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

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

Products of the Research (Type of Data Produced)

The main data in this project are synthetic observations generated from the LFM-helio model. Coronal synthetic data include CME width, speed, source magnetic flux, and orientation/location relative to background corona -- all of which are directly determined from parameters. The LFM-helio model has existing capability to forward model heliospheric imaging data, and so we extract synthetic images at a range of ICME projections. We identify those cases where 3D reconstruction is possible, and apply existing SolarSoft analysis tools to obtain speed, center of mass, and width as a function of distance from the Sun. It is straightforward to extract in-situ observations, such as magnetic strength and orientation, speed, density and temperature, as a function of time. We do so for all model realizations where a synthetic HI ICME can be identified at 1 AU, and for a range of simulated spacecraft crossings.

In addition, we will store a subset of the full datacube results from the LFM-helio model. This output includes information about the eight MHD state vector variables on a computational grid that covers the the inner heliosphere region.

Finally, we will store best fit parameter values for the observational and synthetic climatology, and a distribution of model parameters.

Data Format (Data Organization and File Format)

Both the LFM-Helio model and the synthetic data utilize the HDF4 portable data format. The sizes of the datasets produced by this project vary widely depending on the duration of the runs. The 3D HDF files produced by the low-resolution LFM-helio are approximately 235MB in size, and a typical run produces files on 30-minute to 1-hour cadence. Runs for this proposal will typically have a size of 115GB. The naming convention for the data files will be created to reflect mainly the parameter measured in each file, and the final naming structure will be described and shared via NCAR Search and Discovery system.

Metadata

The technical or file level metadata for each project data file will be recorded automatically through the built-in capability of the HDF file, which includes the following key information: descriptions of the size of the grid and units of the variables as well as the parameters used. Metadata that will help others in searching, discovering, and understanding the project, such as highlights and explanations of unique features of a specific simulation design, will also be documented and made available through NCAR DASH Search and Discovery system.

Access to Data and Data Sharing Practices and Policies

Publications will be the main method for accessing the observational and model distribution parameterization from this project, and the primary papers will be published during the project lifecycle. All files used to support publications will also be made available to members of the scientific community upon request (the instructions regarding how to request access will be provided and shared via NCAR Search and Discovery system). Where possible, visualization and analysis tools will be supplied as well.

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

Other heliophysics scientists who want to do analyses of processes of Interplanetary Coronal Mass Ejections (ICMEs) are the most likely, primary users. However, once the results from this project are shared via publications, NASA's HEC-ASS system, and NCAR's Search and Discovery system, the

results from this project are also open and free for others to use according to NASA and UCAR/NCAR's policies. 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 3D HDF files produced by the high-resolution LFM-helio are approximately 1GB in size and contain the output of one time step. It is not practical to save full 3D data sets for all 15,000 runs per year, so we plan to save synthetic data obtained for all runs (minimal space needed), and 3-5 states from a subset of runs, ~1000, resulting in annual requirement of 5TB of space.

Synthetic diagnostic files along with key frames from the simulation will be transferred to the NASA High-end Computing Capability Archival Storage System (HEC-ASS) for long-term preservation. Along with the output data the input files, code executables, and source code versions will be transferred to the HEC-ASS for storage. The files stored on the HEC-ASS will allow for rerunning the code as needed to reproduce any simulation output.

By storing and archiving the project files with the HEC-ASS, the files are expected to be available and accessible for as long as these archives will remain available. It is also expected that these archives will help in providing the long-term strategy and services for maintaining, archiving, curating, and preserving the files.

Cost of Implementing the DMP

During the project lifecycle, the related data management activities will be integrated as part of project tasks. Beyond the project period and by using NASA's HEC-ASS, resources to provide long term data management/stewardship of the project results will be provided by the archives.

Roles and Responsibilities

Project data will be managed by the project's scientists. These duties will include but might not be limited to 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. Additional consultations regarding best data management practices will be conducted with NCAR's Data Curation & Stewardship Coordinator as needed.