

Daymet Follow-On Proposal: Data Management Plan

R.B. Cook, M.M. Thornton, and P.E. Thornton

Environmental Sciences Division, Oak Ridge National Laboratory

The following Data Management Plan was part of the NASA ROSES 2012 Proposal *Daymet Follow-On: Surface Weather Data with Uncertainty Quantification for Terrestrial Ecosystem Process Models* ([summary](#)) submitted to the Terrestrial Ecology Program. It is presented as an example plan.

Data Management Plan

Information about the data

Meteorological station data used as input to Daymet (daymet.ornl.gov) will be obtained from the GHCN-D collection at NCDC as well as regional meteorological offices. These data will be checked for quality and formatted into a consistent comma separated values format, followed by automated ingest and storage as netCDF formatted files.

The data generated by the proposed project will consist of daily, gridded surfaces of climatological data (temperature, precipitation, humidity, and radiation) for North America, Europe, Australia, and portions of Africa and Asia. The Daymet data will be compiled for distribution in 2 degree tiles in CF-1 compliant netCDF format in an equal area projection suitable for each specific region. The data product has been thoroughly documented (Thornton et al. 1997, 2000; Thornton and Running 1999). In addition, the individual meteorological station data used for each tile will be made available in CSV format. New data sources developed under the proposed project may impose distribution restrictions on the raw station data, but we will not engage new data sources unless permission is granted in advance for open distribution of resulting gridded surface weather products.

Metadata content and format

Metadata will be comprised of two formats—contextual information about the data in a text based document and ISO 19115 standard metadata in an xml file. These two formats for metadata were chosen to provide a full explanation of the data (text format) and to ensure compatibility with international standards (xml format). Daymet will follow best practices for compiling metadata (Hook et al., 2010).

Short-term storage and data management

Data products will be generated twice during this project. For each geographical region, we will produce a test version of the data around the mid-point of the project and a final version at the end of the project. The test version will be made available to modeling and applications teams for evaluation and the final version will be provided to the ORNL Distributed Active Archive for long-term data management. During the course of this project, we will generate daily and monthly backups of the data files, which will be retained by the Environmental Data Science and Systems group at Oak Ridge National Laboratory.

Policies for access and sharing

The test Daymet data will be released to the public as soon as our initial quality checks have been completed and the data has been prepared, now planned for the end of Year 1. There is no period of exclusive use by the data collectors. Users can access documentation and Daymet climate data files via the Daymet Web site (daymet.ornl.gov).

Long-term storage and data management

Our intent is that the long-term high quality data product generated by this project will be available for use by the research and policy communities in perpetuity. The investigators have made arrangements for long-term stewardship and curation at the NASA-funded Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC; see letter of support). The standardized metadata record for the Daymet data will be added to the metadata record database at the ORNL DAAC and NASA Earth Observing System Data and Information System clearinghouse (ECHO), so that interested users can discover the Daymet data along with other related Earth science data. Daymet and the ORNL DAAC have a standardize data product citation (http://daymet.ornl.gov/sites/default/files/Daymet_Reference_CitationDOI_link.pdf) which enables users to cite the source of the data, gives credit to the Daymet team, and enables journal readers to obtain a copy of that product. The ORNL DAAC will include a Digital Object Identifier, which will enable clear citation to the product as well as tracking of use of Daymet.

References

Hook, L. A., Vannan, S. K. S., Beaty, T. W., Cook, R. B., & Wilson, B. E. (2010). Best practices for preparing environmental data sets to share and archive. Oak Ridge National Laboratory Distributed Active Archive Center, USA.
doi:10.3334/ORNLDAAC/BestPractices-2010

Thornton, P. E., H. Hasenauer and M. A. White 2000. Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. *Agricultural and Forest Meteorology* 104: 255-271.

Thornton, P. E. and S. W. Running 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agricultural and Forest Meteorology* 93: 211-228

Thornton, P. E., S. W. Running and M. A. White 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology* 190: 214-251