Plan Overview

A Data Management Plan created using DMPonline

Title: Example_Answers_NTNU

Creator: Ane Gabrielsen

Affiliation: Other

Funder: Science Europe

Template: Science Europe

Project abstract:

This DMP contains example answers. Some of them are from existing DMPs, others are completely made up. They are only intended as guidance and inspiration and have not been reviewed or otherwise quality controlled.

Note that the answers are not representing one specific project.

Questions regarding research data management at NTNU? Please contact Research Data @NTNU.

ID: 110215

Start date: 14-07-2022

End date: 31-12-2022

Last modified: 02-11-2022

Grant number / URL: 666666

Example_Answers_NTNU

Data description and collection or re-use of existing data

How will new data be collected or produced and/or how will existing data be re-used?

Example Answer:
The project will generate Qualitative (phenotypic data of cells and tissues), Quantitative (mRNA and protein levels), Mass spectrometry and Image data.
The project will reuse publicly available RNAseq datasets maintained in the ArrayExpress and GEO databases.
Raw data will be analyzed and expressed as graphs, tables and annotated images. A subset of analyzed data will be published.

What data (for example the kinds, formats, and volumes) will be collected or produced?

Example answer:
Data generated will be in various formats and sizes of datasets, all of which will be accessible using common software allowing easy access and long-term validity during and after the project, thus facilitating data sharing. Total amount of data generated is expected is 1TB and 5TB. The format/types of data include:

1. mRNA abundance data: two types of gene expression data will be collected, (i.) individual gene expression data using Real-Time PCR equipment stored in SDS files and (ii.) whole genome RNAseq data represented by aligned short read counts stored in BAM files
2. mass spectrometry spectra. MS data will be analyzed using public MiST Bioconductor package and SEQUEST/Mascot software to match proteins in the human genome dataset.
3. cell line phenotypic data including cell viability will be collected using automatic plate readers and stored in Excel and GraphPad Prism files. Data on presence distribution and abundance of proteins in cell populations will be collected by flow cytometry and store as FSC and jpeg/tiff files.
4. DNA, RNA and protein sequencing data, which will be saved as abi, fastq or MASCOT files. Recombinant DNA construct maps will be generated using software such as Serial Cloner.
5. Cell images such as phase and fluorescence. Images will be acquired and analyzed using Imaris software and Huygens Professional image deconvolution program will be used to increase the resolution of images. Data will be saved as software-specific files e.g. liff and lg3 files, as well as generic formats such as JPEG, TIFF etc.

Documentation and data quality

What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?

Example answer:
Metadata: Microscope images capture and store a range of metadata (field size, magnification, lens phase, zoom, gain, pinhole diameter etc.) with each image.
These metadata will be recorded according to the REMBI standard - Reccommended metadata for microscopic images.
Organization of data and files during project: Files will be named according to a pre-agreed convention. For proper sorting, date order will be YYYY-MM-DD. The dataset will be accompanied by a README file which will describe the directory hierarchy.
Documentation: Each directory will contain an INFO.txt file describing the experimental protocol used in that experiment. It will also record any deviations from the protocol and other useful contextual information.

What data quality control measures will be used?

Example answer:
Standard protocols will be optimized and used to collect data to ensure they are reliable and consistent. All experiments will incorporate appropriate positive and negative controls to ensure validity; biological and technical replicates will be used to assess consistency of data. Team members will be trained by experienced colleagues to ensure high quality and reproducibility of data. Data and the methods used for collection will be reviewed in lab meetings to assure a high level of confidence in the data generated.

These are routine procedures for conducting high quality research, which allow it to be evaluated and published in fully peer reviewed journals, as well as presented at relevant international conferences.

Storage and backup during the research process

How will data and metadata be stored and backed up during the research process?

Example answer (generic project without sensitive data):
The data generated, processed and stored in this project does not contain any personal or confidential data and will be classified as "internal" According to institutional guidelines provided through the NTNU Storage Guide, data will be stored on the T:-drive, a part of the centralized file storage system with automatic backup, managed by the NTNU IT department.

How will data security and protection of sensitive data be taken care of during the research?

Example answer (confidential data):
Confidential data from business partners will be stored on NICE, a server with added security, multifactor authentication and automatic backup managed by the NTNU IT department. Access will be limited to the PI and primary research team members. All computers used in the project are centrally administered by NTNU IT Department, with firewalls and antivirus software automatically upgraded.
Relevant documents:

- NTNU Policy for information security
- NTNU Storage Guide

Legal and ethical requirements, codes of conduct

If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

Example answer (personal data):
To ensure compliance with GDPR, the project will send a notification form describing all relevant elements of the planned data processing to Norwegian Centre for Research Data (NSD)/SIKT for an assessment before data collection begins.
All personal data will be collected, transferred and stored according to their information classification and NTNU guidelines. The data will be pseudonymized as soon as possible and the code key will be stored separately from the data. Only the PI will have access to the code key. Access to the pseudonymized data will be limited to the PI and primary research team members. The PI will assign access rights and ensure that they are up to date. The code key will be deleted when the project ends and data for archiving will be further anonymized.
A risk assessment of data security will be performed before data collection starts.
Relevant documents:
Collection of personal data for research projects (NTNU)

How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Example answer:
Due to the novelty of the data expected to be generated in the course of the project and its potential for commercialization there will be restrictions to sharing some of the data. The University Technology Transfer office will be consulted prior to dissemination of unpublished data to protect IP.
For data owned by the collaborating company, a contract has been signed by both parties stating the project's aims, methods, ownership rights, and how the project outcomes will be used.
For the material containing personal data, all data will be aggregated and anonymized to be compliant with data privacy laws. Data that cannot be fully anonymized will not be published or will be archived with access restrictions according to the informed consent from the participants.

How will possible ethical issues be taken into account, and codes of conduct followed?

Example answer:
No human participants will be used in this research. Fully anonymized human tissue samples (blood, atheromas) will be obtained from blood banks or collaborators. All experiments involving human samples will be reviewed by the Regional Committee for Research Ethics. No data will be collected from human material prior to approval.

Data sharing and long-term preservation

How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Example answer:
Project data (with accompanying metadata) that constitute the basis for research papers will be made available at the time of publication at the latest.
Other data will be made available at the end of the project period.

How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

Example answer:
Most of the experimental data produced by the project can be shared openly and will be released through the repository NTNU Open Research Data under a Creative Commons public domain dedication (CC0).
The repository is part of DataverseNO, a manually curated Core Trust-certified data archive which provides access to the data for at least 10 years after deposition. A ReadMe-file providing necessary documentation will be included, as well as experimental protocols and information about relevant software and formats. Files in proprietary formats will be provided with open format copies.
De-identified personal data will be archived through the NSD Archive with the informed consent from the participants.
Personal data that cannot be shared due to Norwegian privacy law and the Health Research Act will be deleted when the project ends. Health data will be stored with restricted access for 5 years following requirements from the Regional Ethical Committee.

What methods or software tools will be needed to access and use the data?

Example answer:
A ReadMe-file will contain information about relevant software. Files in proprietary formats will be provided with open format copies.
Code developed through the project will be published through the GitHub/Zenodo-integration and provided with a DOI.
For personal data archived through the NSD Archive, there will be a manual data access request procedure ensuring that the data are only accessed for research use.

How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

Example answer:
Reuse opportunities for the experimental data are vast within the field of [xxx]. For this reason, the project participants aim to allow the widest reuse of the data and will release them through the repository NTNU Open Research Data under a Creative Commons public domain dedication (CC0).
The datasets will be issued Digital Object Identifiers (DOI) through the archive. Metadata will be publicly searchable and discoverable and provide information on how the dataset can be accessed.

Data management responsibilities and resources

Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

Example answer:
Together with the Project Manager, project team members will be responsible for data management, metadata creation, data security and quality assurance. Colleagues at NTNU will assist with quality assurance by providing critical feedback on data presented at group meetings and Departmental seminars.

What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Example answer:
A data steward will be employed in 40 % position to ensure proper data management and curation.