Unidata
Community, Data, Tools, Services
(plus two)

February 2013
ODIP Workshop, Ostend, Belgium
Ben Domenico
Outline

• Unidata community: multidisciplinary research and education
• Internet Data Distribution (IDD): real-time (push) data systems via Local Data Manager (LDM)
• NetCDF with CF conventions encoding
• THREDDS (pull) web data and metadata services
• Integrated Data Viewer (IDV) for analysis and visualization
Unidata

• Mission:
  To transform the geosciences community, research, and education by providing innovative data services and tools

• At the Unidata Program Center, we
  – **Provide access to data** (via push and pull systems)
  – Develop open source tools and **infrastructure** for data access, analysis, visualization, and data management
  – Advance metadata standards for the Earth system science community
  – Support users of our technologies: faculty, students, and researchers
  – Help to build, represent, and advocate on behalf of a community
The Unidata Community

• Unidata is a community of education and research institutions with shared goals:
  – sharing geoscience data and
  – collaborating to develop software tools to access, analyze, and visualize the data.

• Unidata’s primary sponsor is the National Science Foundation AGS Division

• We seek to serve a broad community by fostering partnerships and collaborative opportunities that enhance research and teaching endeavors.

• Working with our community, the Unidata Program Center actively encourages diversity by engaging other scientific groups, including:
  – Atmospheric Science
  – Hydrology
  – Oceanography
  – GIS
  – Cyberinfrastructure
  – International groups
Pushes data from multiple sources using cooperating LDMs

No data center
Providing innovative data services and tools to transform the conduct of geoscience.

Internet Data Distribution

Data Fan-out from a Single Source
(For simplicity, backup links are not shown)
Real-Time Data Flows

• 30 data feeds provide
  – radar,
  – satellite,
  – weather station, buoy, upper air balloon & aircraft observations,
  – lightning,
  – weather forecast model output, ...

• LDM routinely handles 10 GB/hour input, with about 100 products/sec.

• Unidata Community IDD: 520 machines at ~250 sites, most of them multi-disciplinary

• The U.S. National Weather Service uses LDM-6 to collect and disseminate NEXRAD level 2 radar data operationally for over 150 radars
IDD in the USA

Unidata provides innovative data services and tools to transform the conduct of geoscience in the USA.
Global IDD

Unidata IDD
North American data delivery and sharing network

IDD-Brasil
South American peer of North American IDD

IDD-Caribe (planning)
Central American peer of North American IDD

Antarctic-IDD
Support of US Antarctic research community

Participants
United States
Canada
Puerto Rico
Costa Rica
Barbados
Venezuela
Chile
Brazil
Argentina
England
Portugal
Spain
Austria
Russia
Vietnam
China (Hong Kong)
South Korea
Antarctica (incipient)
Multiple International IDDs

**Unidata IDD**
North American data delivery and sharing network

**IDD-Brasil**
South American peer of North American IDD

**IDD-Caribe (a work in progress)** Central American peer of North American IDD

**Antarctic-IDD**
Support of US Antarctic research community

**Participants:**
- United States, Canada, Puerto Rico, Costa Rica, Barbados, Chile, Brazil, Argentina
- England, Portugal, Spain, Austria, Russia
- South Korea, Vietnam, Hong Kong, Macau
- Antarctica (episodic)
LDM Networks Outside the Unidata IDD

• US agencies
  – NOAA: All NEXRAD level-2 is collected and distributed via the LDM
  – NASA, USGS, ACE
• Government Agencies: Australia, Brazil, China, South Korea, Spain, Vietnam
• The TIGGE project (about 15 GB/hour, 24x7)
Collection and Distribution of NEXRAD-2 Data

The above graphic shows the collection and distribution of NEXRAD Level-2 radar data by the National Weather Service. The LDM is used throughout the system. From http://www.roc.noaa.gov/wsr88d/PublicDocs/Level_II/2010Level2_Arch.pdf.
Providing innovative data services and tools to transform the conduct of geoscience.

The above graphic shows the distribution of data among participating TIGGE data-nodes (approx. 15 GB/hr). From http://tigge.ecmwf.int/documents/additional/TIGGE_leaflet_2010.pdf.
NetCDF

• Scientific data systems sage: “Initially I thought of netCDF as a format, but it is really more like a container.”

• NetCDF is for binary data what XML is for text data.

• The 6 following netCDF and Common Data Model slides are hidden.
netCDF

- A simple **data model** for scientific datasets
- A **format** for portable, self-describing data
- A **programming library** that uses efficient direct access and efficient subsetting of multidimensional arrays
- Several programming interfaces: C, Fortran, C++, Java, Python, Perl, Ruby, ...
- Support for appending, sharing, and archiving data
A file has named variables, dimensions, and attributes. Variables also have attributes. Variables may share dimensions, indicating a common grid. One dimension may be of unlimited length.
netCDF 4

Addresses limitations of netCDF-3

- **NetCDF-4** integrates netCDF with HDF5, another major standard format and data model
- Parallel netCDF has proved suitable for high-performance computing
- NetCDF-4 data model (**CDM**) improves interoperability with other scientific data representations
- **NetCDF-Java** has advanced features, including access to remote data
- User-defined compound types (portable structs)
- User-defined variable-length types
- Groups for nested scopes
- Multiple unlimited dimensions
- String type
- Additional numeric types
- Unicode names
- Efficient dynamic schema changes
- Multidimensional tiling (chunking)
- Per variable compression
- Parallel I/O
- Reader-Makes-Right conversion
Common Data Model

- For a common subset of abstractions in OPeNDAP, HDF5, and netCDF-4
- User-defined compound types (portable structs)
- User-defined variable-length types
- Groups for nested scopes
- String type
- Additional numeric types
- Prototype implemented in netCDF-Java
- Attempts a balance between simplicity and power of representation
Common Data Model

Scientific Datatypes
- Point
- Trajectory
- Station
- Radial
- Grid
- Swath

Coordinate Systems

Common Data Access Model
- THREDDS
- OPeNDAP
- HDF5
- GRIB
- netCDF
- ...

Applications
netCDF Java

- 100% Java library has advances compared to C-based interfaces
- Prototype implementation of Common Data Model for access to netCDF-4, OPeNDAP, HDF5
  - Provides netCDF interfaces to other formats: Grids (GRIB1, GRIB2), Radar (NEXRAD, NIDS, DORADE), Satellite (DMSP, GINI), Point Observations (BUFR)
  - Provides uniform coordinate systems layer
- Includes access to THREDDS inventory catalogs
Climate & Forecast Conventions

- A widely used metadata standard for atmospheric, ocean, and climate data, based on netCDF
- Specifies coordinate systems used in models, data cell properties and methods, packing, standard names for quantities, and grid mappings
- CF-aware software can automatically determine space-time location of data variables
- Originally intended for climate model output conventions, but use has broadened to weather and ocean models and observational data
- Community governance structure now in place for maintaining and advancing the CF conventions

- *If netCDF is for binary files what XML is for text files, the CF conventions can be thought of as the binary equivalent of an XML schema*
OPeNDAP
formerly the Distributed Ocean Data System

- DODS was initially developed as a library that enabled applications programs to specify a URL that would enable access to data on remote servers via the netCDF API.
- DODS URLs specifying remote datasets and operations on them were difficult to construct by hand
- The next two slides on OPeNDAP are hidden.
OPeNDAP

- Open-source Project for a Network Data Access Protocol, see www.opendap.org
- A discipline-neutral protocol to get remote scientific data and metadata *(not files)*
- Allows requests for *subsets* and *aggregations*
- Software reference implementations for many kinds of data: netCDF, SQL (databases), HDF, FITS, JGOFS,
- In use in earth sciences, astronomy, medicine, ...
- Serves IPCC model output
OPeNDAP

- Protocol uses URLs and HTTP
- Several OPeNDAP servers available: pyDAP, FDS, GDS, DAPPER, THREDDS Data Server
- OPeNDAP clients include: Ferret, GrADS, Matlab, IDL, ArcGIS, netCDF-Java, IDV, McIDAS-V
- OPeNDAP version 2 now a NASA standard
- Version 4 under development: adds XML, new types, new functions, THREDDS catalogs, SOAP, outputs in HTML and ASCII
- Will add authentication, more server-side processing
THRED DDS Data Server (TDS)

- THematic Real-time Environmental Distributed Data Services
- THRED DDS was initially developed as a way to provide the equivalent of a remote file systems so users could browse OPeNDAP datasets on remote servers
- THRED DDS automated the process of constructing the messy DODS/OPeNDAP URLs
- The next 3 slides on TDS are hidden
THREDDS Data Server (TDS)

- For data providers, implements data catalogs to present to users and applications
- Catalogs are XML documents (metadata) describing and pointing to datasets accessible via client/server protocols (OPeNDAP, ADDE, WCS, HTTP)
- Datasets may be found by discovery centers (master directories, digital libraries, data portals) via catalogs
- Catalog hierarchy provides places to hang common metadata
- Unidata coordinates THREDDS activities, community implements servers
- Many partners as data providers, tool builders, interoperability experts from academia, government, industry
TDS

- Serves data, THREDDS catalogs, and metadata
- Reads and serves several kinds of data through a uniform CDM interface: netCDF, OPeNDAP, HDF5, GRIB, NEXRAD, ...
- Adds Earth-location coordinate systems to data
- Provides OPeNDAP access and subsetting of any data readable with NetCDF-Java library
- An integrated server provides data access through the OpenGIS Consortium Web Coverage Service (OGC/WCS)
- Easy to install, 100% Java, freely available
- Supports dynamic generation of catalogs
providing innovative data services and tools to transform the conduct of geoscience

TDS

HTTP Tomcat Server

THREDDS Server

NetCDF-Java library

Catalog.xml

• OPeNDAP
• HTTPServer
• WCS

hostname.edu

Datasets

Application

IDD Data

unidata

providing innovative data services and tools to transform the conduct of geoscience
IDV

- Unidata’s newest scientific analysis and visualization tool
- Freely available 100% Java framework and reference application
- Provides 2-, 3- and 4-D displays of geoscience data
- Stand-alone or networked application
- Integrates data from different sources
- Provides end-to-end test for technologies
VisAD Data Model

- Designed to support virtually any numerical data
- Metadata can be integrated into each data object
- Supports mathematical operations as well as evaluation and resampling of data
- Supports spatial and temporal co-location of data
- Supports data sharing among different users, different data sources and different scientific disciplines
- May be used independently of the VisAD display model
IDV Examples

FY2 IR/WV Comparison
IDV Examples

GFS Windspeed Isosurface and Temperature Cross Section
IDV Examples
IDV Examples

TREX Field Project – Airplane Tracks
IDV Examples

Mantle Convection, Strain rate, Tomography
GEON-IDV/UNAVCO
IDV Examples
IDV Examples

Multi-channel Products:
Convective, Air Mass, IR Color, Dust
IDV Features

- Client-server data access from remote systems
- Suite of data probes for interactive exploration (slice and dice)
- Animations (temporal and spatial)
- HTML interface for pedagogic materials
- XML configuration and bundling allows collaboration with other educators
- Java-based framework supports extensions built via plug-ins: for example, geosciences network (GEON) solid earth community

providing innovative data services and tools to transform the conduct of geoscience
IDV Features

• Versatile data interaction
  – 3D views of 3D data
  – Probes to interrogate data – time series, vertical profiles, etc.

Model simulation of wind, isentropic potential vorticity and low level moisture flow over the Great Salt Lake basin

Time Height Cross section

Time Series Data Probe
IDV Features

- Resource creation – Color tables, projections, station models
- Highly Configurable
  - Multiple UIs & displays – 3D Map, 2D Map, Globe, Transect
  - Plug-ins
    - New Features
    - Language Support
Summary of Widely Used Systems

- **Unidata community**: multidisciplinary research and education
- **Internet Data Distribution (IDD)**: real-time (push) data systems via **Local Data Manager (LDM)**
- **NetCDF with CF conventions encoding**
- **OPeNDAP/DODS** (the Distributed Ocean Data System)
- **THREDDS** (pull) web data and metadata services
- **Integrated Data Viewer (IDV)** for analysis and visualization
Why add a brokering layer?

- WCS 1.0 Client
- WCS 2.0 Client
- OPeNDAP Client
- CS/W - ISO Client
- CS/W ebRim Client
- OpenSearch Client

- WCS 1.0 Web Service
- WCS 2.0 Web Service
- OPeNDAP Web Service
- CS/W - ISO Web Service
- CS/W ebRim Web Service
- OpenSearch Web Service
Three-tiered Architecture
with Brokering Layer

WCS 1.0 Client

WCS 2.0 Client

OPeNDAP Client

CS/W ebRim Client

OpenSearch Client

WCS 1.0 Service

WCS 2.0 Service

OPeNDAP Service

CS/W ISO Service

CS/W ebRim Service

OpenSearch Service

Common Data Model for Data Access Protocols and Formats

Common Metadata Model for Discovery Protocols and Semantics

WCS 1.0 Accessor

WCS 2.0 Accessor

OPeNDAP Accessor

CS/W ISO Accessor

CS/W ebRim Accessor

OpenSearch Accessor

WCS 1.0 Web Service

OPeNDAP Web Service

CS/W - ISO Web Service

CS/W ebRim Web Service

OpenSearch Web Service

providing innovative data services and tools to transform the conduct of geoscience
NODC Option for SOS

**Brokering Service as a Separate Layer**

<table>
<thead>
<tr>
<th>Service Type</th>
<th>WCS 1.0 Service</th>
<th>WCS 2.0 Service</th>
<th>OPeNDAP Service</th>
<th>CS/W ISO Service</th>
<th>CS/W ebRIM Service</th>
<th>OpenSearch Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Data Model</td>
<td>WCS 1.0 Accessor</td>
<td>WCS 2.0 Accessor</td>
<td>OPeNDAP Accessor</td>
<td>CS/W ISO Accessor</td>
<td>CS/W ebRIM Accessor</td>
<td>OpenSearch Accessor</td>
</tr>
<tr>
<td>Common Metadata Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Web Services**

- **WCS 1.0 Web Service**
- **OPeNDAP Web Service**
- **CS/W ebRim Web Service**
- **OpenSearch Web Service**

**Client Applications**

- **WCS 1.0 Client**
- **WCS .20 Client**
- **OPeNDAP Client**
- **CS/W - ISO Client**
- **CS/W ebRim Client**
- **OpenSearch Client**

**Accesors**

- **WCS 1.0 Accessor**
- **WCS 2.0 Accessor**
- **OPeNDAP Accessor**
- **CS/W ISO Accessor**
- **CS/W ebRIM Accessor**
- **OpenSearch Accessor**
Further Information

http://www.unidata.ucar.edu

support@unidata.ucar.edu