

**Advancement of the NOMADS for Observational Data and Model
Intercomparisons
and the Establishment of a NCDC NOMADS Team and HelpDesk**

A NOMADS Collaboration White Paper
5/5/2003

This white paper is intended to develop a concept of operations for the: 1) advancement of NOMADS for model intercomparison and product development ; 2) a NCDC model “HelpDesk” and 3) an NCDC inter-divisional NOMADS Team. A summary of actions follows this report.

1. Promotion of Model Intercomparison and Climate Products

In order to promote the use of NOMADS as an access instrument for various modeling and product development groups, this white has been developed. The NOMADS distributed philosophy relies on:

- 1) agreed-to, self-describing formats with “adequate” description of the data (the metadata);
- 2) a discipline-neutral data access protocol (OPeNDAP); computer tools (desktop applications and Web servers) that can access, visualize, analyze, and intercompare distributed data sets (LAS, GrADS, Ferret, IDL, Matlab and others);
- 3) quantitative model intercomparison and diagnostic tools (PCMDI); and
- 4) data discovery systems to locate the data (GCMD, THREDDS).

NOMADS is in no sense whatsoever a “stovepipe” solution (a usually proprietary solution limiting the options that users have over access or display of data). NOMADS is a collection of technologies drawn from a broad range of collaborators across several disciplines with much of it already widely deployed. In fact, NOMADS is just the opposite of a “stovepipe”. It is a collaborative, open-source, standards-based network of data servers that integrates into, with, and between existing data systems. Advancement and refinement of necessary metadata, using existing standards (FGDC, DIF, OPeNDAP, etc.) needs to be performed. Much of the necessary capability is inherent to NOMADS. It is using NOMADS, rather than just supporting NOMADS.

No Single Solution -- an “Open Solution”

No less viable options for model to observational intercomparisons are desktop clients such as LAS/Ferret, MATLAB, and IDL and others. For scientists who wish to do custom detailed analysis in these desktop tools the solutions are in place. We must however work to ensure that the model outputs (and inputs for data assimilation) are

served through OPeNDAP. The OPeNDAP consortium places access to in-situ data on the front burner for development. Therefore, any plans that NOMADS makes in this regard must foster this effort.

Currently under NOMADS the COARDS and the CF (extended COARDS) standards, shortfalls in describing some of GFDL's model output exist. GFDL have complex curvilinear coordinate system models not describable under these conventions. However new metadata schema's or extensions to existing standards are being developed and implemented. For example in the early 1990s, the geospatial community began work towards the development of an international standard for geospatial metadata. The International Organization for Standardization (ISO) Technical Committee's [(TC)211] Metadata Standard 19115 (previously known as 15046-15) (see: <http://www.isotc211.org/>) sought to provide "a consistent suite of geographic information schemata that allows geographic information to be integrated with information technology. The goal of this work item is to produce a schema for geographic information metadata." The adoption of the ISO 19115/TC211 geospatial metadata standard took place on April 4th, 2003.

Additional developments to address enhanced metadata requirements include the NetCDF Markup Language (NcML). The Earth System Grid and THREDDS projects have developed NcML as an abstract object model for expressing metadata associated with generic netCDF data. This effort directly supports, or can support, some of the metadata shortfalls regarding the adequate description of GFDL models. Another is a recent NASA AIST award to Unidata (working with NCSA) on a fusion of the next generation of netCDF and HDF. It would essentially be netCDF interfaces on top of the HDF file handling system. Part of this effort is to make many improvements to netCDF that the modeling community has requested for some time.

These new international metadata efforts such as ISO19115, NcML, and fused netCDF/HDF, and of course, FGDC, need to be further explored for adaptability and use into OPeNDAP and NOMADS. To achieve adequacy in metadata for GFDL model for example, GFDL, has already initiated collaborations with several of the aforementioned activities and requires additional work.

Collaborative Directions and Intercomparisons

To develop access capabilities for model-to-model, and model-to-observational intercomparisons, it is necessary to coordinate specific objectives between GFDL, NCDC, and the other NOMADS Team participants (PCMDI, NCAR, and PMEL). Under NOMADS NCDC will make selected observational, and reference datasets available under the NOMADS framework; while GFDL must ensure for the adequacy of the metadata currently in use or expected to be in use under the various metadata efforts listed above (THREDDS, FGDC, ESP, etc.).

Today there are issues surrounding climate model and observational intercomparisons. For example, the accuracy and completeness of any model validation ultimately rests on

the verifying observations and the model-observation data interface, i.e., the technical and scientific comparison of model and observations. In the case of atmospheric general circulation models (AGCMs) applied to climate, the actual observations (irregularly spaced, conventional and satellite-based) are analyzed to a AGCM-compatible grid. These analyses are further processed (e.g., monthly means) to become an “observation” standard against which the quality of the model is assessed. Thus, the conventional practice of AGCM climate model verification is indirect and dependent on the quality of the analyzed observation standard. In addition, attribution of differences between model output statistics and the analyses (“error”) to model deficiencies (e.g., physical parameterization) is difficult if not impossible without a direct comparison to observations of the physical processes as they operate in the models. In both cases (model output and model process validation), the model-observation data interface is the key and the viewpoint through which model performance is judged. The current data interface for climate model validation is simply a difference between model output statistics and statistics of analyzed observations and there is an overarching objective to provide more direct interfaces including data assimilation. The first requirement is technical handling of non-gridded or station data.

Achieving the Goal

In order to promote NOMADS access for Model Intercomparison Project (MIP)-type use, (and not dictate intercomparison methods or processes), we must: 1) position NOMADS as both an interface to traditional NCDC on-line services; 2) develop enhanced metadata capabilities to include climate models; 3) include selected observational datasets into the current NOMADS data holdings, and 4) initiate product development using models. For example comparing modeled and observed changes in max/min temperatures over 20th century.

The focus will be making NCDC and GFDL model *and* observational datasets available under emerging, and community-accepted, distributed-data access philosophy; and under traditional NCDC access methods. This way, commercial users would not be able to bypass NCDC data sales requirements, while permitting research access in a non-restricted, format neutral manner.

The desired model comparison and observational analysis work capabilities requested at the NCDC-GFDL telecom of 4/15/03 are an inherent goal within NOMADS. NOMADS partners already provide interoperable components that permit (or are under immediate development to permit):

1. gridded data (models) to be shared and efficiently accessed over the Internet (OPeNDAP);
2. in-situ data to be shared and analyzed over the Internet (COLA-GDS);
3. desktop analysis and comparison of distributed models and data (OPeNDAP-enabled GrADS, Ferret, IDL, IDV, Matlab);
4. Web browser-based visualization, analysis, and comparison of models and data (PMEL-LAS, NCDC NOMADS, PCMDI-CDAT, etc.).

5. Data discovery and access of both historical and real-time data sets (NCDC-NOMADS, GCMD-NASA, THREDDS-Unidata, COLA, etc.)
6. Experience at hosting model intercomparison projects and computational tools to support it (PCMDI-DOE, GFDL, NCAR); and
7. High speed, secure data transmissions and access to restricted data sets (“the Grid”)

Current climate models, which operate on complex curvilinear coordinate grids, have introduced new “semantic” concepts that need to be addressed in the NOMADS data sharing and intercomparison framework. The standardized metadata that would describe many of these concepts either does not exist, or exists in the “CF” standard (still a draft, formally), but has not yet entered into wide circulation. These issues require further exploration by GFDL, PCMDI, NCAR, Unidata and associated data and modeling groups. Selected modeling groups need to collaborate to refine and expand the metadata schema’s to ensure for interoperability between various modeling efforts and desired functionality (e.g., model output as model input to another model, or plug and play parameterizations across modeling groups). It is recommended that small steps be taken to develop goals and specific functionality, with selected models and data and collaborating modeling Centers. The continued support of the NOMADS effort at NCDC and GFDL will establish the requested functionality

Data Cataloging

During a recent meeting between NCDC and GFDL, the issues surrounding data cataloging, and search engines was raised. While this is a very real concern; the goal of a MIP-type functionality should not be confused with the goal of data search and retrieval services for all possible data users and needs. We do not need, nor should we address these larger issues of data access and cataloging *in toto*. Rather, as the NOMADS Program Plan outlines, we should look to existing large-scale and on-going efforts for solutions to search and cataloging. Since this is such a large problem, it is recommended that we leave this effort to those with a larger stake (and funding). Specifically, these efforts include:

- 1) NASA GCMD and the Earth Science Portal (thru NOMADS);
- 2) NSF THREDDS and NSDL;
- 3) GMU/NASA DIMES; and
- 4) FGDC Clearinghouse efforts.

THREDDS and NSDL are working on a variety of search and discovery mechanisms (from building tools that allow data providers to feed digital libraries and other discovery centers to more fine grain search mechanisms). These mechanisms are currently in, or soon to be moving into, the prototype stage. The framework for dataset inventory and relating various types of metadata standards (DIF, FGDC, Dublin Core, NcML, etc) to those datasets is in place and being used in a number of systems. THREDDS is also working with OPeNDAP to integrate THREDDS catalogs into OPeNDAP servers.

2. The NCDC-GFDL Model “HelpDesk”

With the establishment of a Model HelpDesk, users at all skill level can obtain on-line or phone assistance with models. Users and policy makers must not only have access to data, but the science and information behind the models- and how to properly interpret and use these complex data sets. There is a clear danger, that if users are not properly educated on the science and/or use of a specific model, it may be interpreted incorrectly, or users may use a model in a way that is inappropriate.

A significant payoff under NOMADS is that on-line and staff resources will provide this information. An additional benefit is that Helpdesk staff will help compile user requirements for associated modeling groups (GFDL) to conduct basic climate science research, and for NCDC staff to advance next generation climate products, using model and observational data. The implementation of a Model HelpDesk function will require at a minimum, NCDC and GFDL coordination and training of NCDC personnel on weather and climate model development, their limitations, and use.

The implementation of a HelpDesk function will be developed based upon existing NOMADS capabilities, and therefore after other areas of development outlined in this paper are complete. The NOMADS web page must be enhanced to answer climate and weather model use, and basic access questions. Additional requirements developed by NCDC such as requirements for product development are required.

NCDC and GFDL need to address the following questions for the HelpDesk function:

A. *Who uses GFDL models?*

- IPCC
- US National Assessments
- MIPSs- Model Intercomparison Projects (CMIP, AMIP, ENSMIP, etc.)
- Individual Researchers (Gov’t, University, Private)
 - Modelers, Impacts groups, etc.)

B. *What would they ask for?*

- Data from a particular experiment / integration
- Related to paper or MIP
- Communication of what a data set is a problem
- Subsets of variables and of domain
- Global average, hemispheric average, near a given location (over US, UK, etc.)
- New model-based products such as normals, climate divisions, gridded airport location data, reanalysis in-situ analysis (e.g., CARDS – Reanalysis).

C. *What is needed for a Climate and Weather Model “HelpDesk” Function at NCDC?*

- NCDC Climate model education needed for
 - User access assistance using NOMADS/NCDC access technologies
 - Increase CSD scientific service rep. knowledge base to climate, Wx models (training at GFDL?).

- Development of scientific expertise to integrate model and observational data, and studies for validation and verification of model simulations, short-term and seasonal forecasts and climate change detections processes. Develop collaborations with GFDL, LLNL, NCAR, CDC.
- Tailored NOMADS Web interfaces to NCDC's archives of climate and weather models,
- Development an on-line (phone and web) customer service model archive HelpDesk• function at NCDC and “referral” capability at GFDL for model specific questions.
- Develop web search engines and cataloging techniques to access models at NCDC and across the NOMADS distributed network of data servers.
- Document and maintain all data and [agree to] metadata for all required models and data.
- Communication and documentation of requested products (next generation normals etc.).

D. Center Roles and Responsibilities

The specific tasks for the advancement of NOMADS will be developed through a series of informal meetings. It is envisioned three teams are required:

- 1) The NCDC NOMADS Team (with NOMADS wide coordination)
- 2) The advancement of the CF and COARDS metadata team
- 3) The development of the NCDC/GFDL HelpDesk:
 - GFDL will serve directly the MIP requirements e.g., AMIP, CMIP, PMIP and serve IPCC needs for scenario generation and access to model data;
 - NCDC and GFDL will provide model and data for rest of user community and develop requirements and investigate product development based upon user needs.

While this list is by no means comprehensive, it should provide the initial steps necessary to implement a pilot HelpDesk, and metadata expansion process.

3. The Establishment of a Inter-Divisional NOMADS Team at NCDC

The NCDC NOMADS team will be established to provide inter-Divisional, matrix-style implementation of a NOMADS capability within NCDC. Once internal NCDC capabilities are operational, external collaboration and projects will be available.

The purpose of the team will: 1) develop NCDC-wide knowledge of the NOMADS project; 2) to eliminate duplicative efforts (where they are really duplicative); and 3) initiate and advance the use of NOMADS at NCDC and with GFDL and other modeling groups.

A. The NCDC NOMADS Team and Roles

The NCDC NOMADS team and roles will be:

- 1) Scientific Services Division
 - Glenn Rutledge- NCDC Team Leader, NOMADS PI, 2nd level HelpDesk
 - Russ Vose- Project and Science Oversight
 - Dave Easterling- Science and Project Oversight
 - Danny Brinegar- Systems and Data Implementation; core NOMADS
 - Ted Smith- Systems and Data Implementation; core NOMADS Team
- 2) Data Operations Branch
 - Wayne Fass- Ingest/Archive Oversight
 - Gus Shumbura- Archive Oversight
 - Alan Hall- Ingest processes
 - Dave Bowman- Archive data structures and access
 - Charles Philips- FGDC metadata development and documentation
 - Jeff Arnfield- Station History metadata coordination
- 3) Climate Services Division
 - Neal Lott- Access Oversight: CDO, HAS Interface
 - Doug Ross- HAS Interface (with NOMADS)
 - Vernell Woldu- HelpDesk oversight and development
 - Scott Stephens- HelpDesk development and representative
 - Marc Plantico- Product development, Climate Normals, observational and model data requirements necessary for product development.
- 4) Support Services Division
 - John Fauerbach- ftp services, networking, security
 - Duane Dunston- security

External coordination will be lead by Glenn Rutledge, Marc Plantico, Mike Fiorino and Ron Stouffer.

B. NOMADS Team Areas of Development

These representatives, as appropriate, will conduct regular meetings based upon 6 proposed areas of development (areas of development can be modified based upon team discussions):

- Area 1. User friendly web access system to highly desired data stored on disk.
- Area 2. Addressing the HDSS archive model file structure (and ingest mechanisms).
- Area 3. Archive access to data stored on the HDSS.
- Area 4. Data Visualization options
- Area 5. Observational MIP-type pilots (this white paper).
- Area 6. Access to off-site model/obs data (NOMADS and data cataloging efforts).

C. NOMADS Access and Intercomparison Capabilities: A Phased-in Approach

Specific and proposed tasks to be refined by the Team include:

Phase 1: Provide a user-friendly web access system for model data currently stored on disk at NCDC, which includes GFS (AVN/MRF), ETA and RUC data from Jan 2003. Other data sets to follow during FY03-4. This includes

- NCEP Regional Reanalysis
- NCEP model output: NGM, Ensembles, and two ocean models.
- Surface and U/A observations suggested for access include (post)-CARDS, GHCN and ISH.
- Access to GFDL and other climate modeling groups already possible under NOMADS.

The access system will place data files on the ftp server for user retrieval via links provided through the web access system, in demand mode (i.e., no email yet). Completion: end of April 2003. System will be provided to CSB to review; end of May - have system online to the public via NCDC's web site.

Phase 2: Develop HDSS archive file structures for NWP input and output based upon user need and access processes. Determine effort required to reprocess current NOAAPort archive (2000-present) based on new file structures and determine if current model archive is 1) removed, 2) left as is, or 3) reprocessed into new file structure. Begin reprocessing if determination made to do so, and ensure that HDSS archive from this point forward meets requirements. Completion: Jun 03.

Phase 3: Add HDSS archive access to the web access system, using HAS to access the files. This will include batch mode operations to retrieve files from HDSS tapes, and email the user with links to the files on the ftp server. Doug Ross will assist in this phase. Completion: Aug. 2003.

Phase 4: Add components of data visualization to the web access system, so that the user has a very easy method to retrieve the data, with the option to download the file and

a 2nd link provided to view an image file. Both files would be placed on the ftp server and linked on the final web page provided to the user. Completion: early FY04.

Phase 5: Add web portal capabilities for off-site model data to the web access system, with initial prototype for one off-site location. Ensure for distributed access to other modeling centers via NOMADS technologies. Completion: early FY04.

Phase 6: Begun during Phase one, Phase 6 includes the development and implementation of MIP type data access for climate and weather model diagnostics and verification. This collaborative effort will require coordination between NCDC, GFDL and other selected modeling and analysis centers. Phase 6 requires additional external collaboration and the development and adoption of metadata conventions. Phase 6 will be addressed under separate cover after coordination between the various modeling groups has been completed. Initial coordination: Jun 03. Metadata (COARDS=>cf Team).

Who wrote this document?

This document was written by Glenn Rutledge with input and review by Russ Vose, Neal Lott, Mike Fiorino (PCMDI), Steve Hankin (PMEL), Ethan Davis and Ben Domenico (Unidata) and Ron Stouffer (GFDL).

**Summary of Actions for the Development of a NOMADS Model Intercomparisons
Capability a NCDC-GFDL “HelpDesk”, and a
NCDC NOMADS Inter-Divisional Team
May 6, 2003**

The following list outlines proposed actions to promote the use of NOMADS as an access instrument for various modeling and product development groups; an internal NCDC NOMADS Team,; and Model “HelpDesk” as discussed at the NCDC-GFDL telecon on 4/15/03.

1. Develop the adequacy of
 - o the NOMADS metadata schema’s (CF, NetCDF, GRIB, HDF, etc.) for NOMADS access to promote model intercomparisons; and
 - o the integration of selected observational schemas and data for these models and the existing external large-scale data dictionary and cataloging efforts. Specifically the NSF sponsored THREDDS and NSDL projects, the NOMADS Earth Science Portal, and the DOE’s Earth System Grid (ESG). THREDDS includes the FGDC, OPeNDAP, NcML, and NASA GCMD Data Interchange Formats (DIF); and WMO based GRIB and BUFR metadata standards. **Due:**
 - Initial analysis- July ‘04. Stouffer/Marshall (GFDL), Arnfield/Rutledge (NCDC), Olsen (NASA), Davis (Unidata), Hanlkin, (PMEL), Williams & Fiorino (DOE), Middleton/Buja (NCAR), Lawrence (BADC).

2. Establish a NCDC NOMADS Team for the advancement and understanding of distributed data access; and for the development of a prototype NCDC “Model HelpDesk” that includes on-line FAQ’s and “referral” capability to GFDL from NCDC. **Due:**
 - NOMADS Team: May ’03 Rutledge,Easterling;
 - HelpDesk: Jan04 Rutledge,Woldu, Lott, Stouffer.

3. Under the NOMADS Team, advance an operational NOMADS capability at NCDC that will include Web and ftp access to 1) model and 2) research access to observational and reference data sets, and 3) access to NCDC retrospective data within the archive; and initiate product development for next generation climate products and services. **Due:**
 - Web and ftp capability: June ‘03. Rutledge, Smith, Brinegar
 - Research access to observations: Dec ’03 Rutledge, Lott, CAB
 - NOMADS access to NCDC archive: Sep ‘03 Rutledge, Smith, D. Ross
 - Product Development: Initial Requirements- Sep ’03, Products CY04.