

Report of the 2nd Technical Meeting of BioData.pt Association

2nd of March, 2022

University of Minho, Braga



Organizing Committee

Ana Francisco, Ana Portugal Melo, Jorge Oliveira, and Miguel Rocha

Table of Contents

Introduction	3
Session I: “Building BioData.pt Communities and Platforms”	3
Overview	3
Session II: “My Data are Your Data: open science and FAIR data principles in Portuguese R&I organizations” (open session)	6
Overview	6
Final Remarks	7

Introduction

The 2nd Technical Meeting of BioData.pt Association was held on the 2nd of March 2022, from 11h00 to 17h00, at the University of Minho, in Braga, and counted with the presence of 28 researchers from 9 BioData.pt Institutions (BIP4DAB, CIIMAR, FCUL, IBET, IGC, INESC-ID, ITQB-NOVA, UAveiro, and UMinho), and a guest, UCIBIO, whose membership to BioData.pt Association will be ratified in the General Assembly of March 2022. The day was organised in two sessions: Session I dedicated to setting the bases for the BioData.pt Communities and Platforms activities; and Session II, open to the entire community, dedicated to sharing experiences and best practices of BioData.pt Associates open science and FAIR data principles.

Session I: “Building BioData.pt Communities and Platforms”

Overview

A brainstorm took place to reflect on BioData.pt Communities and Platforms and think of common processes to support their development. Three Communities and a Platform were selected as subjects of this exercise: Plant Sciences, 3D-Bioinfo, Microbial Biotechnology/Systems Biology, and Data/Interoperability. Each Community/Platform met to debate and create slides with its main challenges, goals and action plan. The results of the four discussions were presented and further discussed in plenary, to enable comments and suggestions by peers. The results are summarised in table I.



Figure 1. Manuel Melo (ITQB-NOVA), presenting the conclusions of the 3D-Bioinfo breakout.

Table I - Summary of Communities/Platforms meetings.

Participants	Challenges	Goals	Action Plan
<i>3D Bioinfo Community</i>			
-Manuel Melo -Isabel João -Hugo Oliveira -Jorge Oliveira	-Need for computing resources -Outreach to all interested -Dataset heterogeneity and size	-Broaden and reach community -Capacity building -Create visibility -Create collaborations and joint funding opportunities -Boost student mobility	-Create regular PT meetings (monthly) - April -Surveys on: -Training skills and needs -Tools and needs -Data management skills and needs Review Article on 3D Bioinfo in -Portugal - April -Training needs landscape analysis - May -Organize Workshops - December or earlier -Annual meeting - December
Participants	Challenges	Goals	Action Plan
<i>Microbiotechnology/Systems Biology Community</i>			
-Pedro T. Monteiro -Sílvia Santos -Miguel Rocha -Tiago Paixão -André Machado -Filipe Castro -Vânia Pobre -Ivone Martins	- Heterogeneous data -Multiplicity of mathematical models -Different organisms -Distinct research objectives	-Increase literacy in systems biology - Listing of: tools (how to use?) groups/people (whom to ask?) free access data (where are they? what kind?) Guidelines from: data acquisition/processing analysis / expected results education / training what exists?	Identify: - international stocks?? (google docs ?) - complementary contributions - Dissemination / training - specific training - communication platform (slack channel?) - inter-community forum - break down barriers to entry of procedures/tools - online training resources -
Participants	Challenges	Goals	Action Plan
<i>Data/Interoperability (includes health data and biocuration) Platform</i>			
-Cátia Pesquita -Daniel Faria -João Rafael Almeida	-Data and scientific questions → data availability and data valorisation	-Promote Biodata.pt Dataverse as a solution for data availability and interoperability	-Slack channel/Google Docs to continue: - survey of data producers' needs - survey of solutions created by

<p>-José Luis Oliveira</p> <p>-Teresa Ribeiro</p> <p>-Ricardo Leite</p> <p>-Maria Paola Tomasino</p> <p>-Angela Novais</p> <p>-Joana Pinto</p> <p>-Carolina Costa</p>	<p>-Data standardization /publishing</p> <p>-Data integration /harmonization</p> <p>-Data and Metadata Lifecycle Support</p> <p>-Human resources (CS)</p> <p>-Lack of trained HR in data stewardship, data integration</p> <p>-Job market</p> <p>-Human factors:</p> <p>-High burden for data best practices</p> <p>-Lack of incentives</p> <p>-Lack/Knowledge of data management tools</p>	<p>(dissemination and training)</p> <p>- Federation of DataVerses of partners (UA, UM, etc)</p> <p>- Dataset usage statistics (citations, downloads, likes)</p> <p>-Continue with RDMKit and Ready4BDM initiatives (lack of tools and knowledge)</p> <p>-Promote incentive scheme policies with national institutions (human factors and data valorisation)</p> <p>-Facilitate data sharing, curation and management</p> <p>- automatic methods</p> <p>- Good practices</p> <p>-Collaboration with other national/international infrastructures/initiatives</p>	<p>data scientists/stewards</p> <p>-1 meeting per quarter (zoom)</p> <p>-White Paper on how to publish and promote a dataset</p> <p>- DOI, associated publication</p> <p>-Contribute to a section of the Biodata.pt white paper</p> <p>-Collaboration with the GDI Forum</p>
Participants	Challenges	Goals	Action Plan
<i>Plant Sciences Community</i>			
<p>-Pedro Barros</p> <p>-Diogo Pratas</p> <p>-Ana Francisco</p> <p>-Óscar Dias</p> <p>-Inês Chaves</p>	<p>-Human resources, computational capacity</p> <p>-Lack of annotated data in some species</p> <p>- Dispersion of repositories</p> <p>-Lack of interest in data annotation</p> <p>-Heterogeneity of species of interest</p> <p>-Systems administration</p>	<p>-Improve the automation of annotation processes</p> <p>- Increase knowledge of data management</p> <p>-Training to enrich knowledge in complementary areas</p>	<p>In progress:</p> <p>-Meetings every 2 months to streamline communication</p> <p>-Implementation of data annotation in Dataverse</p> <p>To be organised:</p> <p>-Workshop (2023-2024):</p> <p>- plant (meta)data management</p> <p>- tools under development (INCREASING, BrAPI, Dataverse)</p> <p>-Training in systems administration tools</p> <p>-Promote Staff Exchange actions (national or international)</p> <p>Survey of Repositories with plant data (omics)</p>

Session II: “My Data are Your Data: open science and FAIR data principles in Portuguese R&I organizations” (open session)

Overview

This session aimed to raise awareness for the importance of FAIR and Open Science principles. Pedro Príncipe and Tiago Paixão presented the examples of University of Minho and Instituto Gulbenkian Ciência, two institutions committed to Open Science, that have well assembled data management systems, including user guidelines for data and a long-term roadmap for data management. Cátia Pesquita talked about the importance of good quality data in the scope of health emergencies, as the pandemic we are still suffering from. Finally, Jorge Oliveira showed in a glimpse, the tools and resources BioData.pt makes available to the community to make data FAIR. An enthusiastic discussion, moderated by Ana Portugal Melo, followed the presentations, and counted with an effective participation of the audience, recognizing the challenges of implementing Open Science and FAIR in their organisations, and the need to design and bring into operation institutional Data Management policies. Most of BioData.pt have researchers enrolled in the 1st edition of the “Ready for BioData Management? - Training Data Stewards for Life Sciences”, and are counting that this program will contribute to drive Portuguese life-science research organisations to a new stage, concerning how they manage their research data.



Figure 2. Panel of the round table “My Data are Your Data: open science and FAIR data principles in Portuguese R&I organizations”. From left to right, Jorge Oliveira (IGC), Pedro Príncipe (UMinho), Ana Portugal Melo (INESC-ID), Tiago Paixão (IGC), and Cátia Pesquita (FCUL).

Final Remarks

This technical meeting reminded us of the greater synergies that get established when we meet in person. Even with a long journey for most participants, everybody was excited to see the faces we have only seen virtually and energetically, working to improve the BioData.pt Communities and Platforms and discuss the relevance of data management and FAIR data for Portuguese organisations and worldwide.