**MSMTIP Phase II: Data Management Plan**

D. Huntzinger\(^1\), R. B. Cook\(^2\), Y. Wei\(^3\), C. Schwalm\(^4\), Kevin Schaefer\(^5\), Anna Michalak\(^6\), Yuanyuan Fang\(^7\), and Josh Fisher\(^8\)

\(^1\)Northern Arizona University, \(^2\)Oak Ridge National Laboratory, \(^3\)University of Colorado, \(^4\)Carnegie Institute, and \(^5\)JPL NASA

The following Data Management Plan was part of the NASA ROSES 2013 Proposal North American Carbon Program Multi-scale synthesis and Terrestrial Model Intercomparison Project (MSMTIP) Phase II (Summary; Web site) submitted to the Terrestrial Ecology Program. It is presented here as an example plan.

**INFORMATION ABOUT THE DATA**

**Driver data:** The data generated by the proposed MsTMIP project will consist of model driver data, model output data, and observational data that will be used to compare with model results. The model driver and model output data will be gridded at 0.5° by 0.5° for the globe and 0.25° by 0.25° degrees for North America. For the proposed future simulations as part of MsTMIP Phase II, we will assemble a full suite of environmental driver data (covering 2011 to 2100) that will smoothly transition with those used in MsTMIP Phase I: climate, atmospheric CO\(_2\) concentrations, satellite-based phenology, land use/land cover change, and nitrogen deposition. This will ensure a seamless transition between MsTMIP Phase I and Phase II driver data. All MsTMIP environmental driver datasets from Phase I and Phase II will be archived at ORNL (see sections below on short and long-term data management).

**Simulation output:** As part of MsTMIP Phase II, we will run for each TBM, an ensemble of carbon cycle projections using climate projections from CMIP5 models (http://cmip-pcmdi.llnl.gov/cmip5/) and atmospheric CO\(_2\) concentrations from the IPCC representative concentration Pathways (RCP) scenarios (van Vuuren et al., 2011). In addition, as part of MsTMIP Phase II, we will perform a series of climate perturbation experiments using the 6 models in the JPL Model Farm. Using the Model Farm, we will re-run the historical simulation, but perturbing the MsTMIP Phase I atmospheric CO\(_2\) and climate drivers (temperature, precipitation, etc.) one at a time by ±20%, while leaving the other climate factors unchanged. All MsTMIP simulation output from Phase I and Phase II will be archived at ORNL (see sections below on short and long-term data management).

**Observational data:**

As part of MsTMIP Phase II, we will assemble a variety of observational data constraints, including soil temperature, soil moisture, snow depth and density, LAI, regional impacts of extreme events on carbon stocks and fluxes, and site-level observations of the sensitivity of carbon uptake and release to changes in temperature and atmospheric CO\(_2\) concentrations.

In order to evaluate the simulated physical environment within the models, we will compare model estimates of LAI and well as soil moisture and temperature as a function of depth and snow characteristics (particularly snow depth and density) with available observations (Table DMP1).

In addition, we will also compile reference data products that quantify the impact of extreme climatic events on carbon stocks and fluxes. We will start by collecting information related to Hurricane Katrina, 2000-2004 turn-of-the-century drought in the American Southwest, and the loss of carbon uptake during the 2003 summer European drought. We will add to this dataset as more observational or reference datasets become available.

As part of MsTMIP Phase II, we will also compile observations of Net Primary Production (NPP) from the 6 ecosystem controlled warming experimental sites (Lu et al., 2013), as well as NPP and CO\(_2\)
concentration information from at least 4 of the Free-Air CO\textsubscript{2} Enrichment (FACE) experimental sites (e.g., DeKauwe et al., 2013).

In addition to the observational data to be collected as part of MsTMIP Phase II, several observationally based gridded dataset for several variables have been compiled as part of MsTMIP Phase I (Table DMP2).

All MsTMIP model driver, model output, and observational data will be compiled for distribution in CF-1 compliant netCDF format in a geographic projection. The final MsTMIP data products distributed to users will occupy less than 20 terabytes.

Table DMP1: Physical environment observational datasets for MsTMIP Phase II

<table>
<thead>
<tr>
<th>Variable</th>
<th>Dataset</th>
<th>Domain</th>
<th>Coverage</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil moisture &amp; temperature</td>
<td>International Soil Moisture Network (ISMN)</td>
<td>Global</td>
<td>Site-level; hourly/monthly; 1950-2013</td>
<td>Dorigo et al. (2011)</td>
</tr>
<tr>
<td>Soil moisture</td>
<td>Airborne Microwave Observatory of Sub-canopy and Subsurface (AirMOSS)</td>
<td>North America</td>
<td>Gridded; hourly/monthly; 2012-15</td>
<td>Tabatabaeenejad and Moghaddam (2011)</td>
</tr>
<tr>
<td>Soil temperature</td>
<td>Russian Soil Temperatures</td>
<td>Russia</td>
<td>Regional; monthly; 1898-1990</td>
<td>Zhang et al. (2001a, 2001b)</td>
</tr>
<tr>
<td>Soil temperature, carbon fluxes, LAI</td>
<td>AmeriFlux, FLUXNET</td>
<td>North America, Europe</td>
<td>Site-level; hourly; 1990-2013</td>
<td>Baldocchi et al. (2012); Papale et al. (2012)</td>
</tr>
<tr>
<td>Soil temperature</td>
<td>Boreas &amp; LBA</td>
<td>Regional</td>
<td>Site-level; monthly; duration of campaigns</td>
<td>Sutton et al. (1998); Hinkel (1998)</td>
</tr>
<tr>
<td>Snow characteristics</td>
<td>NOAA Snow Cover</td>
<td>Northern hemisphere</td>
<td>Regional; weekly; 1967-2013</td>
<td>Armstrong and Brodzik (2002)</td>
</tr>
<tr>
<td>LAI</td>
<td>MODIS</td>
<td>Global</td>
<td>Gridded; weekly; 2002-2013</td>
<td>Wang et al., 2001</td>
</tr>
<tr>
<td>NDVI, LAI</td>
<td>GIMMSg</td>
<td>Global</td>
<td>Gridded; monthly; 1982-2011</td>
<td>Tucker et al., 2005</td>
</tr>
</tbody>
</table>
Table DMP2: Gridded observationally based datasets complied under MsTMIP Phase I.

<table>
<thead>
<tr>
<th>Dataset (Time)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net Ecosystem Exchange (1982-2008)</td>
<td></td>
</tr>
<tr>
<td>Total Respiration (1982-2008)</td>
<td></td>
</tr>
<tr>
<td>Sensible Heat (1982-2008)</td>
<td></td>
</tr>
<tr>
<td>Leaf Area Index (2000-2010)</td>
<td>Yuan et al. (2011)</td>
</tr>
</tbody>
</table>

As described above, MsTMIP Phase II proposes to acquire, process, and use model driver and observational data generated by other researchers and held at a variety of data centers. Full credit will be given through citation or acknowledgement to the data archive and to those who initially generated these data (NACP, 2007). This practice will also ensure that future researchers will be able to access the same data used in MsTMIP Phase II.

**METADATA CONTENT AND FORMAT**

Metadata for MsTMIP data products 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). MsTMIP will follow best practices for compiling metadata (Hook et al., 2010).

**SHORT-TERM STORAGE AND DATA MANAGEMENT**

The data products will be made available to the project team and workshop participants for quality checking and initial analysis. During the course of the project, data (e.g., driver data, simulation output, and any subsets of observation datasets used for comparison with the MsTMIP simulations) will be made available via the MsTMIP Data Repository (Theme 3). The final version of all MsTMIP data products will be provided to the ORNL Distributed Active Archive Center (DAAC) for long-term data management after documentation and metadata are prepared.

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 MsTMIP data products will comply with the U.S. Global Change Research Program’s policy of full and open access to data generated by research projects (http://www.gcrio.org/USGCRP/DataPolicy.html). The data will be released to the public as soon as our initial quality checks and preliminary data assessment have been completed and the data has been prepared. There is no period of exclusive use by the data collectors.

The MsTMIP project will have an open access policy, with the expectations that users would submit or provide an "abstract of intent" that is a heads up on how they plan to use the MsTMIP products. Ideally this would occur before the work has started and MsTMIP participants and core team members would have a chance to contribute to the work. As part of this open access policy, MsTMIP will request that manuscripts using MsTMIP products be sent to the MsTMIP core and any relevant MsTMIP participants for review well before they are submitted for publication, so that modelers and the MsTMIP team have
adequate time to provide input on the quality and limitations of MsTMIP products. The goal is to ensure an opportunity for meaningful input in a timely manner.

**LONG-TERM STORAGE AND DATA MANAGEMENT**

Our intent is that the data products generated by the MsMTIP project will be discoverable and available for use by research 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). MsTMIP will prepare for submittal to the ORNL DAAC the driver data, simulation output, and those observational datasets that were modified for comparison with the MsTMIP simulations. For observational data products used in MsTMIP Phase II that are largely unmodified from the products archived elsewhere, MsTMIP will provide a detailed citation / link, so that users can obtain the data used. As part of the DAAC’s curation, the standardized metadata record for data generated by MsTMIP 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 data products along with other related Earth science data. The ORNL DAAC will generate a standard data product citation (Data Citation Policy) that enables users to cite the source of the data, gives credit to the MsTMIP team and modelers, and enables journal readers to obtain a copy of that data product. The ORNL DAAC will include a Digital Object Identifier, which will enable clear citation and linking to the product as well tracking of use of MsTMIP data products.

**REFERENCES**


