Ready for BioData.pt Management?

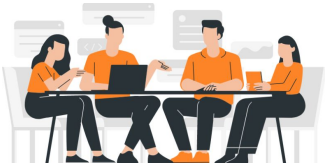
# Intensive Course

## Data Reuse

Daniel Faria, Jorge Oliveira, Gil Poiares-Oliveira

BioData.pt | elixir PORTUGAL

Ready for BioData.pt Management?



Intensive Course

Data Reuse

# I – Hands On

**Learning Outcomes:**

- Understand why data discovery is critical in personalised medicine.
- Identify relevant genomic and phenotypic datasets using:
  - Beacon Network
  - European Genome-phenome Archive (EGA)
- Interpret discovery-level metadata to evaluate dataset relevance.
- Understand access and governance implications of sensitive human data.

BioData.pt | elixir PORTUGAL

# Data Reuse in Personalised Medicine

In personalised medicine, researchers often need to answer:

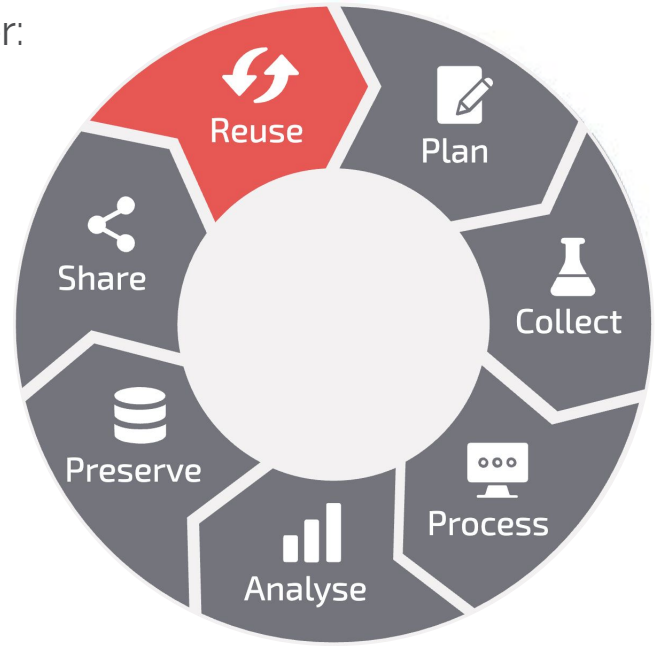
- Does any dataset exist that contains patients with a specific variant, disease, or phenotype?

However:

- Human genomic data cannot be openly shared.
- Data discovery must balance findability and privacy.

Two key tools addressing this:

- Beacon Network → Variant-centric discovery
- EGA → Dataset-centric discovery & controlled access



# Beacon

- Beacon is an API (sometimes extended with a user interface) that allows for **data discovery of genomic and phenoclinic data.**

- Allows researchers to get information about the **presence/absence of a given, specific, genomic mutation** in a set of data, from patients of a given disease or from the population in general



- Currently in version 2 that also includes:
    - More informative queries, like **filtering** by gender or age;
    - An option to trigger the next step in the **data access** process, e.g. who to contact or which are the data use conditions;
    - **Annotations** about the variants found, among which the expert/clinician conclusion about the pathogenicity of a given mutation in a given individual or its role in producing a given phenotype; Information about cohorts.

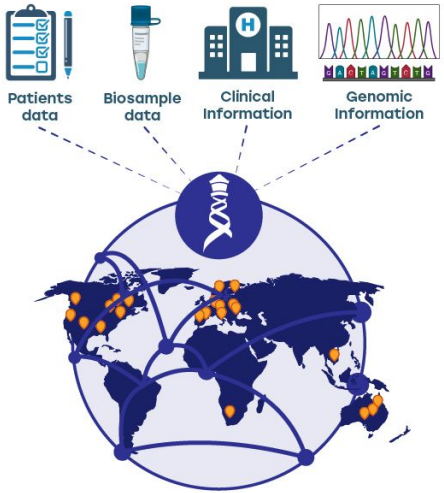
# Beacon

### 1. Create a query

Check for the presence of genomic variants, or look for detailed information through structured queries.


### 2. Beacon API search

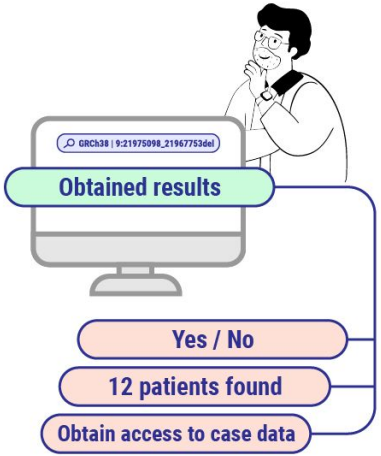
Beacon looks for the response, accommodating a wide range of data types, such as genomic and clinical data.




### 3. Get the response

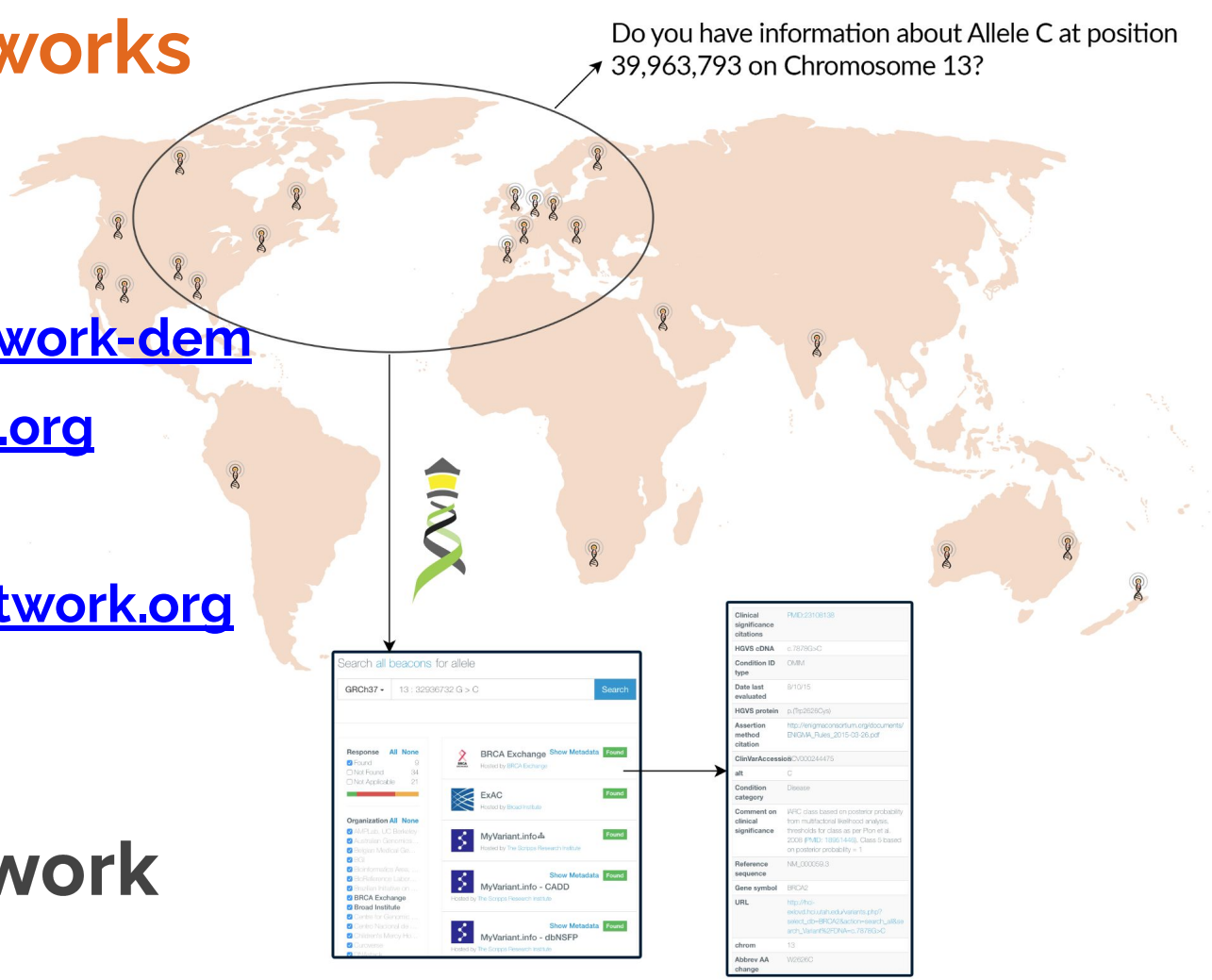
Get the results to the query:

- Boolean (yes/no).
- Count.
- Case-level data.



https://beacon.biodata.pt/

# Beacon Networks




https://beacon-network-demo.ega-archive.org

https://beacon-network.org

Beacon Network

EUROPEAN GENOME-PHENOME ARCHIVE

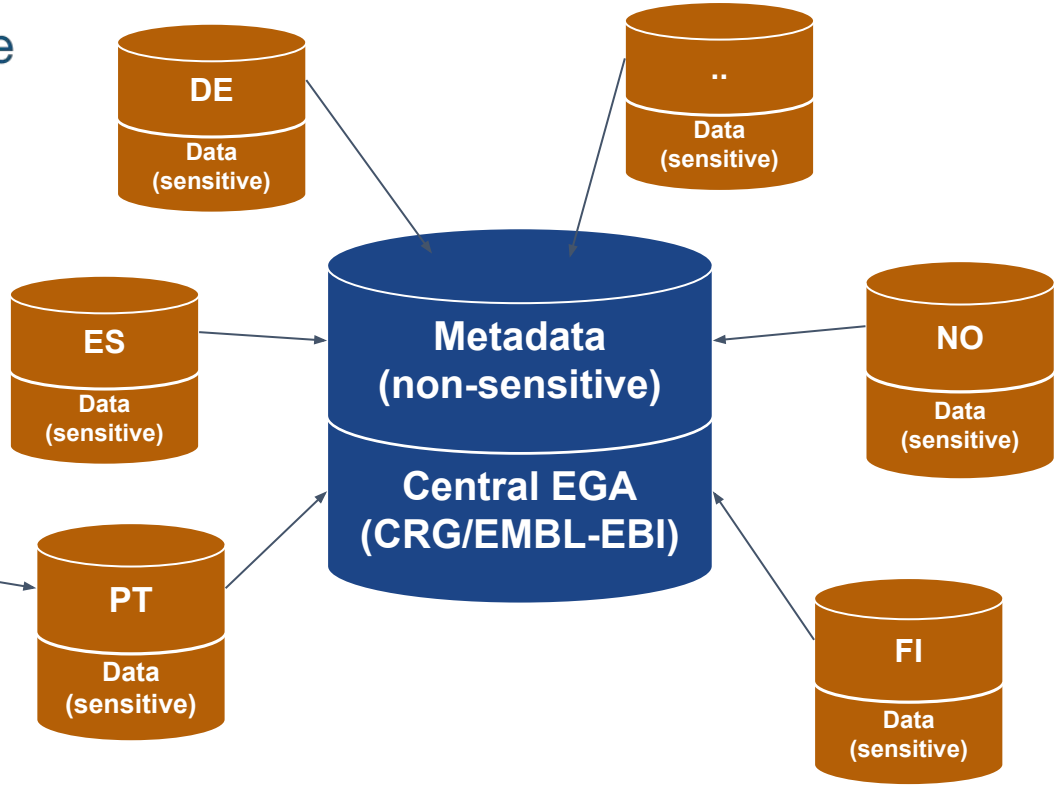
https://ega-archive.org/

- Service for 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.
- Currently federated into National nodes to meets the requirements of the General Data Protection Regulation (**GDPR**).
- Data submitted to the archive is subject to **controlled access**
    - Access to the data only will be granted after a formal application procedure.

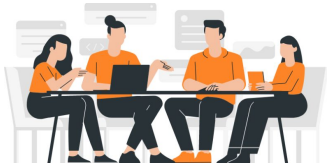


Federated European Genome-phenome Archive Portugal

Archive for **secure** storing and sharing **sensitive** data from Portuguese **research** projects.



Ready for BioData.pt Management?



# Intensive Course

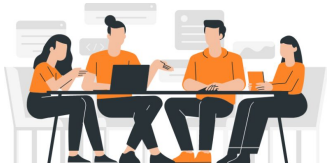
## Data Reuse

# Exercise

- Pick a disease of your choice
- Explore the Beacon networks and search for genomic variants
- Identify relevant datasets in EGA
- Assess their suitability for a research question


BioData.pt | elixir PORTUGAL

Ready for BioData.pt Management?



Intensive Course

Data Sharing & Reuse

# Thank You!

# Questions?

BioData.pt | elixir PORTUGAL