

BioData.pt



BIODATA.PT

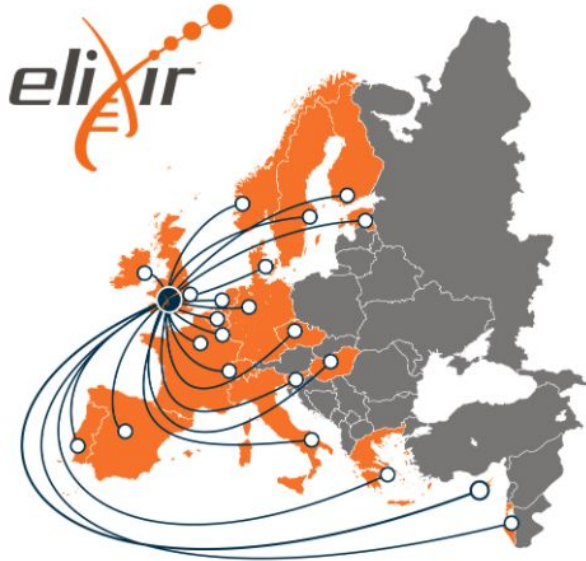
INFRAESTRUTURA PORTUGUESA PARA A INFORMAÇÃO BIOLÓGICA

Ana Portugal Melo e José Pereira Leal

Infraestruturas de investigação – Práticas, políticas e serviços para a gestão de dados e FAIR data
Fórum GDI 2020

Co-financiado por:





- O nó Português do ELIXIR: European infrastructure for biological information
- Uma rede nacional multidisciplinar de peritos em bioinformática e gestão avançada de dados
- A plataforma nacional de dados biológicos



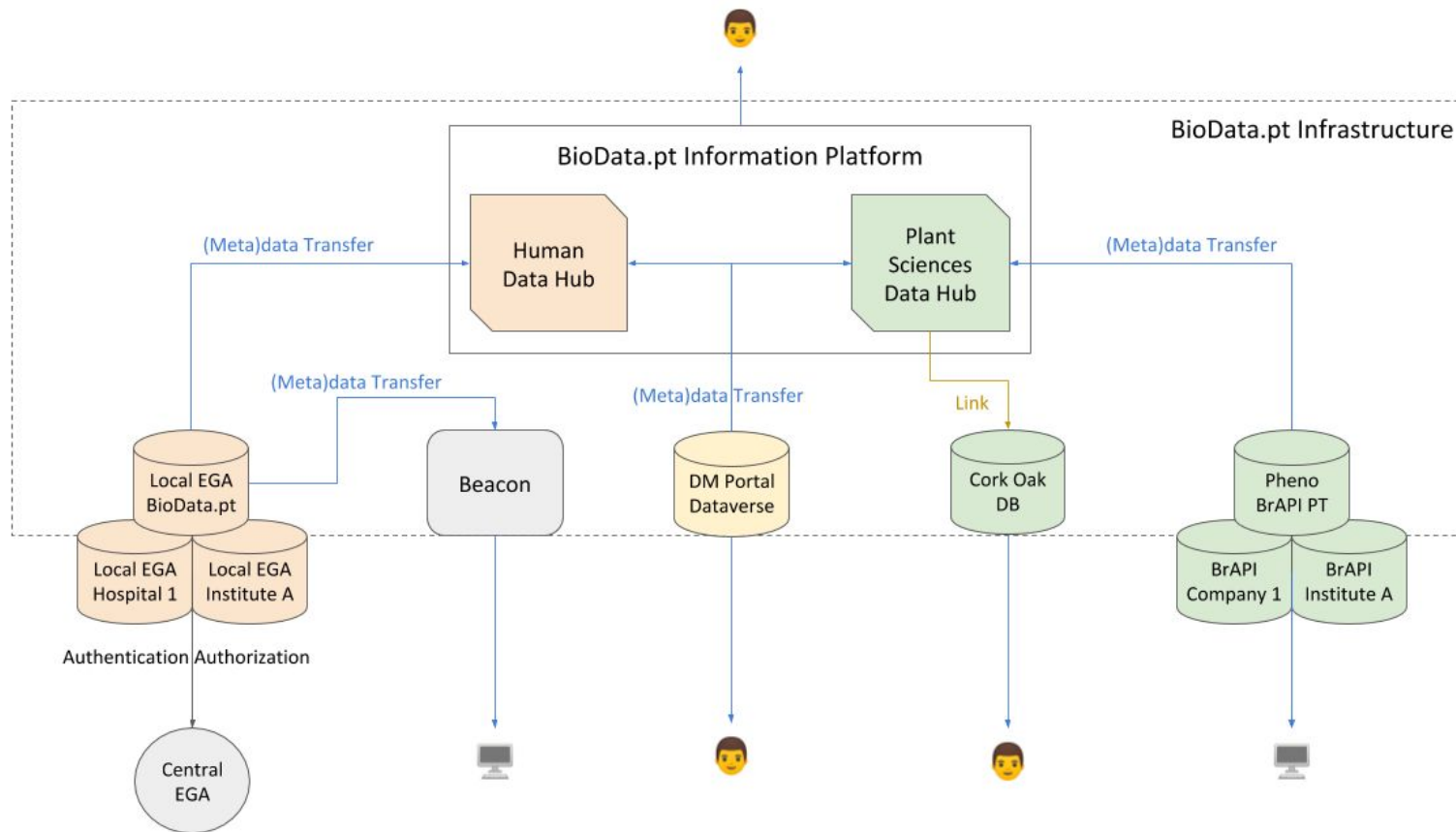
SOBRE O BIODATA.PT

O BIODATA.PT E A GESTÃO DE DADOS

1. O BioData.pt não gera dados. É uma plataforma nacional que agrega os dados da investigação biológica.
2. O BioData.pt apoia a comunidade científica nacional das ciências da vida e da saúde na gestão dos seus dados.
3. O BioData.pt tem parceiros, que geram dados, e estão organizados em comunidades de domínio específico que, ao nível do ELIXIR, definem critérios de organização desses dados.

1.

PLATAFORMA BIODATA.PT PARA A INFORMAÇÃO BIOLÓGICA



2.

PORTAL BIODATA.PT PARA A GESTÃO DE DADOS

BioData.pt Add Data ▾ Search ▾ User Guide Support Sign Up Log In

[Home](#) (biodata.pt)

Metrics 0 Downloads Contact Share

The main Dataverse, in which all others are hosted.

Search this dataverse... Find [Advanced Search](#)

Dataverses (2)
 Datasets (0)
 Files (0)





Dataverse Category
Organization or Institution (2)

Metadata Source
[Home](#) (1)
[Root](#) (1)

Publication Year
2020 (2)

Subject
[Medicine, Health and Life Sciences](#) (1)

1 to 2 of 2 Results Sort ▾

	Human BioDataVerse (biodata.pt) Jun 25, 2020 BioData.pt Human Data Community.	
	Plant BioDataVerse (biodata.pt) Jun 25, 2020 BioData.pt Plant Sciences Community Dataverse.	

2.

SERVIÇOS BIODATA.PT PARA A GESTÃO DE DADOS

BioData.pt Service Hub

Data Solutions for the Life Sciences



BioData.pt 



Home	Services	How to Order	Forum	Terms of Service
------	----------	--------------	-------	------------------

Data Management Services



Project Data Management

We carry out data management planning and data management activities of research projects or project applications, in a package that can include any of the following components, as required:

- Data Management Plan draft/consulting
- Data storage (with access to Data Management Portal)
- Data curation/annotation
- Data Management Plan execution (e.g. data publication after project is complete)
- Training

Request through [Agendo](#)

Price: budget on request



Data Management Portal

An online platform for data management and publication.

You can access the portal [here](#) and register for free.

Request [additional](#) storage or user support through [Agendo](#)

Price: budget on request



Ready for BioData Management

Capacity building programme to empower researchers in Data Management, which includes a range of training offers:

Lectures

- Introduction to Data Management in Science
- Demystifying Data Management Plans

Workshops

- Introduction to Data Management Plans

Courses

- Advanced Data Management Plans
- Day-to-day Data Management

You can consult the [Ready for BioData Management website](#) for more information.

3. AS COMUNIDADES BIODATA E A GESTÃO DE DADOS

- **Plantas** (a maior contribuição de Portugal para o ELIXIR até ao presente) - ITQB, IBET, INESC-ID, FCUL, CEBAL, IGC
- **Microbiologia e Biotecnologia** - ITQB, IBB, UMinho
- **Microbioma** - IGC, CCMAR, IST
- **Dados Humanos** - IGC, INESC-ID

3. AS COMUNIDADES BIODATA E A GESTÃO DE DADOS

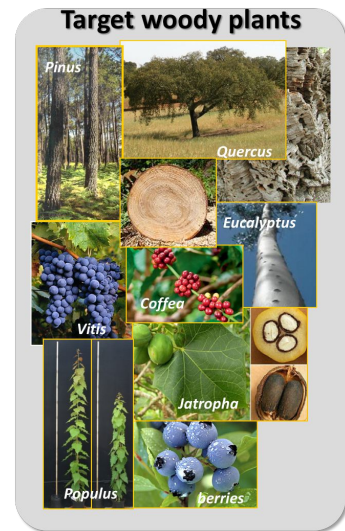
O EXEMPLO DA COMUNIDADE DAS PLANTAS

Problema

Dados fenotípicos heterogéneos, dispersos, pobremente anotados e desligados dos dados genotípicos, e não FAIR.

Solução

Construir uma infraestrutura para suportar a publicação e acesso a dados fenotípicos e genotípicos, anotada de acordo com padrões estabelecidos.



3. AS COMUNIDADES BIODATA E A GESTÃO DE DADOS

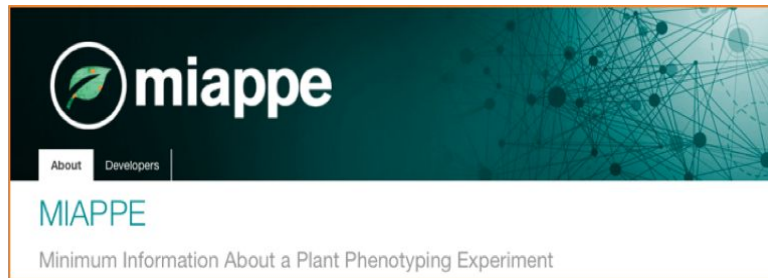
A COMUNIDADE DAS PLANTAS

- Desenvolve **padrões e ontologias** para a **integração dos dados das plantas**.
- Faz **curadoria e integração dos dados das plantas** através de níveis de organização numa única plataforma de dados.
- Prepara **datasets e ferramentas específicas de análise para a comunidade das plantas**.
- Desenvolve ferramentas de anotação de dados e metadados com **interfaces de utilização fácil**.
- **Interage com a indústria** na definição de prioridades de anotação genótipo/fenótipo e demonstrando os recursos desenvolvidos.
- Colabora com a **plataforma de treino** para o treino nos recursos desenvolvidos.

3. AS COMUNIDADES BIODATA E A GESTÃO DE DADOS

MIAPPE

- Extensão do âmbito para acomodar o caso de uso das plantas lenhosas.
- Compatibilidade com frameworks comuns para representar estes tipos de dados, ISA-Tools e o Breeding API (BrAPI).
- Enriquecimento com a disponibilização de definições claras e exemplos para todos os campos, bem como de um modelo de dados explícito.



line #	MIAPPE Check list	Definition	Example	Format	Cardinality
DM-1	Investigation	Investigations are research programmes with defined aims. They can exist at various scales (for example, they could encompass a grant-funded programme of work, the various components comprising a peer-reviewed publication, or a single experiment).			1 per MIAPPE submission
DM-2	Investigation unique ID	Identifier comprising the unique name of the institution/database hosting the submission of the investigation data, and the accession number of the investigation in that institution.	EBI-12345678	Unique identifier	0-1
	Investigation title	Human-readable string summarising the investigation	Adaptation of Maize to Temperate Climates: Md-Density Genome-Wide Association Genetics and Diversity Patterns Reveal Key Genomic Regions, with	Free text (short)	1
Environment					
ENV-1	Non-exhaustive list of Environment Parameters.				
Environment parameters					
Environment parameters		Definition	Example	Format	
Growth facility					
ENV-3	Air temperature	List of hourly air temperature throughout the experiment.	22 °C	Numeric	
ENV-4	Organ temperature	List of hourly organ temperatures throughout the experiment.	18 °C	Numeric	
ENV-5					
Experimental Factors					
TR-1	Non-exhaustive list of Experimental Factors that can be applied.				
Experimental Factors					
Factor type		Definition	Example factor values	Format	
TR-2	Seasonal environment	A plant treatment (EO:0001001) involving an exposure to a given conditions of regional seasons.	Spring season; dry season	Plant Environment Ontology; 'EO_0007038'	
TR-3	Air treatment regime	The treatment involving an exposure to wind/air with varying degree of temperature, which may depend on the study type or the regional environment.	28/25°C (Day/Night)	Plant Environment Ontology; 'EO_0007161'	
TR-4	Soil temperature regime	A physical plant treatment (EO:0007316) involving an exposure to varying degree of temperature, which may depend on regional environment.	27/25°C (Day/Night)	Plant Environment Ontology; 'EO_0007161'	
TR-5					

3. AS COMUNIDADES BIODATA E A GESTÃO DE DADOS

A COMUNIDADE DOS DADOS HUMANOS

Problema: Acesso federado seguro inter-fronteiras a dados clínicos personalizados

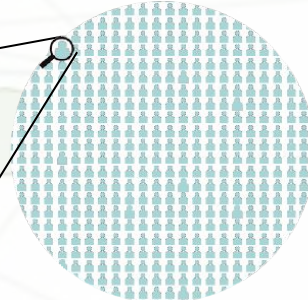
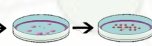
Data (genotype, phenotype, health care, longitudinal)
Scale

Scope
Biospecimens
Patient recall

Access/regulations

Informatics

Interoperability
Clinical data



Patient

Clinician

Researcher



3. AS COMUNIDADES BIODATA E A GESTÃO DE DADOS

A COMUNIDADE DOS DADOS HUMANOS

	Necessary minimal infrastructure component	In development ^a	Implemented at scale ^a
Padrões	Genomics data and clinical information standards geared towards specific disease communities	Yes	No
APIs	Common application programming interfaces to enable remote data discovery and access	Yes	Yes
Clouds federadas, seguras	Computational resources, including secure, federated cloud computing environments that offer secure access across national boundaries to raw data and interoperable results	Yes	Yes
Ferramentas, serviços & workflows	Regulatory frameworks that enable access to and the processing of genomic data across borders, including the management of transnational user access and compliance	Yes	No
	A repository of tools and services, including workflows to analyse deposited data while enabling these analysis workflows to operate on data across national borders. This will contribute towards data reproducibility and provenance, which are of high importance in both research and clinical practices	Yes	Yes
	A training and capacity-building programme to develop the skills and workforce required for genomics and big data in health care as well as shift the culture towards openness and integration of research data across national boundaries	Yes	Yes

^a"In development" and "Implemented at scale" refer to locally defined status within ELIXIR and/or Biobanking and Biomolecular Resources Research Infrastructure.

Enquadramentos legais e regulatórios

Treino

Programa de desenvolvimento de competências

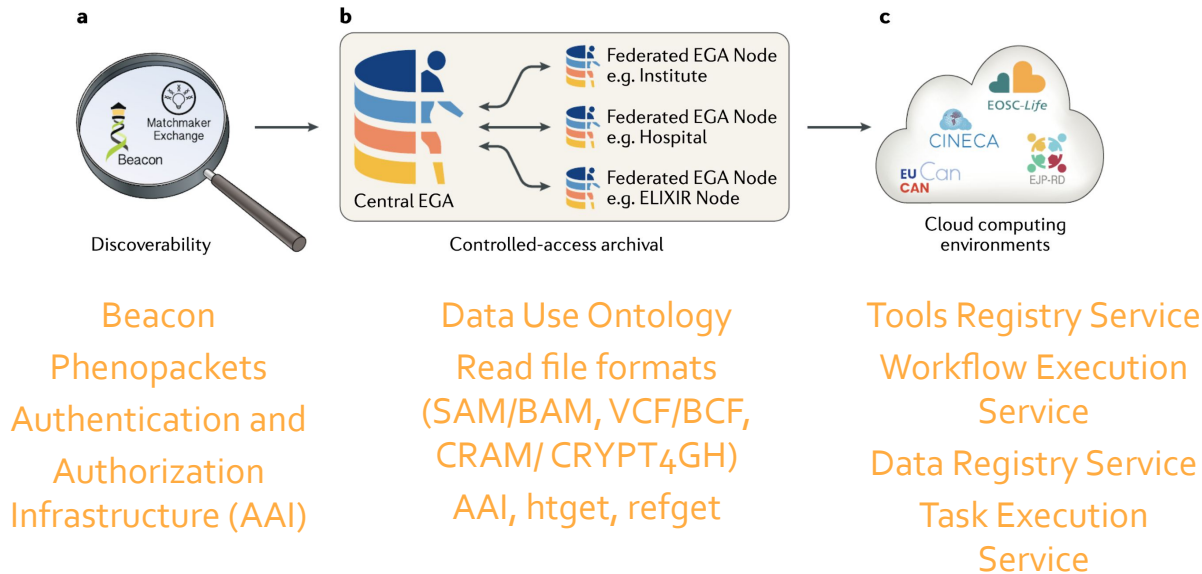
Recomendações mínimas para a infraestrutura Europeia aceder e analisar dados genómicos.

3. AS COMUNIDADES BIODATA E A GESTÃO DE DADOS

A COMUNIDADE DOS DADOS HUMANOS

Desenvolvimento de padrões e condições de utilização

Incentivar Interoperabilidade usando os padrões GA4GH



Take Home Messages

Giram os vossos dados!

- Dados abertos e FAIR são o caminho.

Open and FAIR is money!

- O custo estimado por não termos dados de investigação FAIR, na Europa, é de EUR 10.2 billion, por ano.

As coisas podem ficar mais fáceis!

- Há pessoas a trabalhar em soluções de gestão de dados de utilização fácil — elas precisam de feedback sobre o que pode ser melhorado.

Nós fazemos acontecer!

