National Aeronautics and Space Administration

ORNL DAAC Distributed Active Archive Center for Biogeochemical Dynamics

WINTER 2014

ORNL DAAC News

The ORNL Distributed Active Archive Center (DAAC) is a NASA-sponsored source for biogeochemical data and services useful in environmental research. The ORNL DAAC currently archives and distributes nearly 1,100 products categorized as Field Campaign, Land Validation, Regional and Global, or Model Archive.

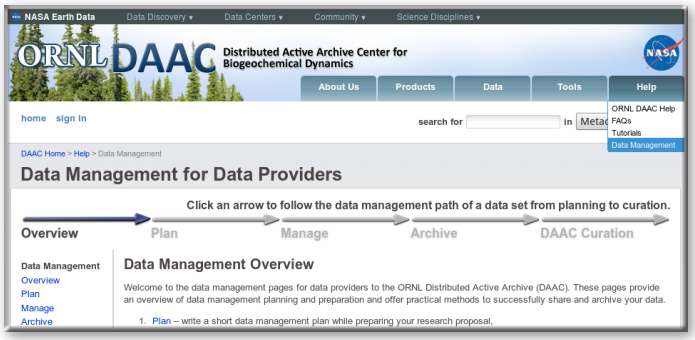
Please visit us online at *http://daac.ornl.gov* for a comprehensive description of data, services, and tools available from the ORNL DAAC. Current and past news can be found at *http://daac.ornl.gov/news.shtml*.

Contents:

http://www.nasa.gov

Data Management Resources for Data Providers

An improved resource for data providers is now available on the website of the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC). The Data Management part of the "Help" section has been expanded into five pages that define active archiving, offer best practices, and guide data providers into efficient data practices and accessible archiving of data sets.



The "Overview" page outlines the process and explains the benefits of good data management practices to save time and increase the use of data sets, while providing for long-term secure storage.

The "Plan" page explains Data Management Plans (DMP). The DMP for a proposal briefly outlines the procedures for handling the data during and after the research. Using a data management plan as the blueprint for following best practices will make data sets well-documented, accessible, and safe – now and for future use. Links to DMP tools and example data management plans are included in this section.

The "Management" page details seven steps to managing data with descriptive data set titles and file names and with consistent organizational techniques. The steps also include measures to protect data such as keeping a "read-only" copy of raw data files and periodically testing files to detect transfer errors or lost data in back-up copies.

Information on the "Archive" page leads data providers through the final steps of preparing their data for submission to the ORNL DAAC. Data providers must provide documentation and create metadata, the "who, what, where, when, and why" of the data. Data Provider questions are presented to help the providers create the metadata needed for active archiving. Steps to perform basic quality assurance and identify the level of maturity of the data products are explained here. Then the data is ready to share with the DAAC.

(continued on p. 2)

Data Management Resources for Data Providers (continued)

The "DAAC Curation" page answers two questions: What does the archive do with my data, and what does active archiving do for my data? Understanding the many procedures the DAAC staff performs – ingesting, publishing, and maintaining data sets – clarifies the partnership between a data provider and the DAAC and expounds on the value of active archiving.

When you are ready to submit a data set to the ORNL DAAC, we ask that you fill out a short archival interest form. This process has been automated. Upon contact, you will be emailed with instructions on how to submit the contents of your data set and how to answer the data provider questions. Upon completion of the data set submission process, your data set will be assessed for completeness and applicability for the ORNL DAAC.

"*In today's research, data are becoming even more important and people need the tools and skills to manage their data and prepare them for archiving. Through these webpages we are trying to do that,*" said Robert Cook, Chief Scientist for NASA's ORNL DAAC for Biogeochemical Dynamics. "*The US government is promoting initiatives to make data more accessible so that means managing and working with an active archive to curate your data has become more important.*"

For more on data management, please visit our page at http://daac.ornl.gov/PI/pi_info.shtml
To discuss archiving your data set at the ORNL DAAC, contact the Users Service Office.

New search tool for data used in published articles.

The ORNL DAAC staff has created a new tool for finding data sets, based on the DAAC data used in published articles. Since 1998, the ORNL DAAC has recommended that authors cite DAAC data sets when the data are used in published articles. This practice allows the data generators (data authors) to get credit –through a data citation– and also enables other researchers to find the exact data used in an article. Since 2007, the ORNL DAAC has also assigned Digital Object Identifiers (DOIs) to each data set to provide an easy and direct way for users to locate the cited data set.

The ORNL DAAC has produced a database that links published articles and the data used in that article. The information in this database enables the DAAC to gauge the impact of data archived at the ORNL DAAC on scientific research– the data citation index– just as an author's impact on science is estimated via a traditional citation index. This database can also allow other users to find data products based on the information identifying the article.

Here's how the tool works. Suppose you are reading an article and would like to find the data that were used in that article. On the Tools menu on the ORNL DAAC homepage (*http://daac.ornl.gov*), you select Search by article DOI and enter the article DOI. The tool looks in the DAAC database for data that were either used in the article or data that are related to that article.

On the Results page of this tool, additional features are available. Users can obtain the article's full citation. Along with the Search Results, showing any data sets used in that article, the DAAC will display the top five data sets that are related to the original article and supply links to these data sets.

In addition to access to the data sets themselves, if the data are geospatial, the tool also has links to advanced visualization and subsetting tools – the Spatial Data Access Tool (SDAT) and the THREDDS Data Server. SDAT is a community standards-based Web application that allows users to interactively view spatial data and download them in a user-specified region, resolution, format, etc. THREDDS is a Web service that provides metadata and data access to entire or parts of scientific data sets without needing to download entire files to their local systems.

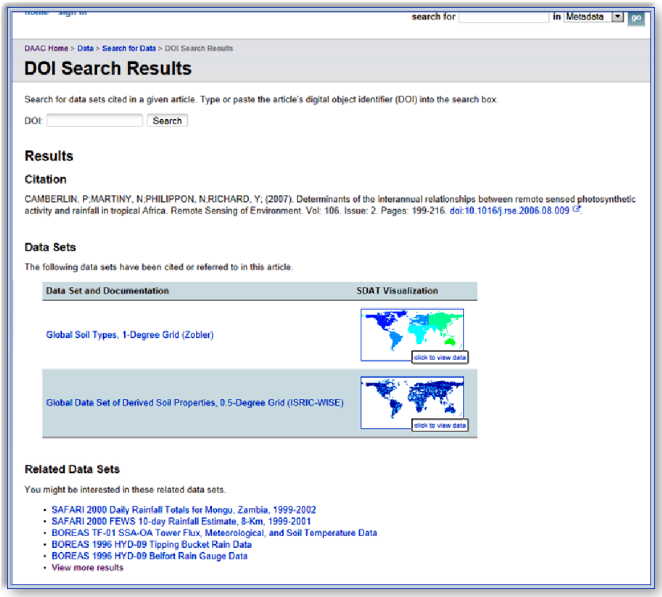
(continued on p. 3)

New search tool for data used in published articles. (continued)

To find data sets:

For more on this tool, please visit http://daac.ornl.gov/doi_search_page.shtml.

The ORNL DAAC and science publisher Elsevier are now linking articles published on Elsevier's ScienceDirect website with the corresponding data sets archived in the DAAC repository. Users are able to access the data sets used in or related to the original article through a direct banner link in the right panel– under Applications and Tools. This banner link directly connects the user to the DAAC's DOI Results landing page for the data set.

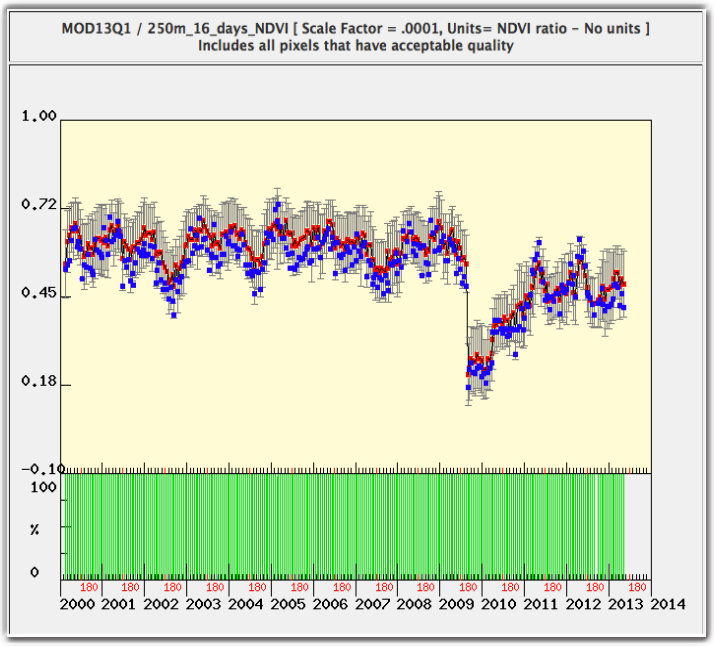


MODIS Land Product Subsets Tool Updates

MODIS Collection 6 release will include some updated and new land products that will be incorporated into the subsetting tools at the ORNL DAAC.

Product Updates:

- Consistent global improved 1-km Gross Primary Productivity (MOD17) from year 2000 to 2012
- Evapotranspiration (MOD16) product cutouts
- LAI/FPAR and GPP in 500m resolution (was 1km in Collection 5)
- BRDF/Albedo products will be produced daily (were being produced every 8-day in Collection 5)
- Daily Surface Reflectance (500m)



MODIS 10.25 x 10.25 km NDVI subset for a location in Angeles National Forest, CA, USA showing the 2009 station fire.

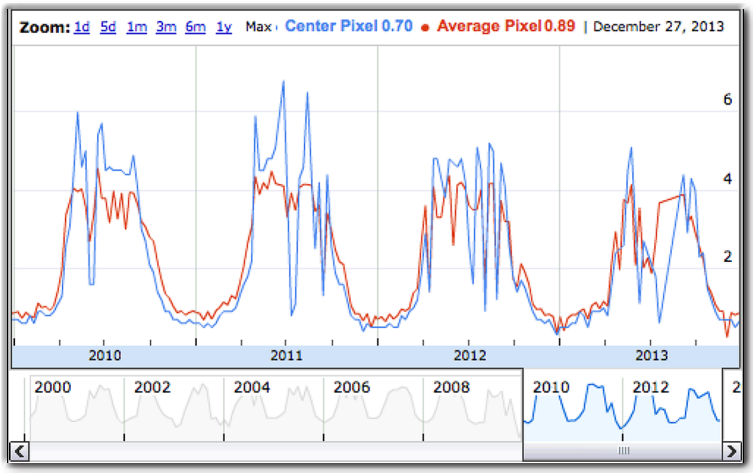
(continued on p. 4)

MODIS Land Product Subsets Tool Updates (continued)

Global Tool Updates

In addition to the product updates, we are adding some new features to the Global Tool. The current version of the Global Tool provides a web-based interface to create MODIS land product subsets for up to 201 x 201 km. The completed order provides visualization of the subset data as stacked time series plots and Interactive Time Series Plots based on Google Visualization.

The subset data and summary statistics provided for download in Comma Separated Value (CSV) files; GeoTIFF image files with reprojection information are also available.



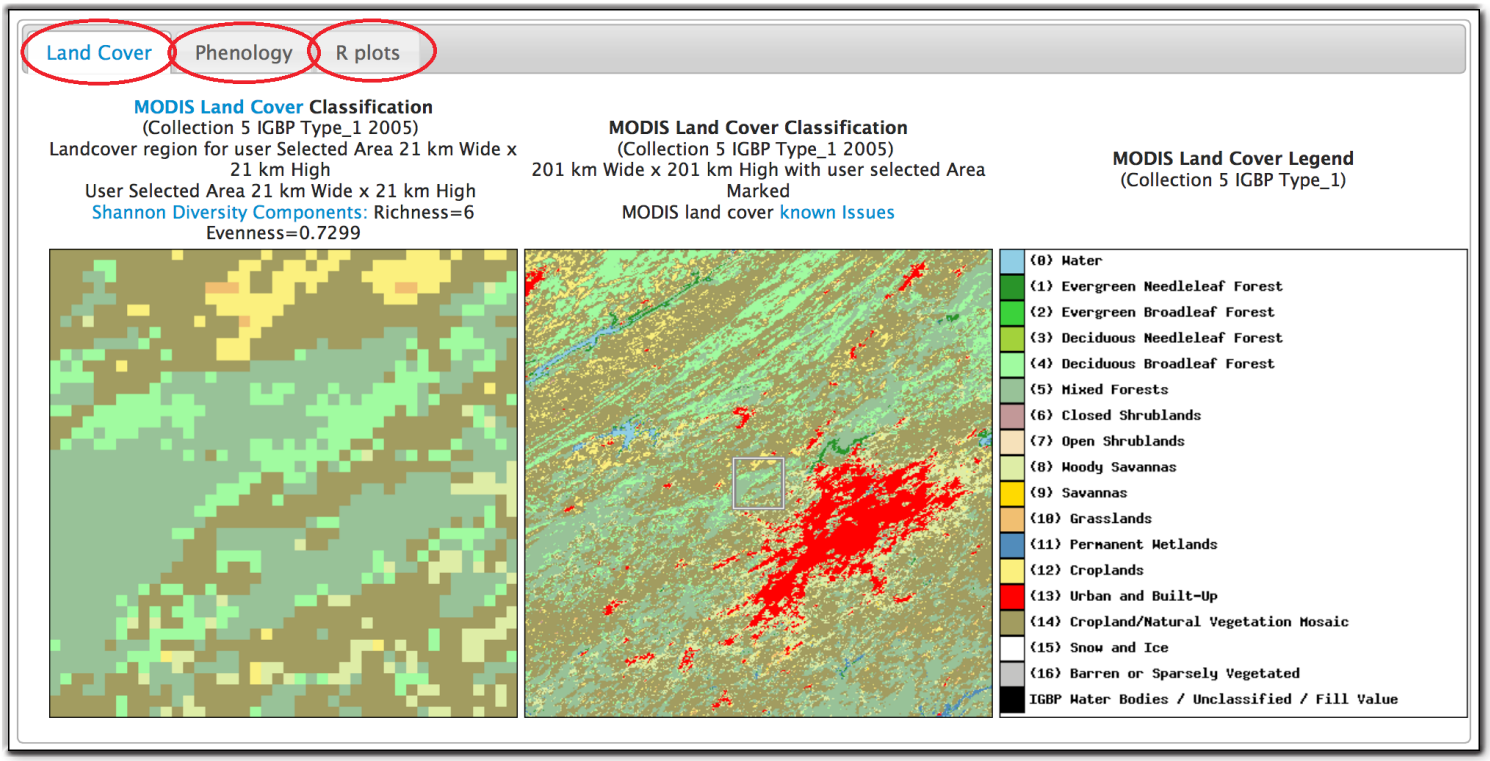
Interactive time-series plots, base on Google Visualization

Planned Updates

Planned updates to the Global Tool include an improved ordering interface and a better results page that organizes the output into tabs.

New features for the Global Tool include

- Ability to create larger subsets (up to 501 x 501 km)
- Ability to create subsets from pre-defined polygonal Region of interest [State, County, Hydrologic Unit Code (HUC) region]
- New visualizations and improved ordering interface



Global Tool results page with output organized into tabs

Recent additions to the ORNL DAAC online archive

Data sets archived at the ORNL DAAC are categorized into four project types: *Field Campaigns*, *Land Validation*, *Regional and Global Data*, and *Model Archive*.

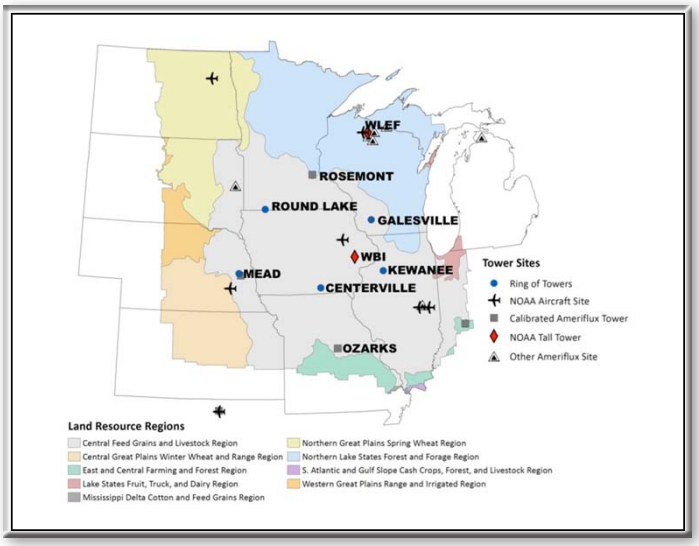
35 new data products were recently added within the *Field Campaign Data* category. Field campaigns combine ground-, aircraft-, and satellite-based measurements of biogeochemical features in specific ecosystems over multi-year time periods. 27 of these data sets were added to *The Large-Scale Biosphere-Atmosphere Experiment in Amazonia (LBA)* project and the remaining 8 from the *North American Carbon Program (NACP)*.

- *NACP Regional: National Greenhouse Gas Inventories and Aggregated Gridded Model Data*
- *NACP MCI: CO2 Flux Tower Measurements, Upper Midwest Region, USA, 2007-2009*
- *NACP Integrated Wildland and Cropland 30-m Fuel Characteristics Map, U.S.A., 2010*
- *NACP Site: Tower Meteorology, Flux Observations with Uncertainty, and Ancillary Data*
- *NACP Site: Terrestrial Biosphere Model and Aggregated Flux Data in Standard Format*
- *NACP Site: Terrestrial Biosphere Model Output Data in Original Format*
- *NACP Regional: Original Observation Data and Biosphere and Inverse Model Outputs*
- *NACP North American 8-km Net Ecosystem Exchange and Component Fluxes, 2004*

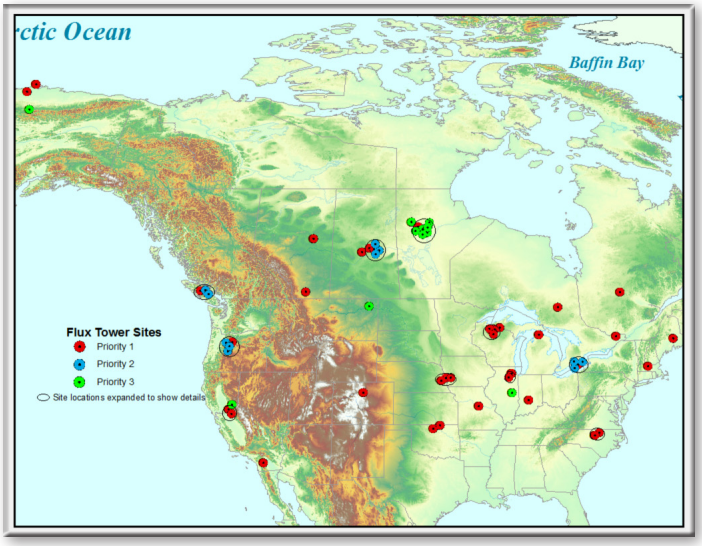
ORNL DAAC expects to archive between 10 and 20 additional NACP data sets in the next few months. NACP aims to measure and understand the sources and sinks of Carbon Dioxide (CO_2), Methane (CH_4), and Carbon Monoxide (CO) in North America and in adjacent ocean regions.

27 new data products were also recently added within the Regional and Global data category, including:

- *Global Fire Emissions Database 3.1*
- *TransCom 3: Seasonal CO2 Flux Estimates from Atmospheric Inversions (Level 2)*
- 25 revised *Net Primary Productivity (NPP) data sets*.



Map of MCI domain located in U.S. Upper Midwest. [*NACP MCI: CO2 Flux Tower Measurements, Upper Midwest*]



Flux Tower Site Locations: First-priority sites (36 sites) and Second-priority sites (11 chronosequence sites) were used for the model-data comparison. [*NACP Site: Tower Meteorology, Flux Observations with Uncertainty, and Ancillary Data*]

ORNL DAAC Announces Membership in ICSU-WDS

The ORNL DAAC is now a regular member of the *International Council for Science - World Data System (ICSU-WDS)*.

Like the ORNL DAAC, the goals of the WDS are to preserve assured scientific data and information and to facilitate open access and promote the adoption of standards. Joining an international community of like-minded data centers is a benefit to everyone.

WDS strives to form a worldwide 'community of excellence' for multidisciplinary scientific data, which ensures the long-term stewardship and provision of quality-assessed data and data services to the international science community and other stakeholders. Its concept aims at a transition from existing stand-alone components and services to a common globally interoperable distributed data system, with searchable common data directories and catalogues that incorporates emerging technologies and new scientific data activities. Disciplinary and multidisciplinary data networks within WDS will play a key role in moving this concept forward.

ORNL Distributed Active Archive Center
1 Bethel Valley Rd., MS 6290
Oak Ridge, TN 37830

ACCESSING ORNL DAAC DATA

Web-based interface:
http://daac.ornl.gov/

Advanced data search:
http://mercury.ornl.gov/ornldaac/

Anonymous FTP browsing:
ftp://daac.ornl.gov/data/

DAAC WebGIS:
http://daac.ornl.gov/mapserver.shtml

MODIS Land Products Subsets:
http://daac.ornl.gov/MODIS/modis.shtml

LBA Project:
http://daac.ornl.gov/LBA/lba.html

DAAC FLUXNET Project:
http://daac.ornl.gov/FLUXNET/fluxnet.shtml

DAAC SDAT:
http://webmap.ornl.gov/wcsdown

All data from the DAAC are free and are available electronically.