



# Unidata

*Providing data services, tools, & cyberinfrastructure leadership  
that advance Earth system science, enhance educational opportunities, & broaden participation*

## Unidata's Involvement in Developing and Supporting Climate Science Infrastructure

•Russ Rew

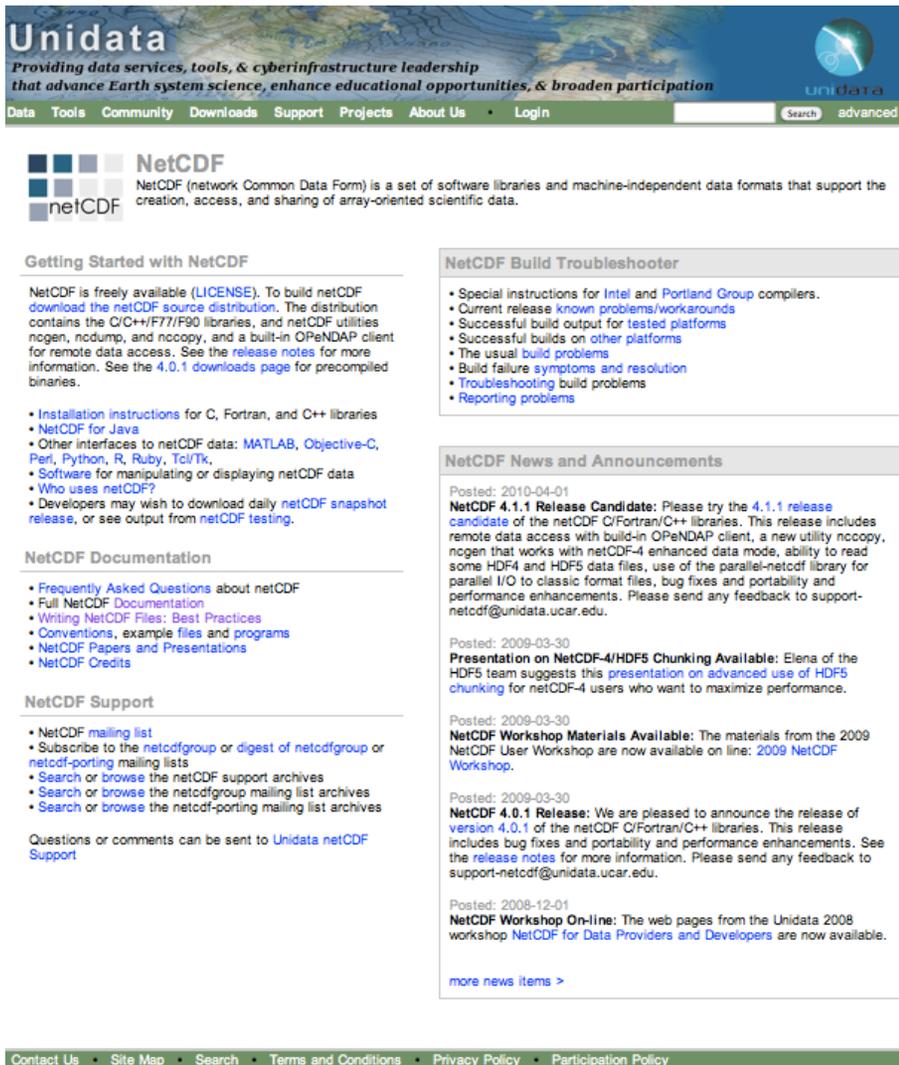
*UCAR Unidata  
April 2010*



# Overview

- Background
  - NetCDF
  - Climate and Forecast Metadata Conventions
- Current involvement
  - CF conventions and governance
  - TDS, CDM, NcML, Libcf, GridSpec
- Upcoming efforts
  - NOAA projects: GIP, NCDC climate portal, CF satellite conventions
  - Proposals under review
- Concluding remarks

# NetCDF: Unidata's first cyberinfrastructure



The screenshot shows the Unidata website with a navigation bar at the top containing links for Data, Tools, Community, Downloads, Support, Projects, About Us, and Login. Below the navigation bar is a search box and a 'Search advanced' button. The main content area is titled 'NetCDF' and includes a sub-header 'Getting Started with NetCDF'. The text describes NetCDF as a set of software libraries and machine-independent data formats. It provides links to installation instructions for various languages (C, Fortran, C++, Java, Perl, Python, R, Ruby, Tcl/Tk) and a frequently asked questions section. There are also sections for 'NetCDF Build Troubleshooter' and 'NetCDF News and Announcements' with several news items dated from 2010 to 2008. At the bottom of the page, there is a footer with contact information and logos for UCAR and the National Science Foundation.

- Portable, self-describing data format, data model, and software libraries supporting creation, access, and sharing of scientific data
- 1990's: achieves widespread use in ocean and climate modeling
- 2002: Java version with OPeNDAP client support
- 2003: NASA funds netCDF/HDF project, Argonne/Northwestern parallel netCDF released
- 2007: netCDF/CF format mandated for CMIP model archive for IPCC AR4
- 2008: netCDF-4 with HDF5 integration, enhanced data mode, parallel I/O
- 2009: netCDF format standard endorsed by NASA
- 2010: OPeNDAP client support for C/Fortran libraries



COMMUNITY PROGRAMS

Unidata is a member of the UCAR Community Programs, is managed by the University Corporation for Atmospheric Research, and is sponsored by the National Science Foundation.  
P.O. Box 3000 • Boulder, CO 80307-3000 USA • Tel: 303-497-8643 • Fax: 303-497-8690

# Climate and Forecast (CF) Conventions

cfconventions.org

The screenshot shows the CF Metadata website interface. At the top, there is a blue header with the text "CF Metadata" and "NetCDF Climate and Forecast Metadata Convention". Below the header is a navigation bar with tabs for "home", "documents", "conformance", "discussion", "governance", and "working groups", along with a "log in" link. A breadcrumb trail indicates the current page: "you are here: home → documents → cf conventions → 1.4 → netcdf climate and forecast (cf) metadata conventions".

The main content area is titled "NetCDF Climate and Forecast (CF) Metadata Conventions" and includes a "Next" link. Below the title, it states "Version 1.4, 27 February, 2009". The "Original Authors" section lists:

- Brian Eaton** (NCAR)
- Jonathan Gregory** (Hadley Centre, UK Met Office)
- Bob Drach** (PCMDI, LLNL)
- Karl Taylor** (PCMDI, LLNL)
- Steve Hankin** (PMEL, NOAA)

The "Additional Authors" section lists:

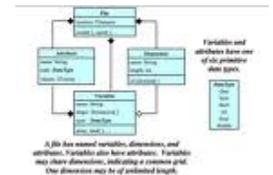
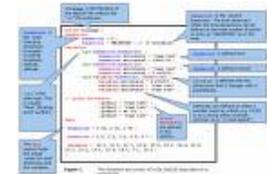
- John Caron** (UCAR)
- Rich Signell** (USGS)
- Phil Bentley** (Hadley Centre, UK Met Office)
- Greg Rappa** (MIT)

A navigation sidebar on the left lists various categories: Home, Documents, CF Conventions (with sub-items 1.4, 1.3, 1.2, 1.1, 1.0), Conformance, Discussion, Governance, and Working Groups.

- Community agreements for earth science metadata interoperability
- Conventional ways to specify
  - Coordinate information needed to locate data in space and time
  - Standard names for quantities to determine whether data from different sources are comparable
  - Additional grid information (e.g., grid cell bounds, cell averaging methods)
- Infrastructure widely used in tools and climate models: AMIP, CCMVal, CEOP, CFMIP, C-LAMP, CMIP3, CMIP5, ENSEMBLES, IOOS, MERSEA, NARCCAP, PMIP, ...

# Unidata's Roles in the CF Conventions

- **Development**
  - Active in resolving CF issues and developing conventions
  - Completing new conventions for observational data
- **Software**
  - NetCDF-Java generates CF metadata when reading data in other formats
  - Libcf helps data providers create CF-compliant data
- **Standards**
  - Unidata-written draft CF standard for NASA under review
  - Comprehensive specification of meaning of CF-compliance
- **Governance**
  - Serving on CF Governance and Conventions committees



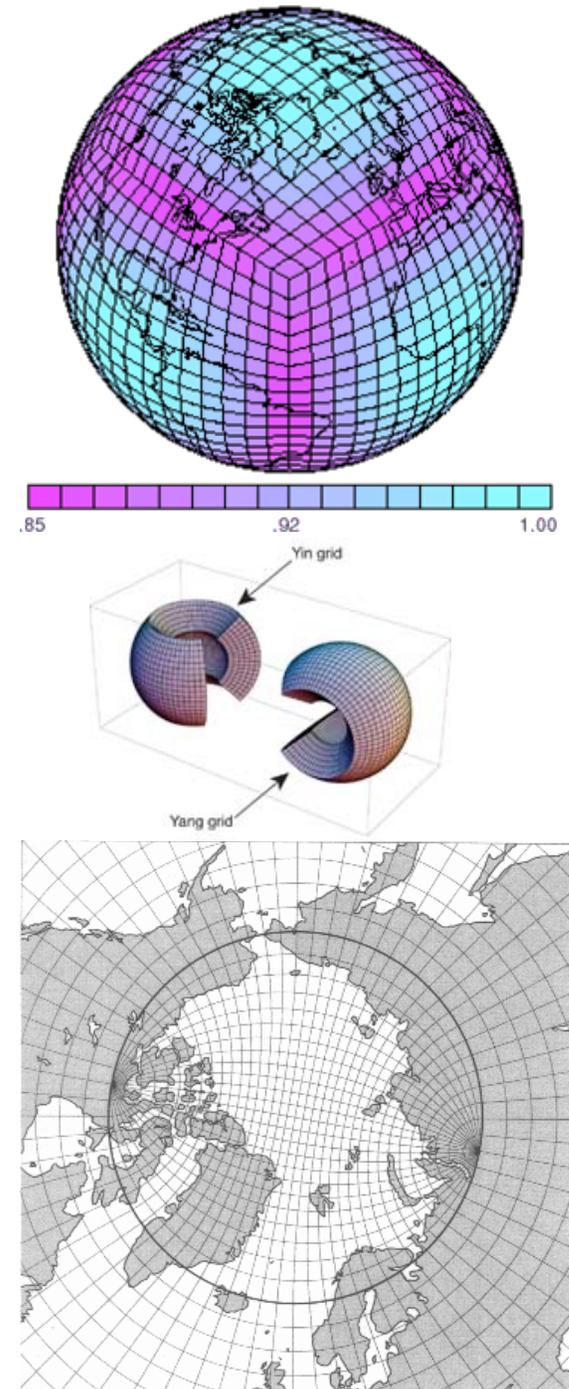
# Unidata's Development of a CF Library



- Goals:
  - Make use of CF Conventions easy for data writers and readers
  - Help ensure CF-conforming files, CF-compliant applications
  - Provide geo coordinate systems for netCDF data
  - Provide advanced features of netCDF-Java for C and Fortran developers
- Status:
  - Available in alpha-release form
  - Includes early release of Gridspec library

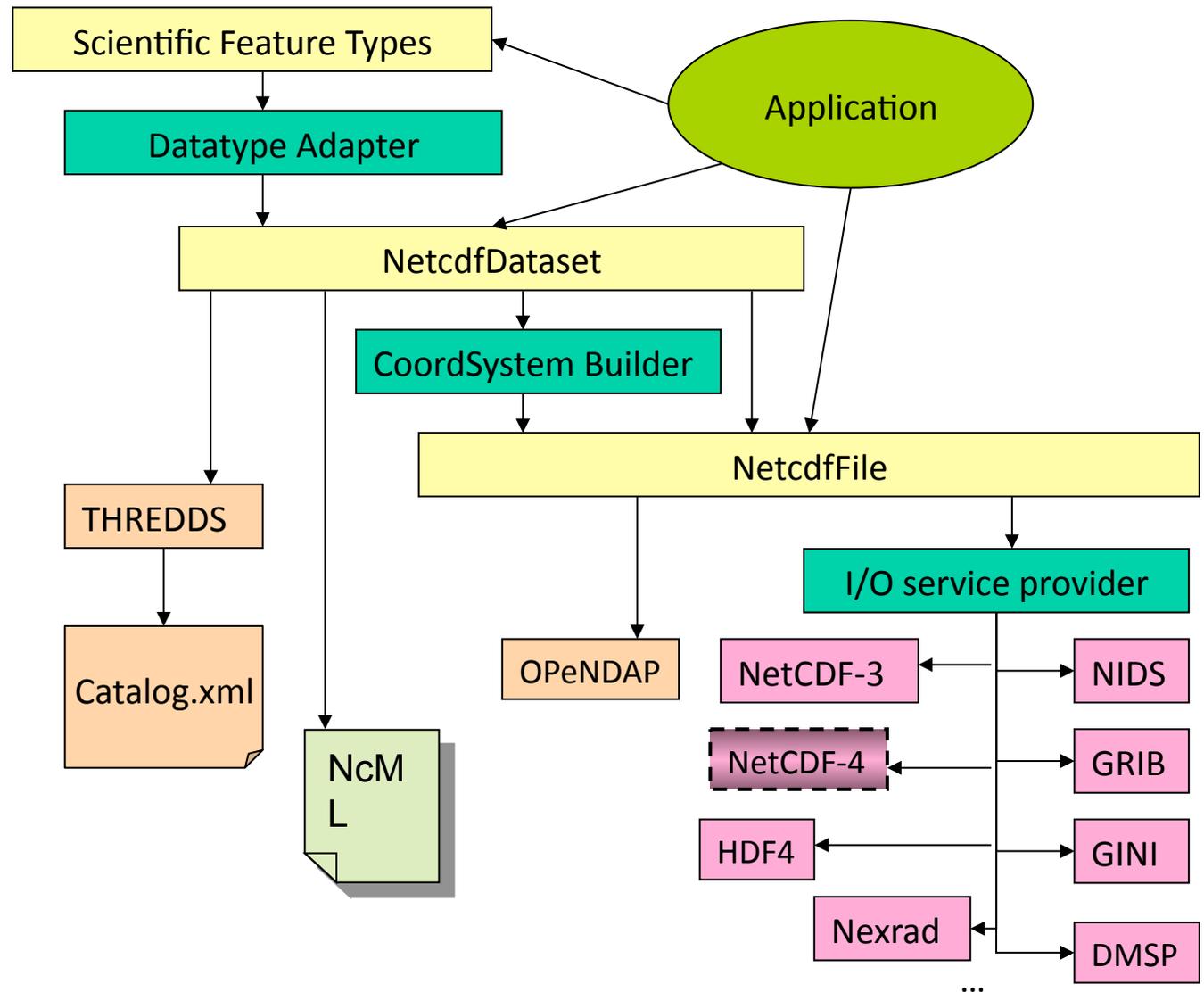
# Gridspec

- A proposed CF standard extension for the description of grids used in Earth System models
- Developed at Princeton GFDL (Balaji, Liang)
- Library API included in Unidata's libcf
- Allows AR5 model output to use model native grids: staggered, nested, cube-sphere, tripolar, yin-yang
- Supports conservative regridding by users
- Unidata's primary role in NOAA Global Interoperability Project
  - Inclusion of Gridspec C in libcf
  - Development of Fortran 2003 API
  - Automated build system and testing
  - Documentation of APIs



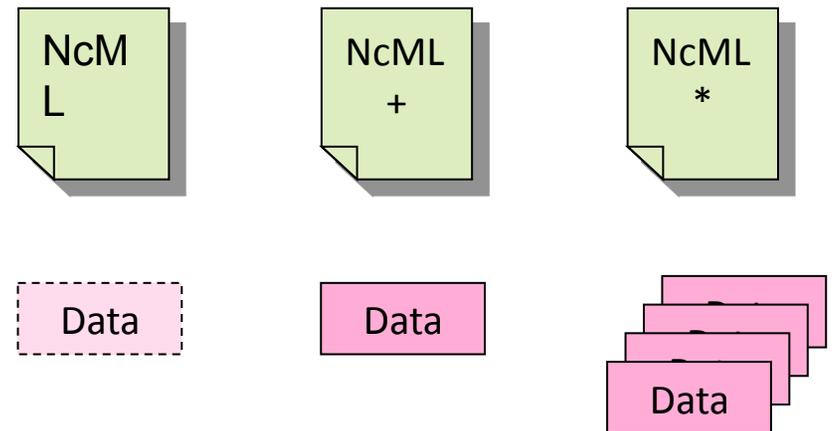
# Unidata's Common Data Model

- A useful merger of netCDF, OPeNDAP, and HDF5 technologies
- Implemented in netCDF-Java
- Reads over 20 data formats through netCDF interface
- Adds coordinate systems layer
- Adds scientific features layer (grids, trajectories, swaths, discrete samplings, ...)



# What is NcML?

- Server-side XML representation of netCDF metadata
- Uses include:
  - creating new netCDF files
  - modifying (“fixing”) existing datasets without rewriting them
  - creating virtual datasets as aggregations of multiple existing files
- Integrated with the TDS



# What is the THREDDS Data Server?



- Web server for scientific data
- Bulk file transfer via HTTP
- Remote access, subsetting CDM files
  - OPeNDAP (any CDM files)
  - Open Geospatial Consortium (OGC) Web Coverage Service (grids)
  - OGC Web Map Service (grids)
  - NetCDF subset service (grids)
  - Experimental data access protocols (any)
- Clients include IDL, MATLAB, ArcGIS, GoogleEarth, IDV, McIDAS-V, NASA World Wind, netCDF-Java applications such as Panoply, ncBrowse, ERDDAP, ...

# The Present Situation

- Climate researchers make use of software and services from Unidata: netCDF, TDS, CDM, NcML, libcf, Gridspec
- Unidata is participating in various climate data and related projects: CF conventions development, OGC standards, preparations for CMIP5 for IPCC AR5
- Unidata is being pulled into new climate initiatives: NOAA National Climate Model Portal, NOAA Global Interoperability Program, CF satellite product conventions

# NCDC March 2010 Workshop on Ensuring Access and Trustworthiness of Climate Observations and Models

- From final report:

*Throughout discussions at the workshop it was clear that both the research and operational Agencies that are involved with the emerging National Climate Service rely on foundational IT infrastructure that has been developed and supported by efforts such as the NSF UCAR Unidata group (e.g., the NetCDF format, CF names and common data model) ...*

# Recent Developments

- Unidata has developed and mostly implemented CF conventions for discrete sampling (formerly “point observations”)
- Unidata has proposed further involvement with related research initiatives, including NSF proposals to SDCI solicitation
- With recent NCDC climate observation meeting, a larger role for Unidata technologies in the National Climate Service is envisioned
  - Funded collaboration to improve model-to-observational data intercomparisons in support of NOAA’s National Climate Model Portal
  - Visit from NCDC developer to learn TDS technologies (IOSPs, NcML, ...)
  - and coordinate developing CF conventions for satellite data products
  - Participation in new Unidata [cf-satellite mailing list](#), with NCDC, SSEC, EUMETSAT, and other collaborators

# Proposal to NSF OCI SDCI:

## *Enhancements for Developers of netCDF Tools*

- Proposed collaboration with C. Zender, UCI, netCDF operator utilities, widely used for analysis of climate model outputs
- Adds new high-level interfaces for tool developers, to be used in NCO for quick feedback loop
- Breaks a chicken-and-egg logjam delaying effective use of recent advances in scientific data models for large and complex collections

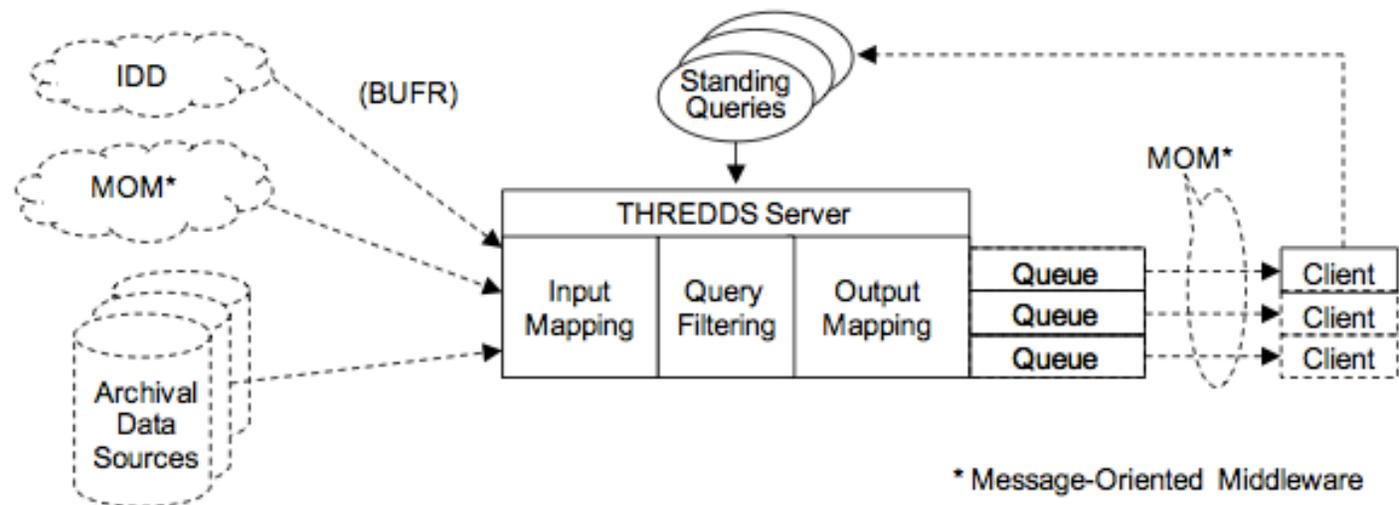


# Second Proposal to NSF OCI SDCI:

*Extensions to Unidata's CDM, IDD, and THREDDS Data Server*

*for Streaming Real-time Data and Large Archived Data Sets*

- Provides ability to stream packets of earth science data based on filters
  - Add user-defined content-based filters
  - Query language based on CDM, specialized for scientific data
- Scalable to very large archives
  - Parallel techniques (Google map-reduce)
  - Asynchronous and synchronous communication
  - Commodity message queues to decouple client and server



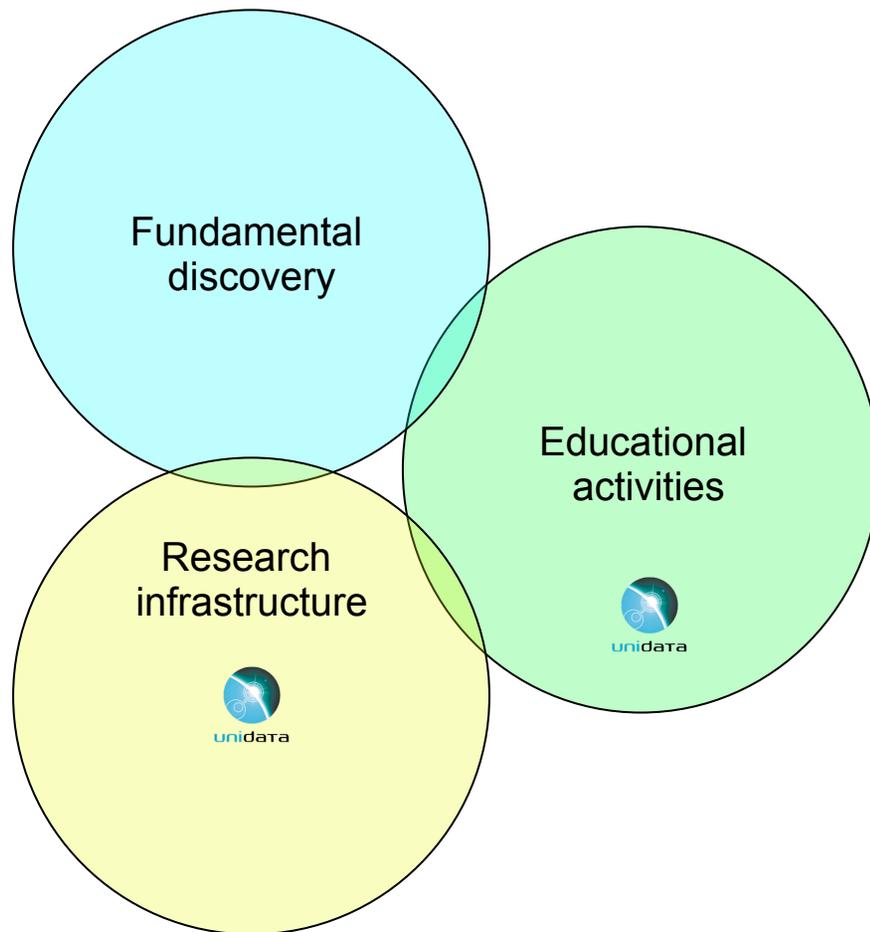
# Concluding Remarks

- Unidata has had significant involvement in climate research infrastructure since the early 1990's
  - netCDF and udunits software
  - early conventions
- Unidata's contributions have been important and sustained
  - CF Conventions, CDM, TDS, OPeNDAP clients and servers
- Unidata is currently being pulled into more climate-related activities and projects
  - Libcf, Gridspec, NCDC climate project, proposals, CMIP5

# Questions and Discussion

Extra slides

# Unidata and NSF Science, Engineering, and Education for Sustainability (SEES)



- Climate change research infrastructure
  - Software
  - Services
  - Standards
  - Support
- Educational activities
  - Engaged multidisciplinary community
  - Experience bringing data to classrooms

# Desirable Outcomes of NCDC/Unidata Project

- Directly support the improvement of some of the required scientific data software infrastructure
- Facilitate model-to-observational inter-comparisons in support of the IPCC's Climate Model Intercomparison Project(s) (CMIP) and their associated Assessment Reports (AR)
- Provide improved scientific data manipulation capabilities using Unidata software and applications (the THREDDS Data Server, TDS) to enable aggregation of long time-series model data

# Primary Goals of NCDC/Unidata Project

- Optimize aggregation capabilities in netCDF-Java library and TDS. Work to ensure that file aggregation works for the very large, very long timescales for ensemble and reanalysis data
- Develop IOSP's or NcML to make specific datasets CF-compliant: Climate Forecast System Reanalysis, Reynolds O/I ST's, NCEP Global Forecast System, NCEP Global Ensemble, Smith-Reynolds ERSST
- Develop and maintain new IOSPs for five ASCII datasets: Global Historical Climatology Network (GHCN) in-situ data, Interactive Global Rawinsonde Archive (IGRA) upper-air in-situ data, USHCN v2, ISD/ISH, ICOADS
- Serve data through TDS using CLASS as backend data store
- Assist in training NCDC software engineers to maintain and extend the work.

# A Long Term Vision

- Leverage NSF's investment in Unidata's proven expertise in developing software and services for the climate change research community
- Balance Unidata's efforts between weather and climate
- Provide cyberinfrastructure for climate change research
- Focus on enabling educators to bring climate change science to the classroom

# Unidata GO-ESSP Participation

- Global Organization for Earth System Science Portals
- Collaboration to develop software infrastructure for distributed access to observed and simulated data from climate and weather communities
  - by developing individual software components
  - by building a federation of frameworks that can work together using agreed-upon standards
- Crosses institutional, agency and international boundaries: Bryan Lawrence (BADC), V. Balaji (GFDL), Michael Lautenschlager (German Climate Computing Centre), Dean Williams (LLNL), Don Middleton (NCAR), Steve Hankin (PMEL)
- Annual meetings include CF Conventions discussions, decisions
- Unidata an invited participant since 2004