Data Management Plan: Plan 2

Version	1
Template	RFK-2020-test
Last modified date	2020-09-03 19:47:01Z
Last modified by	Adil Hasan (adilhasan2@gmail.com)
Last checked OK	2020-09-03 19:47:01Z
Editors	Adil Hasan (adilhasan2@gmail.com)

1. General Project Information

This section covers general details about your project.

1.1 Please provide the name of your project.

Modelling of rainfall in Bergen

1.2 Please provide a description of your project

This project will use micro-climate models to model the rainfall in Bergen. The project will produce simulation data and will make use of observational data to verify the models.

1.3 Which academic subject(s) does your project belong to?

Climate Science

1.4 Please provide the name of the project principle investigator

Harriett Somebody

1.5 Please provide the funding sources for this project.

Norwegian Research Council

1.6 Who will be the Data Officer for your project?

Clarence Curator

1.7 Does your project have the appropriate resources for the management of your data?

The project funds explicitly cover the management of data during the project's lifetime. The funds cover the storage (including backup) and personnel to manage the data. The project will make the data and software are publicly accessible once the project's findings have been published. We will store the data in the NIRD archive that does not charge for archiving public data. The software will be in github and we will actively ensure the code remains

usable whilst the data is relevant. The project will have in place a scheme to ensure a data custodian knowledgeable on the data will exist during the lifetime of the data. Through our data custodian we will do our best, in collaboration with other groups, to make sure the data remain accessible should the data require migration to a new format. In the past, tools have been backwards compatible thanks to the format being open enabling new tools to be developed that can read the existing data.

2. Data

This section covers the data that your project will create or use.

2.1 Please describe how your project will create and/or reuse data.

The project will make use of the microclima package (https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13093) to develop a model for precipitation in Bergen. Our code will be stored in github and will be publicly accessible. The models will be released in incremental versions after review by the project. We will use semantic versioning and will make use of R and Python packages. All dependencies will be described in the software package. Our simulation data will be stored in the community standard formats XXX and YYY. These formats are well described and understood by the community. A body maintains the formats and approves new versions. We will make use of existing observational data. Our project will produce simulation data that is unique as no other projects have used this approach to model precipitation.

2.2 Please describe how you will manage the intellectual property rights and ownership of your data.

We will make use of existing observational data. The data is openly accessible and abides by an open license (Creative Commons BY 4.0 license). We will give correct attribution to the produces of the data that we use. The simulation data we produce will also follow an open license (CC BY 4.0 license) to enable unrestricted reuse of our data. The software will be made openly accessible under the MIT license. We will follow the guidelines of our universities regarding copyright ownership and IPR ownership.

2.3 Describe how you will ensure compliance with legislation and institutional regulation?

Each institution that is a member of the project has a lead person. This person is in contact with their institution research office to make sure the project outputs are aligned with each institutions requirements. The project has an institutional board where any conflicts or discrepancies can be resolved. The contact people for the project are: Sven.Svenson@exampleone.no, Ann.Other@exampletwo.no, Stefan.Else@examplethree.org.

2.4 Please describe any ethical issues that may affect your data.

There are no ethical issues.

3. Documentation and Metadata

This section covers the information that will help you and your colleagues and other researchers to find and reuse your data.

3.1 What metadata and documentation do you plan to provide with your data?

We will follow the community standard metadata schemes (https://example.org/metadata1/description, https://example.org/metadata2/description). We will provide documentation on the software and how to use it including examples. We plan to provide documentation on the data through a publication that will describe how the model was developed and how the data was produced. Within the project we will create a small group responsible for documentation that will review documentation and identify missing documentation.

3.2 What data quality measures will you use for your data?

We will define a series of quality checks based on past experience. We will verify our parameters exist within the physical ranges and will verify the stability of the model by testing it under different conditions (ie different starting parameters) to reduce the systematic errors. Within the project we will define a group to review the process and results.

4. Storage

This section covers how you will store your data.

4.1 Where will you store your data?

NIRD

4.1.1 Please provide your NIRD project ID

NS9999K

4.1.2 Which NIRD services do you intend to use (please select all that apply)?

Computing Resources, Data Storage and Research Data Archive

4.2 What fraction of data do you plan to store each year from 2020-2024?

Storage forecast:

- 2020: 3 TiB, backup = 0%
- 2021: 6 TiB, backup ≤ 100%
- 2022: 6 TiB, backup ≤ 100%
- 2023: 6 TiB, backup ≤ 100%
- 2024: 4 TiB, backup ≤ 100%

4.3 What will you primarily use the storage for?

Computing (including HPC) input and output, Sharing data and Backup

4.3.1 How much data do you intend to transfer to/from the Computing platform?

1 to 10TB

4.4 Please briefly describe how you will ensure the safety of your data.

Our simulations will take a long time to run, so will be expensive to reproduce. We have requested that our data produced on NIRD is backed-up at NIRD. We will regularly checksum random portions of the data to ensure integrity. NIRD provides restricted access to the data and we will make use of this to provide access only to our project partners whilst the project is in progress. We will request audit information from NIRD every 6 months on who has accessed the data to verify no unauthorised access has taken place.

5. End of project

This section covers end of your project where your findings on your data have been published.

5.1 Do you plan to make some/all of your data available to others?

Yes, all/some

5.1.1 How will you make your data accessible?

NIRD Archive

5.1.2 Describe how you will select the data for reuse and ensure it can be reused?

The project will produce simulated data modelling the rainfall in Bergen. The simulated data are in a format that enables access by our peers in the climate science community. We plan to make all our simulated data accessible for reuse including the program configurations used to produce the simulated data. The code will also be accessible via github. As we are using community-wide format for our data a wide variety of tools exist to access the data.

5.1.3 When will your data be available for reuse?

As soon as the data has been analysed.

5.1.4 Apart from a possible embargo period, are there any other restrictions on the reuse of your data?

None

5.2 Please provide any additional information you think is relevant to your plan.

No additional relevant information provided.