

For questions or more information, please contact: datahelp@ucar.edu

Data Management Plan Sample: A Sample Plan for CESM Model-Based Projects
[DASH Preferred Data Management Plan Format]
[Primary Funder Requirements: NSF/NOAA/NASA]
[Solicitation #]

Products of the Research (Type of Data Produced)

The data created by this project will be exclusively model output from two global runs (i.e. data generated by the Community Earth System Model (CESM) component models at run time). In particular, the data is categorized as “CESM Experiment integrations” per the CESM Data Management and Data Distribution Plan (<http://www.cesm.ucar.edu/management/docs/data.mgt.plan.2011.pdf>).

Data Format (Data Organization and File Format)

The model output data will be in netCDF format, and the file size will range from 1GB to 150GB. The data files will be named and structured using the following convention:

CaseName.ComponentName.OutputType.Date.nc

This convention is defined and described by the CESM webpage (under the “CESM1 Model Output Data Filenames” section): (http://www.cesm.ucar.edu/models/cesm1.0/filename_conventions_cesm.html):

Metadata

The technical or file level metadata for each project data file will be recorded automatically through the built-in header section of the netCDF file, which is also compliant to the Climate and Forecast (CF) metadata convention. Metadata that will help others in searching, discovering, and recreating the runs, including input files, versions of codes/scripts used, and data process workflow, will also be documented automatically, and supplemented manually as needed, and be made available through NCAR DASH Search and Discovery system along with the final model output data.

Access to Data and Data Sharing Practices and Policies

During the project lifecycle, all data generated during the project will be stored and available for project use and for other interested parties on a request basis that recognizes proprietary access to the project members as these data are initially the property of the CESM community. The final results of the project will be moved into the public domain and be made publicly available through the Earth System Grid (<https://www.earthsystemgrid.org>). While the public access is free of charge, a registration process with the Earth System Grid will be required. This is to help in tracking the data’s distribution and usage. Additional relevant policy statements that are outlined by the CESM Data Management and Data Distribution Plan, which will be followed by this project, are:

- CESM Experiment data shall be available to members of any CESM Working Group no later than six (6) months following the conclusion of the Experiment Simulation.
- CESM Experiment data shall become Public as soon as a scientific paper on the results has been submitted by the PIs who originated the model run or one (1) year after the end of the simulation, whichever comes sooner.

Policies for Re-Use, Re-Distribution, and Production of Derivatives

It is anticipated that the two global runs under different climates with full biogeochemistry will generate interest in the broader community, but likely only from specialists in the field who would be familiar with model data of a similar nature and volume. There is also likely to be interest in the individual region simulations. Those who are interested in using the project’s final data can obtain access via Earth System Grid as described in the above section. Additionally, the users are expected to cite the project and the corresponding data according to the citations that will be established by the project team.

Archiving of Data (Data Storage and Preservation of Access)

The data produced by this project is specialized, and therefore, has a relatively short expected lifetime, as new developments and increased computational capability tend to make them obsolete within a few years. As a result, there are no long term archival plans. However, the capability will be maintained to re-run any numerical experiment and re-generate data that are requested after the project, provided the computational resources can be found.

Additionally, following the recommendation by the CESM Data Management and Data Distribution Plan, data will be retained for a period of seven (7) years from the date of the end of the simulation.

Cost of Implementing the DMP

Given the expected users and the existing, available practices as established by the CESM Data Management and Data Distribution Plan, the duties will be light and is currently planned to be <10% Full Time Equivalent (FTE). Infrastructure wise, request for computing resources from the NCAR Computational & Information Systems Laboratory will include sufficient data storage over the duration of the project, but for less than 1 year beyond.

Roles and Responsibilities

Project data will be managed by the project's scientists. These duties will include generating thorough documentation, facilitating data distribution, tracking of processed data, performing quality assurance, and ensuring that all data could be readily reproduced by re-running experiments, if necessary.

SAMPLE