

Workshop Feedback

General Feedback

Overall, all groups showed engagement with the Hands-On Data Management Plans exercise, devoting themselves to the task of understanding the data management issues underlying the mock research project and detailing them in their data management plan. There were some interesting discussion points raised throughout the section, including the topic of data privacy and the risk of identifying patients through gut metagenomics due to contamination with mitochondrial DNA.

In general, the information that all groups listed in their data management plan was correct/adequate, with the main differentiating factor between groups being the level of depth and detail they provided. However, most groups also stumbled on a few common issues:

- Mention of the use of proprietary encrypted file formats (e.g. .xls) for dataset publication, whereas open source plain file formats are the best practice for findability, interpretability, and long term accessibility.
- Preference for non-commercial usage licenses for their datasets. This is particularly inadequate in a project related to healthcare, as it would preclude pharmaceutical companies from applying the project's results into solutions for the disease. It also violates the spirit of public funding for scientific research, which is given in the expectation that scientific discoveries will give back to the economy. Unless researchers plan on leveraging their discoveries themselves to start a company (in which case they may want to delay or abstain from publishing their data) they should not restrict others from doing so.
- Identification of only computational resources in the technical resources section, leaving out other critical resources such as sequencing machines.
- Identification of only storage costs in the costs section, leaving out other key costs that apply in this context, such as personnel and material.

The feedback for each group is given in the following pages, overlayed on the data management plan the group submitted.

Hands on DMP session

Group 1

Step 1

DMP

A characterization of the DMP

Project: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Project: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Step 2

Project

Information regarding the project to which the DMP is associated

Project 1: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Project 2: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Technical Resources

Characterization of any technical resources associated with each of the identified datasets

Dataset 1: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Dataset 2: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Metadata

Characterization of metadata associated with each of the identified datasets

Dataset 1: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Dataset 2: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Security and Privacy

Characterization of any security and privacy policies associated with

Dataset 1: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Dataset 2: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Distribution

Characterization of how each dataset is to be distributed

Dataset 1: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Dataset 2: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Host

Characterization of the preservation policies

Dataset 1: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Dataset 2: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

License

Characterization of the licenses associated with each distribution policy

Dataset 1: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Dataset 2: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Contact

A person or institution that is to serve as contact for this DMP

Project: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Funding

Specific information pertaining the funding of the project

Project: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

IM Staff

Characterization of the staff members that are to be listed within this DMP

Project: Sequencing and analysis of the genome of the *Arabidopsis thaliana* plant. The project aims to identify the genetic basis of the plant's resistance to drought and to develop a drought-tolerant plant. The project is funded by the European Union and the Portuguese government.

Group 1 created what we believed to be the most detailed DMP of all groups. They were one of the few groups to mention both software and sequencers in the technical resources section, as well as provide a fairly descriptive metadata section. They also proved to have considered possible ethical issues in the project, by providing a fairly detailed ethical issues description. However, they undermined possible future applications derived from the project by protecting their data with non commercial licences.

Group 2

Project
Information regarding the project to which the DMP is associated

Contact
A person or institution that is to serve as contact for the DMP

Funding
Specific information pertaining the funding of the project

DM Staff
A characterization of the staff members that are to be listed within this DMP

Datasets
A characterization of all the datasets associated with the DMP

Security and Metadata
A characterization of any security and metadata associated with each of the identified datasets

Distribution
Information regarding the policies on how each dataset is to be distributed

Host
A characterization of the preservation policies

License
A characterization of the licenses associated with each distribution policy

Data management Costs
Information of the associated costs of the DMP

Group 2 showed plenty of creativity in their approach to the exercise. They were one of the few groups to mention sequencers as a necessary technical resource. However, their DMP lacked depth and detail overall.

Group 3

Hands on DMP session

Step 1

Step 2

Group 3

Project

Datasets

Security and Privacy

Metadata

Distribution

Host

License

Data Management Costs

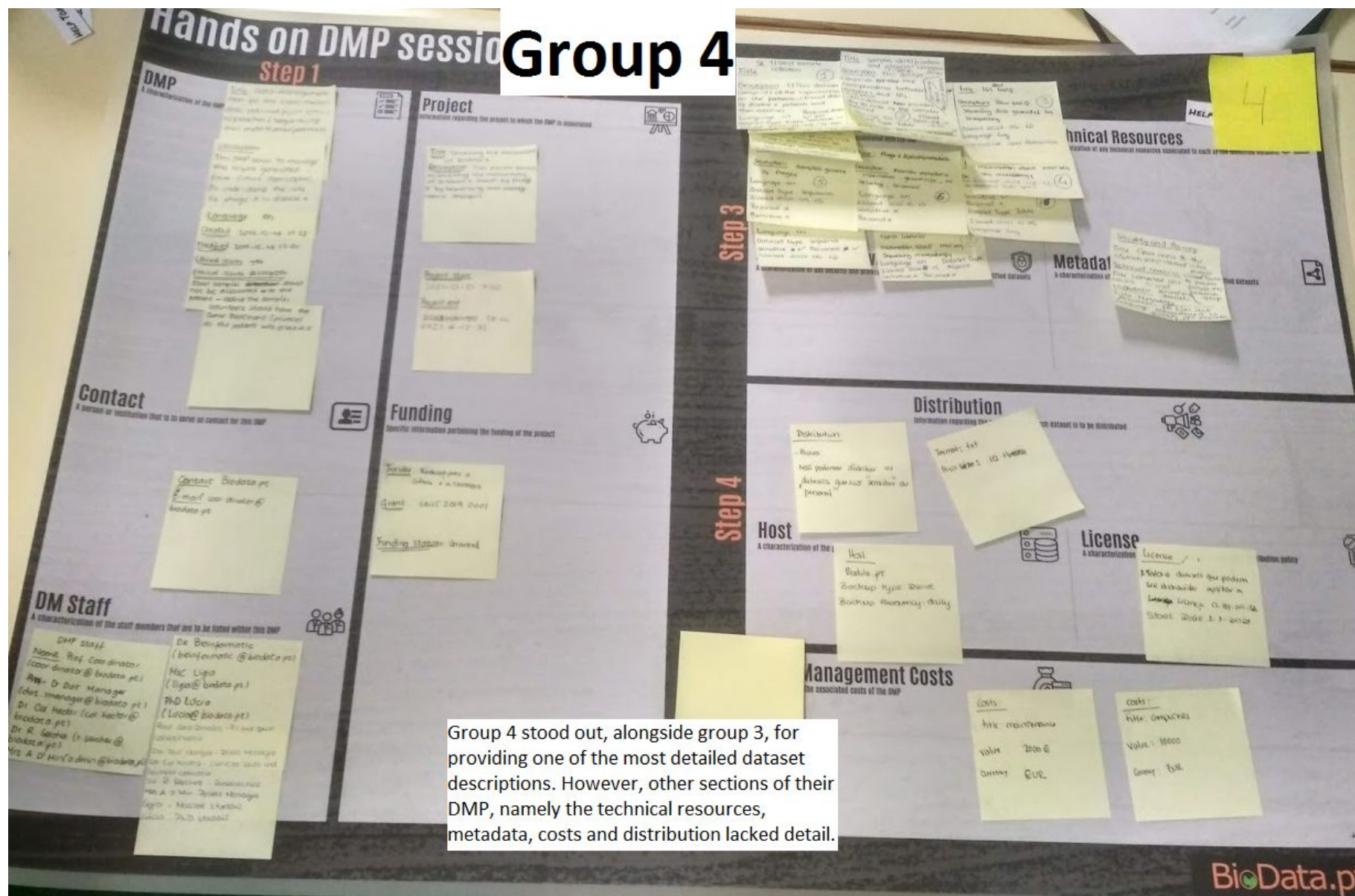
DM Staff

Contact

Technical Resources

HELP TAKEN

BioData.pt



Group 5

Hands on DMP session

Step 1

DMP
A characterization of the DMP

Project
Information regarding the project to which the DMP is associated

Datasets
A characterization of datasets

Technical Resources
A characterization of any technical resources associated with each of the identified datasets

Security and Privacy
A characterization of any security and privacy policies associated with each of the identified datasets

Metadata
A characterization of metadata associated with each of the identified datasets

Distribution
Information regarding the policies on how each dataset is to be distributed

Host
A characterization of the preservation policies

License
A characterization of the licenses associated with each distribution policy

Contact
A person or institution that is to serve as contact for this DMP

Funding
Specific information pertaining the funding of the project

DM Staff
A characterization of the staff members that are to be listed within this DMP

HELP TOKEN

5

Group 5 followed the guide thoroughly, going beyond what was required by adding extra detail to the dataset section. However their remaining sections lacked the level of detail of Group 1.