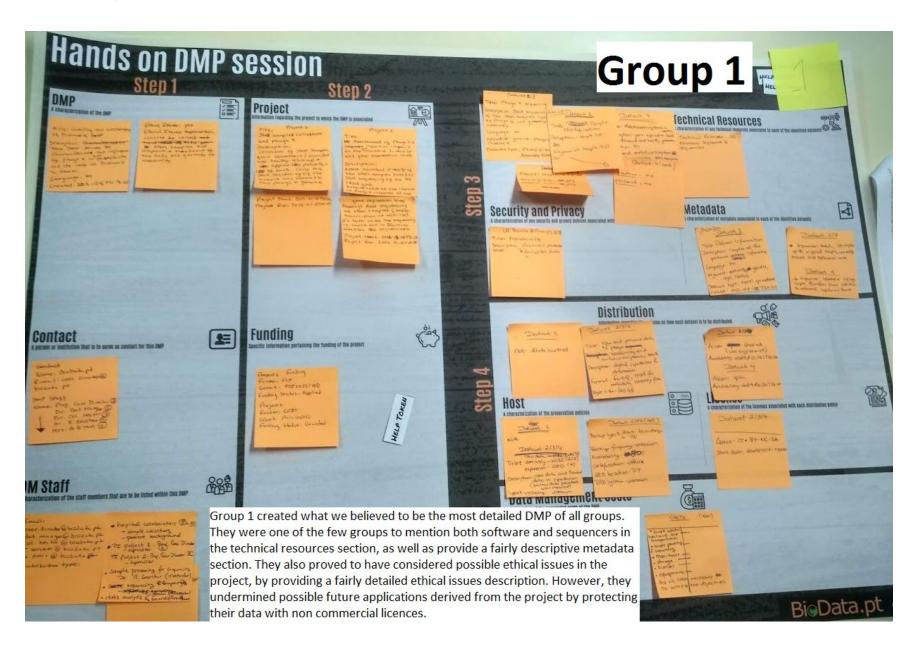# Workshop Feedback

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:
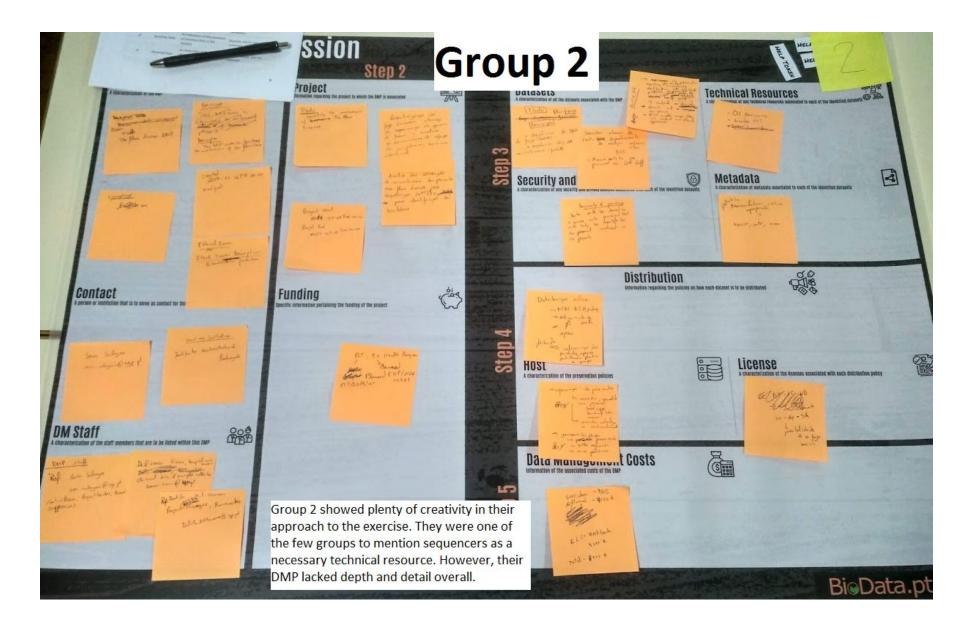
- Mention of the use of proprietary encrypted file formats (e.g. .xls) for dataset publication, whereas open source plain file formats are the best practice for findability, interpretability, and long term accessibility.

- 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.

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 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 provided a detailed DMP, in particular concerning the description of the datasets. However, it lacked depth in the technical resources and metadata sections, and the cost section could be more detailed.

Group 4 stood out, alongside group 3, for providing one of the most detailed dataset descriptions. However, other sections of their DMP, namely the technical resources, metadata, costs and distribution lacked detail.

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.