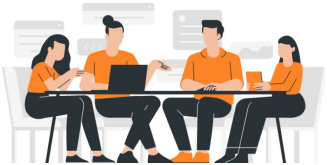
Ready for BioData.pt Management?

Intensive Course

Data Sharing & Reuse

Daniel Faria, Jorge Oliveira, Gil Poiares-Oliveira

BioData.pt | elixir PORTUGAL



I – Challenges

Learning Outcomes:

- Tackle the RDM challenges that arise in data sharing and reuse

Data Sharing & Reuse

- **Data Sharing** is the stage at which the only concern is making data available to others in a manner amenable to **Reuse**
  - We can (and often do) share data earlier in the life cycle with collaborators
    - The data is not final, but the challenges are identical

Data Sharing & Reuse

- **Data Reuse** is not really a stage
  - Making our data reusable should be a concern across the whole life cycle
  - Reusing existing data is a concern during the planning and collection stages
  - The challenges of these scenarios are two sides of the same coin

Data Sharing & Reuse

- **Challenges:**
  - Rights, obligations and restrictions
  - Accessibility
  - Selecting a repository
  - Licensing
  - Security
  - Documentation

Rights, Obligations & Restrictions

- **Rights:** are you the data owner or do you have consent from the data owner to share the data?
- **Obligations:** does your funder or institution require you to share the data?
- **Restrictions:** are there ethical, legal, contractual or IP reasons not to share the data

Accessibility

Type of Access	Who can access the data?	Metadata
Open Access	Anyone	Public
Registered Access	Authenticated users	Public
Controlled Access	Users approved by a Data Access Committee	Public
Access upon Request	Users approved by the data owner	???

Selecting a Repository

- Is there a discipline-specific repository available for your data?
- If not, is there an institutional or general purpose repository you could use?
- Does the repository support your desired type of access?
- Does it support the metadata standard for your domain?
  - Does it allow you to customize the metadata?

Licensing

- If you deposit your data in a public repository (open or registered access), you should add a license to it if able
  - Some repositories don't allow you to choose (data deposited there becomes public domain by default)
- For data under controlled access, a license usually does not make sense, as users will access the data under a contract

Licensing

Licence	Can I copy & redistribute the work?	Is it required to attribute the author?	Can I use the work commercially?	Am I allowed to adapt the work?	Can I change the licence when redistributing?
CC0	✔	✘	✔	✔	✔
CC BY	✔	✔	✔	✔	✔
CC BY-SA	✔	✔	✔	✔	✘
CC BY-ND	✔	✔	✔	✘	✔
CC BY-NC	✔	✔	✘	✔	✔
CC BY-NC-SA	✔	✔	✘	✔	✘
CC BY-NC-ND	✔	✔	✘	✘	✔

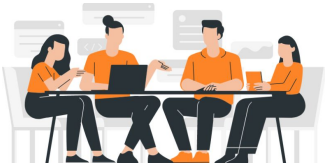
Security

- Are the security practices of a data repository adequate to your needs?
  - Does it implement an established authentication and authorization protocol?
  - Does it encrypt the data?
  - Does it have a reasonable backup policy? (this can be hard to check)

Documentation

- Documentation is again one of the most critical concerns when sharing data, as reusability hinges almost entirely on the quality of your documentation
- Fortunately, if we followed good documentation practices throughout the data life cycle, we don't need to do much at this stage
  - We can export and publish our experiments from the ELN
  - We can publish our data processing and analysis workflow

Ready for BioData.pt Management?



Intensive Course

Data Sharing & Reuse

II – Hands-On

Learning Outcomes:

- Share sensitive data on a data repository

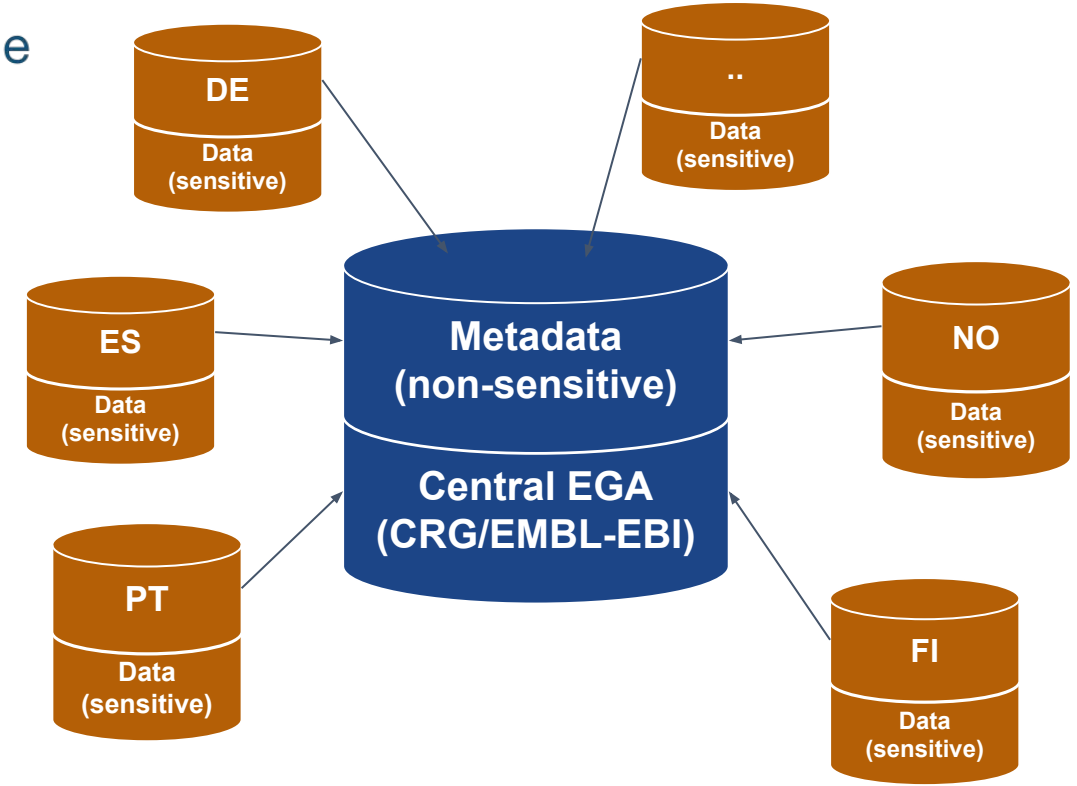
BioData.pt | elixir PORTUGAL

Federated European Genome-phenome Archive Portugal

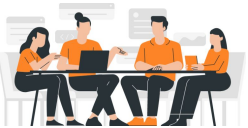
https://fegaportugal.biodata.pt/

- Archive for **secure** storing and sharing **sensitive** data from Portuguese **research** projects.
- National node of the Federated European Genome-phenome Archive (FEGA).
- 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

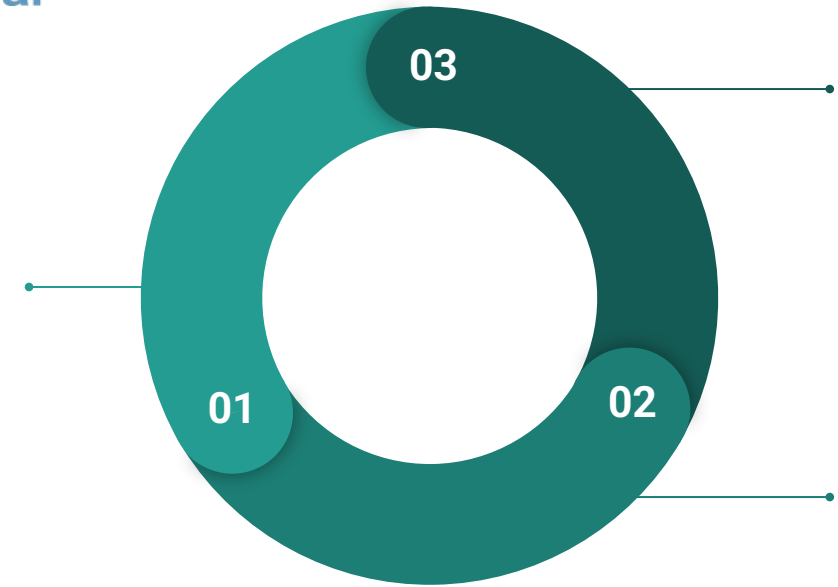


Ready for BioData.pt Management?



Federated European Genome-phenome Archive Portugal

Group Exercise



Upload sensitive data

Prepare your data by encrypting it and upload to the FEGA PT servers

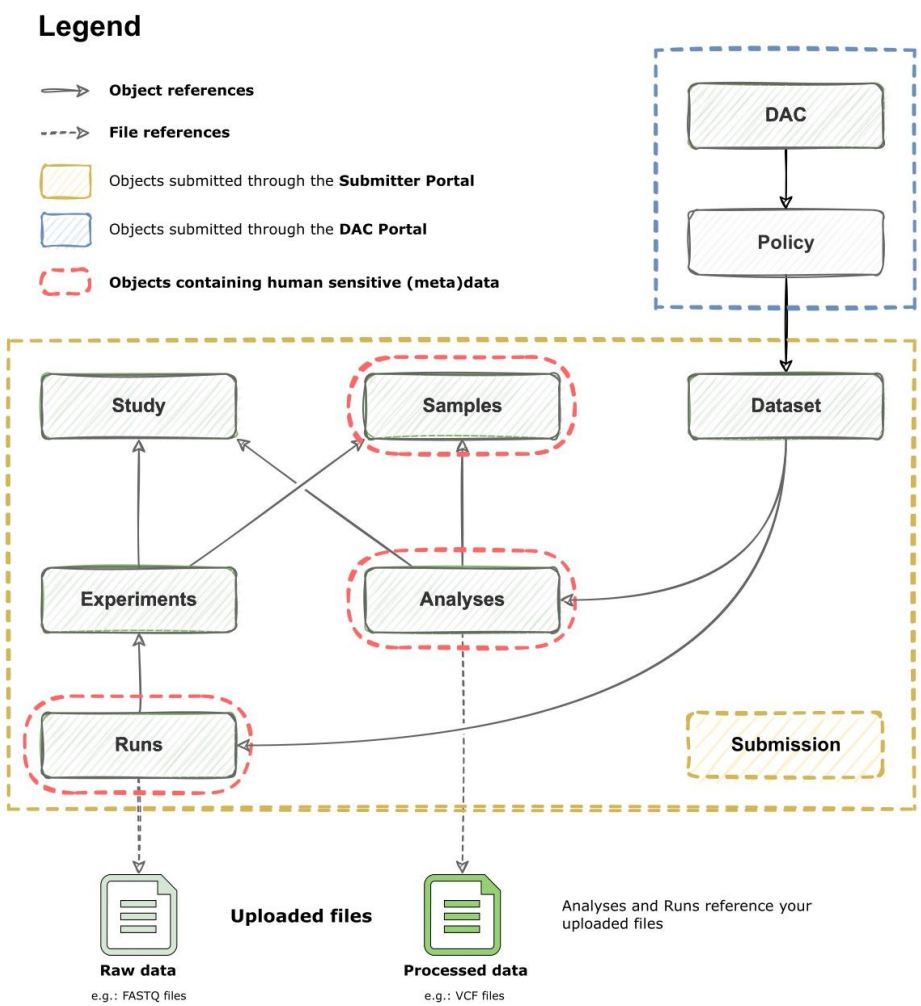
Submit metadata

Prepare and submit your data into the submitter portal of FEGA

Create DAC & Policies

Create a Data Access Committee and Data Policies

Ready for BioData.pt Management?


Legend
Object references
File references
Objects submitted through the Submitter Portal
Objects submitted through the DAC Portal
Objects containing human sensitive (meta)data
DAC
Policy
Study
Samples
Dataset
Experiments
Analyses
Runs
Submission
Uploaded files
Raw data
e.g.: FASTQ files
Processed data
e.g.: VCF files
Analyses and Runs reference your uploaded files

Group Exercise

Part I

https://submission-test.portugal.ega-archive.org/

username: trainingbiodata

pass: FAIR.123

1. Create a mock submission on FEGA PT to deposit the genomic data of this study (Check Data Availability Section)
2. If it helps, check the EGA metadata schema

Group Exercise

Part II

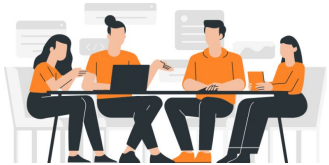
1. Access the Data Access Committee Portal
2. https://dac.test.ega-archive.org/
3.
4. Create a Data Access Committee (DAC)
5. Create a Data Policy and associate it with your DAC
6. Associate them with your EGA submission

Group Exercise

Part III

1. Access the [FEGA Portugal website]
2. Explore the site and learn how to prepare your data to upload to the Local EGA Servers
3. **Do not upload the files** (this is only an exercise)

Ready for BioData.pt Management?



Intensive Course

Data Sharing & Reuse

Thank You!

Questions?

BioData.pt | elixir PORTUGAL