Data Management Plan – Help Sheet

Comments	Examples
1. Data description and collection or re-use of existing data	
1.1 How will new data be collected or produced and/or how will existing data be re-used?	
If you will be performing wet-lab experiments in your project, you should state the types of data-generating experiments that you will be doing. If you will be generating new data by computational methods, also state this here. If you will be conducting analyses that re-use data from public databases, you should state what data types and from what sources you will use. If you will be using open, commercial or custom software that will produce new or processed data, you should also mention it here.	*New data in the project will be obtained by means of confocal microscopy / Western blotting / real-time PCR / mass spectrometry …* *We will obtain large numeric datasets from computer simulations of …* *In this project we will re-use DNA sequences from the NCBI GenBank.* *In this project we will re-use protein structures from the Protein Data Bank.* *A custom computational workflow will be used to calculate the frequencies of …* *The commercial software … will be used to obtain processed data from raw HPLC files.* *The open software … will be used to obtain processed statistical data from raw numerical data.*
1.2 What data (for example the kinds, formats, and volumes) will be collected or produced?	
List the data types and file formats that you will obtain from the procedures described above.	*We will collect: microscopic images (tiff), images of DNA gels (jpg), numerical data obtained from … experiments (spreadsheets in csv format / xlsx format), ...*

By default, you are expected to collect data in formats most convenient for you (e.g. associated with the instrument or software you are using), but to convert them to open formats for data sharing. If you are generating data in commercial formats, explain to what open formats you will export this data for sharing.	*Data processed in Statistica will be in the proprietary sta format but will be exported to the xlsx format before sharing.*
Always give a rough estimate of how much data will the project generate in total.	*We expect to collect up to 1 TB of data in the project.*
2. Documentation and data quality	
2.1 What metadata and documentation (for example the methodology of data collection and way of organizing data) will accompany data?	
Explain where and in what form you will store the data documentation.	*All required metadata, including experimental protocols used for data collection, instrument specifications and settings, and other relevant parameters, will be provided in text format and will accompany the data files.*
If there are some specific community rules you should follow, mention them. For data that will be deposited in a repository, there will likely be requirements at the point of data deposition.	*A dictionary describing the phenotypic categories used for mutant description will be collected in a spreadsheet and will accompany the data.* *Custom computational workflows will be deposited in GitHub.* *All computational workflows will be described in text documents and accompany the data.* *Descriptions of Quantitative Real-Time PCR experiments will follow the MIQE guidelines.* *Metadata accompanying imaging data will follow REMBI guidelines, as required by the BioImage Archive.*

You are encouraged to plan the folder structure and file naming that you will be using during data collection.	*All researchers collecting microscopic images will name their files according to the scheme: YYYYMMDD_mutantname_samplenumber.tiff .*
2.2 What data quality control measures will be used?	
Mention any quality control actions that you normally apply during your research.	*Appropriate data quality will be ensured by following the requirements of the implemented technologies and good scientific practice, including:* *regular calibration of …, use of internal standards for …, using at least three biological and two technical replicates for …, following MIAME guidelines when collecting microarray data, …*
3. Storage and backup during the research process	
3.1 How will data and metadata be stored and backed up during the research process?	
Consider where you and your co-workers will be keeping your data while you work on the project (usually as is the standard practice followed in your research group). Make sure you plan a working copy and at least two backups (preferably at least one backup should be stored in a different location from the other two). If your project involves the collection of large datasets (such as omics datasets), you can mention them separately in this section.	*The collected data will be stored on the hard drives of personal computers and synchronized with a NAS server owned by the research group. An additional backup will be done weekly on external drives.* *Proteomics data will be additionally stored on the archiving system of the IBB MS Lab.* *RNA-seq data will be stored… and backed up…*

3.2 How will data security and protection of sensitive data be taken care of during the research?	
First, please explain how your data will be protected from leakage or access by unauthorized persons. Second, if you are dealing with human (patient) data, or any other sensitive data (e.g. about endangered species, or confidential commercial data), explain what additional safety measures will be applied. If your project does NOT involve any kinds of sensitive data, this section does not apply to you.	*Access to the data will be restricted to authorised project members only: lab computers and external drives will be password-protected and stored in rooms with controlled access. Security upgrades to operating systems and other software will be performed regularly.* *No sensitive data will be collected or processed in this project.*
4. Legal requirements, codes of conduct	
4.1 If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?	
If you are NOT dealing with human (patient) data, this section does not apply to you. If you are dealing with patient data, but you have access only to anonymized data (or pseudonymized, but with no access to decoding information) that you receive from an external collaborator (e.g. a medical university), explain that here. BUT if the unit collecting human data is a project partner instead of being an external collaborator, or if you are dealing directly	*No personal data will be collected or processed in this project.* *Personal data will be collected in the project by external collaborators from the … medical unit. IBB PAS will have access only to anonymized data. The anonymized clinical data generated by the collaborators will be shared with us in an encrypted format to ensure security during transmission and will be stored …*

with personal data concerning human subjects, then please explain in detail how the data will be stored and secured (a medical research unit will have its own standard procedures, which they should describe here).	
4.2 How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?	
The intellectual property rights applicable to your data might include copyright and/or database rights. According to the Regulation of the IBB PAS Scientific Council No. 51/2020 dated May 6, 2020, these rights lie with the Institute. If multiple institutions cooperate in your project, make sure to come to an agreement with each other about data ownership and future sharing of data / access to data. Describe here the rules you agreed upon. If you are using data from external sources (from public databases, or data from collaborators who are not project partners, etc.) – as mentioned in **1.1** – state here who owns rights e.g. to resulting processed data.	*To the extent covered by intellectual property rights, the owner of the data will be IBB PAS.* *The data collected in Task 1 will be owned by the Project Partner… This Partner will be responsible for managing and sharing of this data, in accordance with this Data Management Plan.* *A Consortium Agreement between Partners … will be signed, stating that …* *The input data for the analyses described in Task 3 will be owned by …, and an agreement will be signed with the data owner that will allow us to ….* *All data from public databases that will be used in this project is released under a CC0 waiver, allowing us to process it without restrictions.*

5. Data sharing and long-term preservation	
5.1 How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?	
Data underlying a publication If you don't have any special reasons to restrict access to your data (e.g. commercialization plans or sensitive data), you should make it openly available to the public no later than at the time of publication of an article that uses that data. NCN expects your data to be released under CC0 conditions, but the CC-BY license is also acceptable. Other licensing options are currently not accepted by NCN. If you need to restrict access to data, or to publish your data later than your article, explain why (e.g. because you plan to file a patent application, or because of limitations related to data sensitivity). If you are not sure whether you should or should not openly share anonymized (or pseudonymized) data related to human subjects, please contact rdm@ibb.waw.pl. **Data that is not directly necessary for the validation of your published results**	*Data underlying the findings presented in scientific articles will be made available to the public, without restrictions, at the time of article publication. The data will be shared under a CC0 waiver (https://creativecommons.org/publicdomain/zero/1.0/).* *[or: … under a CC-BY license (https://creativecommons.org/licenses/by/4.0/)].* *Since we are planning to patent the following project results: …, all related data will be withheld until the patenting procedure is secure. Afterwards, the data will be made available to the public, without restrictions, under a CC0 waiver.* *Data associated with human subjects will not be shared openly due to the limitations of the informed consent forms that the study participants will be signing. In accordance with the consent given by study participants, access to this data for external researchers will be possible only for research purposes, on an individual basis, strictly regulated by IBB PAS.*

For data that is not directly necessary for the validation of your published results, but has been collected during the project, you have the choice of either making it publicly available at the end of the project or keeping it for yourself (e.g. because you plan to use it for a further project). If you do not plan to share it, you do not need to mention it here.	*Other data resulting from the project that is considered to carry scientific value of its own will be shared openly at the end of the project.*
5.2 How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?	
Usually you will select for preservation all data underlying published findings (NCN expects you to preserve it for 10 years) and any other data from the project that you find scientifically valuable. All data underlying published articles should be made openly available through repositories, as described above in 5.1. It is **very important** to state which repositories you will use to make your datasets open. If your data fits into a specialized repository (sequencing data, mass spectrometry, structural data, expression data, image data, etc.), please choose such a repository. For other data, e.g. BioStudies is a good choice for Life Sciences data and Pangaea for Earth Sciences. A list of recommended repositories is available here: https://ibb.edu.pl/en/research-data-management/	*We will preserve all data underlying published findings as well as the following data: …* *All microscopy images will be deposited in the BioImage Archive (https://www.ebi.ac.uk/bioimage-archive/). RNA-seq data will be submitted to the ArrayExpress database (https://www.ebi.ac.uk/biostudies/arrayexpress).* *Other data connected to published results will be deposited to the BioStudies database (https://www.ebi.ac.uk/biostudies/).* *Other data underlying publications will be deposited in the Pangaea repository (https://www.pangaea.de/).*

If you have data that will not be shared openly through a repository (for reasons described in **5.1**), but you will store it for the long-term, describe where. If you plan to publish a dataset from the project in a data journal (journal specializing in datasets), mention it here.	*The data … will not be shared openly at the end of the project because …, so it will be archived locally in the institutional IBB PAS long-term archive.* *The microarray data resulting from Task 1 will be published in a data journal, because it will have high scientific value outside of the project.*
5.3 What methods or software tools will be needed to access and use the data?	
Explain here if a user who downloads your data from a repository will need any special, less-known software to use it. Usually all your files will be associated with commonly known software, so you don't need to list it (e.g. you do not need to explain how one opens an xlsx or csv file). If you will be uploading specialized file formats that can be opened only using a less-known software, please specify the software here (and state whether it is free – at least in a read-only version – or not).	*No special software will be required to use the deposited data.* *The data from the … experiment will be only in … format which can be opened using … This is commercial software, but a free read-only version can be downloaded from ….*
5.4 How will the application of a unique and persistent identifier (such us a Digital Object Identifier (DOI)) to each data set be ensured?	
All repositories listed at https://ibb.edu.pl/en/research-data-management/ automatically assign a unique persistent identifier to each deposit (DOI, or NCBI identifier, etc.). If you are planning to use a different repository, make sure it assigns DOIs.	*The data repositories selected above will automatically assign a DOI number to every deposited dataset.*

6. Data management responsibilities and resources	
6.1 Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?	
Typically this will be the project PI, though you can also move this responsibility to someone else participating in the project (e.g. lab manager, senior researcher). If there are many project partners, name one person taking responsibility for overall data management and also persons who will be responsible for data management at each partner institution.	*The PI will be responsible for data management.*
6.2 What resources (for example financial and time) will be dedicated to data management and ensuring the data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?	
Usually you are not including additional resources in your plan (though you should consider in your planning that data management and deposition will take time). If you do need to allocate funding for certain data management activities in your project (e.g. to purchase hardware for data storage, hire a data steward who will prepare your data for deposit, or pay for external storage of very large amounts of data), you should include this in the budget section of the grant proposal, and explain here why you need those funds and for what.	*The necessary time will be dedicated by the researchers collecting the data and the project PI. No additional financial resources will be required.* *New storage space will be required for the collected data, and for this reason we are planning to purchase …, as described in the project budget …*