NACP Data Center for Modeling and Synthesis

A proposal submitted to
NASA Research Announcement NNH05ZDA001N, Research Opportunities in Space and Earth Sciences 2005, A.6. North American Carbon Program

Principal Investigator: Robert B. Cook
Environmental Sciences Division
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6407
Telephone: 865-574-7319
Facsimile: 865-574-4665
Email: cookrb@ornl.gov

Co-Investigators: W.M. Post, III, Environmental Sciences Division, ORNL
B.T. Rhyne, Computer Science and Engineering Division, ORNL
Peter E. Thornton, National Center for Atmospheric Research

Start Date: July 1, 2006
End Date: June 30, 2009

Note: This document contains the proposal submitted to NASA in December 2005. NASA did not fully fund all of the activities described, and also requested that some advanced tools be eliminated.

Notice
Restriction on Use and Disclosure of Proposal Information

The information (data) contained in pages 1-12 of this proposal constitutes a trade secret and/or information that is commercial or financial and confidential or privileged. It is furnished to the Government in confidence with the understanding that it will not, without permission of the offeror, be used or disclosed other than for evaluation purposes; provided, however, that in the event a contract (or other agreement) is awarded on the basis of this proposal the Government shall have the right to use and disclose this information (data) to the extent provided in the contract (or other agreement). This restriction does not limit the Government's right to use or disclose this information (data) if obtained from another source without restriction.

TABLE OF CONTENTS

ABSTRACT ..... **ii**
PROJECT DESCRIPTION **1**
1. Introduction **1**
1.1 Proposal Objective 1
1.2 Background 1
1.3 Role of Proposed MAST-DC 3
1.4 Relevance to NASA's Strategic Objectives 3
1.5 Organization of NACP 4
2. Objectives **4**
3. Technical Approach **5**
3.1 Task 1. Establish specific data management needs of NACP modelers and synthesis groups 7
3.2 Task 2. Prepare and distribute common model input data 8
3.2.1 Input Data Products: Digital Elevation Model Data 9
3.2.2 Input Data Products: Soils 9
3.2.3 Input Data Products: Land Cover 10
3.2.4 Input Data Products: Historical Land Use 10
3.2.5 Input Data Products: Climate / Weather Data 11
3.2.6 Input Data Products: Leaf Area Index / fraction of photosynthetically absorbed radiation 12
3.2.7 Input Data Products: Nitrogen 12
3.3 Task 3. Data management support for model outputs 12
3.4 Task 4. Provide tools for visualization, subsetting, and reprojection 13
3.5 Task 5. Provide data packages to evaluate model output: confronting models with data 14
3.6 Task 6. Support synthesis activities, including data support for workshops 16
3.7 Other data system functions: Acquiring and Distributing Model Source Code 18
4.0 Benefits of Proposed Work **18**
5.0 Management Plan and Personnel **18**
References **20**
Personnel **23**
Time table for activities in MAST-DC **23**
Facilities and Equipment **24**
Resumes **26**
Letters of Support / Collaboration **32**
Acronyms **33**

ABSTRACT

The North American Carbon Program (NACP) is designed to quantify the magnitudes and distributions of carbon sources and sinks, explain the processes controlling them, and produce a consistent analysis of North America's carbon budget. To accomplish these ambitious goals, NACP requires an integrated data and information management system that will enable researchers to access, understand, use, and analyze large volumes of diverse data at multiple thematic, temporal, and spatial scales. In response to a NASA solicitation for proposals to contribute to NACP, we are proposing to establish a Modeling and Synthesis Thematic Data Center (MAST-DC) to be an integral component of the NACP data system and to support NACP by providing data products and data management services needed for modeling and synthesis activities.

The overall objective of the proposed MAST-DC is to provide advanced data management support to NACP investigators and agencies performing modeling and synthesis activities. Based on specific requirements established by NACP, we will provide data products for modeling and synthesis in consistent and uniform grids, projections, and formats.

The specific tasks of the proposed MAST-DC are (1) coordinate data management activities with NACP modelers and synthesis groups; (2) prepare and distribute model input data; (3) provide data management support for model outputs; (4) provide tools for accessing, subsetting and visualization; (5) provide data packages to evaluate model output; and (6) support synthesis activities, including data support for workshops.

MAST-DC will provide data products and services required by NACP in a central location, with common and co-registered spatial projection, in easily converted formats. The MAST-DC will free modelers and those doing the synthesis and integration from having to perform data management functions. Consequently the MAST-DC will enable NACP participants to conduct their work more readily, facilitate the development of new model products needed by models, and assist in gaining new insights into the carbon cycle in North America.

PROJECT DESCRIPTION

1. Introduction

The North American Carbon Program (NACP) is a multidisciplinary and interagency research program to obtain scientific understanding of North America's carbon sources and sinks. The NACP Science Implementation Strategy Plan (Denning et al., 2005) calls for an unprecedented level of coordination among observational, experimental, and modeling efforts for the terrestrial, oceanic, atmospheric, and human components of the carbon cycle. At the heart of the NACP research strategy is an integrated data and information management system that will enable researchers to access, understand, use, and analyze large volumes of diverse data at multiple thematic, temporal, and spatial scales.

To meet these ambitious scientific and data management goals, NACP has established a set of recommendations for the NACP data system, including establishment of a Data Central and a suite of distributed Thematic Data Centers (Fig. 1; NACP 2005). Existing data centers have missions and data holdings (e.g., remote sensing, biomass inventory, atmospheric CO_2 monitoring network, and flux tower data) necessary to meet many of the recommendations for thematic data centers, but not all. In one of the recommendations, NACP specified that a thematic data center for modeling and synthesis would be established to assemble and distribute model input and output data for process-based and inverse modeling, as well as to acquire and distribute model source code and associated documentation. A key part of this recommendation was that the input and output data sets for both process-based and inverse modeling would be served with consistent formats to facilitate use and synthesis activities.

1.1 Proposal Objective

In response to NASA Research Announcement NNH05ZDA001N, Research Opportunities in Space and Earth Sciences 2005, A.6. North American Carbon Program, we are proposing to establish a Modeling and Synthesis Thematic Data Center (MAST-DC) to be an integral component of the NACP data system and to support NACP by providing data products and services needed for modeling and synthesis activities. The overall objective of MAST-DC is to provide advanced data management support to NACP investigators and agencies performing modeling and synthesis activities.

1.2 Background

The North American Carbon Program (Denning et al., 2005) is designed to quantify the magnitudes and distributions of carbon sources and sinks, explain the processes controlling them, and produce a consistent analysis of North America's carbon budget.

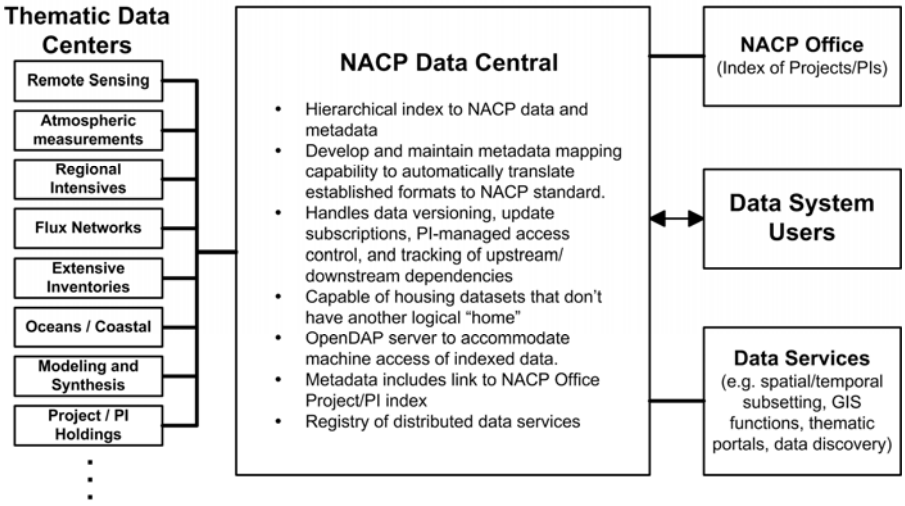

Figure 1 (left). Schematic diagram of the NACP data system components (from NACP 2005), showing the relationship between Data Central and the Thematic Data Centers, including the Modeling and Synthesis Thematic Data Center.

Quantification of the geographical distribution of sources and sinks of CO_2 and attributing them to processes are major undertakings. Observational and modeling strategies to accomplish these goals must take into account spatial and temporal variability that arise from changing environmental conditions and also management practices over areas as small as a 100 km^2. Data assimilation (a.k.a. data fusion) techniques similar to those used by weather forecasting services will combine atmospheric concentration measurements and flux observations to constrain model-based estimates of sources and sinks. These data assimilation techniques have the added benefit that models constrained by observations can provide additional information on the underlying processes (Bennett, 2002).

Models are a key integrating component of the NACP strategy (Wofsy and Harriss, 2002; Denning et al. 2005). Several different modeling approaches will be used in NACP to understand processes, sources and sinks, and provide realistic simulations of future changes in carbon cycling. One modeling approach will involve process-based (bottom-up) models (e.g., BIOME-BGC, Century, etc.) that estimate carbon sources and sinks at relatively fine scale (~1-km resolution). Inverse models (top-down) estimate atmospheric transport of carbon using wind and cloud transport along with observations of atmospheric carbon concentrations and then infer the sources and sinks of carbon on the land responsible for that atmospheric transport. These two approaches—top-down and bottom–up—will be combined to provide improved understanding of carbon fluxes from land surfaces.

A central organizing strategy of NACP is the use of regional intensive field-based studies as test-beds of new approaches (Denning et al., 2005). For example the first test bed—the Mid-Continent Intensive – will reconcile estimates of regional carbon sources and sinks derived from atmospheric models with estimates based on bottom-up models and estimates from field measurements, inventories, regional geographic information, and remote sensing data. This intensive campaign will attribute sources and sinks to ecosystem processes and human activities within the region.

With models playing such an important role, NACP will need to develop a common model framework that specifies the modeling experiments and establishes a consistent and uniform set of input and output data sets. NACP intends to build on a number of past modeling activities that established a model framework for processes to be modeled, the driving variables, and inputs and outputs (VEMAP Members 1995; Ryan 1996a, b; Smith et al., 1997; Amthor et al., 2001; Hanson et al. 2004). Each of the modeling activities reported in these papers established common boundary conditions and driving variables so that differences in model results were derived only from the models and their implementation rather than from differences in inputs. Removing the confounding influences of different input data sets enabled the modeling activities to focus on improving how the models represented various processes, identifying conditions or regions in which model outputs of sources and sinks differ, and providing estimates of model uncertainty. From a practical standpoint, the modeling framework necessitates that data products be assembled in a coordinated way by the proposed Center, rather than have each modeling group prepare each of the input data products (see Technical Approach below).

1.3 Role of Proposed MAST-DC

Our proposed MAST-DC will support the integrated science approach of NACP by working closely with the NACP modeling and synthesis groups. Together we will establish a common framework for modeling and synthesis, including specific requirements for input and output data (data products, formats, geographic projections, grid) as well as tools to assist users in finding, accessing, subsetting, and visualizing the input and output data products. The MAST-DC will compile these input and output products and acquire / develop tools and make them freely available to NACP researchers and the wider research community. In addition, the MAST-DC will support periodic (semi-annual) scientific workshops conducted by NACP (see Task 6 below).

1.4 Relevance to NASA's Strategic Objectives

One of NASA's strategic objectives is to advance Earth observation from space, improve scientific understanding and demonstrate new technologies with the potential to improve future operational systems. Within the Earth-Sun System Division at NASA, programs are carefully designed to address a specific set of questions regarding how the earth is changing, the primary causes of those changes, the Earth's response to natural and human-induced change, the consequences to life and our ability to predict them. For example, one of the questions of the NASA's Carbon Cycle and Ecosystems Focus Area is: How large and variable are the dynamic reservoirs and fluxes of carbon within the Earth system, and how might carbon cycling change and be managed in future years, decades, and centuries? This Focus Area plays a key leadership role in NACP.

The MAST-DC proposes to provide direct and specific data management support to NACP research on carbon sources and sinks, processes affecting those fluxes, and on improving predictions of future changes. As such, the MAST-DC will use remote sensing data products along with other types of data to improve our scientific understanding of the Earth, specifically the carbon cycle.

1.5 Organization of NACP

NACP is a large and complex program supported by eight federal agencies and conducted by hundreds of scientists. This section describes the organization of NACP, how the program will operate, and how the data system, the thematic data centers, and the proposed MAST-DC fit into that organization.

The North American Carbon Program will be conducted by two groups of scientists: agency core and supporting project participants (e.g., scientists from USDA Forest Service and Agricultural Research Service) and individual investigators (e.g., funded extramurally from, for example, NASA, USDA, or DOE). A list of NACP participants was under development by the Carbon Cycle Interagency Working Group (CCIWG) as this proposal was being prepared.

The CCIWG coordinates NACP in the U.S., and NASA is a member of the CCIWG. To provide science leadership of NACP, the CCIWG has appointed a Science Steering Group (SSG) composed of leaders from the group of science investigators and from agency core and supporting projects.

The SSG has established a series of expert Task Forces to provide guidance and feedback to the SSG and the CCIWG. To date the SSG has established three Task Forces, with representation from relevant U.S. agency programs and research communities, including international NACP partners. The SSG with financial support from the CCIWG, will hold annual all-NACP-scientists meetings.

The Mid-Continent Intensive Campaign Task Force is charged to plan and coordinate the field and synthesis activities for the campaign. The Synthesis Task Force is charged to coordinate research and synthesis activities, including identifying the data product requirements and services needed to meet the research goals of NACP. Both of these groups will hold periodic synthesis and evaluation workshops. The MAST-DC will provide direct and specific data management support to both the Mid-Continent Intensive Task Force and the Synthesis Task Force.

The third Task Force is the Data Systems Task Force. This Task Force was charged with developing the Data Systems Recommendations (NACP 2005b) that called for the establishment of a Data Central and a suite of distributed Thematic Data Centers (Fig. 1), including a center for Modeling and Synthesis (the proposed MAST-DC). At the time this proposal was being written (December 2005), the NACP Data System was being planned and had not been implemented. Through the course of NACP, the Data Systems Task Force will provide additional guidance and recommendations for the operation of the NACP data system.

2. Objectives

The overall objective of this proposed work is to provide advanced data management support to NACP investigators and agencies performing modeling and synthesis activities. This objective

can be met through close, real integration and dialogue between the modeling and data systems to both define and implement requirements. The overall objective will be achieved through six tasks described briefly below and expanded upon in Sect. 3.

1. Establish specific data management needs of NACP modelers and synthesis groups.

We will work actively and directly collaborate with NACP modeling groups and the Synthesis Task Force of the SSG to identify the input and output data products and tools and services required to accomplish the goals of NACP.

2. Prepare and distribute model input data.

We will acquire, process, reproject, and distribute a base set of input and driver data in consistent and uniform format. Assembly and distribution of consistent input files will allow all modeling groups to start from common boundary conditions and driving variables, thus eliminating the confounding influence of multiple versions of input files.

3. Provide data management support for model outputs.

We will acquire, process (as needed), and distribute model output products for use by NACP synthesis and modeling groups. These outputs will be critical components of synthesis activities, including the workshops (Task 5).

4. Provide tools for accessing, subsetting, and visualization.

In support of the user community, we will provide tools critically required for accessing, subsetting, and displaying MAST-DC data products. Our approach will be to use existing tools, but make required modifications as requested by the NACP user community.

5. Provide data packages to evaluate model output: confronting models with data

We will acquire, process, and distribute field-based data products that can be used to evaluate the bottom-up and top-down models.

6. Support synthesis activities, including data support for workshops.

The Synthesis Task Force of the SSG will conduct workshops dedicated to planning observation campaigns, modeling, and synthesis activities. The MAST-DC will closely coordinate with the Synthesis Task Force to provide data management support for synthesis activities and to develop the data products required at these workshops.

3. Technical Approach

MAST-DC will develop data products and services required by the NACP community and the Synthesis Task Force, as described in detail below.

The MAST-DC is one component of the distributed data system for NACP (Fig. 2). Data held by MAST-DC will be registered in the hierarchical data index at Data Central, along with data from other distributed Thematic Data Centers. The metadata standard that Data Central will use has not yet been defined as of December 2005, but the MAST-DC team has worked with several metadata formats, including, but not limited to, NASA Global Change Master Directory's Directory Interchange Format (DIF) and the Federal Geospatial Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) used by the National Spatial Data Infrastructure (NSDI). Users will access Data Central's index to query for data products, view the data products available, and download directly from MAST-DC.

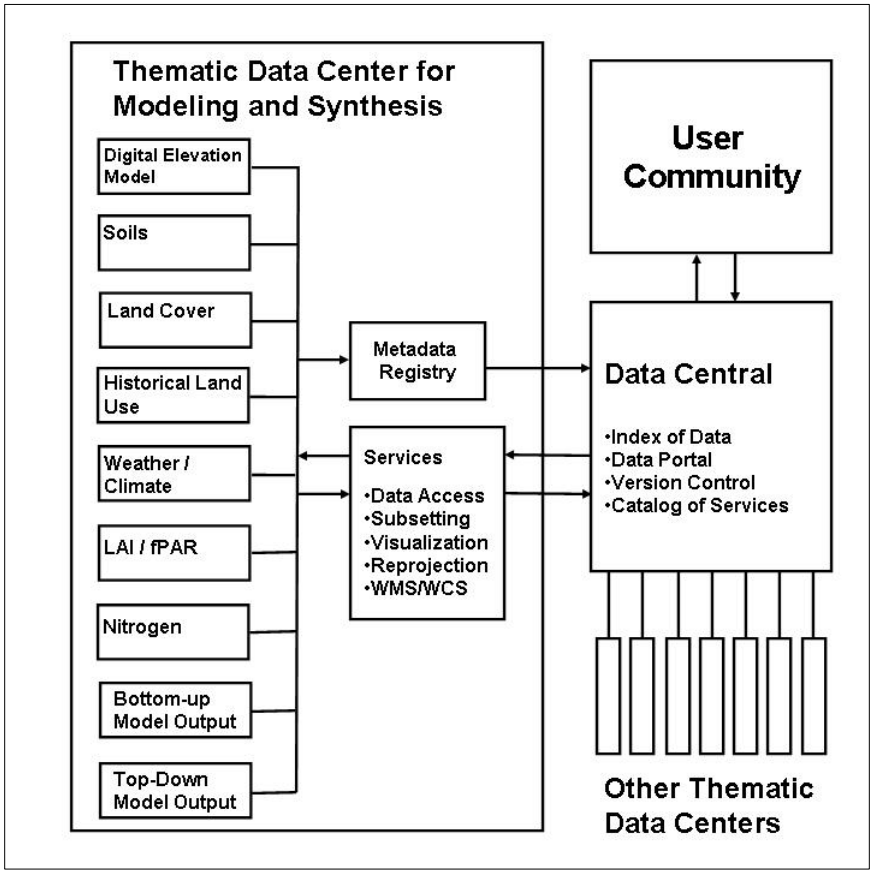

Figure 2 (left). Conceptual diagram of the MAST-DC and its relationship to Data Central and the other thematic data centers. MAST-DC will acquire, process, index, and distribute data products required for NACP modeling. The blocks on the left-hand side represent processed, uniform, and consistent data products required by the synthesis and modeling communities and prepared by the MAST-DC. Users will query the Data Central hierarchical data index to access data products, which will be served from the MAST-DC OPeNDAP system. Services for facilitating access and use of data products will be available from MAST-DC.

.

Data products developed and provided by the MAST-DC will be delivered to a broad community via a collection of services and capabilities. In addition to FTP-based access, MAST-DC will also offer data services based on both Open Geospatial Consortium (OGC-http://www.opengeospatial.org) standards and OPeNDAP, the Open-source Project for a Network Data Access Protocol (http://opendap.org/). OGC Web Map Service (WMS) and Web Coverage Service (WCS) protocols allow visualization and data access, respectively, to

geospatial data. OPeNDAP provides both discovery and delivery, and also enables users to select and access and select a well-defined portion of the OPeNDAP data products. Users select the portion of the data they need (e.g., spatial, temporal, and parameter subset) and the data are extracted from the OPeNDAP holdings and downloaded directly to the user's computer and local applications that they might use to process and analyze the data.

When Data Central is established, we will closely coordinate our activities to comply with metadata, documentation, and other standards established by NACP. We will rely heavily on existing infrastructure to provide data support to fulfill the recommendations for the NACP Modeling and Synthesis Data Center. Data products or services available from other Thematic Data Centers and indexed in Data Central will not have to be made available through MAST-DC. For example, gridded data on carbon sources (e.g., sources associated with power generation, transportation, heating, and industry) are being compiled by individual investigators (e.g., Kevin Gurney, Purdue University) and will be available from the individual investigator (Fig. 1).

As of December 2005, the data system for NACP had not evolved to the point of being operational and is still being planned by the Data Systems Task Force. We will maintain close contact with the Data Task Force, Data Central, and the other Thematic Data Centers to ensure coordination of activities as the NACP Data System is developed.

Also at this time, NACP does not have a data policy in place that would define sharing of data collected under the auspices of the NACP, access to NACP data by the broader scientific community and by the public, and expectations for short and long-term data archive. We expect that there will be some data sets that will have restricted access for a period of time (e.g., student data or prelimary data). We are prepared to establish secure areas for such products so that only authorized investigators can access and download the data.

3.1 Task 1. Establish specific data management needs of NACP modelers and synthesis groups

This proposed project will actively and directly collaborate with modeling teams and synthesis groups to ascertain the exact data support and services required. Preliminary information about the data products, support, and tools needed have been documented in the Science Implementation Strategy Plan (Denning et. al, 2005), the Report from the Data Systems Planning Workshop (NACP 2005a) and the Phase 1 Data System Recommendations (NACP 2005b). However, these guidelines do not provide the level of detail required by the data users and producers. We will work closely with NACP modelers, the SSG's Synthesis Task Force and Data Systems Task Force through monthly teleconferences, email list servers, and annual workshops to establish the modeling and synthesis data requirements needed by NACP, including assisting in NACP synthesis workshops. During the annual NACP All-Scientists meetings, MAST-DC and the Synthesis Task Force will convene a break out session specifically devoted to Task 1 and establishing plans for synthesis activities.

General data issues that will be addressed in Task 1 include the spatial gridding system for North America, geographic projection, geographic datum for elevation and area, data file formats, and

file-naming and metadata conventions. We will maintain a Web site that lists the specifications and requirements established by this task for NACP.

3.2 Task 2. Prepare and distribute common model input data

The purpose of Task 2 is to provide a base set of input and driver data in a uniform and consistent format, thereby freeing the modeling groups from having to prepare these common data sets themselves. In addition, assembly and distribution of consistent input files will allow all modeling groups to start with a common set of conditions and driving variables, thus eliminating the confounding influence of multiple versions of input files.

NACP has identified soil, land cover, land use change / disturbance history, and weather / meteorological data as the key products that need to be developed for the NACP modeling community (Table 9; Denning et al., 2005). The activities conducted in Task 1 will further refine the data requirements and data products needed by NACP modelers and synthesis groups.

Some of these data products have already been prepared at resolutions, projections, format, and quality that will be directly useful for NACP modeling and synthesis activities. These will not require additional processing; users will be able to locate and acquire these data products through Data Central. Examples of these products are gap-filled and aggregated flux tower data available through AmeriFlux and FLUXNET Canada, land cover change / disturbance / recovery data for the 1970s through the present from LEDAPS (Landsat Ecosystem Disturbance Adaptive Processing System, J. Masek, NASA, Principal Investigator), and specialized MODIS (Moderate Resolution Imaging Spectroradiometer) remote sensing products available through the recently funded project "Improving access to Land and Atmosphere science products from Earth Observing Satellites: Helping NACP investigators better utilize MODIS data products" (J. Morisette, NASA, Principal Investigator).

Other key data products required for synthesis and modeling have either not been prepared or are available only in formats, resolutions, and projections that are not directly suitable for NACP. In Task 2, we propose to acquire, assemble, process, and document these key data products required by synthesis and modeling groups so that the products are consistent and uniform and thus suitable for broad use for synthesis and modeling.

We propose to provide data products for modeling and synthesis in NACP grid, projection, and format. We will develop model input data products at 1-km grid in both native projection and in Albers Equal Area Conic Projection (NAD83). We will provide all data in netCDF format, and we will provide sufficient instructions so that this format can be converted using common GIS tools (e.g., ArcInfo) to other formats used in synthesis and modeling activities (e.g., HDF-5, ASCII Grid for ArcInfo, and other GIS formats). The proposed data projection and format will be revised / approved in Task 1 by the Synthesis Task Force and the Data Systems Task Force prior to implementation.

Based on information in the NACP Science Implementation Strategy (Table 9 in Denning et al., 2005), we have developed a list of data products that will likely be necessary for NACP. Other

data products and exact data requirements needed by NACP modeling and synthesis groups will be identified in Task 1.

The following describes some of the data products likely required for NACP and how those data will be acquired and processed by MAST-DC for use by synthesis and modeling groups within NACP.

3.2.1 Input Data Products: Digital Elevation Model Data

We will use the Shuttle Radar Topographic Mission (SRTM) data for North America for our Digital Elevation Model (DEM) data. The SRTM data (seamless and finished) are available at 1 arc-second (30 m) resolution for the United States and at 3-arc-second (90-m) resolution for the Globe. The data are available from the USGS as tiles in DTED format (Digital Terrain Elevation Data, a uniform matrix of elevation values indexed to specific points on the ground). For a consistent North American product, we will use the 3 arc-second resolution product, prepare a continental scale product, and place into Albers Equal Area Conic projection. We will also acquire the 1 arc-second (30-m) product for North America, convert it to Albers Equal Area Conic Projection, and make it available for use in intensively studied areas.

We will aggregate the 90-m product into coarser resolution (1-km) NACP grid for use in modeling and synthesis. In addition to topography, we will determine slope and aspect of the 1-km product and make that available in both native format and Albers Equal Area Conic Projection.

3.2.2 Input Data Products: Soils

For use in intensive campaigns located in the continental U.S. a multilayer soil characteristics dataset for the conterminous United States (CONUS-SOIL) that specifically addresses the need for soil physical and hydraulic property information over large areas is available (Miller and White 1998). The State Soil Geographic Database (STATSGO) served as the starting point for CONUS-SOIL. Geographic information system and Perl computer programming language tools were used to create map coverages of soil properties including soil texture and rock fragment classes, depth-to-bedrock, bulk density, porosity, rock fragment volume, particle-size (sand, silt, and clay) fractions, available water capacity, and hydrologic soil group at 1 km spatial resolution. CONUS-SOIL is available in several formats and may be accessed via the World Wide Web.

For North America, we will use the mid-latitude soils data set (Cryosol Working Group 2003). Data are in the US soil classification system and includes the distribution of soil types (%) within a map unit (polygon). The soil data product will include area of each soil type, bulk density, thickness of predominant soil horizons, depth to bedrock, soil texture and rock fragments, water holding capacity, and soil organic carbon. We will convert the polygon-based soil data to the NACP grid to facilitate the products use in models.

3.2.3 Input Data Products: Land Cover

We will process and offer several different land cover data products in the NACP grid, projection, and format. The primary land cover product will be the Vegetation Continuous Fields product from both AVHRR (Advanced Very High Resolution Radiometer) and MODIS sensors at 1-km resolution. As new Vegetation Continuous Fields (VCF) products are released from MODIS onboard Terra (e.g., reprocessing for Collections 4 and 5), we will acquire, process, and incorporate those into MAST-DC. Coverage of these products for North America will be obtained from the University of Maryland's Global Land Cover Facility, mosaiced, and converted into Albers Equal Area Conic Projection and netCDF format.

Another land cover product needed for modeling and synthesis is the classified land cover. For this product, we will acquire and process the MODIS land cover (IGBP, UMD, and Leaf Area Index (LAI) Classifications) into an Albers Equal Area Conic Projection and netCDF format.

Other land cover data products such as the GLC2000 and AVHRR GLCF land cover data set for modeling will be investigated and made available. For intensively studied areas in the US, we will also acquire the USGS's National Land Cover Data (NLCD), which includes a more detailed classification scheme at 1-km resolution.

3.2.4 Input Data Products: Historical Land Use

Recent efforts to understand the role of disturbance in carbon dynamics have resulted in a number of historical land use products that will be useful for NACP. Ramankutty and Foley (1999) have developed a global croplands data product at 0.5 degree resolution for the period 1700 to 1992. The LEDAPS (Landsat Ecosystem Disturbance Adaptive Processing System) project is funded to prepare land cover / disturbance maps for North America based on LANDSAT era remote sensing (1970s to the present). We will monitor that activity so that their products can be incorporated into NACP modeling activities when they are completed. Because LEDAPS's stated goal is to "work with carbon-cycle science community to tailor processing to meet science needs," we do not expect that processing or formatting of these data will be required for synthesis activities.

Waisanen and Bliss (2002) developed a database of changes in population and agricultural land, primarily from digital sources. The county-level data provide more spatial detail than was previously available. The purpose was to provide data on the timing of land conversion as an input to dynamic models of the carbon cycle, although a wide variety of applications exist for the physical, biological, and social sciences. The spatial data represent the appropriate county boundaries for each census year between 1790 and 1997, and the census attributes are attached to the appropriate spatial region. The resulting time series and maps show the history of population (1790–1990) and the history of agricultural development (1850–1997).

Natural disturbances also influence the carbon dynamics of terrestrial ecosystems by reducing biomass below equilibrium levels or even result in changed land cover and therefore could be handled in the same fashion as historical land-use change to obtain more accurate estimates of

carbon exchange with the atmosphere. The most important natural disturbance for carbon dynamics is wildfire.

A global fire emissions product, based on MODIS (2000 – present) has been compiled by Jim Randerson (University of California-Irvine). This product will be acquired, subsetted for North America, and offered in both native and Albers Equal Area Conic projections.

The National Wildland Fire Information System in Canada has a historical database available for 1959-1999 and weekly fire statistics are currently generated by the National Forestry Database Program

During the course of NACP, we will monitor the development of new land use history data products, and work to make them available via the MAST-DC when they are completed.

3.2.5 Input Data Products: Climate / Weather Data

Weather is the most important control on vegetation processes. Carbon dynamics and flux estimates from process models depend strongly on daily and subdaily weather conditions. Model behavior over time depends on the history of these weather conditions (climate). Key climate / weather parameters are precipitation (rain, snow), temperature, relative humidity/ water vapor pressure, wind speed, and radiation.

- Daily maximum temperature (°C)
- Daily minimum temperature (°C)
- Daylight average temperature (°C)
- Daily total precipitation (cm)
- Daylight average partial pressure of water vapor (Pa)
- Daylight average shortwave radiant flux density (W/m2)
- Daylength (s)

3.2.5.1 Daymet

A core requirement for many forward modeling approaches is surface weather fields, including temperature, precipitation, humidity, and incident solar radiation. Models differ in their exact requirements for these fields, depending on the model's native time step and the level of process complexity included. One common obstacle to model implementation over continental-scale regions is the difficulty of obtaining the relevant surface weather observations from *in situ* networks, and producing spatially interpolated (gridded) surfaces of the necessary weather fields at the appropriate spatial and temporal resolution. One approach that has been developed to overcome this obstacle is Daymet (Thornton et al. 1997, Thornton and Running 1999, Thornton et al. 2000), a numerical method for producing gridded surfaces of subdaily temperature (daily maximum and minimum temperature), and daily precipitation, humidity, and radiation over large regions of complex terrain, using daily surface weather observations and an accurate elevation grid as input. The Daymet method has already been used to produce a high-quality daily data set over the conterminous United States for the period 1980-2003, and this data set has been made freely available to the research community (www.daymet.org), currently hosted from a server at NCAR.

We propose to provide a high-quality gridded surface weather product over North America for input to NACP process modeling studies by expanding on the conterminous U.S. Daymet domain to include Alaska, Canada, and Mexico. Input data requirements are already in hand to produce the Daymet product for Alaska, and we will work with the appropriate contacts in Canada and Mexico to obtain the necessary surface observations for those countries, within the context of the mechanisms being established for international cooperation on NACP research efforts. We also propose to move the current and expanded Daymet data sets to the MAST-DC for future service of these data products to the NACP research community. Once the data streams for Canadian and Mexican data have been addressed and integrated into the current operational Daymet software package, the package will be transferred from NCAR to the MAST-DC, where periodic updates to the continental dataset will be implemented as new years of surface observations become available.

3.2.6 Input Data Products: Leaf Area Index / fraction of photosynthetically absorbed radiation

Several models to be used in NACP rely on the 1-day and 8-day LAI / fPAR (fraction of photosynthetically absorbed radiation) data products from MODIS as inputs (Myneni et al., 1997). We will provide a metadata link to global composite products at Boston University (HDF-EOS format and Sinusoidal Projection) and also acquire these products and subset for North America. These subsetted products will be offered in native and Albers Equal Area Conic projections at 1-km resolution.

3.2.7 Input Data Products: Nitrogen

Several models that will be used in NACP incorporate biogeochemical processes for nitrogen. For these models, we will provide estimates of nitrogen in atmospheric deposition and estimates of nitrogen additions in fertilizer.

We will acquire and distribute the Galloway et al. (2004) nitrogen deposition data, which is provide for 1860, 1993, and 2050 and can be made available at 1 degree resolution (F. Dentner, pers. comm.) . We will also acquire and distribute N-deposition data will be derived from Community Atmosphere Model (and atmospheric chemistry model Mozart-2) (Lamarque et al., in press), which is provided at 3.5 degree resolution for 1890, 2000, and 2100; these data are available directly from Jean-Francois Lamarque (pers. comm.).

The history of nitrogen fertilizer application in agricultural areas of North America will be obtained from the International Fertilizer Industry Association compilations for the US, Canada, and Mexico.

3.3 Task 3. Data management support for model outputs

The activities conducted in Task 1 will define the model output data products and requirements needed by NACP modelers and synthesis groups. In Task 3, we will acquire, process (as needed), and distribute model output products for use by NACP synthesis and modeling groups.

The output products will be used for integration and analysis during the course of NACP. For example the sources and sinks of carbon from land estimated from inverse modeling will be compiled for quantitative comparison and evaluation against forward model (bottom-up) estimates of sources and sinks. These outputs will be critical components of synthesis activities, including the NACP workshops that MAST-DC will support (see Task 6).

As needed, the model output will be placed into a consistent and uniform format and projection so that the data products can be readily used by NACP for comparison with other model outputs and synthesis data products. Through the collaborative work with NCAR, CDP systems will be deployed in order to expose hierarchical, searchable directories of data products, each with a collection of data services layered upon it. These services will include OPeNDAP capabilities with provisions for format-hiding, along with capabilities for subsetting and aggregation that leverage enhanced metadata descriptions of the data products. CDP software will also provide facilities for catalog federation that may be used to establish autonomous connectivity to Data Central functions, as they are specified. It is also anticipated that there will be data products that require a time-phased authorization profile. A typical example of this would be a dataset that is initially available to a small, limited group that over time expands to include any individual registered with MAST-DC.

As model output data products are assembled at the MAST-DC, we will place the data products in an OPeNDAP server to provide access to the NACP user community. In addition we will register the data products with the NACP Data Central, which will house the data portal consisting of a hierarchical index.

3.4 Task 4. Provide tools for visualization, subsetting, and reprojection

Task 1 will establish the requirements that the NACP modeling and synthesis groups and the Synthesis Task Force have for data tools and services. During the course of this proposed project, we will maintain close coordination with this community to ensure that we meet their evolving needs.

MAST-DC will provide subsetting, visualization, and reprojection services for local data via its web site. .

The MAST-DC will offer web-based GIS services for subsetting and reprojection through the use of both the new ESRI product suite and Minnesota MapServer (an open-source web map server). ESRI's ArcIMS 9.2 (currently in beta release) will include format support for netCDF and should be available by year 2 of this project. These web-based GIS services will enable users to reproject and subset spatial data that is not originally stored in the format or projection they need. Members of the MAST-DC team have deployed both of these web GIS systems in another application in order to take advantage of the strengths of both systems (such as

Minnesota MapServer's WCS support and ArcIMS's advanced user interface). In Task 2, we will prepare model input data in the format that benefits the most modelers (standard format). In Task 4, we will provide services to enable the other modelers to convert from the standard format into other formats as they may require (for example, netCDF to ASCII or HDF-5). We will also assist NACP modelers that want to use the tools we will use internally, such as the Geospatial Data Abstraction Library (GDAL), which is an open source library used by a number of geospatial tools.

We will make available services for browsing, visualizing, regridding, and subsetting three-dimensional atmospheric transport (wind and cloud) data for use by inverse modelers, recognizing that a wide range of uses is likely for these data. It will be important in some NACP activities to be able to select a limited spatial area during a short period, for example to analyze data collected during an intensive field campaign. It will also be important to other users to obtain global analyses over a number of years to provide large-scale boundary conditions for regional analyses. MAST-DC will support both of these needs. The Minnesota MapServer provides specialized subsetting and regridding tools that preserve important properties of the original data (e.g., mass conservation) using its WCS capability.

As a browser-based visualization tool for OPeNDAP data, we will employ the Live Access Server (LAS) OPeNDAP client. LAS is a highly configurable Web server designed to provide flexible access to geo-referenced scientific data. LAS uses Ferret, a data analysis and visualization package available from NOAA/PMEL, to produce plots of OPeNDAP datasets. Ferret is designed to analyze large and complex gridded data sets.

3.5 Task 5. Provide data packages to evaluate model output: confronting models with data

An important component in the modeling framework for NACP is the comparison of model results to field data (Denning et al., 2005). The Synthesis Task Force of the SSG has recently been charged with designing a workshop to address the issues of model uncertainty and multi-model evaluations. As part of this task, we propose to work with the modeling and field investigator communities and the SSG and Synthesis Task Force to develop field-based data products that can be used to evaluate the models in a synthesis workshop setting. These data products will include site-based measurements of carbon dynamics from the USDA Forest Service's Forest Inventory Analysis (FIA) data, and flux tower data from the AmeriFlux and FLUXNET Canada networks. We will provide an evaluation package for the hydrological components of forward models, based on national river discharge monitoring networks. In addition, we will prepare data evaluation packages from atmospheric sampling sites to be used to evaluate forward models.

These evaluation data products will enable modelers and others evaluating results to compare estimates of carbon dynamics at specific locations with the results of bottom-up and top-down models. For example, the FIA provides estimates of forest biomass at over 100,000 sites across US. We will provide the FIA estimates at sites (crudely located to maintain the security of the exact FIA sites, but within an NACP pixel) along with the model drivers and input data products

for that pixel. The models can then be run and the output compared to the FIA data. These comparisons can be used to help improve model processing / algorithms and to gain an understanding of model uncertainty.

For each evaluation site / pixel, we will compile a set of model driver and input data (e.g., products 1-7 from Task 2) and organize on a hierarchical FTP area, in addition to compiling the specific field-based data described below.

1. FIA Evaluation Package

We will work with the FIA Data Management Group and the Modeling and Synthesis group to identify a subset (1000s) of the FIA data for "fuzzed" locations. For these locations we will provide estimates of biometric information (species, land cover, biomass, age structure, forest health) over time as well as the package of model driver and input data.

2. Flux Tower Evaluation Package

We will identify sites in the US and Canada that have daily flux tower data from 1997 to the present. We will contact the national networks (AmeriFlux and Fluxnet Canada) to ensure that gap-filled, aggregated data files are available for this period, including fluxes of water vapor, carbon dioxide, and sensible heat, as well as ancillary measurements available from the network archives. We will acquire the input and driver data for the 1-km pixel from the above 7 data products.

3. River Discharge Network Evaluation Package

Virtually all forward / process models have a water balance component that includes precipitation, evapotranspiration, and runoff. A critical check on the water balance component of these models, particularly on estimates of evapotranspiration, is to compare model-based estimates of runoff against discharge measured at river monitoring stations. For this evaluation package we will compile records for selected river discharge monitoring stations from the USGS river monitoring network, with selection made in coordination with the modeling community and the Synthesis task force. We will use the SRTM DEM and GIS tools to determine the watershed area contributing to discharge measured at selected gauging stations. This watershed area will then be extracted from all of the spatial data layers (data products 1-7 under Task 2) and packaged as a single netCDF file. These data products will be made available on a hierarchical FTP area.

4. Atmospheric trace gas evaluation package

In addition to their use for inverse modeling of regional fluxes, atmospheric mixing ratio data can be used by investigators as an evaluation tool for forward models of surface exchange, provided the researcher couples the flux model to an atmospheric transport model. Systematic long-term measurements of atmospheric carbon gases and related tracers by flask sampling are

conducted by NOAA Global Monitoring Division (formerly the Climate Monitoring and Diagnostics Laboratory). Their data are easily accessible and well-documented at http://www.cmdl.noaa.gov/ccgg/. In addition, there are at least 11 stations in the USA and southern Canada that are now making continuous calibrated measurements of CO_2 mixing ratio (see Table 1). Campaign-based aircraft sampling and in-situ analyses have been performed by a number of investigators in recent years (CO_2 Budget and Rectification Airborne Study (COBRA), Lin et al, 2004, http://www.cmdl.noaa.gov/hotitems/cobra.html; Airborne Carbon in the Mountains Experiment (ACME), http://www.ucar.edu/communications/newsreleases/2004/acme.html; Stephanie Vay, personal comm.). These tower and aircraft data could be a powerful tool for evaluation of forward models of surface carbon cycling and transport, but are difficult to obtain and use because they reside with and are controlled by individual investigators. MAST-DC will work with each of these programs and investigators to obtain quality-controlled mixing ratio data that can be placed in the public domain. We will document these data and make them available for download from the Center for evaluation of forward models.

3.6 Task 6. Support synthesis activities, including data support for workshops

As modeling tools and datasets mature through the NACP, an important element of synthesis and integration will be a series of model intercomparison and evaluation workshops to be organized by the science leadership with the support of the CC IWG. These workshops, not yet defined or scheduled, will involve multiple models being used to perform standardized experiments and to calculate the same quantities (carbon pools and/or fluxes). These workshops will enable the intercomparison of the results of these calculations to assess uncertainty, and the evaluation of these results by quantitative comparison to observations.

MAST-DC will work closely with the NACP-SSG and especially with the Synthesis Task Force to define model intercomparison and evaluation workshops, to prepare model input data sets in support of standardized experiments, to collect and archive model output, and to provide tools to facilitate efficient comparison of models to models and observations at the workshops.

Table 1. Tower-based observatories of continuous CO_2 mixing ratio.

Site	Latitude	Longitude	Principal Investigator
NOBS (Manitoba Old Black Spruce; BOREAS NSA - Old Black Spruce Canada)	53° 59’ N	105° 7’ W	Barr, Alan
BERMS Saskatchewan Flux Station - Western Boreal Forest (southern ecotone)	54° 29’ N	105° 49’ W	Barr, Alan
Fraserdale (Groundhog River, Ontario,Canada)	48° 13’ N	82° 9’W	McCaughey, Harry
Metolius - intermediate (69 yrs) ponderosa pine, Oregon USA	44° 27' N	121° 33' W	Law, Beverly E.
WLEF (Park Falls/WLEF, Wisconsin)	45° 57' N	90° 16' W	Bakwin, Peter S.
Harvard Forest (Massachusetts)	42° 32' N	72° 10' W	Wofsy, Steven C
Maine (Howland Forest - main tower, Maine USA)	45° 12' N	68° 44' W	Hollinger, David Y.
Niwot Ridge Forest, Colorado	40° 02' N	105° 33' W	Monson, Russ K.
Atmospheric Radiation Monitoring (ARM) Site in Oklahoma	36° 36' N	97° 29' W	Torn, Margaret S.
KETK Texas	31° 19' N	97° 19' W	Andrews, Arlyn
Morgan Monroe State Forest	39° 19' N	86° 25' W	Schmid, Hans Peter

Previous model intercomparison activities (e.g., VEMAP, TransCom) have shown that provision of standard data products for model inputs and model output products in easily digestible form lead to increased participation even by modeling groups who have not been funded by the original intercomparison activity. In the case of TransCom, convenient and well-organized model inputs and outputs on the web led eventually to participation by every major inverse modeling group in the world in the activity. MAST-DC will work closely with the scientific and agency leadership to provide this kind of support for model intercomparison and evaluation for NACP.

Example workshops:

- carbon fluxes, yields, and fate of organic material simulated by agroecosystem models to be compared to USDA agricultural census data
- simulations and analyses of forest biomass and carbon stocks to be compared to FIA and NRI data
- simulated/analyzed lateral flow of water, carbon, and nutrients in rivers to be compared to USGS discharge and water quality data

- forward models of surface fluxes and atmospheric transport, to be compared to campaign-style airborne sampling data.

3.7 Other data system functions: Acquiring and Distributing Model Source Code

One of the recommendations from the Data Task Force and in the NASA ROSES Solicitation (A.6. NACP) was that the Modeling and Synthesis Thematic Data Center would "acquire and distribute model source code and associated documentation" (NACP 2005b). This function for a data center is relatively new (Thornton et al., 2005) and is currently being conducted by the NASA's Distributed Active Archive Center (DAAC) for biogeochemical dynamics at Oak Ridge National Laboratory (ORNL DAAC). MAST-DC will rely on the ORNL DAAC to perform this function (see attached letter from R. L. Graham, ORNL DAAC Manager). MAST-DC will not request funds to archive models, but will link to the ORNL DAAC's model archive, so that the NACP community can access, query, and obtain the archived models.

4.0 Benefits of Proposed Work

The proposed MAST-DC will provide data and services that the modelers need in a central location, with common and co-registered spatial projection, in easily converted formats. The MAST-DC will free modelers and those doing the synthesis and integration from having to perform data management functions and thereby enable NACP participants to conduct their work more readily. It will also facilitate the development of new model products needed by models and assist in gaining new insights into the carbon cycle in North America.

The MAST-DC data services will result in the best available data sets being used by modelers and facilitate utilization of common data and significantly enhance the inter-comparison of results. The availability of high quality data sets and advanced tools for their utilization will enable more participation by modelers in synthesis activities. In addition, NACP synthesis activities will benefit from the community-based infrastructure provided by the MAST-DC for handling model results and model code. This will greatly enhance our capability to diagnosis model results and to intercompare model outputs.

5.0 Management Plan and Personnel

This project will be a collaborative effort between environmental scientists, computer scientists, and data managers at ORNL, and biogeochemical modelrs at NCAR. Like NACP, this data project is multidisciplinary and multi-institutional in order to meet the program's objectives. The team assembled for MAST-DC is uniquely qualified to carry out the proposed work, and brings a wide range of personal expertise and institutional strengths in the scientific, computer science, and data base management arenas required by NACP.

Robert Cook, ORNL, will have overall responsibility of the project and will coordinate the activities of the other co-Is. He will work with the SSG, the Synthesis Task Force, and the Data

Systems Task Force to establish the data product and service requirements. When Data Central has been designed and implemented, Cook will engage the Data Central personnel to identify requirements and the way in which MAST-DC will complement the efforts of Data Central. At ORNL, he will oversee the assembly of required data products (model input data, model output data, evaluation data packages) and data management and services. *Tim Rhyne*, ORNL, will be responsible for overseeing the systems operation and maintenance, including the interface to Data Central and Web site and services external users. Tim will provide technical coordination between ORNL and the computer science component of Task 4. *Mac Post*, ORNL, will provide scientific guidance on the model input data products (Task 2).

Peter Thornton, NACR, will be responsible for overseeing production of the DAYMET data output at NCAR and to provide support for the transfer of the code base into an operational mode at the MAST-DC. Post and Thornton will also work on Tasks 1, 3, 5, and 6 to provide specific guidance on implementing the data requirements that arise out of discussions with the modeling and synthesis groups and the Synthesis Task Force (Task 1) and provide guidance regarding placing the model output data products into the MAST-DC for both top-down and bottom-up models (Task 3). In addition, these two Co-Is will assist in supporting the model intercomparison and model-data comparisons workshops, including providing guidance on preparing data products including uncertainty, to facilitate quantitative comparison of different carbon accounting approaches for NACP.

Given the scope of the project, and the number of participants, active coordination of everyone's efforts will be essential. Teleconferences will be held bi-weekly among all project participants to summarize progress, list open issues, and review upcoming milestones. We will communicate via email and also develop an internal project web page /Wiki to report status and distribute preliminary products.

References

Amthor, J.S., J.M. Chen, J.S. Clein, S.E. Frolking, M.L. Goulden, R.F. Grant, J.S. Kimball, A.W. King, A.D. McGuire, N.T. Nikolov, C.S. Potter, S. Wang, and S.C. Wofsy. 2001. Boreal forest CO2 exchange and evapotranspiration predicted by nine ecosystem process models: intermodel comparisons and relationships to field measurements. Journal of Geophysical Research 106: 33,623-33,648.

Bennett, A. F., 2002: Inverse modeling of the ocean and atmosphere, Cambridge University Press, 346 pp.

Denning, A.S. and others. 2005. The Science Implementation Strategy for the North American Carbon Program. NACP Implementation Strategy Group. Available on-line at http://www.nacarbon.org/nacp/documents.html

Galloway JN, FJ Dentener, DG Capone, EW Boyer, RW Howarth, SP Seitzinger, GP Asner, C Cleveland, P Green, E Holland, DM Karl, AF Michaels, JH Porter, A Townsend and C Vörösmarty. 2004. Nitrogen Cycles: Past, Present and Future. Biogeochemistry 70: 153-226.

Gerbig, C., J.C. Lin, S.C. Wofsy, B.C. Daube, et al.. Toward constraining regional-scale fluxes of CO2 with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms, J. Geophys. Res., 108(D24), 4756, doi: 10.1029/2002JD003018, 2003.

Gerbig, C., J.C. Lin, S.C. Wofsy, B.C. Daube, et al., Toward constraining regional-scale fluxes of CO2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, J. Geophys. Res., 108(D24), 4757, doi: 10.1029/2003JD003770, 2003.

Gurney, K. R., R.M. Law, A.S. Denning, P.J. Rayner, B. Pak, and the TransCom 3 L2 modelers, 2004. Transcom 3 Inversion Intercomparison: Control results for the estimation of seasonal carbon sources and sinks. Global Biogeochemical Cycles, 18, G 1010, doi: 10.1029/2003GB002111.

Hanson PJ, Amthor JS, Wullschleger SD, Wilson KB, Grant RF, Hartley A, Hui D, Hunt ER Jr., Johnson DW, Kimball JS, King AW, Luo Y, McNulty SG, Sun G., Thornton PE, Wang S, Williams M, Baldocchi DD, Cushman RM (2004) Oak forest carbon and water simulations: model intercomparisons and evaluations against independent data. Ecological Monographs 74(3):443-489.

Lamarque, J. F., J. Kiehl, G. Brasseur, T. Butler, P. Cameron-Smith, W. D. Collins, W. J. Collins, C. Granier, D. Hauglustaine, P. Hess, E. Holland, L. Horowitz, M. Lawrence, D. McKenna, P. Merilees, M. Prather, P. Rasch, D. Rotman, D. Shindell, and P. Thornton, in press: Assessing future nitrogen deposition and carbon cycle feedbacks using a multi-model approach. Part 1: Analysis of nitrogen deposition. Journal of Geophysical Research.

Lin, J.C., C. Gerbig, B.C. Daube, et al., An empirical analysis of the spatial variability of atmospheric CO2: implications for inverse analyses and space-borne sensors, Geophysical Research Letters, 31 (L23104), doi:10.1029/2004GL020957, 2004.
Miller, D. A. and R.A. White 1998. A Conterminous United States Multilayer Soil Characteristics Dataset for Regional Climate and Hydrology Modeling. Earth Interactions 2: 1-26.

Myneni, R. B., R. R. Nemani and S. W. Running. 1997. Estimation of global leaf area index and absorbed par using radiative transfer models, IEEE Transactions on Geoscience and Remote Sensing, 35, 1380-1393.

NACP. 2005a. Data Management for the North American Carbon Program. Workshop Report, January 25-27, 2005, New Orleans, LA. Available on-line at http://www.nacarbon.org/nacp/documents.html

NACP. 2005b. Phase 1 Recommendations for a North American Carbon Program Data System. NACP Data System Task Force. Available on-line at http://www.nacarbon.org/nacp/documents.html

Nicholls, M.E., A.S. Denning, L. Prihodko, P.-L. Vidale, K. Davis, P.Bakwin, 2004: A multiple-scale simulation of variations in atmospheric carbon dioxide using a coupled biosphere-atmospheric model. Journal of Geophysical Research, 109, D18117, doi: 10.1029/2003JD004482.

Ramankutty, N. and J. Foley. 1999. Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global Biogeochemical Cycles, 13, 997-1027.

Ryan, M.G., E.R. Hunt Jr., R.E. McMurtrie, G.I. Ågren, J.D. Aber, A.D. Friend, E.B. Rastetter, W.M. Pulliam, R.J. Raison, and S. Linder. 1996a. Comparing models of ecosystem function for temperate conifer forests. I. Model description and validation, pp. 313-362 in A.I. Breymeyer, D.O. Hall, J.M. Melillo, and G.I. Agren, editors. Global Change: Effects on coniferous forests and grasslands, John Wiley & Sons, New York.

Ryan, M.G., R.E. McMurtrie, G.I. Ågren, E.R. Hunt Jr., J.D. Aber, A.D. Friend, E.B. Rastetter, and W.M. Pulliam. 1996b. Comparing models of ecosystem function for temperate conifer forests. II. Simulation of the effect of climate change, pp. 363-387. in A.I. Breymeyer, D.O. Hall, J.M. Melillo, and G.I. Agren, editors. Global Change: Effects on coniferous forests and grasslands, John Wiley & Sons, New York.

Smith et al. (1997) A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. Geoderma 81:153-225.

Thornton, P.E., R.B. Cook, B.H. Braswell, B.E. Law, W.M. Post, H.H. Shugart, B.T. Rhyne, and L.A. Hook. 2005. Archiving numerical models of biogeochemical dynamics. Eos, Transactions of the American Geophysical Union 86(44): 431.

Thornton, P. E., H. Hasenauer, et al. (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., S. W. Running, et al. (1997). "Generating surfaces of daily meteorological variables over large regions of complex terrain." Journal of Hydrology 190: 214-251.

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.

USDA. 1994. State soil geographic (STATSGO) data base. Miscellaneous Publication Number 1492. US Department of Agriculture, Soil Conservation Service, Washington, DC.

VEMAP Members. 1995. Vegetation/ecosystem modeling and analysis project: comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate change and CO_2 doubling. Global Biogeochemical Cycles 9:407-437.

Waisanen, P.J. and N.B. Bliss. 2002. Changes in population and agricultural land in conterminous United States counties, 1790-1997. Global Biogeochemical Cycles 16, 1137-

Wofsy, S.C. and R.C. Harriss. 2002. The North American Carbon Program Plan (NACP). A Report of the Committee of the U.S. Carbon Cycle Science Steering Group. Available on-line at http://www.esig.ucar.edu/nacp/

Personnel

Robert Cook has extensive experience managing large environmental and data management projects, including serving as ORNL DAAC Scientist for the past 8 years. Cook was Co-Chair of the NACP Data Management Planning Workshop, conducted in January 2005, and serves on the NACP SSG and the Mid-Continent Intensive Task Force. Cook is the co-chair of the NACP Data Systems Task Force.

Tim Rhyne serves as the Systems Engineer at the ORNL DAAC and directs the project activities of the technical staff. He also serves as ORNL DAAC technical contact for NASA's work on the use of OpenGIS Consortium standards for Web Mapping Servers and Web Content Servers. Rhyne was the PI for development and implementation of Mercury, a Web-based metadata search and data retrieval system for distributed data that relies heavily on XML.

Mac Post is a senior scientist at the Environmental Sciences Division, ORNL, and has over 15 years of experience in biogeochemical modeling and carbon cycle synthesis activities.

Peter Thornton is a Scientist II at the National Center for Atmospheric Research (NCAR) Climate & Global Dynamics Division, with seven years experience in biogeochemical dynamics modeling and carbon cycle studies. Thornton was Co-Chair of the NACP Data Management Planning Workshop, conducted in January 2005, and serves on the NACP SSG. Thornton is co-chair of the Synthesis Task Force and co-chair of the NACP Data Systems Task Force.

Time table for activities in MAST-DC

Task 1 (Establish specific data management needs of NACP modelers and synthesis groups), will occur throughout the course of this proposed project. MAST-DC will initially seek input from all NACP modelers, through sessions at the All-investigator meeting (tentatively scheduled for Summer 2006). We will follow-up through periodic teleconferences, in which the status of the products as well as quality and other issues are identified.

Task 2 (Provide data management support for model inputs) represents the bulk of the work for this project. It will be conducted for all three years, with various data products being generated throughout the project. We will work directly with the NACP community to prioritize our efforts, but we have provided a rather detail plan for Task 2 in this proposal. Our tentative timeline contains the following items. The climate and weather data will be generated by MAST-DC during each of the three years. During the first year, we proposed to develop the NACP grid, projections, and data formats. Also, in this first year we will compile soils, land cover, and some of the historical land use data. In year 2, we will complete the historical land use data and also compile LAI / fPAR, and nitrogen data. In year 3, we will acquire and process data products developed during the first two years of NACP, such as MODIS products from reprocessing activities (Collection 5) as well as other regional and continental scale projects developed during Project 2006 – 2007 for NACP. The exact schedule for Task 2 data products

will be determined with direct input from the SSG, the Synthesis Task Force, and the NACP community.

Task 3 (Provide data management support for model outputs) will similarly be conducted during the three years of this project. We expect that the level of effort will be somewhat lower for the first year, although some modeling activities have been previously funded and are on-going. As NACP modeling activity funded by the ROSES 2005 call (A6. NACP) begins in earnest, we expect to see some increase in modeling results in Project Years 2 and 3.

Task 4 (Provide tools for accessing, subsetting, and visualization) will continue throughout all four years of the project. During the first year, we will establish a web-based GIS services for subsetting and reprojection. During subsequent years, we will populate the WMS / WCS with data sets contributed by MAST-DC. We will also assist the community as they access and use the open source tools at the MAST-DC. ……

Task 5 (Provide data packages to evaluate model output) will be conducted during all three years. During Years 1 and 2, the primary efforts will working directly with the Synthesis Task Force and the NACP modeling community to define the exact types of data evaluations packages that would be most beneficial for testing models. We will compile these site-based data packages. In years 2 and 3, we will pull together driver data at each of the sites / pixels.

Task 6 (Support synthesis activities, including data support for workshops) will occur over the course of the three-year project. During the NACP All-Scientist meetings (summer 2006 and annulally thereafter) we will work directly and closely with the Synthesis Task Force as they hold initial break-out sessions to identify and plan the NACP synthesis activities. As the synthesis activities / workshops evolve we will participate in planning meetings, teleconferencesms, and the workshops themselves to ensure that data management support is provided and the specific data products and services are available.

Facilities and Equipment

The MAST-DC will be physically co-located with the ORNL DAAC. The MAST-DC will leverage the DAAC environment without compromising DAAC capabilities.

In year one of the project, MAST-DC will acquire its own storage space in the form of an expansion of the DAAC's Dell/EMC Storage Area Network (SAN). Additional Fibre Channel RAID Array enclosures will be acquired for an initial MAST-DC storage in the range of 4Tb. An additional 2-4Tb will be added to the SAN each following year, as needed. By adding on to the DAAC's existing SAN we will reduce the cost of the system. Each Array will be mounted in one of the DAAC 19" racks located in ORNL's Building 5600 computer room. Even though the use of this room is free, ORNL provides a raised computer room floor, air conditioning, fire suppression system, 24 hour access control, and UPS.

Processing power for MAST-DC is not expected to be overly demanding. We will use cycles available on the DAAC's Dell PowerEdge 6650 quad CPU system. However, MAST-DC

software will be completely isolated from DAAC systems by installing VMWare ESX server software and running the MAST-DC system(s) in virtual machines. MAST-DC will provide an 8Gb memory expansion for the DAAC Dell system to support its virtual machines. MAST-DC technical staff will determine the best configuration of VMs to support MAST-DC software (OpenDAP, LAS, Minnesota MapServer, Apache, etc.) based on empirical study after upgrades are made. MAST-DC will use the DAAC's ArcIMS installation for web map services.

Resumes

Resumes for the PI and Co-Is are provided on the following pages.

1. Robert B. Cook, ORNL
2. B. Timothy Rhyne, ORNL
3. W.M. Post, ORNL
4. Peter E. Thornton, NCAR

Robert B. Cook

PRESENT POSITION:

Senior Staff, Environmental Data Systems, Environmental Sciences Division, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6407
Phone: (865) 574-7319; FAX (865) 574-4665; E-mail: cookrb@ornl.gov

EDUCATION:

Eisenhower College, Seneca Falls, New York	Chemistry	**B.A.**	1974
Columbia University, New York, New York	Geology	**M.A.**	1976
Columbia University, New York, New York	Geology	**M.Ph.**	1977
Columbia University, New York, New York	Geochemistry	**Ph.D.**	1981

RELEVANT PROJECT EXPERIENCE:

- Scientist for the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC), NASA, from 1996 to present
- Principal Investigator, Metadata Clearinghouse for the National Biological Information Infrastructure, USGS, 2003-present
- Principal Investigator, FLUXNET DIS Project, NASA, 2003 – 2005

PROFESSIONAL ACTIVITIES

- Member, Terrestrial Carbon Observations Panel of the Global Terrestrial Observing System (GTOS) (2003 – 2005).
- Associate Editor, Biogeochemistry (1996 - present).
- Co-Chair, Steering Committee, Data Planning Workshop, North American Carbon Program (January 2005)
- Member, Science Steering Group, North American Carbon Program (2005 – present)
- Member, Mid-Continent Intensive Task Force, North American Carbon Program (2005 – Present)
- Co-Chair, Data Systems Task Force, North American Carbon Program (2005 – Present)
- Member, CyberInfrastructure Planning Group, Long-Term Ecological Research Network (2005 – 2006)

SELECTED PUBLICATIONS (out of over 40 publications)

Five Most Relevant Publications

Cook, R.B., R.J. Olson, P. Kanciruk, and L.A. Hook. 2001. Best practices for preparing ecological and ground based data sets to share and archive. Ecol. Bulletins 82: 138 - 141.

Olson, R. J., S.K. Holladay, R.B. Cook, E. Falge, D. Baldocchi, L. Gu. 2004. FLUXNET: Database of Fluxes, Site Characteristics, and Flux-Community Information. ORNL Technical Memorandum, Oak Ridge, TN.

Denning, S., L. Dilling, S. Doney, L. Heath, D. McGuire, B. McKee, C. Sabine, R. Oren, K.Paustian, J. Randerson, J. Reilly, S.W. Running, R. Stallard, M. Torn, S.C. Wofsy, R.B. Cook, and others. 2005. Science Implementation Strategy for the North American Carbon Program. Carbon Cycle Interagency Working Group and Science Steering Group. http://www.nacarbon.org/nacp/documents.html

Cook, R.B. and P.E. Thornton, co-chairs. 2005. Data Management for the North American Carbon Program, Report from a Workshop, New Orleans, LA. http://www.nacarbon.org/nacp/documents.html

Thornton, P.E., R.B. Cook, B.H. Braswell, B.E. Law, W.M. Post, H.H. Shugart, B.T. Rhyne, and L.A. Hook. 2005. Archiving numerical models of biogeochemical dynamics. Eos, Transactions of the American Geophysical Union 86(44): 431.

Five Other Publications

Bolin, B. and R.B. Cook (Eds.). 1984. The Major Biogeochemical Cycles and their Interactions. John Wiley and Sons. Chichester, United Kingdom. 532p.

Cook, R.B., K.A. Rose, A.L. Brenkert, and P.F. Ryan. 1992. Systematic comparison of ILWAS, MAGIC, and ETD watershed acidification models. 3. Mass balance budgets for acid neutralizing capacity. Environmental Pollution. 77: 235-242.

Cook, R.B., J.W. Elwood, R.R. Turner, M.A. Bogle, P.J. Mulholland, and A.V. Palumbo. 1994. Acid-base chemistry of high-elevation streams in the Great Smoky Mountains. Water, Air, and Soil Pollution 72: 331-356.

Cook, R.B., J.D. Joslin, S.M. McLaughlin, C. Egar, J. Knoepp, and C. Trettin. 1998. Effects of Acid Deposition on Forests. In: 1996 Integrated Assessment Report to Congress. National Acid Precipitation Assessment Program, Washington, D.C.

Cook, R.B., G.W. Suter, and E.R. Sain. 1999. Ecological Risk Assessment in a Large River-Reservoir: 1. Introduction and Background. Environmental Toxicology and Chemistry 18: 581 - 588.

B. Timothy Rhyne

PRESENT POSITION:
Oak Ridge National Laboratory (ORNL), Computational Sciences and Engineering Division, Data System Sciences and Engineering
Phone: (865)574-7447; Fax: (865)574-4665; Email: rhynebt@ornl.gov

KEY SKILLS:
Scientific Data Management
Metadata and Interoperability Standards
Extensible Markup Language (XML) and Applications
Personal Computer Hardware and Software
Relational Database Design
Bar Code Technology
Technical Project Management

EDUCATION:
M.S. in Computer Science, University of Tennessee at Knoxville (1988)
B.S. in Industrial Engineering, University of Tennessee at Knoxville (1979)

EXPERIENCE:

Tim Rhyne serves as the Systems Engineer for the ORNL Distributed Active Archive Center (DAAC), a NASA-funded ecological data center operating out of Environmental Sciences Division. In this capacity, Tim directs the project activities of the DAAC's technical staff. He also serves as technical point-of-contact with NASA on numerous subjects, traveling to meetings and participating in frequent telecons. Included in this NASA work is the use of OpenGIS Consortium (OGC) standards for Web Mapping Servers (WMS) and Web Content Servers (WCS). The DAAC also uses Mercury (described in the next paragraph) in several ways.

Mr. Rhyne is the Principal Investigator (PI) for and designer of Mercury, a Web-based metadata search and data retrieval system for distributed data that relies heavily on XML. Mercury supports numerous projects, including ones for NASA, DOE, EPA, and USGS; providing access to data in numerous countries around the world. The Mercury project incorporates several national and international standards, including the Content Standard for Digital Geospatial Metadata (CSDGM), the Web Services Description Language (WSDL), and Universal Description, Discovery and Integration of Web Services (UDDI). The Mercury team has won a cash prize and a Certificate of Recognition from NASA.

Rhyne is the PI for ORNL's bioinformatics support to the USGS National Biological Information Infrastructure (NBII). ORNL operates NBII's metadata clearinghouse that includes nearly 50 partner locations. ORNL has also written or set up various web services for NBII including thesaurus, specimen, gazetteer, and OGC Catalog Services.

In his career, Mr. Rhyne has also served in line management, directing the Microcomputer Applications Group for eight years. He has authored various reports, proposals, and posters and has given numerous talks in the US and other countries.

Clearance: DOE Q (Top Secret) Clearance

Wilfred M. Post
Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6335
E-mail: wmp@ornl.gov

Senior Scientist, Environmental Sciences Division, Oak Ridge National Laboratory, and Adjunct-Associate Professor, Ecology and Evolutionary Biology, University of Tennessee.

EDUCATION:

B.S. 1971, University of Wisconsin Madison, Mathematics
M.S. 1975, University of Wisconsin Madison, Botany
Ph.D. 1978, University of Tennessee, Ecology

PUBLICATIONS: (Selected, 90+ total)

Emanuel, W. R., G. G. Killough, W. M. Post, and H. H. Shugart. 1984. Modeling terrestrial ecosystems in the global carbon cycle with shifts in carbon storage capacity by land-use change. Ecology 65:970--983.

Pastor, J., and W.M. Post. 1986. Influence of climate, soil moisture, and succession on forest carbon and nitrogen cycles. Biogeochemistry 2:3-27.

King, A. W., W. R. Emanuel, and W. M. Post. 1992. Projecting future concentrations of atmospheric CO2 with global carbon cycle models: The importance of simulating historical changes. Environmental Management 16:91-108.

Post, W. M., J. Pastor, A. W. King, and W. R. Emanuel. 1992. Aspects of the interaction between vegetation and soil under global change. Water, Air, and Soil Pollution 64:345-363.

Shugart, H. H., T. M. Smith, and W. M. Post. 1992. The potential for application of individual-based simulation models for assessing the effects of global change. Annual Reviews of Ecology and Systematics 23:15-38.

Post, W. M., A. W. King, and S. D. Wullschleger 1997. Historical variations in terrestrial biospheric carbon storage. Global Biogeochemical Cycles 11:99-109.

King, A. W., W. M. Post, and S. D. Wullschleger 1997. The potential response of terrestrial carbon storage to changes in climate and atmospheric CO2. Climatic Change 35:199-227.

Paustian, K., E. Levine, W. M. Post, and I. M. Ryzhova 1997. The use of models to integrate information and understanding of soil C at the regional scale. Geoderma 79:227-260.

Gu, L., W.M. Post, and A.W. King. 2003. Fast labile carbon turnover obscures sensitivity of heterotrophic respiration from soils to temperature: a model analysis. Global Biogeochemical Cycles 18:GB1022, doi:10.1029/2003GB002119.

Jain, A.K., T.O. West, X. Yang, and W.M. Post. 2005. Assessing the Impact of Changes in Climate and CO2 on Potential Carbon Sequestration in Agricultural Soils. Geophysical Research Letters 32, L19711, doi:10.1029/2005GL023922.

Peter E. Thornton

Contact information:

Address: National Center for Atmospheric Research
1850 Table Mesa Dr.
Boulder, CO 80305

Ph.D. 1998. The University of Montana, Missoula, MT

Advisor and Department: Steven W. Running, School of Forestry

Research Focus: Terrestrial biogeochemistry, biometeorology, and visible, near-infrared, and thermal-infrared remote sensing.

Dissertation: *Regional ecosystem simulation: combining surface- and satellite-based observations to study linkages between terrestrial energy and mass budgets*.

Current Employment:

August 2004 – present: **Scientist II**, Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO.

Recent Publications (relevant to this proposal)

Thornton, P.E., R.B. Cook, B.H. Braswell, B.E. Law, W.M. Post, H.H. Shugart, B.T. Rhyne, L.A. Hook, **2005**. Archiving numerical models of biogeochemical dynamics. *Eos - Transactions of the American Geophysical Union* 86 (44): 431-432.

Thornton, P.E., N. Rosenbloom, **2005**. Ecosystem model spin-up: estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model. *Ecological Modeling* 189 (1/2): 25-48.

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.

Hanson, P.J., J.S. Amthor, S.D. Wullschleger, K.B. Wilson, R.F. Grant, A. Hartley, D. Hui, E.R. Hunt Jr., D.W. Johnson, J.S. Kimball, A.W. King, Y. Luo, S.G. McNulty, G. Sun, P.E. Thornton, S. Wang, M. Williams, D.D. Baldocchi, R.M. Cushman, **2004**. Oak forest carbon and water simulations: model intercomparisons and evaluations against independent data. *Ecological Monographs*, 74(3): 443-489.

Hasenauer, H., K. Merganicova, R. Petritsch, S. Pietsch, P.E. Thornton, **2003**. Validating daily climate interpolations of complex terrain in Austria. *Agricultural and Forest Meteorology* 119: 87-107.

McKenzie, D., D.W. Peterson, D.L. Peterson, P.E. Thornton, **2003**. Climatic and biophysical controls on conifer species distributions in mountain forests of Washington State, USA. *Journal of Biogeography* 30: 1093-1108.

Letters of Support / Collaboration

1. Jeff Masek, Goddard Space Flight Center, NASA
   Principal Investigator, Landsat Ecosystem Disturbance Adaptive Processing System

2. Kevin Gurney

3. Jeanne Behnke
   Manager, Science Operations Office, Earth Science Data and Information Systems Project

4. Robin Graham
   Manager, ORNL Distributed Active Archive Center (DAAC)

Acronyms

ACME - Airborne Carbon in the Mountains Experiment
AGU - American Geophysical Union
ARM - Atmospheric Radiation Measurement (Program)
ArcIMS – Software product for delivering GIS data and services via web.
ASCII - American Standard Code for Information Interchange
AVHRR - Advanced Very High Resolution Radiometer
BERMS - Boreal Ecosystem Research and Monitoring Sites
BOREAS - Boreal Ecosystem Atmosphere Study
CCIWG – Carbon Cycle Interagency Working Group
CDF – Common Data Form
CDP – Community Data Portal
CMDL - Climate Monitoring and Diagnostics Laboratory
COBRA – CO2 Budget and Rectification Airborne Study
CSDGM – Content Standard for Digital Geospatial Metadata
DAAC – Distributed Active Archive Center
DAYMET – Daily Meteorology Software
DEM – Digital Elevation Model
DIF – Directory Interchange Format
DTED – Digital Terrain Elevation Data
DOE – Department of Energy
ECMWF – European Centre for medium-Range Weather Forecasting
EMC - A company specializing in information storage and management.
ESRI – Environmental Systems Research Institute
ESRL – Earth System Research Laboratory
ESX - Software that allows one machine to act as several virtual machines
FIA – Forest Inventory Analysis
FGDC – Federal Geospatial Data Committee
fPAR – fraction of photosynthetically active radiation absorbed by vegetation
FTP - File Transfer Protocol
GDAL – Geospatial Data Abstraction Library
GIS – Geographic Information System
GMD - Global Monitoring Division
GSD – Global Systems Division
HDF – Hierarchical Data Format
KETK - TV station's tower in Texas
LAI – Leaf Area Index
LANDSAT - Land Remote-Sensing Satellite (System)
LEDAPS – Landsat Ecosystem Disturbance Adaptive Processing System
LAS – Live Access Server
MAST-DC – Modeling and Synthesis Thematic Data Center
MODIS – Moderate Resolution Imaging Spectroradiometer
NACP – North American Carbon Program
NASA - National Aeronautics and Space Administration
NCAR - National Center for Atmospheric Research

NCEP – National Centers for Environmental Prediction
NLCD – National Land Cover Data
NOAA – National Oceanic & Atmospheric Administration
NOBS - Naval Observatory
NRCS - Natural Resources Conservation Service
NSA - northern study area in BOREAS
NSDI – National Spatial Data Infrastructure
ODC - OPeNDAP Data Connector
OGC – Open Geospatial Consortium
OPeNDAP – Open-source Project for a Network Data Access Protocol
ORISE – Oak Ridge Institute for Science and Education
ORNL – Oak Ridge National Laboratory
PMEL - Pacific Marine Environmental Laboratory
RUC – Rapid Update Cycle
SAN – Storage Area Network
SRTM – Shuttle Radar Topographic Mission
SSG – Science Steering Group
UMD - University of Maryland
USDA – United States Department of Agriculture
USGS – United States Geological Survey
VCF – Vegetation Continuous Fields
VEMAP - Vegetation/Ecosystem Modeling and Analysis Project
VM - Virtual Machine
WCS – Web Coverage Service
WMS – Web Map Service
WLEF - TV station's tower in Wisconsin