



# Poster session for the Scientific Advisory Board assessment of BioData.pt | ELIXIR PT

## Abstracts

### 1 - Microbial Biotechnology and Systems Biology Community

merlin is a simple, graphical and user-oriented solution for the reconstruction of genome-scale metabolic models. It will guide you along the model reconstruction, providing several tools that help to improve and curate the model throughout the whole process. The Transport Systems Tracker (TranSyT) is a tool to identify transport systems and the compounds carried across membranes, based on the annotations of the Transporter Classification Database (TCDB). BioISO is aimed at evaluating either the biomass or metabolic network formulation of a given constraint-based metabolic model.

MEWpy is an integrated Metabolic Engineering Workbench for strain design optimization. It offers methods to explore different classes of constraint-based models (GECKO, sMOMENT, ETFL), kinetic models and microbial community models. It also provides Omics integrative methods to enhance phenotypic predictions. SDDb, the Strain Design Database, is a WEB application that gathers designs from the literature and generated by MEWPy, and allows to simulate genetic modifications and visualize how they impact metabolic pathways.

### 2 - YEASTRACT+: new developments

YEASTRACT+ (<http://yeastract-plus.org/>) is a portal for the analysis, prediction and modelling of transcription regulatory data at gene and genomic levels in yeasts, focused on *Saccharomyces cerevisiae*, pathogenic yeasts of the *Candida* genus, and non-conventional yeasts of biotechnological relevance.

In its newest release, YEASTRACT+ offers upgraded information on transcription regulation for eleven yeast species. It further integrates the Community Yeastract (<http://yeastract-plus.org/community/>) database, that offers the community a platform for the creation, use, and future update of YEASTRACT-like databases for any yeast of the users'

choice. In addition, a new set of tools, currently implemented for *S. cerevisiae* and *C. albicans*, is further offered, aiming the combination of regulatory information with genome-scale metabolic models to provide predictions on the most promising transcription factors to be exploited in cell factory optimization or to be used as novel drug targets.

### 3 - Data management in Plant Sciences: update on the BioData.pt Plant Sciences community

To address the challenge of sustainable production of food and non-food plant products (e.g. wood, cork, paper), particularly in the scope of climate change, many plant sciences researchers aim to study the interplay between plant development, genomic features and environmental signals.

The Plant Sciences Community aims to tackle the challenge of managing plant genotypic and phenotypic data according to the FAIR data principles. Aligned with ELIXIR Europe, this effort is determined to guarantee efficient integration and reuse of plant phenotyping and genotyping datasets, so that full value can be extracted from the increasing wealth of data collected by researchers and producers worldwide. In Portugal, this community is coordinated by ITQB NOVA and also includes several researchers from iBET, FCUL, U. Minho and CEBAL.

### 4 - Marine Data Resources Community

The Marine Data Resources Community (MDRC) was started in 2022 to help catalogue and organise biological marine data in Portugal. The Community was formed from the marine component of the ELIXIR Marine Metagenomics Community when it was reconfigured to broaden its operational scope, and was renamed the Microbiome Community. We believe that marine science in Portugal is especially important given the country's maritime history, its stewardship over the largest marine environment of any European country, and its reliance on the oceans not only as a source of food, but also as a source of national pride and aesthetic. Consequently, we believe there is ample opportunity for marine data scientist to develop collaborative projects within the scope of BioData.pt|ELIXIR.pt.

The initiation of the MDRC has started a process of reorganisation to define its core of operations. This strengthening approach has been based on a first attempt to expand the number of members and institutions participating in the community. A first step has been to map the expectation the members have for the community and define key area targets for growth.

### 5 - SeagrassTraitDB: a seagrasses traits database

Seagrasses are the only monocotyledonous flowering plants (angiosperms) which grow in marine environments. Seagrasses are found from shallow salty to brackish waters and inhabit environments from the tropics to the Arctic.

With this web service we aim to develop the first database integrating, world-wide, seagrass trait data following community standards for metadata description and plant ontology definitions.

The service currently supports the upload of 206 seagrass traits and associated metadata. Users are responsible for their data and access to their data (<http://seagrasses.ccmr.ualg.pt/>).

## 6 - Roadmap of 3D-BioInfo-PT, the BioData.pt community of computational structural biology researchers

**Inception:** the 3D-BioInfo-PT project is born, partially as a re-branding under Biodata.pt of the community involved in the EJIBCE meetings – Encontros de Jovens Investigadores em Biologia Computacional Estrutural.

**Materialization:** Planned roadmap presented and debated at the Biodata.pt Technical Meeting.

**Teaching:** the first 3D-BioInfo-PT workshop takes place at ITQB NOVA. It is an event with 35 participants (out of 62 applicants) hailing from 7 different institutions. Main support from the Oeiras Valley initiative and Wallfuture Lda.

**Continuity:** maintain the 3D Biotalks seminar cycle and expand its reach.

**Networking:** maintain the EJIBCE tradition of the annual meeting. Organize a satellite workshop.

**Collaboration:** bring together the 3D-BioInfo-PT coordinating groups to write a perspective article on computational structural research in Portugal.

## 7 - Data and Interoperability Platform

The mission of the Data and Interoperability Platform of BioData.pt is to drive the use, re-use and value of life science data in Portugal. It aims to do this by providing users with robust, long-term sustainable data resources and encouraging the life sciences community to adopt data management and sharing best practices, such as standardised file formats, metadata, vocabularies and identifiers.

## 8 - Ready for BioData Management?

ELIXIR Portugal | BioData.pt, is strongly committed to ELIXIR's mission, particularly with respect to the development and adoption of best practices in data management and open science principles. "Ready for BioData Management?" is a capacity building program in data management for the life sciences to empower researchers and institutions in managing their data more effectively and efficiently. To guide researchers' efforts BioData hired the 1<sup>st</sup> Data Steward for Life Sciences in Portugal. The work has been carried out at the Instituto

Gulbenkian de Ciência, where the Data Steward assists researchers in research data management policies and practices and help to define policies, workflows, and procedures of the institute.

## 9 - The BioData.pt Compute Platform

The BioData.pt Compute Platform provides a full stack of computational resources to the Portuguese life science research community, including a dedicated user support. The BioData.pt computational infrastructure, implemented over a hybrid cloud and high performance computing resources, supports access, usage and storage of digital objects and is coordinated by INESC-ID.

This infrastructure was maintained and expanded in 2021 and includes CCMAR, IGC and Técnico ULisboa as cloud service providers, and constitutes the backbone of most of BioData.pt's activities.

Additionally, as a member of the Galaxy Pulsar Network, the BioData.pt continues to contribute with computing capacity to Galaxy Europe, enabling researchers to use the Pulsar Network resources.

## 10 - The Portuguese Local European Genome-Phenome Archive (EGA)

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.

## 11 - Management Best Practices in Support of Research Infrastructure Operations

BioData.pt, the Portuguese Infrastructure of Biological Data and the Portuguese Node of ELIXIR, is a joint venture of many Portuguese researchers from several national organisations, since 2010. In 2021, for sake of excellence in governance and management to support bioinformatics, data management and computing resources and services for researchers in Portugal, 12 Portuguese R&I entities founded Associação BIP4DAD to be the legal entity of BioData.pt.

BioData.pt has adopted tight Management Control Systems, developed in the frame of ELIXIR-Converge and the Executive Masters in Management of Research Infrastructures to effectively support its daily operations. It has also started implementing a thorough communications workflow.

## 12 - “To be or not to be”... How can your RI be impactful?... “That is the question”

Sustainability of research infrastructures (RIs) is a big challenge for funders, stakeholders and operators, as is the adoption of adequate management tools including tools for monitoring and evaluating their performance and impact. The foundations of BioData.pt, the Portuguese RI of Biological Data and Portuguese node of ELIXIR, were laid under the “Building BioData.pt” project.

During this project, performance and impact indicators were collected and analysed under the light of international guidelines for assessing the performance and impact of European RIs. The case presented herein benefitted from several exchanges with BioData.pt and ELIXIR colleagues to facilitate a collective reflection on how to adapt and adopt European frameworks and recommendations to the national landscape and specificity of activities, to support the management of national RIs.