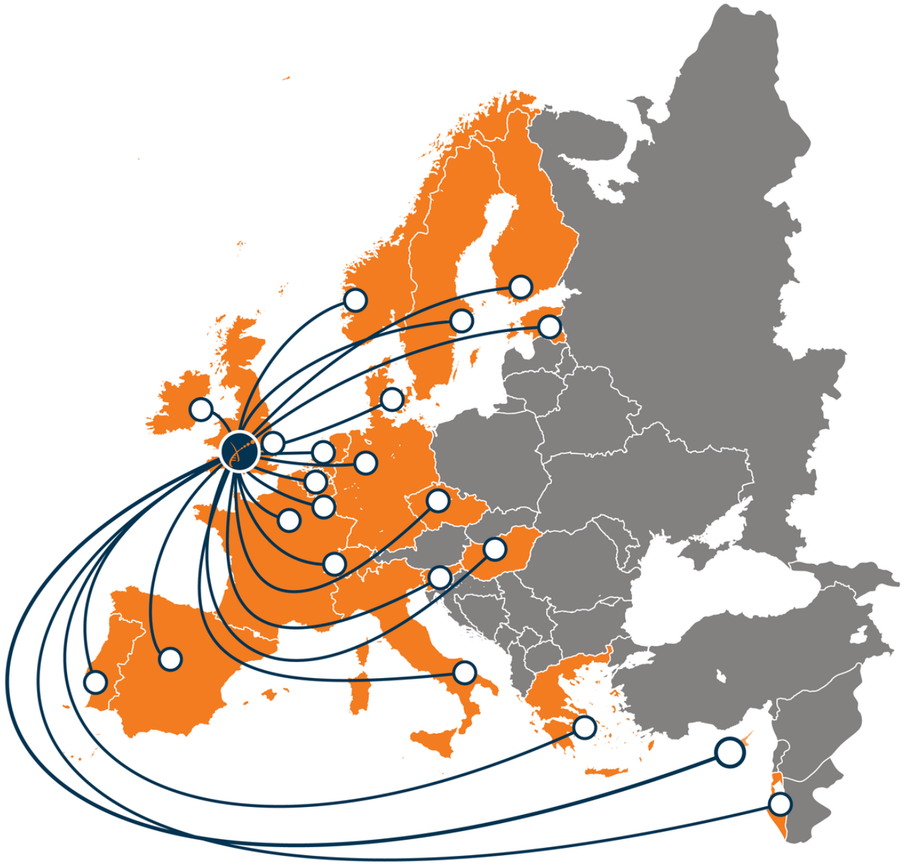
ELIXIR PT & IMPACT

Ana Portugal Melo

ELIXIR Workshop “Impact Assessment at the National Node Level”

Co-financiado por:

INESC-ID, Lisboa, Portugal. 30th January 2020



- The Portuguese distributed e-infrastructure of biological data
- A national expert multidisciplinary network in bioinformatics and advanced data management
- The Portuguese Node of the European network of experts and resources for biological data- ELIXIR

ABOUT BIODATA.PT

WHAT'S IN IT FOR YOU

BioData.pt elixir PORTUGAL

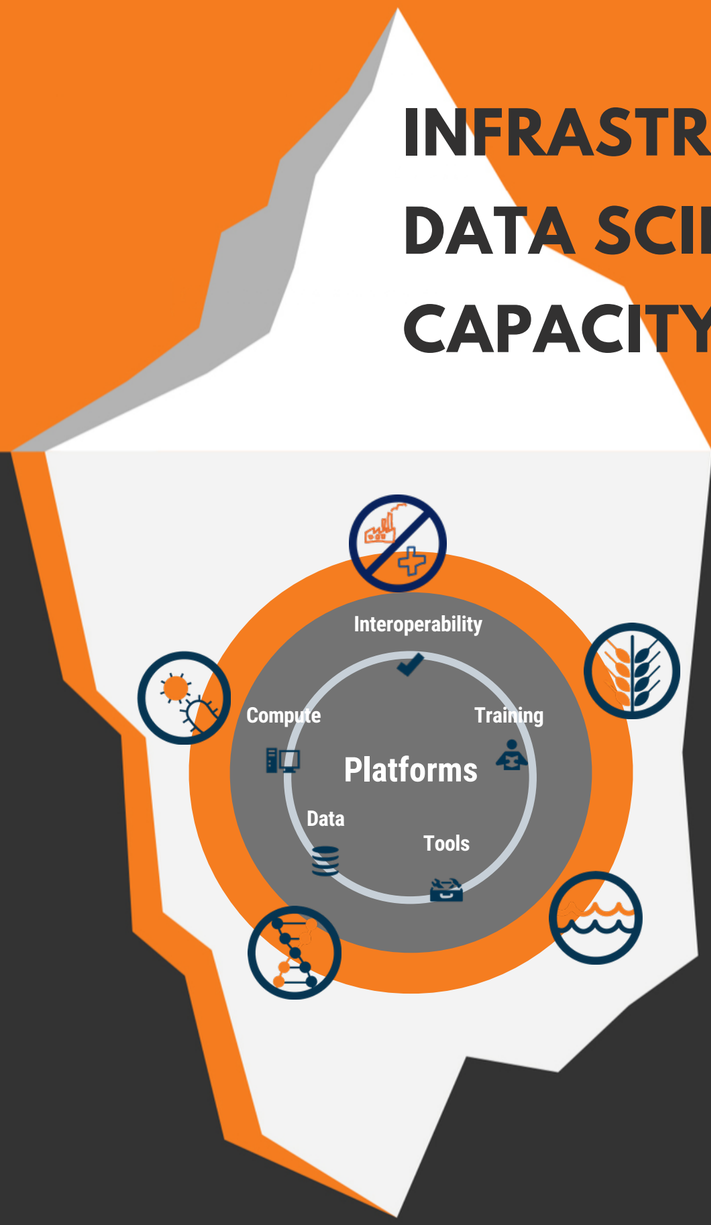
CREATING VALUE

CREATING KNOWLEDGE

INFRASTRUCTURE

DATA SCIENCE

CAPACITY BUILDING



BioData.pt

elixir PORTUGAL

SERVICES

UNLEASHING YOUR CAPACITY

Data Analysis

Data Management

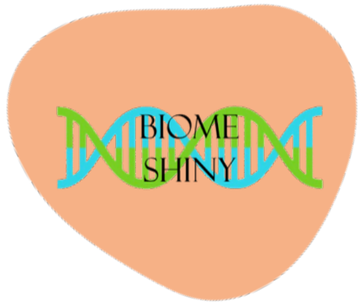
Computing

Consulting and Training

BioData.pt elixir PORTUGAL

DEVELOPMENT

TO EMPOWER THE RESEARCHER



Biome-Shiny
(Crash-course)

PVA

Python Video Annotator

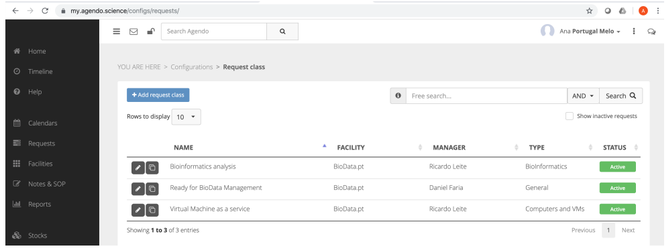
D-Cell

D-Cellerate
(Crash-course)

Research Problem:
Analyse mutations
Need computing power

BioData.pt

Request @ Agendo



Virtual Machine as a service

VIRTUAL MACHINES

TO EMPOWER THE RESEARCHER

Predicted mutations																		
seq id	position	mutation	D84C1	D84C2	D84C3	D84C4	D84C5	D84C6	D84C7	D84C8	D84C9	D84C10	D84C11	D84C12	D84C13	D84C14	annotation	gene
1	1,498	G→T						5.3%									intergenic (-305/-35)	*CABENIGC_00001* ← / → *CABENIGC_000*
1	1,507	C→A		6.9%	9.1%	12.3%		7.4%									intergenic (-314/-26)	*CABENIGC_00001* ← / → *CABENIGC_000*
1	1,536	G→T					6.5%										E2* (GAA→TAA)	*CABENIGC_00002* →
1	1,555	G→T					8.2%										R8I (AGA→ATA)	*CABENIGC_00002* →
1	1,567	T→A				5.6%											L12* (TTA→TAA)	*CABENIGC_00002* →
1	1,587	G→T			18.8%												D19Y (GAT→TAT)	*CABENIGC_00002* →
1	29,414	C→G		10.9%													intergenic (+144/-261)	*CABENIGC_00025* → / → *CABENIGC_000*
1	29,426	G→T	13.2%	25.9%	26.6%	16.6%	14.5%	25.5%	9.5%	8.9%	12.1%	6.5%	13.5%	11.7%	14.3%	11.2%	intergenic (+156/-249)	*CABENIGC_00025* → / → *CABENIGC_000*
1	29,459	C→A			13.3%												intergenic (+189/-216)	*CABENIGC_00025* → / → *CABENIGC_000*
1	74,873	C→T	39.6%	43.7%	43.3%	42.3%	39.8%	44.6%	45.2%	45.8%	44.7%	40.3%	48.1%	46.4%	42.9%	42.8%	L423L (CTG→TTG)	*CABENIGC_00066* →
seq id	position	mutation	D84C1	D84C2	D84C3	D84C4	D84C5	D84C6	D84C7	D84C8	D84C9	D84C10	D84C11	D84C12	D84C13	D84C14	annotation	gene
1	74,911	T→C	29.8%	33.0%	32.3%	32.8%	33.7%	35.3%	36.7%	32.6%	31.2%	26.9%	34.4%	35.3%	29.2%	35.4%	I435I (ATT→ATC)	*CABENIGC_00066* →
1	74,956	G→A	19.7%	19.5%	21.7%	14.9%	20.8%	25.5%	22.5%	18.7%	15.6%	11.6%	20.4%	20.2%	17.3%	19.0%	T450T (ACG→ACA)	*CABENIGC_00066* →
1	75,734	A→G	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		100%	100%	100%	A200A (GCA→GCG)	*CABENIGC_00067* →
1	75,746	T→C	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		100%	100%	100%	G204G (GGT→GGC)	*CABENIGC_00067* →
1	75,785	G→T	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		100%	100%	100%	T217T (ACG→ACT)	*CABENIGC_00067* →
1	75,848	C→T	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		100%	100%	100%	T238T (ACC→ACT)	*CABENIGC_00067* →
1	75,860	G→A	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		100%	100%	100%	Q242Q (CAG→CAA)	*CABENIGC_00067* →
1	75,869	G→A	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%		100%	100%	100%	K245K (AAG→AAA)	*CABENIGC_00067* →

BioData.pt elixir PORTUGAL

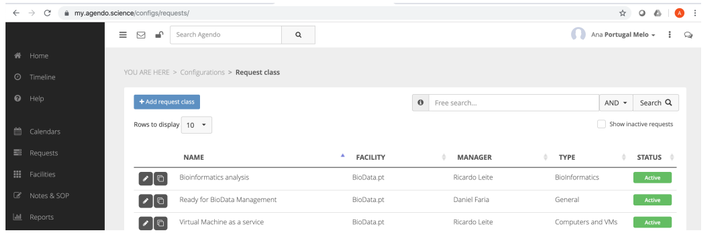
BIOINFORMATICS CONSULTING

TO EMPOWER THE RESEARCHER

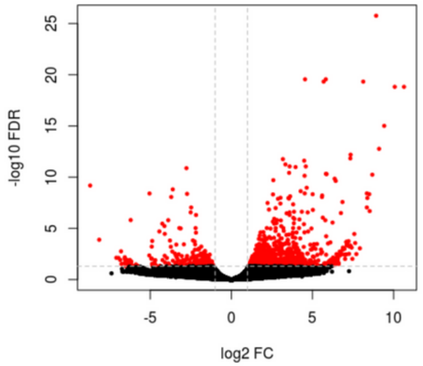
Research Problem:
Analyse big amounts of omics data
Need bioinformatics expertise

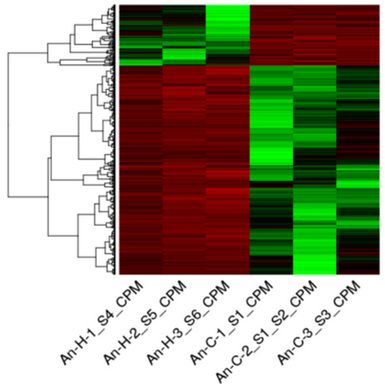
BioData.pt

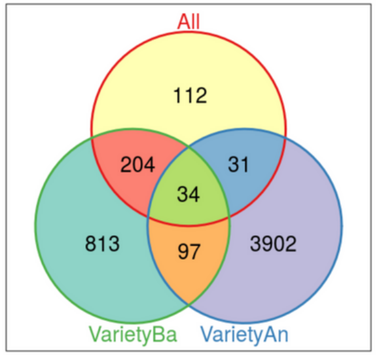
Request @ Agendo



Bioinformatics analysis







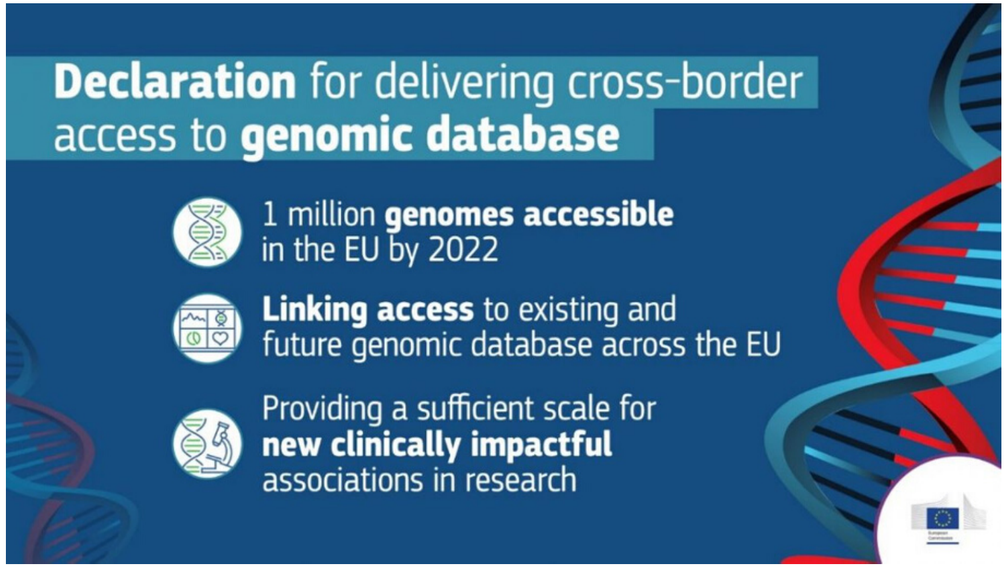
BioData.pt elixir PORTUGAL

1MG

One Million Genomes

FLAGSHIP PROJECTS

TO BE ENGAGED WITH YOU



Deploy Genome Annotation Workflow to BioData.pt Cloud

Set up a national Local EGA end-point to host genomic data

BioData.pt elixir PORTUGAL

FLAGSHIP PROJECTS

TO BE ENGAGED WITH YOU

CONVERGE - A Toolkit for Data Management

Connect and align ELIXIR Nodes to deliver sustainable FAIR life-science data management services

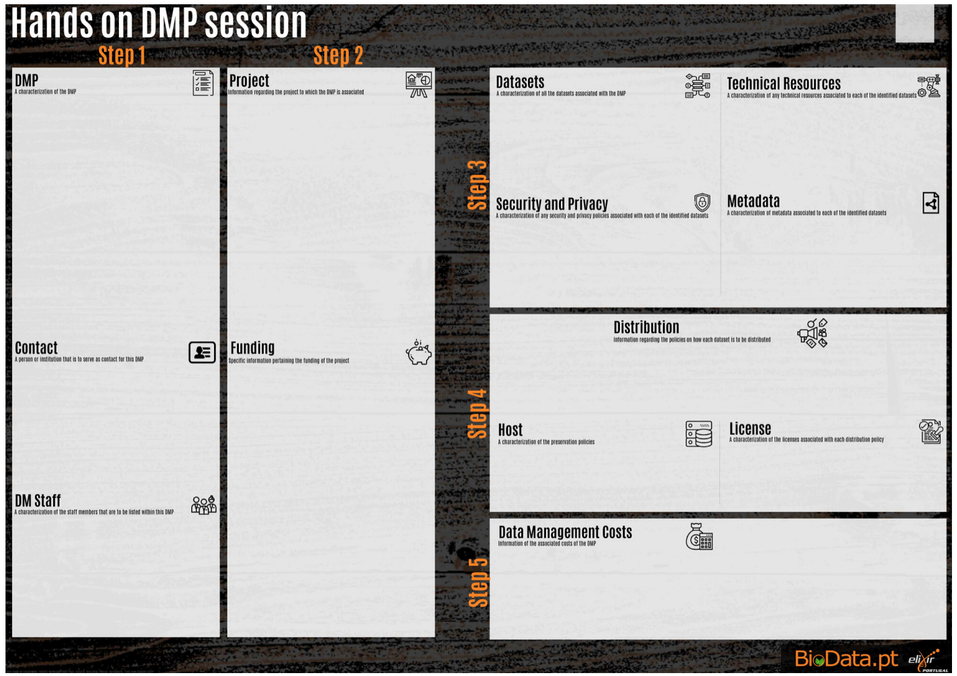
Develop, identify and promote key data management tools and resources for enabling FAIR data in key domains such as the plant sciences

BioData.pt elixir PORTUGAL

CAPACITY BUILDING

SHARING A VISION WITH YOU

Ready for BioData Management?



Intro level workshop

DMP course

Seminar

BioData.pt elixir PORTUGAL

REPORT

TO GET YOUR ATTENTION

REPORT 2017–2019

BioData.pt

THE PORTUGUESE INFRASTRUCTURE FOR BIOLOGICAL DATA

HIGHLIGHTS

June 2017-September 2019

COURSES

To be at the forefront of Science with Bioinformatics expertise.

29
386 participants

18

INTERNATIONAL PUBLICATIONS

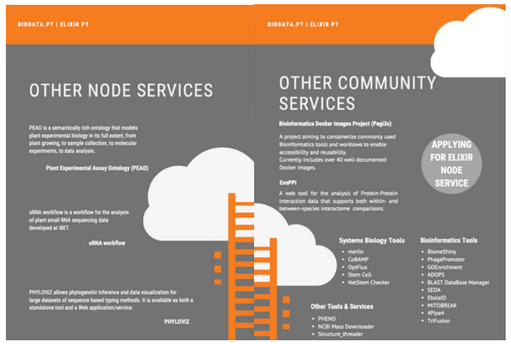
Achieving international recognition.

EVENTS

Sharing a vision towards value creation from biological data management and innovative tools.

4
196 participants

51 Communications	**7** Academic Thesis	**1** New Projects	**1** Communication Awards
17 Computing Applications	**18** Models	**3** Compute Infrastructure Nodes	**7000+** New Website Users



ELIXIR NODE SERVICE

CorkOakDB aims to integrate the knowledge generated from fundamental and applied studies about *Quercus suber, with a focus on genetics.*

TOOLS

- BLAST
- Genome Browser
- Feature Expression Visualization
- Data Searches
- Direct Downloads

Apr - Dec 2019

337 Users	7 Countries

CorkOakDB features the first draft genome of Quercus suber, released in 2018 by the GENOSUBER consortium, and allows genome browsing and gene search.
It also incorporates other types of data from cork oak scientific research, including gene expression data from publicly available datasets.

The knowledge of the genetic structure of cork oak is essential for the future development of innovative breeding and production strategies.
From here on it will be possible to identify and study genes involved in the acquisition of traits of interest, such as cork production or resistance to pests. The genome sequence now available will also serve as a reference for sequencing other trees and identifying genetic
variability related to characteristics of interest. This portal is intended to become a repository of data originated by scientific research in multiple areas.

ELIXIR NODE SERVICE

ELIXIR PT COMPUTING SERVICES

Coordinated by INESC-ID, the BioData.pt computing infrastructure includes IST, IGC and CCMAR as cloud service providers. The service provides virtual machines on demand, complemented by a user-support service.
Additionally, the infrastructure provides computing capacity to Galaxy Europe through the Pulsar network, which enables Portuguese researchers to use the infrastructure when submitting jobs to Galaxy Europe.

Report Period

785 vCPUs	1.8 TB RAM	156 TB Storage	321 Active VMs

PYTHON VIDEO ANNOTATOR

APPLYING FOR ELIXIR NODE SERVICE

Credits: NASA/Dominic Hart

An open source graphical application for analyzing and annotating events captured in video, PythonVideoAnnotator was developed in concert with a video-capturing rig, with the aim of helping neuroscience and ethology researchers study animal behaviors.
Ultimately, the goal of this software tool and supporting set-up is to bring Open Science practices to the study of animal behavior and promote reproducibility and standardization.
During the period of this report, PVA was **downloaded more than 2700 times**.

BioData.pt elixir PORTUGAL

@BioData_PT

@BioData.PT

@BioData-PT

www.BioData.pt

THANK YOU

STAY TUNED