Distributed Online Data Access and Analysis

Ruixin Yang
George Mason University

Slides from SIESIP Partners and from NOMADS PI,
Glenn K. Rutledge of US NCDC on NOMADS

SIESIP: Seasonal-to-Interannual Earth Science Information Partner
NOMADS: NOAA Operational Model Archive and Distribution System
SIESIP

SIESIP is a Distributed Data and Information System among GMU, COLA, GDAAC, and UDel

SIESIP developed a distributed data and information system that addresses the research needs of S-I climate model and related communities, TRMM, process studies such as SCSMEX, and interdisciplinary Earth scientists.
SIESIP Federation

DODS

GMU

Others

COLA

SIESIP Client

Internet

Exchange Protocols

Data Ingest

Data Archiving

Data Orders

Other Data Sources (e.g. NOAA)

Data Delivery

Data Orders
(OPeNDAP) DODS

OPeNDAP (Open-source Project for a Network Data Access Protocol)
DODS (Distributed Ocean Data System)

- Client/server architecture
- http protocol
- Generic, flexible data model
- Metadata distribution capability
- “Open your data set with a URL instead of a filename”
Information Technology: GrADS

Visualizing
Maps, Charts, Animations

Interactive
Quantitative

INTEGRATED
USER INTERFACE

Analyzing
Expressions, Functions of Original Variables

User Definable, Extensible

Accessing, Subsetting
General slices of { 4D Grids In Situ Obs Images }

Arbitrary Domains
Optimized for Typical Queries
GDS: GrADS Data Server

OPeNDAP *(Open-source Project for a Network Data Access Protocol)*
“Open your data set with a URL instead of a filename” (DODS)

+ 

GrADS *(Grid Analysis and Display System)*

GDS

Data server with power of analysis
Data Access/Interoperability/Analysis

• **Level 0:** FTP & basic Web capability — limited functions but can provide **quick results** to relatively new users.

• **Level 1:** DODS server concept where server serves data and some metadata in a general way; supports subsetting. Client can support **data interoperability**.

• **Level 2:** Analysis server. Uses GrADS unique encapsulated analysis capability. Analysis results cached on server for further use.

  *Example:*
  *Calculate – at the server! – sea level pressure anomaly over N. America when tropical Pacific SSTA > 1.0. Return the result – as data ! – to the desktop.*

  *Size of data sets: 3GB*
  *Data processed at server: 5 MB*
  *Returned to client: 10KB*
An XML-Based Distributed Metadata Server (DIMES)

Metadata Integrated Data Analysis Server (MIDAS)

Contributors to DIMES/MIDAS

Ruixin Yang, Yujie Zhao, Menas Kafatos, X. Sean Wang, Yixiang Nie, Xinhua Deng, Changzhou Wang

Center for Earth Observing & Space Research (CEOSR)
George Mason University (GMU)
MIDAS?

Who is MIDAS (virtual)

What is MIDAS (real)
Gold (GDS Data) Inside (?)

DIMES (metadata) + GDS (data) => Metadata
Integrated Data Analysis Server
(MIDAS)

GDS Data
Metadata Clients

OPeNDAP Clients

Metadata request/result (XML)

Internet

Request Server List

Internet

Data request/result (OPeNDAP)

GDS URL Generator

Ingest Tool Box

DIMES

Data Server (GDS)

DATA

MIDAS

GDS URL Generator

Ingest Tool Box

DIMES

Data Server (GDS)

DATA

MIDAS
MIDAS Usage

Data set: TRMM

Temporal Coverage:
to GMT Dec 31 00:00:00 1999
Temporal Resolution: Monthly
Spatial Resolution: Longitude 1.0 degree
Latitude: from -40
Spatial Coverage: Longitude
Latitude
Contact Information:
e-mail: chiu@...

Set Tool Range: Tropical
Tool Color: Black
Grid Spacing: 1.0 X 1.0

Time range selected from 1/1/1998 to 12/31/1999

Temporal Coverage Selection: 1/1/1998 to 12/31/1999

Function Selection: Time Series Analysis
Parameter Selection: rain, 'rain' selected

Air Pressure Selection:

Generate String:

http://esip.gmu.edu:9080/zo4r/cespr_[term] {clloop [varname, lat=-78, lat=-12, lat=2]} [0]
The NOAA Operational Model Archive and Distribution System (NOMADS)

Glenn K. Rutledge
NOMADS Program Manager
National Oceanic and Atmospheric Administration
National Climatic Data Center

7/8/2008 e-SDDC Workshop
The goals of NOMADS are to:

- provide access to models,
- promote product development,
- foster research within the geo-science communities (ocean, weather, and climate) to study multiple earth systems using collections of distributed data,
- expand institutional participation via distributed technologies.
NOMADS simplifies scientific data networking, allowing simple access to high volume remote data, unifying access to Climate and Weather models:

**Data access (client)**
- Access to remote data in the users normal application
  - IDL / IDV / Matlab / Ferret
  - GrADS (GRIB/BUFR w/ GDS)
  - Netscape / Excel / http (wget)
  - CDAT (PCMDI)
  - Any netCDF application (i.e., AWIPS)
- Don’t need to know the format in which the data are stored.

**Data publishing (server)**
- Can serve data in various formats
  - netCDF / GRIB / BUFR / GRIB2
  - HDF (3-5) / EOS
  - SQL / FreeForm
  - JGOFS / NcML
  - DSP
  - ascii, others...

Spatial and temporal sub-setting and host side computations on the fly.
Collaborating Programs

CAP   Climate Action Partnership
CDP   Community Data Portal
CEOS  Committee on EO Satellites
CEOP  Coordinated Earth Obs Period
EPA   Air Quality Models
ESP   Earth Science Portal
NASA GCMD
NERC DataGrid
NSF Cyberinfrastructure
NSF LEAD & Geo-Science Tech Forum (GTF)
NVODS / US GODAE / GOOS
Unidata THREDDS, NSDL, DLESSE
WCRP World Climate Research Program

DOC DOE EPA State Dept
NCAR
NOAA Representative
NOAA Representative
(in progress)
Founding Member
Science Advisory Board
Advisory Committee
Member
Data / Planning Committee
Data Provider
Data Provider
JSC/CLIVAR Briefings
# The Partnerships

<table>
<thead>
<tr>
<th>CDC</th>
<th>BADC (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COLA co-PI</td>
<td>CEOP</td>
</tr>
<tr>
<td>FSL</td>
<td>CEOS</td>
</tr>
<tr>
<td>GFDL co-PI</td>
<td>LEAD &amp; GTF (NSF)</td>
</tr>
<tr>
<td>LLNL</td>
<td>NASA GCMD</td>
</tr>
<tr>
<td>NCAR</td>
<td>NERC DataGrid (NDG)</td>
</tr>
<tr>
<td>NCDC PI</td>
<td>NSSL w/UW/SSEC</td>
</tr>
<tr>
<td>NCEP co-PI</td>
<td>THREDDS</td>
</tr>
<tr>
<td>PMEL co-PI</td>
<td>United Nations FAO</td>
</tr>
<tr>
<td>Unidata</td>
<td></td>
</tr>
</tbody>
</table>
NOMADS Users

- **Data Philosophy and Retention**
  - Data are free.
  - NWP forecast data are retained for five years.
  - Analysis, Reanalysis, observations, and GDAS model input are retained for long term stewardship.

- **Data Users**
  - Resolution of IP addresses indicate a broad range, and consistent use of NOMADS available data:
    - U.S. Agencies, Academic Institutions: K-12 to Research
    - International governments, (Italy, Japan, countries within South America and Africa. Many others).
    - Private Sector and Non-Government Organizations NGO’s
    - World Bank, United Nations (FAO), others.
NCDC and NCEP Data

• NCDC NOMADS Archive
  • NWP from NCEP
    • POR: 2002 to Real-Time
    • Eta (12km); GFS (1 degree); GDAS; NARR 12km 30yrs
    • RUC-II 20/40km; Ocean and Ice WAVE Models
    • NCDC Reference Data Sets (Reynolds SST’s, GHCN...)
    • NCDC Mirror site to NCEP NOMADS for Eta & GFS

• NCEP Real-Time NOMADS
  • Global Forecast System GFS 1 degree
  • Hourly Eta at 12km
  • Regional Spectral Model (RSM) and Ensembles
  • Climate Data Assimilation System (CDAS)
  • AMIP Climate Monitoring, Climate Forecast Model
  • NCEP/NCAR Global Reanalysis 1&2
- Model input data assimilation fields for Regional model initialization (e.g., WRF, MM5,) and Regional Climate Models.
- Analysis of historical NWP for operational forecaster training.
- Third-world internet access to NWP for forecast operations.
- Subsets of high volume NWP and GCM avbl over the Internet.
Desktop Weather Forecasting

NCEP Global Weather Forecasts

COLA GrADS-DODS Server

Region-Specific Lateral BCs

WWW

PC-Based Regional NWP