MsTMIP Phase II: Data Management Plan

D. Huntzinger[1], R.B. Cook[2], Y. Wei[2], C. Schwalm[1], Kevin Schaefer[3], Anna Michalak[4], Yuanyuan Fang[4], and Josh Fisher[5]

[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 (MsTMIP) 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_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

Variable	Dataset	Domain	Coverage	Reference(s)
Soil moisture & temperature	International Soil Moisture Network (ISMN)	Global	Site-level; hourly/monthy; 1950-2013	Dorigo et al. (2011)
Soil moisture	Soil Moisture Active Passive (SMAP)	Global	Gridded; weekly; 2014-15	Cow et al., (2010)
Soil moisture	Airborne Microwave Observatory of Sub-canopy and Subsurface (AirMOSS)	North America	Gridded; hourly/monthly; 2012-15	Tabatabaeenejad and Moghaddam (2011)
Soil temperature	Russian Soil Temperatures	Russia	Regional; monthly; 1898-1990	Zhang et al. (2001a, 2001b)
Soil temperature, carbon fluxes, LAI	AmeriFlux, FLUXNET	North America, Europe	Site-level; hourly; 1990-2013	Baldocchi et al. (2012); Papale et al. (2012)
Soil temperature	Boreas & LBA	Regional	Site-level; monthly; duration of campaigns	Sutton et al. (1998); Hinkel (1998)
Snow characteristics	NOAA Snow Cover	Northern hemisphere	Regional; weekly; 1967-2013	Armstrong and Brodzik (2002)
Snow characteristics	North American Snow Course Network	North America	Regional; annually; 1930-2000	Brown, 2000; NRCS, 2007
LAI	MODIS	Global	Gridded; weekly; 2002-2013	Wang et al., 2001
NDVI, LAI	GIMMSg	Global	Gridded; monthly; 1982-2011	Tucker et al., 2005

Table DMP2: Gridded observationally based datasets complied under MsTMIP Phase I.

Dataset (Time)	Source
Gross Primary Productivity (1982-2008)	Jung et al. (2011)
Net Ecosystem Exchange (1982-2008)	
Total Respiration (1982-2008)	
Sensible Heat (1982-2008)	
Aboveground Biomass (1995-2005)	Saatchi et al. (2011)
Total Living Biomass (1995-2005)	
Total Soil Carbon (1960-2000)	Liu et al. (2012)
Leaf Area Index (2000-2010)	Yuan et al. (2011)
Evapotranspiration (1986-1995)	Fisher et al. (2008)

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

Armstrong, R.L., and M.J. Brodzik (2002), "Northern Hemisphere EASEGrid Weekly Snow Cover and Sea Ice Extent Version 2 [CD-ROM]," Natl. Snow and Ice Data Cent., Boulder, Colo.

Baldocchi, D., M. Reichstein, D. Papale, L. Koteen, R. Vargas, D. Agarwal, and R. Cook, (2012), "The role of trace gas flux networks in the biogeosciences Tracking subtle shifts in ecosystem-atmosphere gas exchange has taken an international effort and an interdisciplinary network," *Eos*, 93: 217-218.

Brown, R.D. (2000), "Northern Hemisphere snow cover variability and change, 1915 – 97," *J. Clim.*, 13(13), 2339 – 2355, doi:10.1175/1520-0442(2000)013<2339:NHSCVA>2.0.CO;2.

Crow, W.T., D.G. Miralles and M.H. Cosh (2010), A quasi-global evaluation system for satellite-based surface soil moisture retrievals, *IEEE Trans. Geosci. Rem. Sens*., 48(6), 2516-2527.

De Kauwe, M.G., Medlyn, B.E., Zaehle, S., Walker, A.P., Dietze, M.C., Hickler, T., Jain, A.K., Luo, Y., Parton, W.J., Prentice, et al. (2013), "Forest water use and water use efficiency at elevated CO_2: a model-data intercomparison at two contrasting temperate forest FACE sites," *Global Change Biology*, 19: 1759–1779. doi: 10.1111/gcb.12164.

Dorigo, W., P. Van Oevelen, W. Wagner, M. Drusch, S. Mecklenburg, A. Robock, and T. Jackson (2011), "A new international network for in situ soil moisture data", *Eos* 92 (17), pp. 141-142.

Fisher, J.B., K.P. Tu, and D.D. Baldocchi (2008), "Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites," *Remote Sensing of Environment*, 112, 901−919.

Hinkel, K.M. (1998), "Soil Temperatures for Happy Valley and Barrow, Alaska, USA," http://nsidc.org/data/arcss038.html, Natl. Snow and Ice Data Cent., Boulder, Colo., (Updated in 2004.).

Hook, L.A., S.K. Santhana Vannan, T.W. Beaty, R.B. Cook, and B.E. Wilson (2010), "Best Practices for Preparing Environmental Data Sets to Share and Archive," Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. doi:10.3334/ORNLDAAC/BestPractices-2010.

Jung, M., M. Reichstein, H.A. Margolis, A. Cescatti, A.D. Richardson, M.A. Arain, Arneth, A., C. Bernhofer, D. Bonal, J. Chen, D. Gianelle, N. Gobron, et al. (2011), "Global patterns of land-atmosphere

fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations," *J. Geophys. Res.*, 116, G00J07, doi:10.1029/2010JG001566.

Liu, S., Y. Wei, W.M. Post, R.B. Cook, K. Schaefer, M.M. and Thornton (2013), "The Unified North American Soil Map and its implication on the soil organic carbon stock in North America," *Biogeosciences*, 10, 2915-2930, doi:10.5194/bg-10-2915-2013.

Lu M., X. Zhou, Y. Luo, et al. (2013), "Responses of ecosystem carbon cycle to experimen- tal warming: a meta-analysis," *Ecology*. doi: 10.1890/12-0279.1

Natural Resources Conservation Service (NRCS) (2007), "Snow Course Data," http://www.wcc.nrcs.usda.gov/snowcourse/snow_rpt.html, Washington, D. C.

North American Carbon Program (NACP) (2007), NACP Data Policy. Available on-line [http://www.nacarbon.org/nacp/documents/6A-NACP_Data_Policy_20070211.pdf], Carbon Cycle Interagency Working Group, Washington, DC. Accessed, July 26, 2013.

Papale, D., D.A. Agarwal, D. Baldocchi, R.B. Cook, J.B. Fisher, and C. van Ingen, (2012), Database Maintenance, Data Sharing Policy, Collaboration pp 411-436. In: Aubinet, Marc; Vesala, Timo; Papale, Dario (Eds.) Eddy Covariance. A Practical Guide to Measurement and Data Analysis, Springer.

Saatchi, S.S., N.L. Harris, S. Brown, M. Lefsky, E.T.A. Mitchard, W. Salas, et al. (2011), "Benchmark map of forest carbon stocks in tropical regions across three continents," *Proceedings of the National Academy of Sciences*, 108, 9899–9904, www.pnas.org/cgi/doi/10.1073/pnas.1019576108.

Sutton, D., M.L. Goulden, A. Bazzaz, B.C. Daube, S.M. Fan, J.W. Munger, and S. Wofsy (1998), "BOREAS TF-03 NSA-OBS Tower Flux, Meteorological, and Soil Temperature Data," http://daac.ornl.gov/BOREAS/ guides/TF03_Flux_Met.html, Oak Ridge Natl. Lab. Distrib. Active Arch. Cent., Oak Ridge, Tenn.

Tabatabaeenejad, A., and M. Moghaddam (2011), "Retrieval of Surface and Deep Soil Moisture and Effect of Moisture Profile on Inversion Accuracy," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 477-481.

Tucker, C.J., J.E. Pinzon, M.E. Brown, D.A. Slayback, E.W. Pak, et al. (2005), "An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data," *Int. J. Remote Sens.*, 26, 4485–4498.

van Vuuren, D.P., J. Edmonds, M. Kainuma, K. Riahi, A. Thomson, K. Hibbard, G. Hurtt, T. Kram, V. Krey, J.F. Lamarque, T. Masui, M. Meinshausen, N. Nakicenovic, S.J. Smith, and S.K. Rose (2011), "The representative concentration pathways: an overview," *Climatic Change*, 109: 5-31, doi: 10.1007/s10584-011-0148-z.

Wang, Y.J, Y.H. Tian, Y. Zhang, N. El-Saleous, Y. Knyazikhin, E. Vermote, and R.B. Myneni (2001), "Investigation of product accuracy as a function of input and model uncertainties - Case study with SeaWiFS and MODIS LAI/FPAR algorithm," *Remote Sensing of Environment*, 78(3), 299-313

Yuan, H., Dai, Y., Xiao, Z., Ji, D., Shangguan, W., (2011), "Reprocessing the MODIS Leaf Area Index Products for Land Surface and Climate Modelling," *Remote Sensing of Environment*, 115(5), 1171-1187. doi:10.1016/j.rse.2011.01.001

Zhang, T., R. Barry, and D. Gilichinsky, compilers (2001a), "Russian historical soil temperature data," National Snow and Ice Data Center, digital media, Boulder, CO.

Zhang, T., R.G. Barry, D. Gilichinksy, S.S. Bykhovets, V. A. Sorokovikov, and J.P. Ye (2001b), "An amplified signal of climatic change in soil temperatures during the last century at Irkutsk," *Russian Clim. Change*, 49, 41– 76.