

---

## Plan Overview

*A Data Management Plan created using DeiC DMP*

**Title:** DTU Compute - PhD example

**Creator:** Tommy Sonne Alstrøm

**Principal Investigator:** Tommy Sonne Alstrøm

**Data Manager:** Tommy Sonne Alstrøm

**Affiliation:** Danmarks Tekniske Universitet / Technical University of Denmark

**Template:** DTU data management plan

### **Project abstract:**

This template is designed to outline a minimum volume DMP and can be used as template for smaller projects, such as a single PhD project.

**ID:** 2311

**Last modified:** 30-06-2023

### **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

# DTU Compute - PhD example

---

## Data Collection

**Describe the data that will be collected.**

### Primary material

The digital raw data acquired in this project will be images of dead animal cells, acquired by an instrument "Biosense Oscilloscope". The images will be stored as files in the lossless image format; TIFF.

The images are analysed using computer code that is coded in Python 3.7.

### Secondary material

Secondary material produced is figures and tables, which are created using the computer code.

### Estimated amount

The estimated amount of research data is 500 GB.

### Data structure and versioning

The digital raw data will be unversioned flat files. The files will be organized as:

<data folder>/<dataset>/<yyyymmdd>/files

The computer code will be stored and versioned using Git

### DMP revision

Once exact file names are known the DMP will be revised with details filled in.

**Describe any restrictions to the data.**

Since this project is a PhD scholarship with no partners or commercial interest, and the data is not covered by GPRD, there are no restrictions on the data.

## Data Storage

**Describe the IT infrastructure to be used.**

### Digital Raw Data

The digital raw data will immediately (or shortly after) be uploaded Perforce (<https://archive.compute.dtu.dk>) under the folder

//comp-data/projects/PhDTitle/

where the data will automatically be replicated to the [DTU Compute file servers](#) under the path

/dtu-compute/comp-data/projects/PhDTitle/

### Computer code

Computer code will be stored at <https://lab.compute.dtu.dk>/<student initials>/

### Secondary material

Secondary material will be stored either in Perforce, Git or on networks drives. Only secondary material that goes into a research publication is managed

### Access control

The PhD student will manage access to the digital raw data using Perforce sharing module, and access to the computer code will be managed using gitlabs sharing module.

## Documentation

**Describe the metadata to be associated with the data.**

Metadata will be generated as described on [www.openml.org](http://www.openml.org). As a minimum, the following will be defined for the digital raw data:

Number of instances, number of features, missing values, number of classes, and majority class percentage.

All other research data created will be either self-documented (computer code) or documented in research papers (tables and figures)

**Describe the types of documentation that will accompany the data.**

The raw digital data will be documented as acquired in an electronic labbooks. Each experiment will have a raw text file (located in the same folder as the raw digital data) that describes the experimental settings that were used during the experiment, and when applicable, a document describing the experimental design.

### Reproducibility

For each publication, a full paper package will be created that describes the full computational environment that was used to create the secondary material that went into the publication. The recipe X from <https://crdp.compute.dtu.dk> will be used.

## Data Sharing

**Describe which data will be shared.**

Research data that goes into a publication, either as primary material or secondary material will be shared. The computer code will be shared under the [Apache 2](#) license and all other data will be shared under the [Create Commons CC-BY](#) license.

Software that will be required to use the data will be described per publication basis

### Revision

Every time a full research publication is released, the DMP will be updated with a direct link to the research data on <https://data.dtu.dk>.

**Describe how the data will be shared for possible reuse.**

The digital raw data will be shared using a DOI on <https://data.dtu.dk>.

For publications that contain research data for less than 1 GB, the full research data set (both primary and secondary materials) will be uploaded directly to data.dtu.dk. For research publications where uploading to data.dtu.dk is not feasible, the exact paths to Perforce and Git will be described instead.

## **Long-term Preservation**

**Describe how data will be archived beyond the scope of the research project.**

**All unpublished research data** will be archived for at least five years.

**All published research data** will have a description on data.dtu.dk and long-term preservation will be handled by DTU Library. This data should be preserved for at least 10 years.

**Readability of the data** is not ensured as such as commercial software packages are used and it is assumed these file formats are still readable over the next 10 years.