

NBIS: document

Data management plan for support projects

Introduction

Research data management plans (DMP) has become a crucial part of life science research. Good data management practices help to organise, analyse, store, publish and re-use data. DMPs are not only a good practice, but may and will be required to consider and document under the policy of various research funders, e.g. VR. This document contains some key points to consider when having data analysed with NBIS.

Overview

1. Project title
2. Project issue number
3. Project contact person
4. Data contact person
5. Data management contact person
6. Is DMP available? Yes/No. If yes, provide details. If no, provide the information below. Refer to [extended online version of this document](#) for sections examples and more information.

Data description

1. Primary data, *incl. a) what type of data will be generated? b) from which type of samples? c) how many samples? d) from what technical platform?*
2. Additional data, *incl. a) what other data will you need to perform the project?, b) do you have the necessary permission(s) to use that data? c) how does your re-use complies with terms and conditions?*

Ethical & Legal aspects

1. Data owner(s), *incl. a) which host university/universities owns the data?*
2. Intellectual property rights (IPR)/Copyrights, *incl. a) are there IPR or copyright issues to consider? b) will the data potentially be used to generate IPR?*
3. Legal Agreements (if applicable), *incl a) what are the agreements with other stakeholders? b) what agreements are needed to regulate relationships between collaborators. c) do you need Data Processing Agreements between collaborators for personal data?*
4. Sensitive Human Data (if applicable), *incl. a) does your project have approval by an ethics committee?, b) what informed consents exist for the samples?, c) are there any consent codes that can be assigned to the samples?*

Data documentation

1. Sample metadata documentation, *incl. a) what metadata will be provided with the collected/generated/reused data? b) are there metadata standards that you can use?*
2. Dataset documentation, *incl. a) which file formats and data types will be used for the data?, b) what is the estimated total size of the data?*

Data storage & backup

1. Storage, *incl. a) how, b) where, and c) for how long will the data be stored during the analysis phase of the project?*

2. Backup, *incl. a) how, b) where and c) at what intervals will the data be backed-up? d) how will data be recovered in the case of a data loss incident?*
3. Security, *incl. a) how will you ensure that only authorized persons have access to the data?, b) how will sensitive data be protected, if applicable?*

Data publication & archiving

1. Long-term storage, *incl. a) how and b) where will the data be stored after the project's completion?*
2. Data publishing, *incl. a) will you deposit your data to a trusted data repository? b) if so, when during the project will you submit the data to the archive? c) in what formats will you submit the data to the repository?, d) will your data receive a persistent identifier?*
3. Data access, *incl. a) will your data be available Open or Controlled Access? b) when will the data become accessible? c) if Controlled Access, who controls access?, d) will all data and metadata, or only parts of it, be published?, e) under what licence(s) or terms will you share your data and code?, f) Are there any restrictions that prevents the publication of all the material? and g) if so, what actions must be taken before the material can be made available?*

Costs

1. Are there costs you need to consider to buy and manage specific software or hardware?
 2. What are the costs you need to consider for storage and backup?
-

Ultra-short general recommendations

1. If you have no specific insight, include a data management and archiving cost of 5-10% of the project budget for a large-scale omics project, provided that you can largely rely on SNIC systems. This does not include any personnel cost to analyse the data.
 2. If you cannot use SNIC systems for your data analysis/storage, you need a major budget post to cover for this, and you should investigate this carefully.
 3. Human sequencing data are classified as sensitive data under GDPR, and needs special attention. More information under <https://nbis.se/support/human-data.html>
 4. Submit data to public repositories early in the project, i.e. under embargo, to ensure an extra backup.
-

More information

- Extension of this document with examples and including [DMP knowlege hub](#)
- SNIC resources for compute and storage during active phase of the project <http://snic.se>
- Working with sensitive data: <https://nbis.se/support/human-data.html>
- A list of public deposition databases can be found on [Elixir Deposition Databases List](#)
- A Quick Guide to Organizing Computational Biology Projects [Noble, 2009]

References

William Stafford Noble. A Quick Guide to Organizing Computational Biology Projects. *PLOS Computational Biology*, 5(7):e1000424, jul 2009. URL <https://doi.org/10.1371/journal.pcbi.1000424>.