8.1 THE NOAA OPERATIONAL MODEL ARCHIVE AND DISTRIBUTION SYSTEM (NOMADS)

Glenn K. Rutledge*[1], Ronald Stouffer[2], Neville Smith[3], and Bryan Lawrence[4]

[1]National Climatic Data Center
Scientific Services Division
Asheville, NC 28801

[2]NOAA Office of Atmospheric Research
Geophysical Fluid Dynamics Laboratory
Princeton, NJ 08542

[3]Bureau of Meteorology Research Centre
Bureau of Meteorology
Melbourne, Victoria 3001, Australia

[4]British Atmospheric Data Centre
Rutherford Appleton Laboratory
Chilton, Didcot, OX11 0QX, U.K.

1.0 Introduction

To address a growing need for real-time and retrospective General Circulation Model (GCM) and Numerical Weather Prediction (NWP) models and data, the National Climatic Data Center (NCDC) along with the Geophysical Fluid Dynamics Laboratory (GFDL) and the National Centers for Environmental Prediction (NCEP) have initiated the collaborative NOAA Operational Model Archive and Distribution System (NOMADS) (Rutledge, et al., 2002). This new system allows access to weather and climate model data sets. A new paradigm for sharing data among climate and weather modelers is evolving. It takes advantage of the Internet and relatively inexpensive computer hardware.

In this new framework, scientists put their data into a computer on the Internet Software running on the computer allows outside users to see not only their local data but also data on other computers running this same software. This framework is also known as “Grid” computing (see Section 2.1). NOMADS uses this framework and is a network of data servers using established and emerging technologies to access and integrate model and other data stored in geographically distributed repositories in heterogeneous formats. NOMADS enables the sharing and inter-comparing of model results and the comparing of model results with observations. It is a major collaborative effort, spanning multiple government agencies and academic institutions. The data available under the NOMADS framework include model input and NWP gridded output from NCEP, and GCM and simulations from GFDL, NCAR, and other leading institutions from around the world. The effort has gained many international partners, as the need for a convergence of emerging yet similar distributed data efforts become implemented.

Corresponding author address: National Climatic Data Center, Asheville, NC 28801; *E-mail:* Glenn.Rutledge@noaa.gov Tel: (828) 271-4097

The goals of NOMADS are:

- Provide access to NWP (weather) and GCM (climate, including ocean related) models.
- Provide the observational and model data assimilation products for Regional model initialization and forecast verification for use in both weather and climate applications.
- Develop linkages between the research and operational modeling communities and foster collaborations between the climate and weather modeling communities.
- Promote product development and collaborations within the geo-science communities (ocean, weather, and climate) to improve operational weather and climate forecasts by allowing more users to interact with the model data.
- Foster inter-disciplinary research to study multiple earth systems using collections of distributed data under sustainable system architectures.
- Ensure permanent stewardship of select agreed upon model data sets.

NOMADS will provide retrospective access to model and observational data by a wide variety of users via the Internet and eventually the Next Generation Internet (NGI) or Internet -2. NOMADS is an inter-operable network architecture with fully integrated data access and manipulation tools using a distributed, format independent client-server methodology. NOMADS benefits from existing and emerging technologies to provide distributed access to models and data.

To enable universal user access and system inter-operability, NOMADS has four primary data servers: 1) Climate Data Analysis Tools (CDAT), (Williams, et al., 2002); 2) the Open source Project for a Network Data Access Protocol (OPeNDAP), (formally called the Distributed Oceanographic Distribution System (DODS)), (Davis, et al., 1999); 3) the GrADS Data Server (GDS); (Doty, et al., 2001); and 4) the Live Access Server (LAS), (Hankin, et al., 2001). Traditional on-line data services through standard Web-based File Transport Protocol (FTP) will also be available. Users with commonly available

desktop data manipulation tools such as Ferret, MatLAB, IDL, GrADS and even typical Web browsers can access data in their preferred format. To provide for this format neutral data access, NOMADS uses the EXtensible Markup Language (XML) (Bray, et al., 1998).

Both researchers and policy-makers now expect national data assets to be easily accessible and interoperable with each other, regardless of their physical location. As a result, an effective interagency distributed data service requires the coordination of data infrastructure and management extending beyond traditional organizational boundaries.

With NOMADS and its collaborators, users at any skill level will be able to obtain weather and climate information. This will allow the users to make better, informed decisions about how nature will impact their future, either in their life, or business decisions.

2.0 Background

A major transition in weather and climate prediction is now occurring, one in which real-time and retrospective NWP and GCM research is spreading from a handful of national centers to dozens of groups across the country. This growth of global and regional scale NWP and GCM model development is in part, now possible due to:

- The availability of low-cost multiprocessor workstations.
- The availability of regional scale models that run on these workstations (e.g., MM5).
- The availability of climate simulations, analysis and forecast grids from NCEP, GFDL, NCAR and other institutions.

NOMADS addresses model data access needs as outlined in the U.S. Weather Research Program (USWRP) Implementation Plan for Research in Quantitative Precipitation Forecasting and Data Assimilation to "redeem practical value of research findings and facilitate their transfer into operations." The NOMADS framework was also developed to facilitate climate model and observational data inter-comparison issues as discussed in documents such as the Intergovernmental Panel on Climate Change (IPCC 1990, 1995, 2001) and the U.S. National Assessment (2000). NOMADS is being developed as a unified climate and weather archive so that users can make decisions about their specific needs on time scales from days (weather), to months (El Nino), to decades (global warming).

Currently, NCEP NWP output are available in real-time through a number of different channels. Historical data, the data needed for research and collaboration, are more difficult to obtain and the variety of products are much more limited. There is no single repository or effective distribution method for both NWP and GCM data. Local piecemeal approaches at government laboratories, data centers, and universities fall short of filling the need for retrospective model data. Even in cases where the data are available, there exists no mechanism to redeem practical value of research findings and facilitate these findings back into operations.

The primary U.S. National responsibility for the archive and service of weather and climate data rests with the NCDC. However, as the temporal and spatial resolution of models increase, the volume and varied formats of data presented for archive at NCDC using current communications technologies and data management techniques are inadequate. The scientific modeling community is a vast intellectual resource. This community is extremely interested in obtaining both weather and climate products for historical cases, and for operational and research purposes.

The NOMADS is actively partnering with existing and development activities including the Comprehensive Large Array Stewardship System (CLASS); the National Oceanographic Partnership Program's (NOPP) National Virtual Ocean Data System (NVODS); the Department of Energy's Earth System Grid (ESG); and the Thematic Real-time Environmental Data Distributed Services (THREDDS) project being developed through the National Science Foundation and Unidata. To ensure that agency and institutional requirements are being met, the NOMADS collaborators have established sience and technical expert teams. These newly established teams, comprised of NOMADS members will ensure system and data inter-operability; and develop data archive requirement recommendations to NOAA.

2.1 Grid Computing

The "Grid" refers to the expanding network of computational and physical scientists and leading organizations (business, government, and academia) from around the world that have agreed to pursue large-scale distributed processing and access across the Internet and Next Generation Internet (NGI) or Internet-2. A feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualization back to the individual user scientists.

The Grid, and NOMADS is an architecture proposed to bring all these issues together and make a reality of such a vision. There are many developing grid technology projects. The Globus Project (Foster, et al., 2001), approach to the Grid defines the Grid as "*an enabler for Virtual Organizations: an infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources*."

Collaborations under the NOMADS umbrella have been reached with several Grid computing projects including:

- The Earth System Grid (ESG) project http://www.earthsystemgrid.org
- Natural Environment Research Council Datagrid: http://umbriel.dcs.gla.ac.uk/NeSC/action/projects/project_action.cfm?title=82
- The Committee for Earth Observation Satellites (CEOS) Data Grid (new)
- The Globus Project http://www.globus.org

For detailed information on Grid technologies see http://www.globus.org/research/papers/anatomy.pdf.

3.0 NOMADS Benefits

NOMADS fosters system inter-operability by integrating legacy systems with existing and emerging technologies and metadata conventions. NOMADS relies on local decisions about data holdings by combining legacy systems, while developing new ways to support data access. Distributed or "Grid" computing allows NOMADS to work on the cutting edge of data access methods. In this effort, no one institution carries the weight of data delivery since data are distributed across the network, and served by the institutions that developed the data. The responsibility for documentation falls on the data generator; with the NOMADS expert teams ensuring inter-operability, accuracy and precision of the data. Further, NOMADS in no way precludes the need for national centers to maintain and support long-term archives. In fact, NOMADS and secure data archives are mutually supportive and necessary for long-term access and research. A primary science benefit of the NOMADS framework is that it enables a feedback mechanism to tie Government and university research back to the modeling community.

Under the NOMADS framework improvements to NWP models will lead to longer lead times and improved accuracy for forecasts over all time scales. This results in fewer lives lost and reductions in property damage. The NOMADS will allow researchers to more easily access and understand data thereby increasing scientific productivity.

NOMADS provides a capability for scientists to assess the potential impact of long term climate change on the Nation using increasingly detailed and complex models since various models will be more readily available for inter-comparison. Climate model development and diagnostics will also be enhanced since NOMADS provides access to GFDL and other climate model results through collaborating institutions, and provides the tools to inter-compare these models and observations. NOMADS will allow for various model diagnostics and the coordination of multi-institutional model development such as forcings, feedbacks and uncertainty analysis as recommended by the IPCC Chapter 8 "Climate Change 2001: The Scientific Basis" IPCC Working Group I (2001). GFDL and NCAR models are already available under the NOMADS framework and described under sections 5.3 and 5.4. In addition, the Atmospheric Model Intercomparison Project (AMIP), and the Coupled Model Intercomparison Project (CMIP) model results are available via NOMADS at the Department of Energy (DOE) Lawrence Livermore National Laboratory (LLNL), Program for Climate Model Diagnosis and Intercomparison (PCMDI).

A significant benefit of the NOMADS framework is that it will reduce data center costs by integrating local efforts to archive and gain access to real-time and retrospective climate and weather models and data. The sheer number of institutions collaborating under the NOMADS framework helps to develop new collaborations and ensures for the long-term viability of the project itself. The NOMADS project leverages off many person-years of software development within NOAA and elsewhere. Heavy dependence is made upon technology transfer and re-use of previously developed and successful web portal technologies.

NOMADS allows an option for the integration of large data sets by performing data aggregation, manipulation, sub-setting, and compression on the servicing hosts rather then transmitting large amounts of data across the Internet at relatively slow speed. NOMADS available data sets can be expanded to include other data sets such as satellite and radar using XML to describe the data using NOMADS sub-setting capabilities for efficient distribution of these high volume data.

4.0 System Architecture

The concept of distributed computing is not new. During the 1970's, AT&T's Bell Laboratories and the UNIX operating system championed capabilities for distributed computing with multi-system inter-operability. Later, Sun Microsystems Inc., corporate slogan included the phrase "...the network is the computer". Indeed for simple text based retrieval, the concept finally escalated with the Internet and a protocol called HTTP. Distributed computing had taken center stage for the search and retrieval of text-based documents and images. Distributed scientific data access is the next step in this progression and allows browsing, obtaining, and manipulating scientific

data on geographically distributed data as easily as Internet Web browsing of today.

To provide inter-operability across multiple operating systems and known user formats, NOMADS currently provides several client-server data servers. Currently available NOMADS core servers include the OPeNDAP, the GDS, the CDAT, and the LAS. These servers provide access to collections of data. A DODS client requests data from the servers using URL's to describe the desired data and data location. DODS servers translate data into format neutral network objects, allowing participating modelers to serve data in their native file formats. Clients once re-linked with DODS can access these format neutral objects through standard Applications Program Interfaces (APIs). Client connectivity is built upon multiple API's, including standards such as netCDF and HDF, the Binary Universal Form for the Representation of meteorological data (BUFR), and the GRIdded Binary (GRIB) format, enabling NOMADS to integrate with standard desktop tools, custom models, GrADS, Matlab, IDL, idv, Ferret, and other desktop display and analysis applications.

The NOMADS framework can support the following data transport protocols:

- Open source Project for a Network Data Access Protocol (OpenDAP) formally known as the DODS protocol
- Common Object Request Broker Architecture (CORBA)
- Java™ Remote Method Invocation (Java/RMI)
- Internet Inter-Orb Protocol (IIOP)
- Hypertext transport protocol (HTTP)
- FTP
- TCP/IP

NOMADS users can use and exchange the following data formats and desktop applications:

- ASCII
- BUFR
- Consolidated PREODB (ECMWF version)
- ECMWF NetCDF in METVIEW
- ESML data forms
- Ferret/IDL/MatLAB/GrADS
- GrIB / GrIB (2)
- HDF
- Prepda (NCEP BUFR for data assimilation)
- Naval Research Laboratory METQC
- NetCDF
- Various spreadsheets, Web browsers
- XML

The NOMADS allows DODS enabled transactions but also has the capability for enhanced server side data manipulation using CDAT, GDS, or LAS. In GDS, GrADS expressions can be provided in the body of the URL request to manipulate data on the server providing only the resultant back to the user. History has shown that network connectivity continually lags behind desktop processing speeds. NOMADS will allow Internet users access to information they would not normally consider possible.

Since most weather model data are stored in the GRIB and BUFR data formats and most climate model data are stored in the NetCDF data format, NOMADS will provide for the comparison between these two formats. Converting between these two data formats has been a very difficult task in the past. The NOMADS software allows users to work in whatever data format is the most advantageous to them. The combination of quality control routines, and independent data format will, for the first time, provide users with a seamless interface to models and associated data. Climate modelers could access weather models, and weather modelers could access climate models. Never before has this capability existed outside a given institution.

NOMADS allows for growth of data sets and for changing data formats. If the current list of applicable formats no longer applies or is no longer in use, the NOMADS framework allows the XML to conform to the new data "standard." Although XML is supposed to be a generic and general language, in real life it requires the prior agreement on a vocabulary by all parties. Thus, full coordination is required for a successful inter-operability of systems for data sharing. This coordination already exists among the NOMADS collaborators as is evident by the sheer number of collaborating institutions contributing their expertise and collections under the NOMADS framework. The NOMADS distributed client-server architecture has the potential to bridge the gaps of inter-operability between systems and thus provide users with a metadata of geo-sciences information, visualization and analysis tools, and research opportunities- potentially reaching beyond the physical sciences.

4.1 Open source Project for a Network Data Access Protocol: *OPeNDAP*

The Open source Project for a Network Data Access Protocol (OPeNDAP- formally known as the Distributed Oceanographic Data System (DODS)) began as a joint effort between staff and scientists at the University of Rhode Island, Graduate School of Oceanography and at the Massachusetts Institute of Technology, Department of Earth Atmospheric and Planetary Science. DODS is a software framework that simplifies all aspects of scientific data networking, allowing simple access to remote data. Local data can be made accessible to remote locations regardless of local storage format by using DODS servers. Existing,

familiar data analysis and visualization applications can be transformed into DODS clients, i.e., applications able to access remote DODS served data. DODS provides a protocol for requesting and transporting data across the web. The current DODS Data Access Protocol (DAP) uses HTTP to frame the requests and responses. For details on the DODS DAP, see DODS Data Access Protocol (DAP 2.0) at www.unidata.ucar.edu/packages/dods/design/dap-rfc-html.

The DODS involves a community of users working together to use, improve, and extend the DODS protocol and software. The DODS design principles are based on two considerations:

- data are often most appropriately distributed by the individual or group that has developed them;
- the user will in general like to access data from the application software with which s/he is most familiar.

This has resulted in a highly distributed system that allows users to control the distribution of their own data and the way they access data from remote sites. The currently available DODS servers include data formats for:

- DODS
- FreeForm
- GDS (GrADS-DODS Server)
- HDF 4
- JGOFS
- Matlab
- NetCDF and
- DODS Aggregation Server (AS)

The DODS AS, is part of the Java/DODS Servlet library, allowing physical files to be logically aggregated and served as a single DODS dataset The AS also presents all available datasets in an integrated THREDDS Catalog, and is also a netCDF file server. The AS is currently in beta testing. For more information on the DODS AS see www.unidata.ucar.edu/projects/THREDDS/tech/DODSAggServer.html.

For more information on DODS see: www.unidata.ucar.edu/packages/dods.

4.2 Live Access Server

The Live Access Server (LAS) is a highly configurable Web server designed to provide flexible access to geo-referenced scientific data. It can present distributed data sets as a unified virtual database through the use of DODS networking. Ferret is the default visualization application used by LAS, though other applications (Matlab, IDL, GrADS) can also be used.

LAS enables the Web user to visualize data with on-the-fly graphics and users can request custom subsets of variables in a choice of file formats. LAS can access background reference material about the data (metadata) and compare (difference) variables from distributed locations. LAS enable the data provider to unify access to multiple types of data in a single interface and create thematic data servers from distributed data sources. LAS also offers derived product generation “on-the-fly”, and has the capability to remedy metadata inadequacies, e.g., poorly self-describing data. LAS also offer the ability to produce unique products, e.g. visualization styles specialized for the data, for scientific exploration. For more information on LAS see http://ferretwrc.noaa.gov/Ferret/LAS/ferret_LAS.html.

4.3 GrADS-Data Server

The Grid Analysis and Display System (GrADS) Data Server (GDS) (ftp://grads.iges.org/pub/gadev/gds7.doc) is a data server that provides sub-setting and analysis services across the Internet These services can be provided for any GrADS-readable dataset The sub-setting capability allows users to retrieve a specified temporal and/or spatial sub-domain from a large dataset, eliminating the need to download everything simply to access a small relevant portion of a dataset The analysis capability allows users to retrieve the results of an operation applied to one or more datasets on the server. Examples of analysis operations include basic math functions, averages, smoothing, differencing, correlation, and regression. The GDS supports operations that can be expressed in a single GrADS expression. Figure 1 provides an example of the processes initiated during a typical GDS user request under the GDS NOMADS framework.

There is currently an extensive GDS user base using GrADS as its primary data manipulation client. It is expected that many users of NWP and model input data will access NWP data via NOMADS using GrADS. For further information on GDS see www.iges.org/grads/gds.

GDS allows for variable comparison. For example, a GDS running at NCAR (http://motherlode.ucar.edu:9090) is distributing a set of ensemble members from the "Climate of the 20th Century" runs of the COLA atmospheric general circulation model.

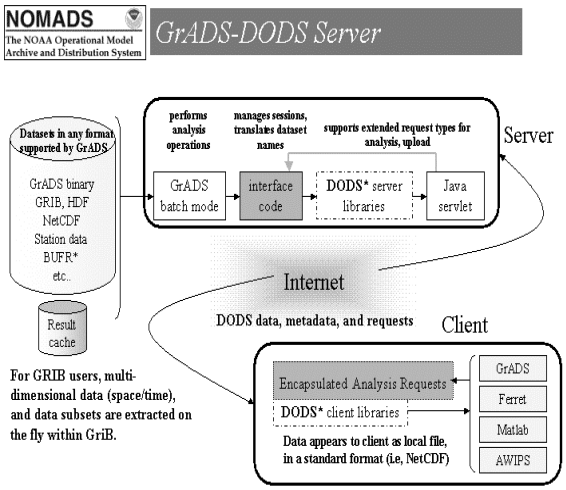

Figure 1. Example path of a GDS user request.

One can easily compare the relative humidity "rh" from the first two datasets, namely "C20C_A" and "C20C_B". If one wants to find a global time-average of the difference at the 1000 mb level in 1960 GrADS can be used as the client to open the following URL as follows

ga-> sdfopen http://motherlode.ucar.edu:9090/dods/_expr_{C20C_A,C20C_B} {ave((rh.1-rh.2),time=1jan1960,time=1dec1960)} {0:360,-90:90,1000:1000,1nov1976:1nov1976}
ga-> display result

The analysis results are returned in the variable "result" in the opened dataset Note that the world coordinate boundaries specified in the third set of curly braces fix the time to 1Nov1976. This can be set to any arbitrary time because the time dimension specification is overridden by the GrADS expression that tells the server to average over the period from January 1960 to December 1960.

In order to facilitate the use of these various servers, a library of GrADS scripts are being developed.

4.4 Climate Data Analysis Tool

Under the NOMADS framework, and a collaboration between PCMDI, GFDL, and PMEL, the LLNL has developed a new web access capability that merged the LAS with the CDAT (Williams, et al., 2002) suite of access and quality control programs. The URL to view CDAT-LAS is http://esg.llnl.gov/las. Currently, CDAT-LAS is serving up AMIP, CMIP, and NCAR's Parallel Climate Model (PCM) data sets. CMIP and PCM data sets are restricted and only accessed with the proper user name and password. The AMIP data sets are unrestricted and can be accessed by the general public. For more information of the PCMDI effort see www-pcmdi.llnl.gov.

The open nature of the CDAT system will permit any member of the climate community to contribute to the system on an equal footing with the members of PCMDI. With this philosophy, the general goal is to develop a consistent and flexible tool for everyone. NCDC will implement this package during 2002. The collaborations under NOMADS, such as THREDDS and PCMDI with other on-line packages (GDS, LAS, DODS), are to increase collaboration among climate and weather research scientists. NOMADS will also act as the technology that will allow NOAA to collaborate under the Earth System Grid (ESG) project.

CDAT is:

- portable open source software (free)
- incorporates modules
- exceptions and error-handling
- dynamic typing (for very fast prototyping)
- supports classes; very clear syntax
- extensible in C or C++ and other languages (i.e., FORTRAN)
- access local or remote database servers containing data files in various data file formats
- data extraction, grid transformation, and computation support
- quick and easy way to browse through terabytes of data
-

For more information on CDAT or ESG see http://esg.llnl.gov/cdat.

4.5 The Thematic Real-time Environmental Data Distributed Services Project

The Thematic Real-time Environmental Data Distributed Services (THREDDS) project is a system to make it possible for educators and researchers to publish, locate, analyze, and visualize a wide variety of environmental data in both their classrooms and laboratories. Just as the World Wide Web and digital-library technologies have simplified the process of publishing and accessing multimedia documents, THREDDS will provide needed infrastructure for publishing and accessing scientific data in a similarly convenient fashion.

THREDDS will establish both an organizational infrastructure and a software infrastructure. A team of data providers, software tool developers, and metadata experts will work together to develop a software framework that allows users to publish, find, analyze, and display data residing on remote servers.

The THREDDS software framework, based on a concept of publishable data inventories and catalogs, will tie together a set of technologies already

in use in existing, extensive collections of environmental data: client/server data-access protocols from the University of Rhode Island and the University of Wisconsin-Madison; Unidata's real-time Internet Data Distribution system; the discovery system at the Digital Library for Earth System Education (DLESE); and an extensive set of client visualization tools. For more information regarding the THREDDS see www.unidata.ucar.edu/projects/THREDDS.

The newly established National Science Digital Library (NSDL) will focus on "Womb to Tomb" education via the Internet, expand data distribution and sharing, aggregation, and cataloging- across many sciences (biologic, physical, math, etc.). NSDL has a direct link to THREDDS, and THREDDS to NOMADS since NOMADS will function as one of the THREDDS data sources.

4.6 Distributed Metadata Server

Under the NOMADS and THREDDS partnership, various database search engines are being explored to locate the many distributed collections of data across the Internet One such effort is being developed by the George Mason University and is called the Distributed Metadata Server (DIMES) (Yang et al., 2002). In contrast to most other standard metadata systems, DIMES employs a flexible metadata structure, linked data providers (nodes), and supports a wide variety of metadata forms with a minimum of semantic rules. DIMES also provides a software framework to search and browse the metadata. DIMES has been integrated with GDS to create a scientific data information "super-server" to support both data and metadata access consistently. One such example is running at GMU. For more information see: http://spring.scs.gmu.edu:8099/servlet/SiesipDataTree.

5.0 NOMADS Data Availability

The distributed NOMADS framework will provide data manipulation capabilities and real-time and retrospective access to NWP model input and output, ensembles, reanalysis, GCM's and observational data both upper-air and surface based, including oceanographic measurements being served by OPeNDAP enabled hosts. NCEP model output will be ingested at NCDC though NCDC's NOAAPort Data Access and Retrieval System (NDARS, Rutledge, et al., 2000) in real-time. These data have been archived at NCDC since October 1999. Real-time NWP data will also be available through NCEP's real-time NOMADS data services effort (Alpert, et al., 2002).

NCEP model output grids that will be available in 2003 include the Global Forecast Model (GFM, formally called AVN and MRF), the Rapid Update Cycle (RUC), Eta, mesoEta, NGM and the Weather Research Forecast (WRF) model when available. Also available under NOMADS will be NCEP's model input data and "run history" information including model spectral coefficients. GFDL, NCAR, and other climate simulations and diagnostics will also be available. For currently available NWP and other data see the preliminary PMEL NOMADS site at: http://ferretwrc.noaa.gov/nomads and the newly established NCDC site at http://nomads2.ncdc.noaa.gov. For GCM data currently available see the GFDL NOMADS site at http://nomads.gfdl.noaa.gov. Near real-time GDAS model input data will be available for testing and limited user access by the spring of 2003.

5.1 NCEP Model Input Data

The NCEP Global Data Assimilation System (GDAS) analysis files will be ingested through NCDC Load Balanced System and will be available under NOMADS from the NCDC archive. Data are currently documented in Federal Geographic Data Committee (FGDC) format as NOAA required. This documentation can be accessed at NCDC under Tape Deck (TD) No. 6172 at: www4.ncdc.noaa.gov/ol/documentlibrary/datasets.html.

The GDAS dataset consists of the minimum set necessary to re-generate NCEP analysis and forecast products (model re-start and initialization files). GDAS includes the Global Spectral Forecast Model (GSM), and the Spectral Statistical Interpolation (SSI) Cycling Analysis System (CAS) with triangular truncation (T) 170 and 28 sigma levels. To start the CAS, model spectral coefficients are provided on gaussian grid in a sigma vertical coordinate system. These data represent the model's "ground truth", and the best estimate- in terms of analyzed fields- for scientific study. Data that are restricted may not be available.

"Post" is a FORTRAN program is available from NCEP that will convert spectral coefficients to gaussian grid, sigma to pressure, and gaussian to latitude and longitude. Work continues to couple Post to the NOMADS user interface for source and executable downloads. The GDAS dataset under NOMADS will include the Global Spectral Forecast Model (GSM) and the Spectral Statistical Interpolation Cycling Analysis System (SSI-CAS) binary files and contains ~2.5Gb per day (4 cycles/day: 00Z, 06Z, 12Z, and 18Z). The binary files are raw data, which are acted on by NOMADS servers to produce useful grids. The archived analysis data sets serve as model verification as well as the source for model reruns and retrospectives. Including the observations allows for cycling analysis systems to re-analyze the observations. Never before has this model input data and information been available to the public. A partial list of NOMADS planned available observations (with associated data format) include:

Analysis Bias Corrected Information (ASCII)
Ship / Buoy Observations (BUFR)
Guess prep / guess output (BUFR)
Observational Toss List (ASCII)
Bogus Observations (BUFR)
ACARS and Aircraft (BUFR)
Wind Observations (BUFR)
Analysis Ready Obs. (prepBUFR)
Surface Analysis Restart Files (BUFR)
Surface and Upper-Air observations (BUFR)
Fixed Snow Field (GRIB)
Previous 6 hour forecast (BUFR)
"Post" Guess Output (spectral binary)
Profiler (BUFR) / SST's (GRIB)
MSU 14 and HIRS 14/15/16 (IEEE)
SSMI Satellite obs (BUFR)
NOAA (satellite) 15/16 AMSU - A/B
TOVS 1B Radiances (IEEE)
TOVS Satellite Obs (BUFR) / GOES Satellite Obs (BUFR)
O3 Sat Obs (binary) and ERS Sat obs
SBUV: Satellite Wind Observations
Radar VAD Winds (BUFR)

The formats of these data sets are generally dictated by the necessity to run models efficiently on modern computers. NOMADS converts the formats and structure to the users requested form but also allows the raw data to be directly accessed.

5.2 The Weather Research and Forecast Model

NOMADS is a technology for collaboration and model development as exemplified by the Weather Research and Forecast (WRF) model. The contribution of the scientific community to the WRF development effort requires the transfer of data sets of model retrospective results for testing and comparison. NOMADS enables both retrospective and real time access to the suite of digital products from reanalysis and operational results to give a range of users, from commercial to university scientist, model comparison access and supply operational grade initialization for forecast model study. For more information regarding the emerging WRF model see www.wrf-model.org/documentation_main.html

5.3 GFDL GCM Data Availability

General Circulation Models available under the NOMADS framework include the GFDL R-30 climate model. R-30 is a coupled Atmosphere-Ocean General Circulation Model (AOGCM). Its four major components are an atmospheric spectral GCM, and ocean GCM, and relatively simple models of sea ice and land surface processes. The name "R30" is derived from the resolution of the atmospheric spectral model (rhomboidal truncation at wave number 30). The R30 model is identified as GFDL_R30c in the Third Assessment Report (TAR) of the Intergovernmental Panel on Climate Change (IPCC). See Chapter 9, Table 9.1 of "Climate Change 2001: The Scientific Basis" IPCC Working Group I (2001).

The model output that is stored on the GFDL NOMADS server are taken from six experiments conducted at GFDL using the GFDL_R30c model. Data files produced by one long-running control integration (one with no changes in external forcings, e.g., constant CO_2) and five climate change scenario experiments have been made accessible to interested researchers. More information regarding the individual experiments is found in the Control & Transient Forcing Experiments section of the GFDL NOMADS Web site. The GFDL R-30 numerical model was developed and the experiments conducted by members of the Climate Dynamics and Prediction Group at the GFDL in Princeton, New Jersey. For further information on current available GCM data see the NOMADS site at GFDL at http://nomads.gfdl.noaa.gov.

5.4 NCAR Data Availability

NCAR also has a distributed computing effort collaborating under NOMADS. Working with UCAR and NCAR, NOMADS will partner with the forward-looking pilot project called the Community Data Portal (CDP). The CDP is targeted directly at elevating NCAR's collective ability to function as a data provider with a coherent web-based presence. Under the CDP it is expected that portions of Community Climate System Model (CCSM), and Parallel Climate Model (PCM), NCEP reanalysis and other data will be available under the NOMADS framework. A NOMADS GDS server has been established at NCAR and can be reached at http://motherlode.ucar.edu:9090/dods/.

5.5 Climate Diagnostic Center

NOAA's Climate Diagnostic Center (CDC) a NOMADS collaborator with an extensive array of distributed data sets for Web access. CDC is currently has one of the largest inventories of client-server listing in the country. CDC is well known for "one-stop shopping" for NCEP reanalysis, and many other NWP and GCM models and observations. For more information on the CDC data sets see www.cdc.noaa.gov/PublicData/data_descriptions.html .

Additional, CDC continues to advance NOMADS capabilities through user support services on the GrADS list server. CDC serves as a role model for distributed data access, by fulfilling their own unique data access needs while permitting external user access to these data sets under the NOMADS distributed processing philosophy.

5.6 NASA's Global Change Master Directory

NASA's Global Change Master Directory (GCMD) is a NOMADS collaborator and provides descriptions of Earth science data sets and services relevant to global change research. The GCMD database includes descriptions of data sets covering agriculture, the atmosphere, biosphere, hydrosphere and oceans, snow and ice, geology and geophysics, paleoclimatology, and human dimensions of global change. The DODS portal at the GCMD can be reached at http://gcmd.gsfc.nasa.gov/Data/portals/dods/index.html.

The GCMD through the PCMDI, GFDL, and PMEL collaboration has agreed to develop an XML-based database search engine for NOMADS implementation.

6.0 The Collaborators

The current participating collaborators under the NOMADS framework now include:

- NOAA National Climatic Data Center (Project Lead)
- National Weather Service, National Centers for Environmental Prediction
- NOAA Geophysical Fluid Dynamics Laboratory
- National Center for Atmospheric Research
- NOAA Climate Diagnostics Center
- NOAA Forecast Systems Laboratory
- NOAA Pacific Marine Environmental Laboratory
- Center for Ocean-Land-Atmosphere Studies
- University Consortium for Atmospheric Research Unidata Program
- National Severe Storms Laboratory (collaborating with the University of Wisconsin, Space Science Engineer Center)
- NASA Global Change Master Directory
- NASA Seasonal-to-Interannual Earth Sciences Information Partner
- LLNL Program for Climate Model Diagnosis and Intercomparison
- DOE Accelerated Climate Prediction Initiative
- DOE Earth System Grid
- George Mason University
- University of Alabama, Huntsville
- University of Washington
- Committee for Earth Observing Satellites

The University of Alabama is leading a collaborative effort that shares distributed data using similar and NOMADS compatible XML based metadata scheme called the Earth Science Markup Language (ESML), (Ramachandran, et al., 2001). For more information on ESML see http://esml.itsc.uah.edu/index2.html. It is expected that satellite and Radar and Geographical Information System (GIS) capable applications will be incorporated into the NOMADS framework using this schema in the near future.

A related and much larger effort is the DOE Earth System Grid, and the Globus project. Working with Argonne National Laboratory, Lawrence Livermore National Laboratory, the University of Southern California, Oak Ridge National Laboratory, and Lawrence Berkeley National Laboratory, NOMADS will deploy an operational component in support of terascale/petascale climate research. These projects are a significant opportunity to advance research and collaboration in the scientific community, and it has already drawn substantial interest that extends into the international community. The ESG project can be found at http://www.earthsystemgrid.org and the grid computing GLOBUS project is at http://www.globus.org.

As so many established U.S. based institutions are collaborating under the NOMADS framework, both U.S National and International interest in the collaboration have rapidly expanded. International NOMADS collaborators now include:

- The British Atmospheric Data Center (BADC)
- E-Science: ClimatePrediction.net
- E-Science: NERC DataGrid
- Hadley Centre

Finally, NOMADS has been selected as one of five programs that are being implemented as coordinated through the Departments of Commerce, State, Energy and the Environmental Protection Agency (EPA) for a new U.S. – Australia “Climate Action Partnership” (CAP). For further information on this partnership see: http://www.state.gov/r/pa/prs/ps/2002/8545.htm.

7.0 The Users

Currently, GFDL has one of the larger NOMADS sites for climate data. On the GFDL NOMADS server there is model data available from various climate studies, from the seasonal-interannual to decadal and longer time scales. There are more than 700 GB of model data available in more than 4,000 files. From October, 2000 to August 2002, there was more than 480 GB of data downloaded by users. There are about 75 registered users, 80% are from the United States and 20% are international users. Of the 75 users, about half are from private groups and corporations, with about a 25% from universities and about 15% from government users. Most of these users are scientists interested in obtaining GFDL model data.

The users of the GFDL data range from people interested in performing studies of how the climate of Arctic region changes due to changes in the greenhouse gases to the IPCC Data Distribution Center. A typical user finds the model data via "word of mouth". The user then searches the GFDL site, for the data needed. At present, this process is rather difficult and is representative of an area of active software development for NOMADS. The user then downloads the model data via ftp to a local computer where the user performs some analysis using the model data. The model data is stored in a netCDF data format, so the analysis process is normally straightforward if the user has experience using netCDF data.

In the future as the software becomes more user friendly, we expect students and policymakers to be able to use the NOMADS servers. At present, it is likely that most would find the current generation software too difficult to use. This is an area of active software development and represents some of the "cutting edge" web page development envisioned under the NOMADS umbrella.

8.0 Future Requirements

Even with the emergence of new web based services, the comparison of GCM results with the observational climate record is still difficult for several reasons. One limitation is the global distributions of a number of basic climate quantities, such as precipitation or clouds, are not well known. Similarly, observational limitations exist with model re-analysis data. Both the NCEP/NCAR (Kistler, Collins, Kalnay, et al., 2001) and the ECMWF (Gibson, et al., 1997) re-analysis eliminate the problems of changing model analysis systems but observational data also contain time-dependant biases by changing observational networks, station moves, and the assimilation of various remotely sensed data (Rutledge, et al., 1991) using differing sensor instruments, or calibrations. These changes in input data are blended with the natural variability making estimates of true variability uncertain. The need for data homogeneity is critical to study questions related to the ability to evaluate simulation of past climate. One approach to correct for time-dependant biases and data sparse regions is the development and use of high quality "reference" data sets (Karl, et al., 2000).

Beyond the ingest and access capability being implemented with NOMADS are the challenges of algorithm development for the inter-comparison of large-array data (e.g., satellite and radar) with surface, upper-air, and sub-surface ocean observational data. The implementation of NOMADS will foster the development of new quality control processes by taking advantage of distributed data access. One major challenge facing the scientific community is the development of methodologies for the inter-comparison of large-array observational data sets with model simulations. In the near future, NOMADS would include the development of algorithms to blend remotely sensed data with in-situ surface and upper-air data, then to use these blended fields for verification and validation of both weather prediction and climate models in both time and space (Rutledge, et al., 2001).

Finally, the interface between the NOMADS spinning disks, and hierarchical tape archive systems such as NCDC's mass storage systems, will be developed during 2003 so users may obtain access to data sets destined for long-term storage.

8.1 U.S. National Model "HelpDesk"

Under President Bush's new Climate Change Research Initiative (CCRI), access to and understanding of climate models and data is a high priority. Several Agencies including the U.S. Departments of Agriculture, Commerce, Energy, and the EPA and NOAA, are currently collaborating to monitor emissions, foster international partnerships for research, uncertainty reduction in climate models, and develop technologies for access to climate models for large-scale systematic approaches to model evaluation. A large and growing number of impacts groups and carbon cycle researchers will require access to, and understanding of these models and data. NCDC is charged with the archive of these data, however users at all skill levels will require on-line assistance with easy to navigate Web interfaces to model data. There is a danger that if users are not properly educated on the science or use of a specific model, it may be interpreted incorrectly, or users may use a model in a way that is inappropriate. Users must not only have access to data but the science and information behind the models, and how to properly use these complex data sets.

In support of these activities and the NOMADS project itself, NCDC has proposed the formation of a "U.S. National Climate and Weather Model *HelpDesk*" to assist these model users. The *Helpdesk* will:

- Provide a tailored Web interface to NCDC's archives of climate and weather models.
- A capability to service users with the information they need, not just model output and data.
- Develop distributed database search engines.
- Provide for the long-term stewardship of these high volume and complex data and metadata at NCDC and elsewhere in a distributed framework.

9.0 Acknowledgments

The authors gratefully acknowledge the combined efforts of all the NOMADS Team members for the growing success of NOMADS. Thanks also to Don Hooper, Jennifer Adams, Jerry Potter, and Don Middleton, and Howard Diamond. Pilot funding has been provided by ESDIM 02-444E.

10.0 Bibliography

Alpert, J.C., 2002, Rutledge, G.K., Williams, D., Stouffer, R., Buja, L., Doty, B, Hankin, S., Domenico, B., Kafatos, M., 2002, "The Plan to Access Real-Time NWP Operational Model Data Sets using NOMADS", Proceedings of the 18th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, American Meteorological Society, Orlando FL. 73 pp.

Bray, T., Paoli, J., Sperberg-McQueen, C.M., 1998, "Extensible Markup Language (XML) 1.0 Specification", W3C REC-xml-19980210: www.w3.org/TR/1998/REC-xml-19980210.

Davis, E. R., J. Gallagher, J., 1999, "Using DODS to Access and Deliver Remote Data", Proceedings of the 15th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, American Meteorological Society, Dallas, TX. 571 pp.

Doty, B.E., Wielgosz, J., Gallagher, J., Holloway, D., 2001, "GrADS and DODS", Proceedings of the 17th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, American Meteorological Society Albuquerque, NM. 385 pp.

Foster, I., Kesselman, C. and Tuecke, S. 2001, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", International Journal of High Performance Computing Applications, 15 (3). 200-222.

Gibson, J. K., P. Kallberg, S. Uppala, A. Noumura, A. Hernandez, and E. Serrano, 1997, "ERA Description. ECMWF Re-Analysis Project Report", Series 1, ECMWF, Reading, UK, 77 pp.

Hankin, S., J. Callahan, and J. Sirott, 2001, "The Live Access Server and DODS: Web visualization and data fusion for distributed holdings", 17th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, American Meteorological Society, Albuquerque, NM. 380 pp.

Karl, T., D. Easterling, P. Groisman, et al., March 2000, "Observed Variability and Trends in Extreme Climate Events: A Brief Review", Bulletin of the American Meteorological Society, 81, 417 pp.

Kistler, R., W. Collins, E. Kalnay, R. Reynolds, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, J. Janowiak, M. Kanamitsu, K. Mo, C. Ropelewski, R. Jenne, D. Joseph and M. Fiorino, 2001: The NCEP/NCAR 50-year Reanalysis: Monthly-means CD-ROM and Documentation. Bulletin of the American Meteorological Society, 92, 247 pp.

Ramachandran R., M. Alshayeb, B. Beaumont, H. Conover, S. Graves, X. Li, S. Movva, A. McDowell and M. Smith, 2001: "Earth Science Markup Language: A Solution for Generic Access to Heterogeneous Data Sets," Earth Science Technology Conference, Maryland

Rutledge, G.K., E. Legg, and P. Menzel, 1991, "Operational Production of Winds from Cloud Motions", Palaeogeography, Palaeoclimaotology, Palaeoecology, Vol. 90 No. 1-3 (Global and Planetary Change Section), 141 pp.

Rutledge, G.K., A. Stanley, E. Page, L. Spayd, and J. Brundage, 2000, "Implementation of the NOAAPort Data Archive and Retrieval System (NDARS) at the National Climatic Data Center", Proceedings 16th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Long Beach, American Meteorological Society, 492 pp.

Rutledge, G.K., T. Karl, D. Easterling, L. Buja, R. Stouffer, 2001 "Evaluating Transient Global and Regional Model Simulations: Bridging the Model/Observations Information Gap", (Invited), American Geophysical Union, Spring Meeting Boston MA S35 pp.

Rutledge, G.K., D. Williams, R. Stouffer, J. Alpert, L. Buja, B. Doty, S. Hankin, B. Domenico, M. Kafatos, 2002, "The NOAA Operational Model Archive and Distribution System (NOMADS)", Proceedings 13th Symposium on Global Change and Climate Variations, American Meteorological Society, Orlando FL. J76 pp.

Williams, D.N., R.S. Drach, P.F. Dubois, C. Doutriaux, C.J. O'Connor, K.M. AchutaRao, and M. Fiorino, 2002: Climate Data Analysis Tool: An open software system approach. 13th Symposium on Global Change and Climate Variations, American Meteorological Society extended abstract volume, Orlando, Florida, J71pp.

Yang, R., X. Deng, M. Kafatos, C. Wang and X. Wang, 2001, "An XML-Based Distributed Metadata Server (DIMES) Supporting Earth Science Metadata", Proceedings of the 13th International Conference on Scientific and Statistical Database Management, Institute of Electrical and Electronics Engineers, Computer Society, Fairfax VA. 251 pp.

11.0 Acronyms

ACARS	Aircraft Communications and Reporting System
AOGCM	Atmospheric-Ocean General Circulation Model
API	Application Program Interface
AMIP	Atmospheric Model Intercomparison Project
AS	DODS Aggregation Server
ASCII	American Standard Code for Information Interchange
AVN	NCEP Aviation Model
AWIPS	NWS Advanced Weather Interactive Processing System
BADC	British Atmospheric Data Center
BUFR	Binary Universal Form for the Representation of meteorological data
CAS	NCEP Cycling Analysis System
CCRI	Climate Change Research Initiative
CCSM	NCAR Community Climate System Model
CDAT	PCMDI Climate Data Analysis Tools
CDC	NOAA Climate Diagnostics Center
CDP	NCAR Community Data Portal
CGI	Common Gateway Interface
CLASS	NESDIS Comprehensive Large Array Stewardship System
CMIP	Coupled Model Intercomparison Project
COARDS	Cooperative Ocean-Atmosphere Research Data Standard
COLA	Center for Ocean-Land-Atmosphere Studies
CORBA	Common Object Request Broker Architecture
DAP	Data Access Protocol
DIMES	Distributed Metadata Server
DOA	U.S. Department of Agriculture
DOC	U.S. Department of Commerce
DOE	U.S. Department of Energy
DODS	Distributed Oceanographic Data System
ECMWF	European Center for Medium-Range Weather Forecasting
EOS	Earth Observing System
EPA	Environmental Protection Agency
ETA	NCEP Eta Model
ERS	European Remote Sensing Satellite
ESG	DOE Earth System Grid
ESML	Earth Science Markup Language
FGDC	Federal Geographic Data Committee
FTP	File Transport Protocol
GCM	General Circulation Model
GCMD	Global Change Master Directory
GDAS	Global Data Assimilation System
GDS	GrADS Data Server
GFDL	Geophysical Fluid Dynamics Laboratory
GIS	Geographical Information System
GOES	Geostationary Operational Environmental Satellite
GrADS	Grid Analysis and Display System
GRIB	GRIdded Binary
GSFC	Goddard Space Flight Center
GSM	Global Spectral Forecast Model
GUI	Graphical User Interface
HDF	Hierarchical Data Format
HDF-EOS	Hierarchical Data Format - EOS
HIRS	High-Resolution Infrared Radiation Sounder
HTML	Hyper Text Markup Language
HTTP	Hypertext transport protocol
IDL	Interactive Display Language
IEEE	Institute of Electrical and Electronics Engineers
IIOP	Internet Inter-Orb Protocol
IPCC	Intergovernmental Panel on Climate Change
JGOFS	Joint Global Ocean Flux Experiment
LAS	Live Access Server
LLNL	Lawrence Livermore National Laboratory
MIT	Massachusetts Institute of Technology
MSU	Microwave Sounding Unit
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCDC	National Climatic Data Center
NCEP	National Centers for Environmental Prediction
NDARS	NOAAPort Data Access and Retrieval System
NERC	Natural Environment Research Council
NetCDF	NETwork Common Data Format Data Access Protocol
NGI	Next Generation Internet
NGM	NCEP Nested Grid Model
NOAA	National Oceanic and Atmospheric Administration
NOPP	National Oceanographic Partnership Program
NSDL	National Science Digital Library
NSF	National Science Foundation
NVODS	NOAA Virtual Ocean Data System
NWP	Numerical Weather Prediction
NWS	National Weather Service
OPeNDAP	Open source Project for a Network Data Access Protocol
PCM	NCAR Parallel Climate Model
PCMDI	Program for Climate Model Diagnosis and Intercomparison
PMEL	Pacific Marine Environmental Laboratory
RUC	NCEP Rapid Update Cycle Model
SSI	Spectral Statistical Interpolation

SSMI	Special Sensor Microwave Instrument
SST	Sea Surface Temperature
TCP/IP	Transmission Control Protocol/Internet Protocol
THREDDS	Thematic Real-time Environmental Data Distributed Services
TIROS	Television Infrared Observation Satellite
TOVS	TIROS Operational Vertical Sounder
UCAR	University Corporation for Atmospheric Research
URI	University of Rhode Island
URL	Uniform Resource Locator
USWRP	U.S. Weather Research Program
WRF	Weather Research Model
WWW	World Wide Web
XML	Extensible Markup Language