
Plan Overview

A Data Management Plan created using DeIC DMP

Title: Genomics Platform

Creator: Magali Michaut

Principal Investigator: Magali Michaut, Adrija Kalvisa, Heike Wollmann

Data Manager: Magali Michaut, Adrija Kalvisa

Project Administrator: Heike Wollmann

Affiliation: Københavns Universitet / University of Copenhagen

Template: UCPH Data Management Plan template

ORCID ID: 0000-0003-2002-2277

Project abstract:

The Genomics Platform offers next generation sequencing (NGS) support and service.

ID: 5186

Last modified: 13-05-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Genomics Platform

0. Project information

0.1 Workzone Case Number

NA

0.2 Project Members (including collaborators)

Magali Michaut, bioinformatics specialist
Adrija Kalvisa, bioinformatics specialist
Heike Wollmann, sequencing specialist
Carina Emmerson, lab technician

0.3 Other Project Documentation

The Novo Nordisk Foundation Center for Stem Cell Medicine, reNEW, is supported by a Novo Nordisk Foundation grant number NNF21CC0073729.

1. Data description

1.1 Describe what material and data will be collected, observed, generated, created or reused in the project. For the different types of research data, address their:

- **Origin / Source**
- **Estimated size / Volume**
- **Expected format(s)**

Sequencing data generated in-house by Illumina and Nanopore sequencers from sequencing libraries derived from non human samples or human cell lines (anonymous non sensitive data) from all our user groups.

Formats

- raw data in BCL files directly from the sequencers
- fastq files generated by us from the raw data
- processed data (e.g. bam files) generated by our pipelines

As of today, we have approx 45,3T of raw data coming from 755 runs (a run represents approx 60Gb of raw data).
We estimate our current data production around 300 runs (approx 18T raw data) per year.

1.2 Describe any material and data that contain sensitive information, in particular:

- **Personal data**
- **Human biological material, including biobanks**
- **Classified information**
- **Confidential (business) information**
- **Any other material or data that must be protected to safeguard the security of individuals, organisms, communities, organizations, etc.**

Generally, we do not handle sensitive data. In very rare occasions, we are receiving sequencing library pools originating from human sample DNA. We rely on the project lead classification of the data and on the specific instructions for processing needs.

2. Rights to research data

2.1 Address whether there are any access restrictions to material and data during the project. If so, describe who can have access to the material and data during the project, under which conditions, and in what timeframe.

Only KU researchers have access to the material and data until publication of the results.

2.2 Describe any material and data in your project that are subject to intellectual property rights.

IPR is with KU.

2.3 List any agreements or contracts set up in your project that contain provisions on rights to material and data, such as research collaboration agreements, non-disclosure agreements, material transfer agreements or license agreements.

We as the platform do not have any.

2.4 List any legislation, policies, guidelines or requirements that govern research data management in your project.

KU rules.

2.5 Describe whether, when and how research data may be used for other purposes (e.g. other research projects) and what arrangements will be made if a project member leaves the project and/or UCPH before the end of the project.

The data we generate are handled by the research groups. The platform staff retain no rights on the data if leaving the platform.

3. Ethical and legal approvals

3.1 Describe any ethical considerations and approvals necessary for the collection, processing or use of material and data in your project.

Responsibility of each research group for each of their projects.

3.2 Describe any legal agreements or approvals necessary for the collection and use of material and data in your project.

Responsibility of each research group for each of their projects.

4. Collection, processing and documentation

4.1 Indicate what methods will be employed in the project to ensure the consistency and quality of the material and data.

- Reproducible pipelines maintained in Git

4.2 Describe how the material and data will be organized and structured.

- Each run is identified by its unique ID (starts with the date e.g. 221205_VH00834_68_AACH37VM5).
- RAW and PROCESSED data organised in different folder, ordered by run ID.
- Run metadata are associated to run ID.

4.3 Describe how the collection, processing and analysis of the material and data will be documented.

- Each request has a submission id, and the bioinformatics processes are described and stored in a file named with the submission id in our git repository
- <https://bitbucket.org/mmichaut/snippets/src/master/Notes/>

4.4 Describe what the approach will be for naming and versioning of data files and material.

- Scripts are on git
- Data files are not modified

4.5 Indicate what metadata will be associated with the material and data.

- Request information and workflow in Jira
- Bioinformatics notes in text files on git

5. Storage and information security

5.1 Describe where and how the material and data will be stored and backed up during the project.

- On unicph network drives and datasets

5.2 Describe how the material and data will be shared with collaborators during the project (if any).

- On unicph network drives and datasets

5.3 For projects in which personal data are processed (including biobanks), please indicate whether a GDPR risk assessment and Data Protection Impact Assessment (DPIA) have been carried out.

Not at the platform level. We rely on the group leader to perform their projects evaluations.

5.4 For all research data types, describe what security measures will be established to prevent breaches of confidentiality. How will unauthorized access be prevented?

Unicph drive access control as well as general and sensitive dataset where only platform staff have access.

5.5 For all research data types, describe what security measures will be established to prevent loss of integrity. How will data and material be safeguarded against loss or modifications?

The folder shared with the users is read-only and the datasets are accessible only to platform staff.

5.6 For all research data types, describe what security measures will be established to prevent reduced availability of material and data. How will the continued accessibility of data and material to the relevant project members be assured?

Unicph network drive and datasets.

6. Data sharing

6.1 Describe which material and data will be made openly available for reuse.

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

6.2 Are there any material or data that cannot be shared openly? Explain why.

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

6.3 Describe any agreements required for sharing material and data.

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

6.4 Findable: What metadata will be created to allow the discovery of the material and data? How will others be able to discover the metadata?

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

Findable: Will the material and data receive persistent identifiers?

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

6.6 Accessible: Where will the material and data be deposited (e.g. in which repository)? How will others be able to access and retrieve the material and data?

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

6.7 Accessible: Will there be any conditions or restrictions for access to the material and data?

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

6.8 Interoperable: Will digital data be shared in file formats that others can easily open and reuse? Will information be provided on how files can be read and processed?

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

6.9 Interoperable: Will standards for metadata (including vocabularies and ontologies) be applied?

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

6.10 Reusable: What documentation is required for others to understand the material and data? How is this documentation provided?

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

6.11 Reusable: Are there any conditions for the reuse of the material and data by others? How are these conditions communicated? Will you apply standard data usage licenses?

This depends on the projects and is up to the group leaders. In general, when possible, data will be uploaded to community standard repository following the FAIR principles like GEO/ArrayExpress including raw, processed and metadata.

7. Long term preservation

7.1 Describe which material and data will be preserved after project end. Note that:

- Unless regulated otherwise, research data underlying published results must be retained for at least five years.
- Projects may be required to register and possibly deposit research data with the Danish National Archives at project end.

RAW, PROCESSED and METADATA that the platform generates is kept on unicph network drive or datasets for at least 5 years.

Describe how the material and data will be preserved after project end:

Where will the material and data be stored? In which formats? For how long? What documentation and metadata will be associated?

As a minimum, describe how (a copy of) the digital data and corresponding documentation will be made available to research managers and/or supervisors at UCPH.

RAW, PROCESSED and METADATA that the platform generates is kept on unicph network drive or datasets for at least 5 years.

7.2 Indicate who will have access to the material and data after project end.

The genomics platform staff and users.

8. Resources and responsibilities

8.1 Describe what costs are associated with the management of material and data during the project. How will these costs be covered?

Covered by the platform staff (and their salaries).

8.2 Describe what costs are associated with the preservation material and data after the project (according to the preservation plan and retention period outlined in question 7.2)? How will these costs be covered?

Covered by the platform staff (and their salaries).

As of today, storage/backup on unicph network drive does not give additional cost to the platform.

8.3 Indicate who will carry out the different tasks for managing the material and data during the project, and for long term preservation.

The platform staff.

8.4 Indicate who will be the main person(s) responsible during and after the project for:

- Allocating costs and resources.
- Obtaining approvals and ethical assessments.
- Meeting legal and contractual obligations.
- Controlling access to material and data.
- Maintaining the integrity and availability of published and preserved material and data.

Kim Jensen as Head reNEW Copenhagen for all that relates to the platform staff tasks and each group leader associated to the specific projects for all the rest

.