Ready for BioData Management?

BioData.pt eligir

Carolina Ventura-Costa^{1,2}, Jorge S. Oliveira^{1,2}, Tiago Paixão^{1,2}, Ricardo Leite^{1,2}, Daniel Faria^{1,3} and Ana M. P. Melo^{1,4} 1. BioData.pt | ELIXIR-PT, Oeiras, Portugal; 2. Instituto Gulbenkian de Ciência, Oeiras, Portugal; 3. LASIGE, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal; 4. INESC-ID, Portugal

1st BioData.pt of All Hands 3rd Technical Meeting CCMAR, 28-30/09/2022



Capacity building program in data management for the life sciences

Introduction to Data Management Plans

Full day hands-on course aiming to teach the basics of making and using Data Management Plans (DMP) in the context of research activities



Advanced Data Management Plans

One-day hands-on course where participants prepare a DMP in the platform DMPOnline according to the H2020 template

More than 300researchers trained on building DMP

November

Training Data Stewards for Life Sciences (February - November 2022)

The 1st training program for data stewards in Portugal with 24 participants from 14 institutions and > 63 h of training

February

DS Intro course

Data Stewardship

Ciências ULisboa GULBENKIAN CIÊNCIA

- FAIR & Open Science Funders policies
- Data Management Plans...

Assignment + DS Lab1

- Institutional DM policy exercise •Institutional DM Plan exercise
- Feedback to participants &
- Review

DS Lab2 Assignment

- Dataset annotation and submission to DMPortal
- Feedback & Review
- Deeply explore metadata and ontologies

Assignment

 Develop guidelines for a specific stage of the Data Life Cycle, based on

- the RDMkit Feedback & Review
- 3-Day Face-2-face course Assignment presentation by

Consolidation course

- trainees Deep-dive into the stages of
- the data life cycle Invited specialist speakers

Outcomes



Institutions reflecting DM policy



Institutions drafting **DMPs**



FAIR Datasets in **DMPortal**

1St Data Steward for Life Sciences in Portugal

Survey current practices and requirements regarding Research **Data Management** at the Instituto Gulbenkian de Ciência (IGC)

research groups

researchers interviewed

16 Lab managers

- Postdocs
- Technicians PhD students
- MSc students

gaps identified between KNOWLEDGE and SKILLS on the topics of data

storage, management and documentation following the FAIR

principles

Electronic Laboratory Notebook (eLab)

eLab is not being used or is used incorrectly

Gaps Research Data Management (RDM)

absence of data management plans (no structure or folder organisation; no labelling convention, no metadata standards)

Informatic infrastructure awareness (ITA)

knowledge gap regarding the informatic procedures, space restrictions...

Support was provided to different groups on:

eLab

RDM

ITA

Electronic Laboratory Notebook



helpdesk support; optimise the use of eLab; create support documentation; testing software and fixe problems

Documentation page: tutorials, tools & FAQs are now available in a dedicated page on the intranet of IGC

backups information

Support via Agendo: dedicated portal to report a problem, make a suggestion, ask a question or talking through requirements

Requests received/addressed via Agendo

 $14_{/14}$ Support from the DS

Features required by IGC members and created by the software developer

Research Data Management



data storage & management support; folder organization; labelling convention; metadata structure

- Support on DMPs: in-person support and share available tools to create Data Management Plans using the BioData.pt Data Stewardship Wizard online tool
- Personalized RDM consulting for research labs: face-to-face support and customised guidance for folder structure metadata collection and labelling convention

File naming convention: {a description of the file naming structure adopted} ProjectAcronym (two capital letters) TaskNumber (one number from 1 to 9) User ID (e.g., João Costa = JC) Strains/Organoids Subjects: Aplical Nuclear Migration (e.g., ANM) Condition: Wild-Type (WT) / Control (C)

20220629JC_SXXXX_ANM_WT.xxx

/file path/PROJECT ACRONYM — "ProjectLevel_ReadMe" {this document} └─ Task Number {this section does not have to be available for users} └─ Subject 1: Aplical Nuclear Migration ├─ Aquisition Method 1: LigthSheet - Raw Data ── Processed Data

├─ Aquisition Method 2: Genomic Facility

- Raw Data └─ Processed Data └─ Aquisition Method 3: Other └─ Subject 2 └─ Subject 3

IT infrastructure awareness



filling the gaps between researchers and IT infrastructure

Increase awareness among researchers regarding IGC server space restrictions

Training (in preparation)



workshops, training sessions

One-day hands-on course: aiming at disseminating the "IGC approach" in the context of participants' data management requirements

Contacts

Name: Carolina Ventura-Costa E-mail: carolinacosta@igc.gulbenkian.pt Website: www.biodata.pt









