

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

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

Products of the Research (Type of Data Produced)

During the proposed project, a series of climate model simulations will be carried out, and as a result, a final, complete reanalysis dataset featuring global hourly analyses on a 40-km horizontal grid, with 28 vertical levels covering the period of 1985-2005 will be produced. In order to create this final reanalysis dataset, the following openly available and publicly accessible input files will be used:

- Initial Land Surface Conditions: NASA 1 degree Global Land Data Assimilation System (GLDAS) - Version 1 (GLDAS - 1: GLDAS_NOAH10SUBP_3H) by Rodell et al. 2004 and retrieved from <http://disc.sci.gsfc.nasa.gov/uui/datasets?keywords=Hydrology>
- Initial Sea-Surface Temperatures (SSTs): National Centers for Environmental Prediction (NCEP) Version 2.0 by Reynolds et al. 2002 and retrieved from <http://www.ncdc.noaa.gov/oisst>
- Terrain File USGS 30-second terrain packages with MM5's TERRAIN interpolated using the Cressman objective option with a radius of influence of 1.5 times the grid increment for smoothing.

These input files will be processed primarily using the Pennsylvania State University (PSU)/NCAR mesoscale model (MM5) and its sequel, the Weather Research and Forecasting (WRF) Model. Additional models and software will be created throughout the proposed project to process the dataset as needed. Further, the performance of the dataset will be verified against selected, previously recorded weather phenomena. The final versions of the models and software used during the project and the validation results will be documented and shared along with the dataset as appropriate.

Data Format (Data Organization and File Format)

The reanalysis dataset will be in the netCDF format in order to accommodate the need to have gridded fields (the dataset will include 0.4 degree grid increment and 28 vertical levels) and easy data access with common tools in the user community. Since the dataset will include hourly data from 1985 to 2005, there should be 183,960 files (21 years x 365 days x 24 hours) with a total dataset size of approximately 27 Terabytes. The file naming convention is designed to reflect the hourly nature of the data files. The proposed file name structure is as follows: CFDDA_YYYYMMDDhhmmss.mdv.nc, where: YYYY= year, MM = month, DD = day, hh = hour, mm = minute, and ss = second (UTC).

Metadata

Leveraging the self-describing capability of the netCDF format, the reanalysis dataset will be primarily documented using netCDF's built-in metadata structure. Specifically, the dataset's longitude, latitude, layer index, number of surface layer, time information, and all of the variables will be described in the header of each data file from the dataset. The dataset will also aim to follow the Climate and Forecast (CF) Conventions for its variable names. Additionally, the project team intends to generate a user manual in Word or PDF format to accompany the final dataset and to further document the project with additional textual metadata.

Access to Data and Data Sharing Practices and Policies

Due to the original conditions of the project sponsorship, the reanalysis dataset will need to observe an initial embargo period of three years. During this time, only the direct team members can have access to the dataset and its related components, which will be stored on NCAR Research Applications Laboratory's internal server with password protection and security measures implemented. After the embargo period, the rights to the dataset will be transferred to UCAR/NCAR, and based on prior

agreement and arrangement that were obtained, the dataset will be made available through NCAR Research Data Archive (RDA - rda.ucar.edu), which is managed by NCAR Data Support Section under the Computational and Information Systems Laboratory. The web-based, public access to the dataset through RDA is free of charge with a valid user account (registration for the RDA user account is also free). The data files along with the related software and documentation of the dataset will be obtainable using the download capabilities provided by RDA.

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

Through NCAR RDA, the reanalysis dataset will be made publicly available and open accessible to other researchers working with climate models, and for use by the general public as they see fit. The dataset should be properly cited when being used/re-used for the purposes of creating new/derivatives products. The recommended citation format will be made available as part of the dataset's final documentation.

Archiving of Data (Data Storage and Preservation of Access)

When the reanalysis dataset and its associated documents and software are archived with NCAR RDA, all components will be made publicly available and openly accessible for at least five years. After the initial five years, RDA can review and update as needed the availability and accessibility of the dataset per RDA's resources. At the time of the project proposal, RDA does not have any foreseeable challenges for providing the necessary infrastructure, including storage space and backup capabilities, and personnel resources to maintain and manage the dataset for beyond five years. If the dataset will need to be relocated to ensure future storage and preservation of access to the dataset, RDA will be responsible in assessing the appropriate archive/repository and providing redirection to the new archive/repository.

Cost of Implementing the DMP

The sponsor program and program funder will be mainly responsible in assisting with the cost of implementing the data management plan during the project. For instance, the documentation, creation of the associated software, and production of the final dataset will all be part of the routine project activities. After the dataset is archived with NCAR RDA, the infrastructure and the personnel resources needed to provide continuing management and maintenance of the dataset files as well as the related documents and the software will be part of the RDA's ongoing operational costs (please see the letter of support from the RDA).

Roles and Responsibilities

During the project, the PIs of the project will primarily be responsible for overseeing and reviewing the progress of the data management plan. The PIs will also work with the project manager to coordinate necessary effort with the project team members. The PIs intend to review the data management plan every 3 months during the project's duration and to provide updates/record completion of tasks as needed. At the end of the project, the project team will also be responsible in documenting the status and accomplishments according to the final data management plan. During the period when the dataset is being transferred to NCAR RDA, the PIs and the project manager will ensure RDA receives all the relevant data files and any relevant documentation/software for the dataset, and will validate the dataset presentation on the dataset homepage (landing page) in the RDA. Once the dataset is transferred to RDA, RDA and its staff will be responsible for the ongoing management and maintenance of the dataset.

Additional Comments

The project team intends to acquire a digital object identifier (DOI) for the final, complete version of the reanalysis dataset; this DOI is expected to be part of the citation that will be recommended for the dataset.

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