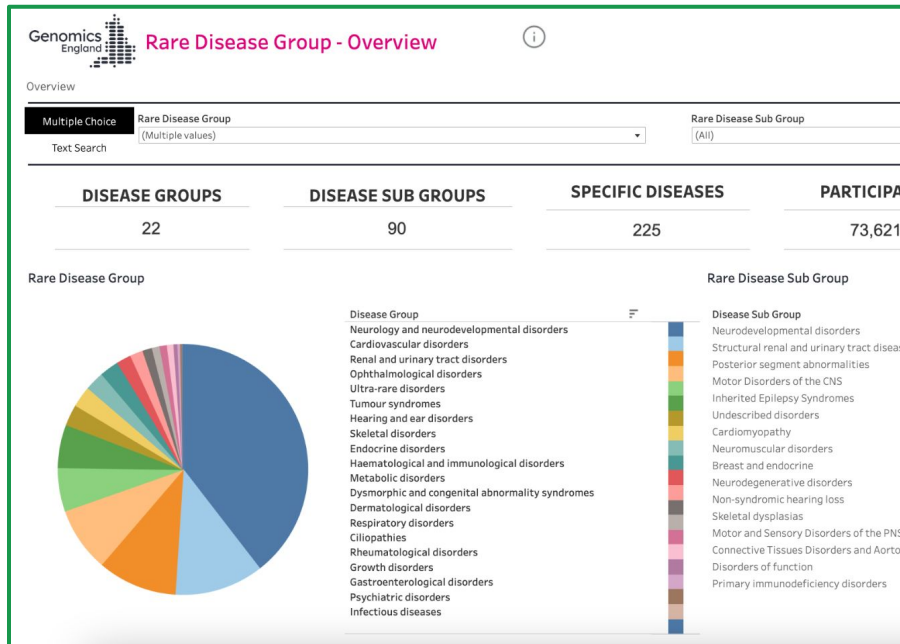


Data access
management

How can I access data?
How can I prevent unauthorised access?

Data Access Bring analyses to the data

- Trusted research environments (TREs) / Secure Data Environments (SDEs)



biobank

Research Analysis Platform

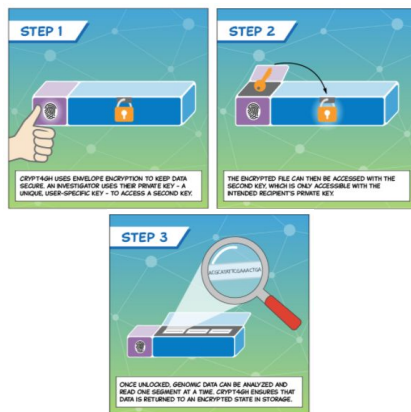
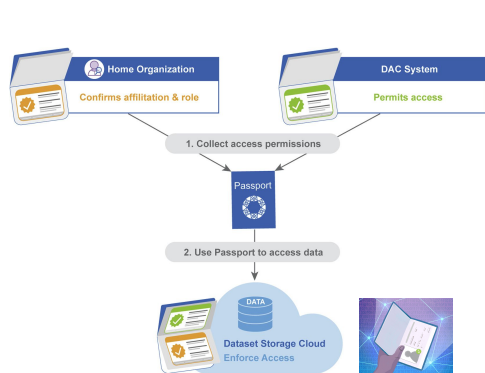
Enabling scientific discoveries that improve human health

CREATE AN ACCOUNT

UK Health Data Research Alliance white paper on Building Trusted Research Environments

0% 2% 4% 6% 8% 10% 12% 14% 16% 18% 20% 22% 24% 26% 28% 30% 32% 34% 36% 38% 40% 42% 44% 46% 48% 50% 52% 54% 56% 58% 60% 62% 64% 66% 68% 70% 72% 74% 76% 78% 80% 82% 84% 86% 88% 90% 92% 94% 96% 98% 100% % of Total Participants

Data Access Adopt modern, automated, scalable processes for federating data access



Standardised researcher identities and authorised data

GA4GH Passports / Visas

Secure file container format

GA4GH Crypt4GH

Machine computable data use conditions and restrictions

GA4GH DUO

Interoperable authentication and authorisation systems

ELIXIR Life Science Login / GA4GH AAI

Genomics & healthcare is a global effort



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

European 1+ Million Genomes Declaration

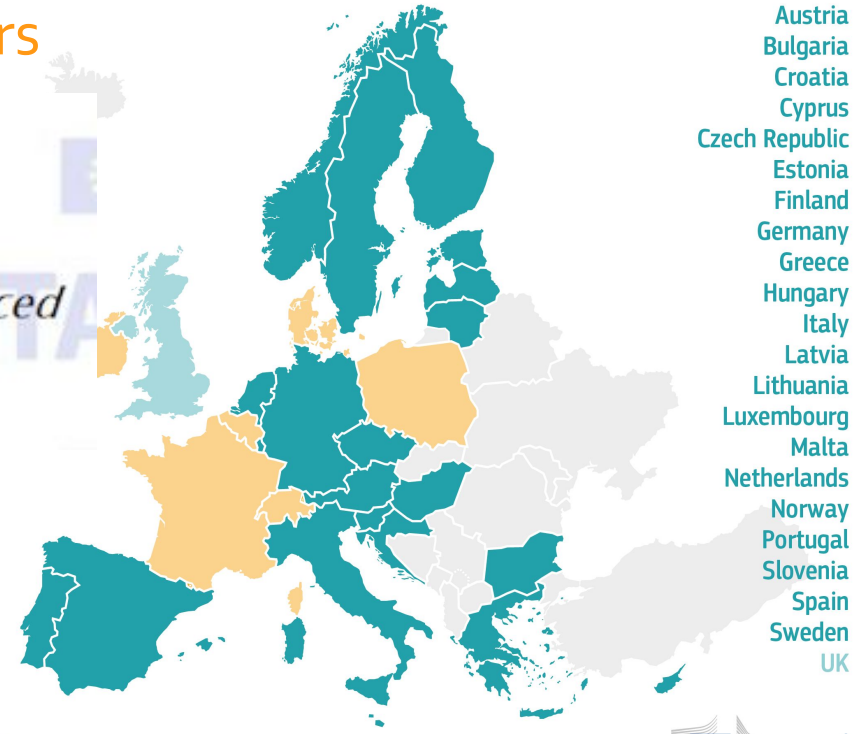
Sharing genomic data across borders

DECLARATION OF COOPERATION

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

Mariya Gabriel, Commissioner for Digital Economy and Society welcomed this new step forward:

"EU cooperation on linking genomic health data becomes ever more important to advance personalised health care, better prevention and excellence in research. We are happy to see Latvia joining the other 18 signatory Member States and look forward to enhance this European collaboration. " "



● Observer countries

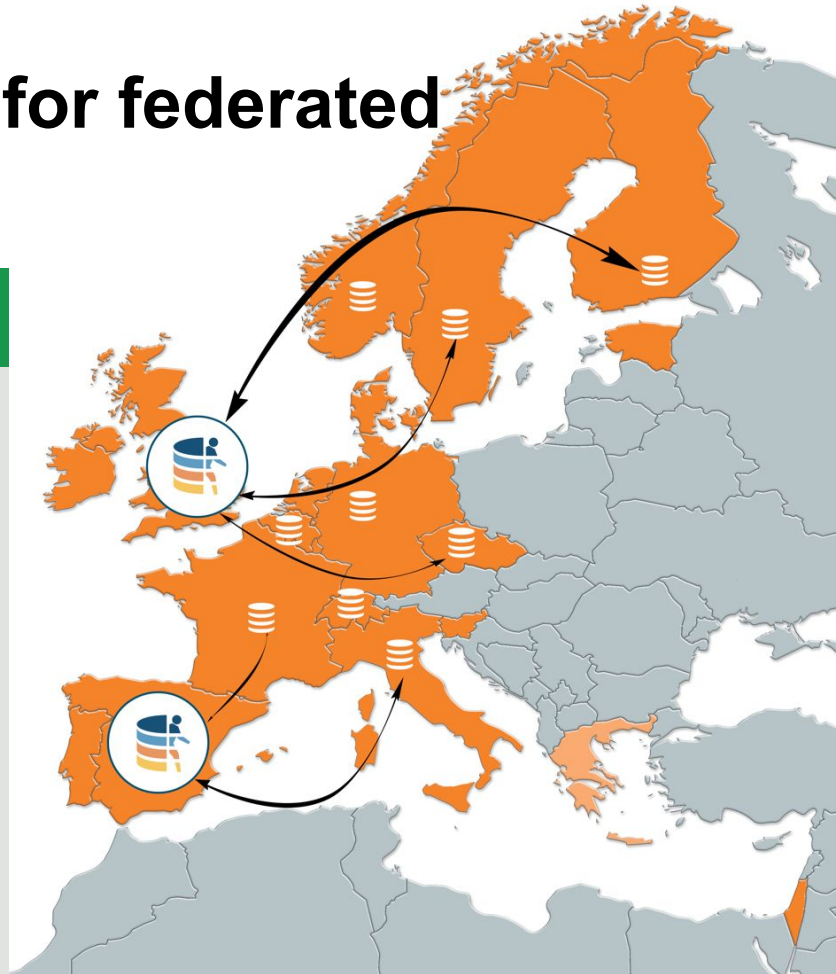


Infrastructure & coordination for federated human data access

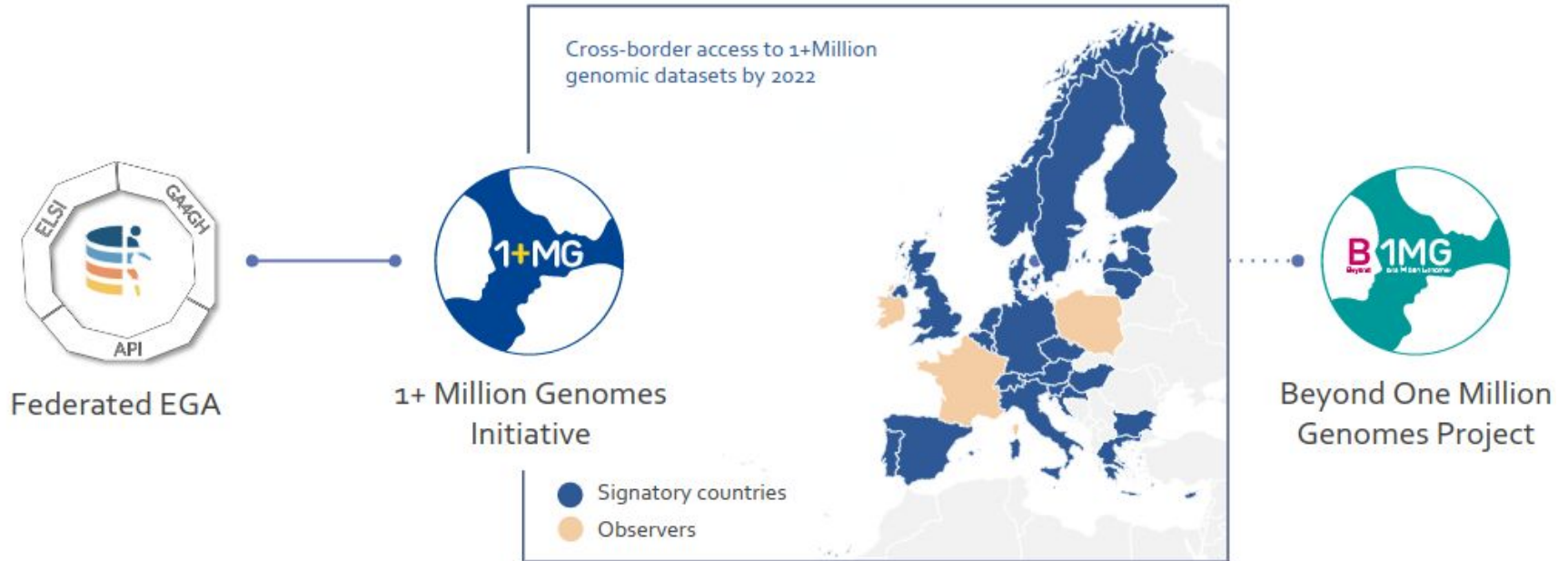
Federated EGA Network provides:

- Federated network of human data repositories and regional hubs
- Platform for discovery, access, distribution of human data
- Coordination and standards for discovery and access
- Community of experts to build capacity across Europe

ELIXIR Nodes engaged through the ELIXIR Federated Human Data Community

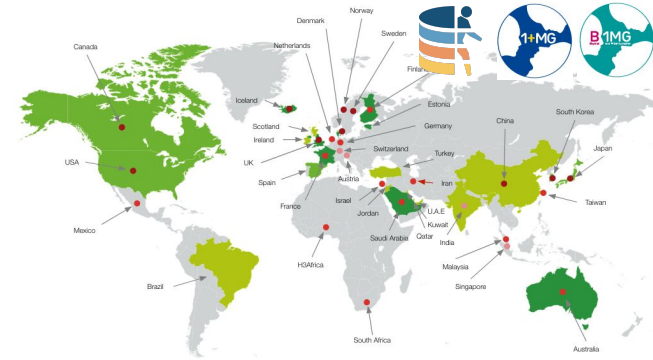
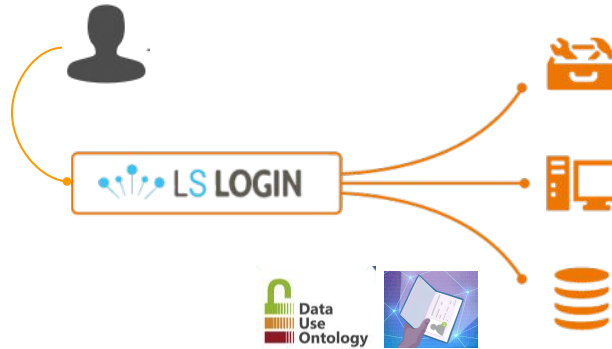
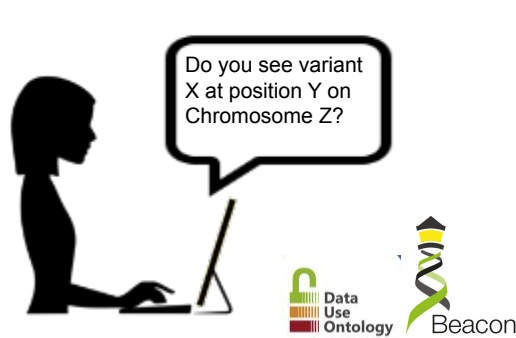


Federated Human Data initiatives in Europe



Building towards genomic data infrastructure

Solutions for managing controlled access data



Data discovery
 FAIR data, Beacon, DUO

Data access management
 TREs, Passports/Visas, Crypt4GH, DUO, AAI

Regional data protection regulations
 1+MG, Federated EGA, B1MG

How do I find data relevant to me?
 How do I know these data are useful?

How can I access data?
 How can I prevent unauthorised access?

What do I do if my data cannot be shared outside regional borders?

Thank you



Thanks to:

- EGA teams at EMBL-EBI and CRG
- Collaborators at ELIXIR and GA4GH
- Collaborators with Federated EGA



[@emlebi](#) [@EGAarchive](#) [@GA4GH](#) [@ELIXIREurope](#)



