

The Portuguese Local European Genome-Phenome Archive (EGA)

Jorge S. Oliveira^{1,2}, Fernando Mira da Silva^{1,3}, José Luís Oliveira^{1,5}, Mário Silva^{1,3,4} and Ana M P Melo^{1,4}

1 - BioData.pt | ELIXIR-PT, 2 - Instituto Gulbenkian de Ciência, 3 - Instituto Superior Técnico, 4 - Universidade de Aveiro, 5 - INESC-ID



Summary

BioData.pt has been participating in EGA related projects since 2018, namely through the ELIXIR Beacon Community, Federated Human Data Community and ELIXIR-CONVERGE WP7. We have been early adopters of Beacon and Federated EGA.

All of these projects align with National Initiatives, namely the mirror groups of the Beyond One Million Genomes (B1MG) project and the Portuguese Initiative for Genomic Medicine National strategy, led by the National Institute of Health, with whom we have been collaborating closely, specially in the technical aspects of these projects.

Both in ELIXIR and B1MG we participated in pilot maturity/maturation models, aiming to access the level of participating nodes.

Local EGA & Beacon

Federated EGA will allow the distribution of Genomic metadata globally. The system includes a mechanism of communication between the central EGA and the Local nodes, allowing request of data access, that is consequently evaluated by a Data Access Committee (DAC).



The Beacon network consists in a distribution platform for genomic variants through an API in each node. We have an ambitious plan which consists in integrating the Beacon pipeline with the Local EGA ingestion system, thus automatically processing each new genome, generating its variants and inserting it in the Beacon database.

1 Implementation



	Local EGA	Beacon v2
Providing feedback as deployers	✓	✓
Deployed docker prototype of local EGA ingestion system	✓	
Connection to real central EGA	✓	
Performed live test of data (and metadata) upload	✓	
Currently testing components with synthetic data	✓	✓
Deploying and testing API	✓	✓
Integrating variant calling pipeline	✓	✓

2 GDI Project

BioData.pt, together with Instituto Superior Técnico and University of Aveiro participate in the Genomic Data Infrastructure (GDI) Project, where we committed to achieve the highest level of maturity (Operational) by the end of the project.

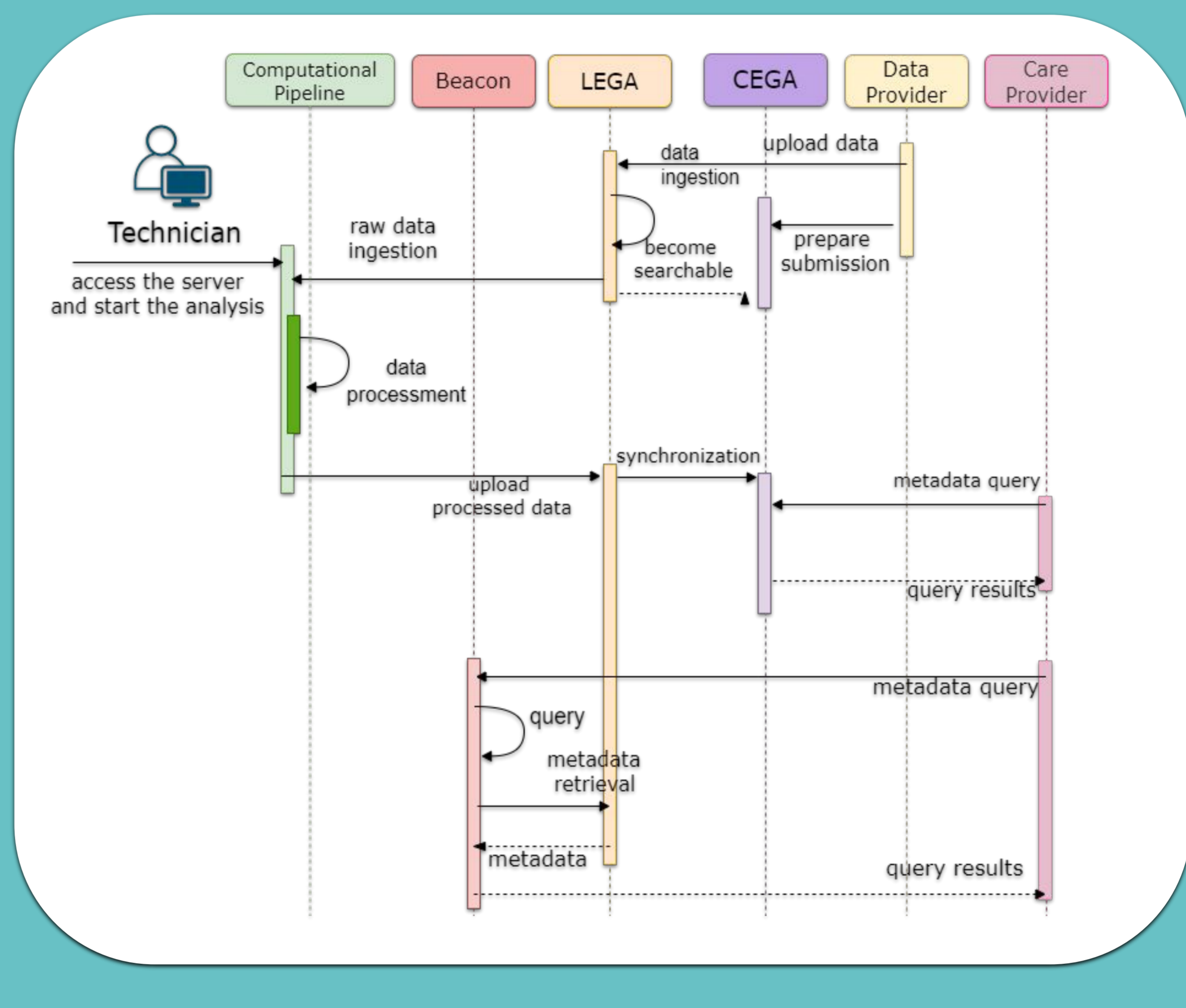
Our participation is in the following work packages:

- WP3 - Deployment of 1MG National Nodes
- WP4 - European Level Operations
- WP6 - Data Management

Data Flow-Chart

3

- Data is uploaded to the Local EGA server, that communicates with the Central EGA to accept the submission of the metadata.
- The LocalEGA system ingests the file, this includes the encryption and backup of the files.
- Data is processed, this includes the generation of genomic variants to be added to the Beacon v2 database
- Once the processing is done, all the information will become available for queries both in the EGA and Beacon websites.



4 Alignment with other Initiatives



- Portuguese Government has committed with **1 Million Genomes** declaration
- Portugal participates in **B1MG** led by INSA; Mário Silva chairs Data Interoperability and Sharing (WG5 / WP4) national mirror group.
- In discussions with the Ministry of Health Shared Services to set-up connection between their repositories and the localEGA foreseen to be managed by BioData.pt | ELIXIR-PT
- **Portuguese Initiative of Genomics Medicine** (national initiative)
- Genomics Data Infrastructure (GDI) European Project

Final Remarks

The implementation of a system to ensure secure access to genomics and the corresponding clinical data across Europe is essential for better research, personalised healthcare and health policy making. The complexity of such a system, abiding to all the legal aspects of GDPR and widely accepted across all members has been a remarkable effort, which ELIXIR-PT is proud of been part of. At the present moment, many initiatives and projects have aligned and converged into a solution that is ready to being implemented. All the knowledge and lessons acquired in this journey are going to be transformed into training material, in order to facilitate other nodes to implement their own systems.

Contacts

Name: Jorge Oliveira
E-mail: cto@biodata.pt
Website: www.biodata.pt

