
HPC environment FMHAT

A Data Management Plan created using DMPonline

Creator: Sissal Erenbjerg

Affiliation: Københavns Universitet / University of Copenhagen

Template: National Research council

ORCID iD: 0000-0001-7619-4049

Project abstract:

This project is to setup and continue a HPC center in the Faroe Islands. For this particular project the focus is on Faroe shelf 3d ocean modelling with FarCoast and ROMS.

Last modified: 04-11-2022

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

HPC environment FMHAT

Data Collection

Simulated physical ocean parameters and tracers from the Faroese shelf from 2010-2020. Including but not limited to current velocities, salinity, temperature etc.

This will be about 5-10 TB of data, and will be saved in netcdf format. The netcdf format is comprised and easy to store longterm.

The data will be collected by running a longterm computer simulation of FarCoast160, on the local HPC environment at Fiskaaling. The data will be named by date and saved in a netcdf format with meta data. As in the standards used by the Regional Ocean Model System (ROMS). The entire dataset will be validated against existing data at Fiskaaling.

The files will be located at offline servers at Fiskaaling, but available upon request for non-commercial use.

Documentation and Metadata

The data will be stored in netcdf format and the netcdf library is necessary in order to read and use the dataset. The means that each file both contains the data and the metadata, each day will be saved in a separate file including 24 one hour timestep averages. Each file will also contain the information of model version, period, time-step, and included model options, dimensions of the grid and a short description of all variables (including a short name, a long name, units and variable dimensions). This is the standard with in the Regional Ocean Model System (ROMS).

Ethics and Legal Compliance

Since this data only contains simulated physical variables, in a simulated environment, forced by publicly available data, at least for non-commercial use, there should not be any ethical concerns to our knowledge. The entire dataset can be shared with out restrictions, for non-commercial use. This project does not use any private or personal data (GDPR does not apply).

The team at Fiskaaling will have a two year privilege for possible publication of the data, afterwards the data will be freely available for anyone that wishes to use the data for non-commercial use. The entire model (ROMS), forcing (WRF, Copernicus) and dataset is open source.

Storage and Backup

The data will be stored offline on servers at Fiskaaling. Where the data will also be backed up. The responsibility of backup is within the team of Fjord dynamics at Fiskaaling. The data will also be possible to recover by rerunning the model, potentially on another machine. The ability of creating restart files within FarCoast and ROMS ensures faster data recovery as well.

The access to the data will be at Fiskaaling since the data will mainly be stored offline, however the access can be gained with an ssh tunnel into the backup system.

Selection and Preservation

The model setup itself does not require a large space, >100GB (excluding forcing) whereas the dataset will become rather larger, dependent on number of tracers that will be included (at least 3). However the dataset will be kept off-line at Fiskaaling. The number of tracers will also depend on users in the general research environment of the Faroe Islands. Where the data set can be used as hydrodynamical input for sea-lice dispersion, wave and coastal simulation, resuspension of POM, fish welfare in fish farming cages and other.

The data repository will stay offline at Fiskaaling due to the large size of the data set. This is included in the budget of Fiskaaling.

Data Sharing

The data will be validated in scientific papers and thus be available for others to use in scientific prospects. The data will be available for other researchers 2 years after the data is created. Some parts of the data can be shared through GitLab, as well as parts of the code formulation can be shared on GitHub.

This will ensure that the data set is findable, accessible, interoperable and re-usable (FAIR).

We will require a data sharing agreement with any users, as we wish that the data set remains open and freely available and is not used for commercial purposes.

Responsibilities and Resources

The main responsible for this DMP is Sissal Erenbjerg, PHD student at Fiskaaling, this DMP will be reviewed and revised annually by her, and the colleagues at Fiskaaling including Trondur Tummason and Jóhannus Kritsmundson at the Fjord oceanography department and Gunnvør á Nórði and Birgitta Andreassen at the ecology department. If funding for this project is provided, this DMP will be publicly available online at dmponline.deic.dk.

For Backup 10 TB of offline storage will be purchased at a cost of about 7000 DKK.