TIGGE, an International Data Archive and Access System

Doug Schuster
(NCAR)

Steve Worley, Dave Stepaniak, Nathan Wilhelmi, Luca Cinquini, Hannah Wilcox (NCAR)

CISL
Computational and Information Systems Laboratory

Topic Outline

- TIGGE Archive Centers and Data Providers
- Data Format
- Status Snap Shot
- Technical Challenges
- User Interface
- Analysis Tools
- Brief Status and Contrast with Partner Centers
- Future Plans

International Foundation

- WMO World Weather Research Programme THORPEX
  - *THe Observing system Research and Predictability Experiment*
  - *Weather research leading to an integrated Global Interactive Forecast System*
    - Integrated across multiple international NWP Centers
  - *THORPEX Interactive Global Grand Ensemble Archive supports research*

NCAR

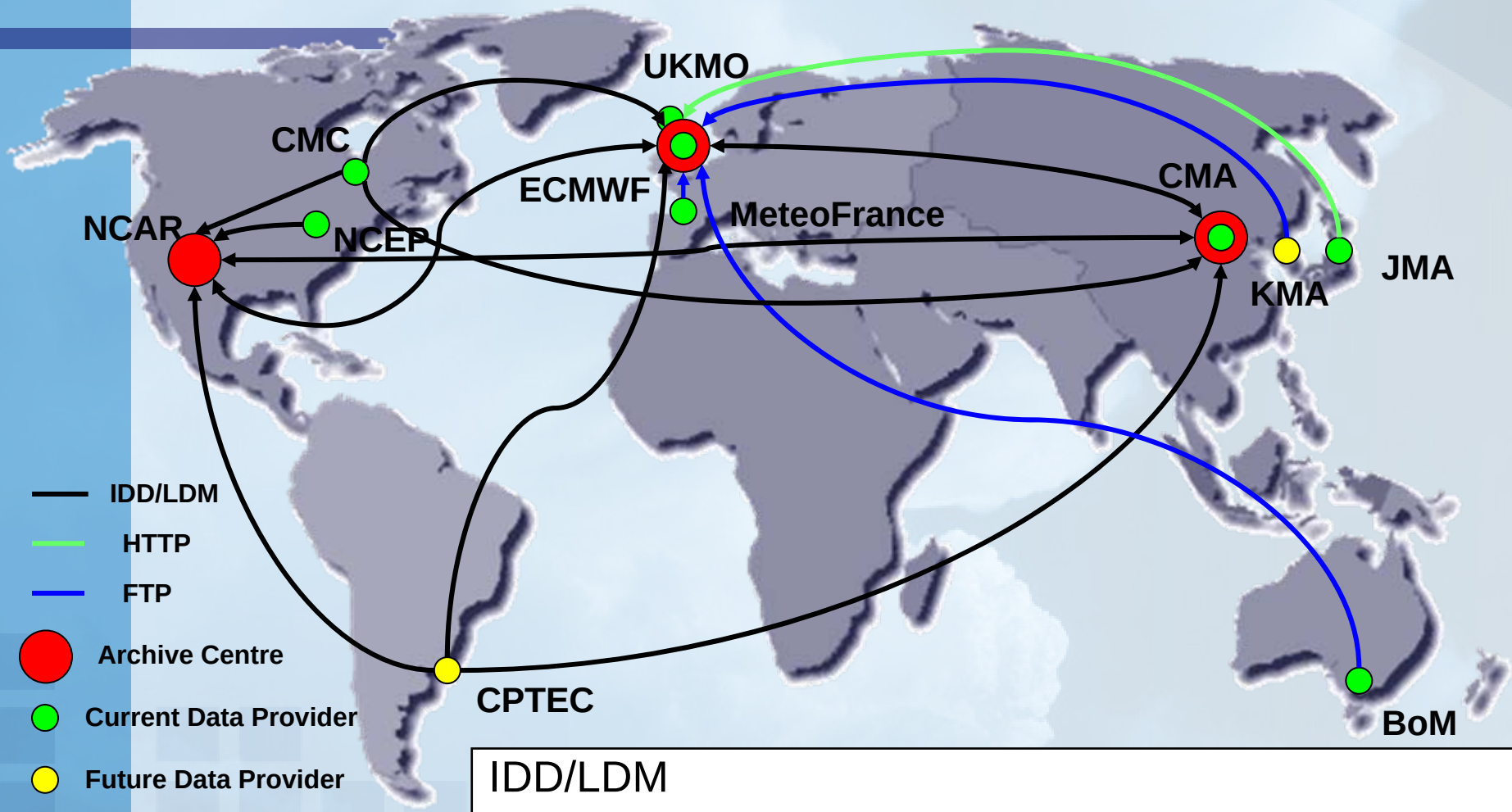
CISL
Computational and Information Systems Laboratory

Why Three International Archive Centers?

- Security and mutual back up at distributed mirrored sites
- Centralization creates a focus data service point for users
  - *Easy for users*
- Use existent proven data handling capability at experienced centers
- Allow most NWP centers to focus on providing data, not additional user service burden
- Note: Future TIGGE system is envisioned to be fully distributed - Phase II
  - *NWP centers could provide their own data service*

TIGGE Archive Centers and Data Providers

NCAR



CISL
Computational and Information Systems Laboratory

IDD/LDM

Internet Data Distribution / Local Data Manager

Commodity internet application to send and receive data

Standardized TIGGE Data Format

- Enables systematic data management
  - *GRIB2 file format*
  - *Consistent parameter encoding across all data providers.*
  - *Field compliancy - standard parameters, units, and pressure levels*
- Enables convenient multi-center multi-model comparison
- Outstanding challenges - anomalies between centers
  - *Native horizontal resolution*
  - *Number of ensemble members*
  - *Number of forecast initialization times (1x, 2x, 4x daily)*
  - *Forecast length*
  - *Number of fields provided*
  - *Internal file compression (e.g. jpg) was not specified*

Summary of Current Data Providers

Center	Conforming Parameters	Ens. Members	Model Res.	Fcst Length	Fcsts/ Day	GB/ Day	Fields/ Day	Files/ Day
ECMWF (ecmf)	70/73	51	N200 (Reduced Gaussian)	10 day	2	115	289,734	328
ECMWF (ecmf)	70/73	51	N128 (Reduced Gaussian)	10-15 day	2	24	138,978	160
UKMO (egrr)	70/73	24	1.25 x 0.83 Deg	15 day	2	21	175,680	488
JMA (rjtd)	61/73	51	1.25 x 1.25 Deg	9 day	1	7	113,192	74
NCEP (kwbc)	59/73	21	1.00 x 1.00 Deg	16 day	4	10	316,596	1040
CMA (babj)	60/73	15	0.56 x 0.56 Deg	10 day	2	28	72,510	82
CMC (cwao)	56/73	21	1.00 x 1.00 Deg	16 day	2	8	163,674	260
BOM (ammc)	55/73	33	1.50 x 1.50 Deg	10 day	2	8	147,972	164
MF (lfpw)	62/73	11	1.50 x 1.50 Deg	2.5 day	1	.15	7,558	33
Total					18	221	1,425,894	2,629

Status Snap Shot

Parameter	ECMWF	UKMO	JMA	NCEP	CMA	BoM	MF	CMC
10 meter U-velocity	x	x	x	x	x	X	x	x
10 meter V-velocity	x	x	x	x	x	X	x	x
Convective available potential energy	x			x			x	
Convective inhibition				x				
Land-sea mask	x	x	x	x	x	x	x	
Mean sea level pressure	x	x	x	x	x	x	x	x
Orography	x	x	x	x	x	x	x	x
Skin Temperature	x	x	x	x			x	x
Snow Depth Water Equivalent	x	x	x	x	x		x	x
Snow Fall Water Equivalent	x	x			x	x	x	
Soil moisture	x	x	x					
Soil temperature	x	x				x	x	x
Sunshine duration	x				x		x	x
Surface air dew point temp	x	x	x	x	x		x	x

Technical Challenges

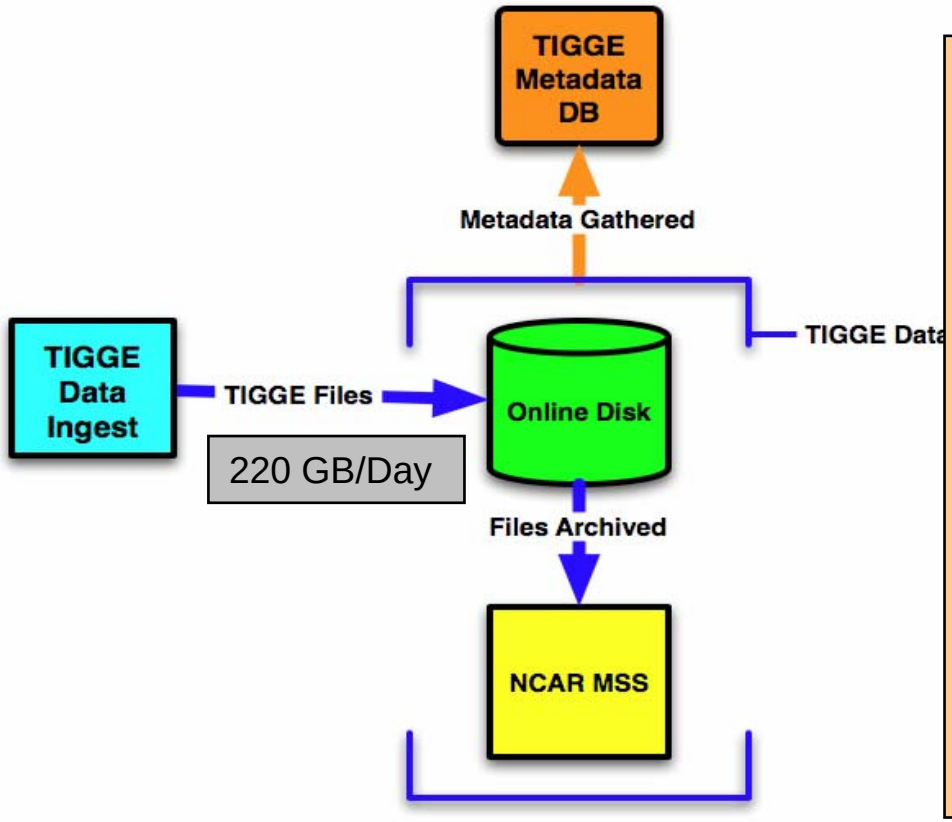
- Building a research file structure
  - *Receive over 1.4 million GRIB2 messages per day*
  - *NCAR doesn't have operational services so we handle TIGGE with methods common in science research - i.e in files*
    - Quite different from ECMWF
    - Create files based on Center, initialization date/time, forecast step, and level type
  - *Outcome - we manage over 2600 files per day*
    - Satisfactory approach with acceptable impact on the NCAR MSS

Technical Challenges

- Building a research file structure
  - *4 files per model timestep (if available)*
    - Single Level
    - Pressure Level
    - Potential Vorticity Level
    - Isentropic Level
  - *Each file includes all parameters and ensemble members for the specified level type and forecast time.*

Technical Challenges

Coordinated Online and MSS data



TIGGE Metadata DB Functions

Currency of all TIGGE data

- Location of all online files
- Location of all MSS files
- Pointers to all online GRIB records within files
- Constantly updated
- Drives display and access at the user interface
  - More discussion later

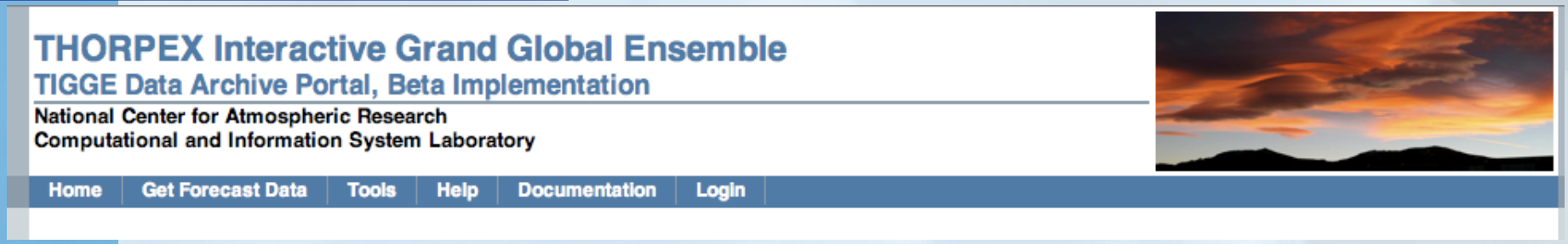
User Interface/Portal

Address: http://tigge.ucar.edu

Main Features

- *Registration and Login*
- *Get Data*
- *User Tools*
- *Documentation*
- *Technical and Community Supported Help*

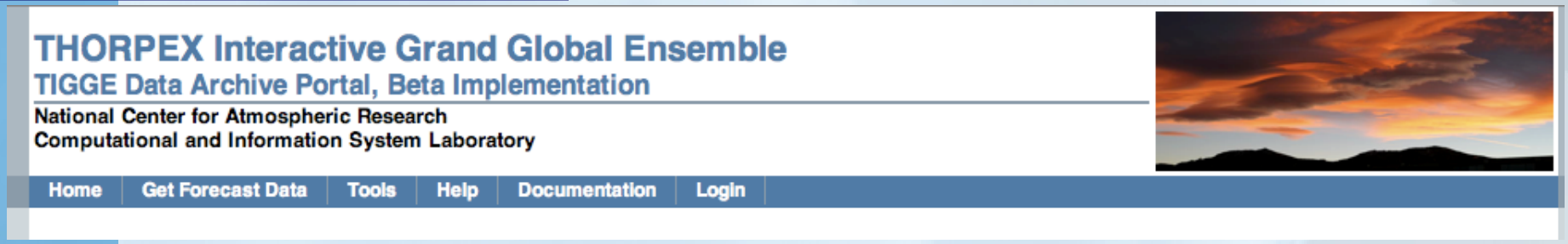
User Interface



Registration and Login

- *Required per international agreement*
  - Users electronically accept conditions for usage
  - Primarily, for education and research
  - 48-hour delay, except by special permission granted by IPO
- *We capture metrics for*
  - Name, email, organization name, organization type (univ., gov.,), and country
  - Who , what, when files were downloaded

User Interface



Get Forecast Data

- Two Selection Interfaces
  - *File Granularity*
    - Developed First
  - *Parameter Granularity*
    - Added Summer 2007

NCAR

CISL
Computational and Information Systems Laboratory

Dates

Center

File Type

Forecast Time

Forecast Duration

Data Request Selection

Start and End Dates

Start Date	End Date
2007-07-31	2007-08-01

Select Centers

China Meteorological Administration	European Center for Medium-Range Weather Forecasts	United Kingdom Met Office	National Centers for Environmental Prediction (USA)	Japan Meteorological Agency
☑ BABJ	☑ ECMF	☑ EGRR	☑ KWBC	☑ RJTD

Select Level Type

Center	Single Level Parameters (includes surface)	Pressure Level Parameters	Potential Vorticity Level Parameters	Potential Temperature Level Parameters	
BABJ	☐	☐			All Clr
ECMF	☐	☐	☐	☐	All Clr
EGRR	☐	☐	☐	☐	All Clr
KWBC	☐	☐	☐		All Clr
RJTD	☐	☐			All Clr

Select Initial Forecast Time

Center	00z	06z	12z	18z	
BABJ	☐		☐		All Clr
ECMF	☐		☐		All Clr
EGRR	☐		☐		All Clr
KWBC	☐	☐	☐	☐	All Clr
RJTD			☐		All Clr

Select Forecast Times

Center	1 (0-24)	2 (24-48)	3 (48-72)	4 (72-96)	5 (96-120)	6 (120-144)	7 (144-168)	8 (168-192)	9 (192-216)	10 (216-240)	11 (240-264)	12 (264-288)	13 (288-312)	14 (312-336)	15 (336-360)	16 (360-384)	
BABJ	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐							All Clr
ECMF	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		All Clr
EGRR	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐		All Clr
KWBC	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	All Clr
RJTD	☐	☐	☐	☐	☐	☐	☐	☐	☐								All Clr

Next

CISL
Computational and Informatio
Systems Laboratory

Get Forecast Data

Two User Interfaces

NCAR online file archive

- Selection options
  - Center(s)
  - Date
  - File type (sl, pl, etc)
  - Initialization time
  - Forecast length

Real Time

User customized files

- Selection options
  - **Same as for files, plus**
  - Parameter
  - Regridding
  - Spatial subsets
  - Formats, GRIB2 netCDF

Delayed Mode

Download Options

- Point and click using browser, one file at a time
- Script to run on local machine
  - User and password encrypted '**wget**' commands
  - background process to access all files

NCAR

CISL
Computational and Information Systems Laboratory

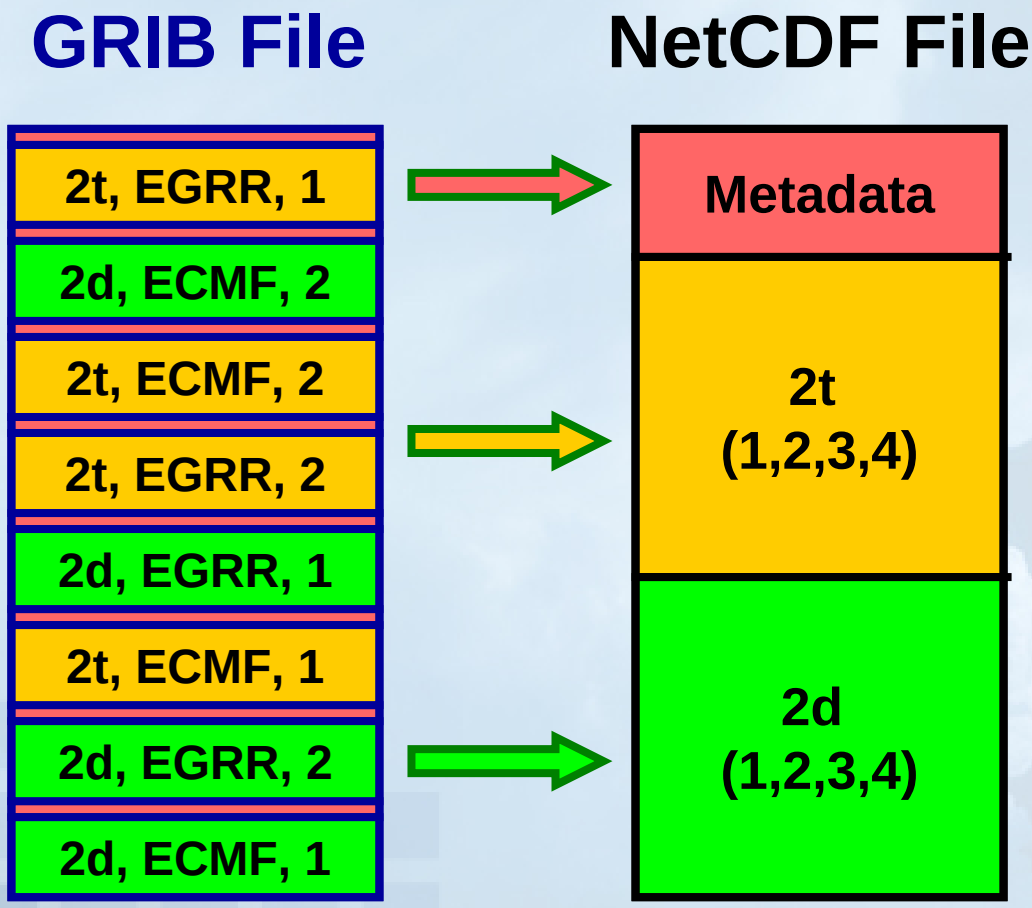
Data handling challenges and solutions

- Fast field extraction from a large GRIB archive
  - *Use a dynamic DB the holds address information for individual fields*
- Deriving user specified horizontal grids when no two native grids are the same
  - *Brute force, use specialized software and sufficient background computing (Spherepak and EMOS)*
- Inform users about delayed mode processing
  - *Have online queue so users can check status of their request*
- Minimize user repetitive interface input
  - *Archive user requests and seed online forms during subsequent visits (to be implemented)*
  - *Submit request as a subscription service (tbi)*

Tools

- Challenges
  - *New format, WMO GRIB2*
  - *Additional dimension, 5th, "ensemble member id"*
- Collection of tools with growing maturity
  - *Contributors*
    - NCAR
    - ECMWF
    - NOAA
    - Unidata
- Forthcoming
  - *NCAR and ECMWF staff are collaborating (ECMWF Consultancy) to develop a GRIB2 to NetCDF conversion tool*
    - Broad application, TIGGE and others
    - Initial development will leverage the ECMWF GRIB API
    - Complimentary to NCAR/NCL GRIB2 ingest capability

Tools: GRIB to NetCDF conversion



Gather metadata and message locations

Create NetCDF file structure

Populate NetCDF parameter arrays

(1,2,3,4) represents ensemble member id (Realization)

NCAR

CISL
Computational and Information Systems Laboratory

Tools: GRIB to NetCDF conversion

- NetCDF File format
  - *Based on available CF conventions*
  - *File organization built according to Doblas-Reyes (ENSEMBLES project) proposed NetCDF file structure*
  - *Provides grid/ensemble specific metadata for each member*
    - Data Provider
    - Forecast type (perturbed, control, deterministic)
  - *Allows for multiple combinations of initialization times and forecast periods within one file.*
    - Consistent across all data providers.
  - *All data on same Lat/Lon GRID.*

- Designed to work with GRIB-1 or GRIB-2 data using ECMWF GRIB-API functionality.

Tools: GRIB to NetCDF conversion

- NetCDF Parameter structure (5 dimensions):
  - *Time (Time id)*
  - *Realization (Ensemble member id)*
  - *Level*
  - *Latitude*
  - *Longitude*

- “Coordinate” variables are use to describe:
  - *Time*
    - Allows for multiple initialization times and forecast periods to be contained within one file
  - *Realization*
    - Provides metadata associated with each ensemble grid.

NCAR

CISL
Computational and Information Systems Laboratory

Tools: GRIB to NetCDF conversion

- Coordinate variables
  - *Provide ensemble member metadata*
    - Institution
    - Forecast Type

2d (1,2,3,4)

Realization	**1**	**2**	**3**	**4**
Institution	**ECMF**	**ECMF**	**EGRR**	**EGRR**
Forecast Type	**cf**	**pf**	**cf**	**pf**

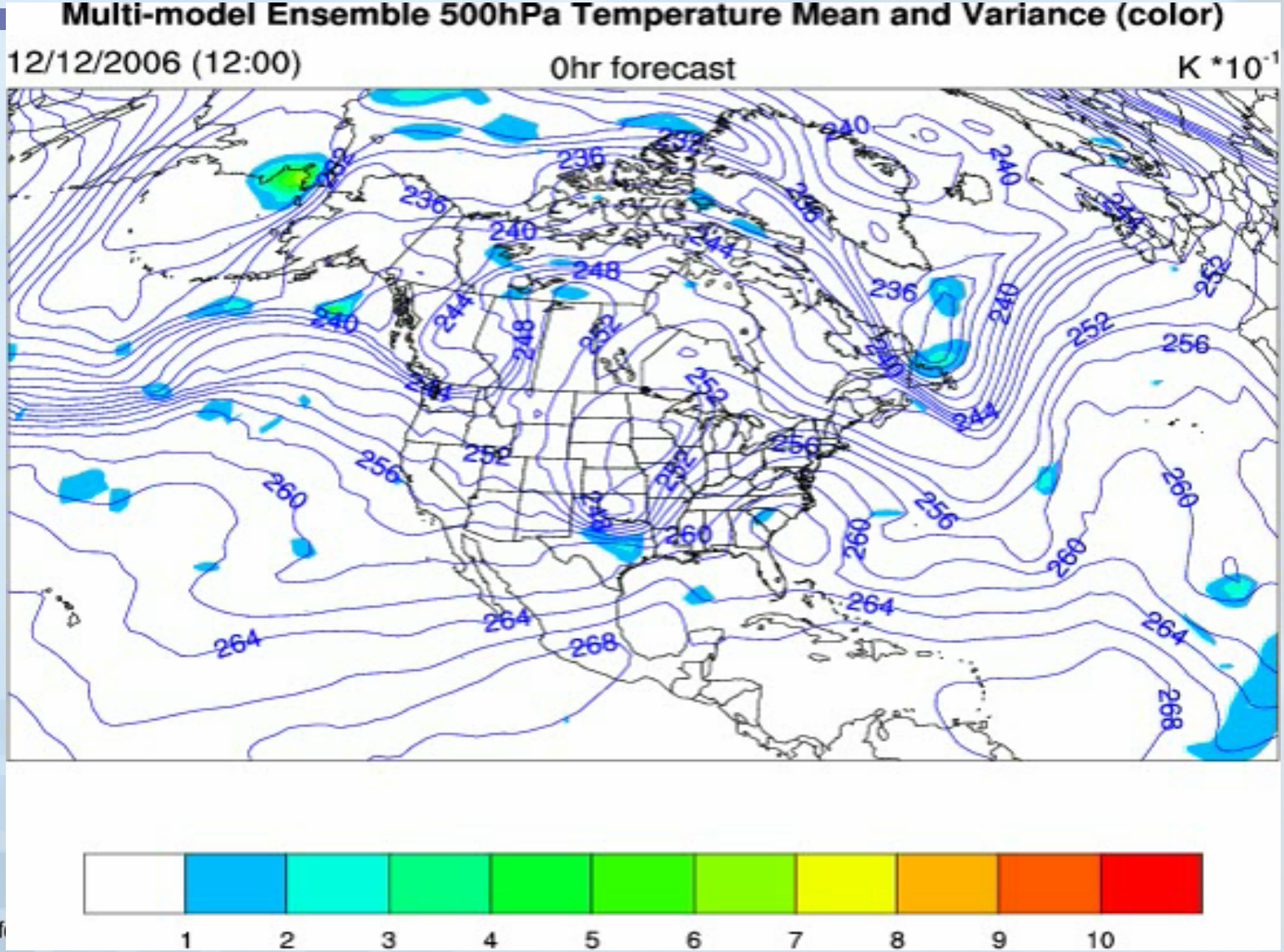
Tools: GRIB to NetCDF conversion

- Coordinate variables
  - *Provide time description*
    - Initialization time
    - Forecast Hour

2d (1,2,3)

Time	**1**	**2**	**3**
Initialization	**2007101212**	**2007101300**	**2007101300**
Forecast Hour	**24**	**12**	**18**

Tools, example; NCAR NCL



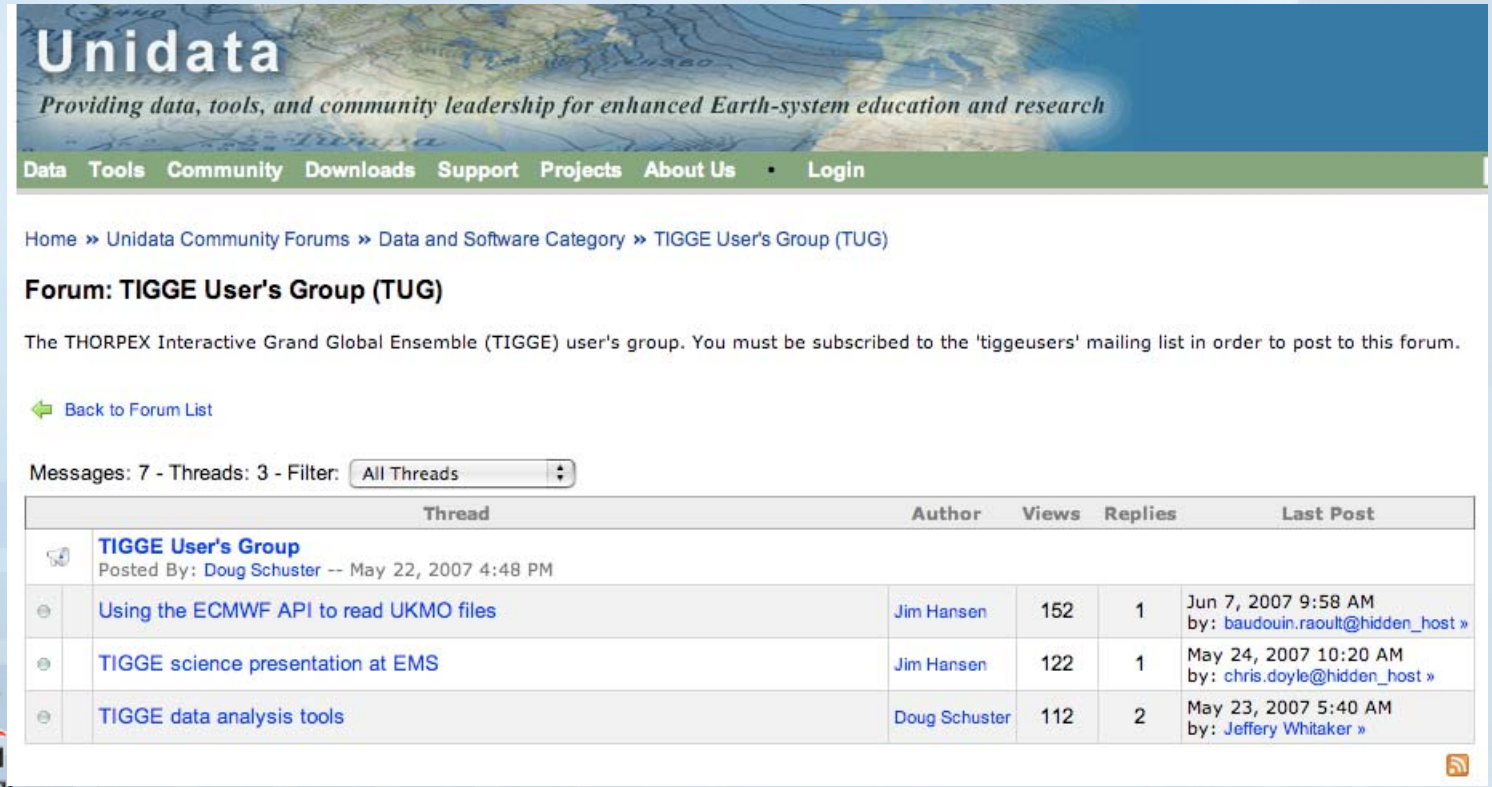
CISL
Computational and Information Systems Laboratory

NCAR

User Help

Two modes

- Technical assistance directly from TIGGE staff at NCAR via email
  - *Could originate from the portal*
- Open community website forum, including subscription email
  - *Enrollees can post questions, give answers, and share ideas and experiences*
  - *Provided by Unidata*

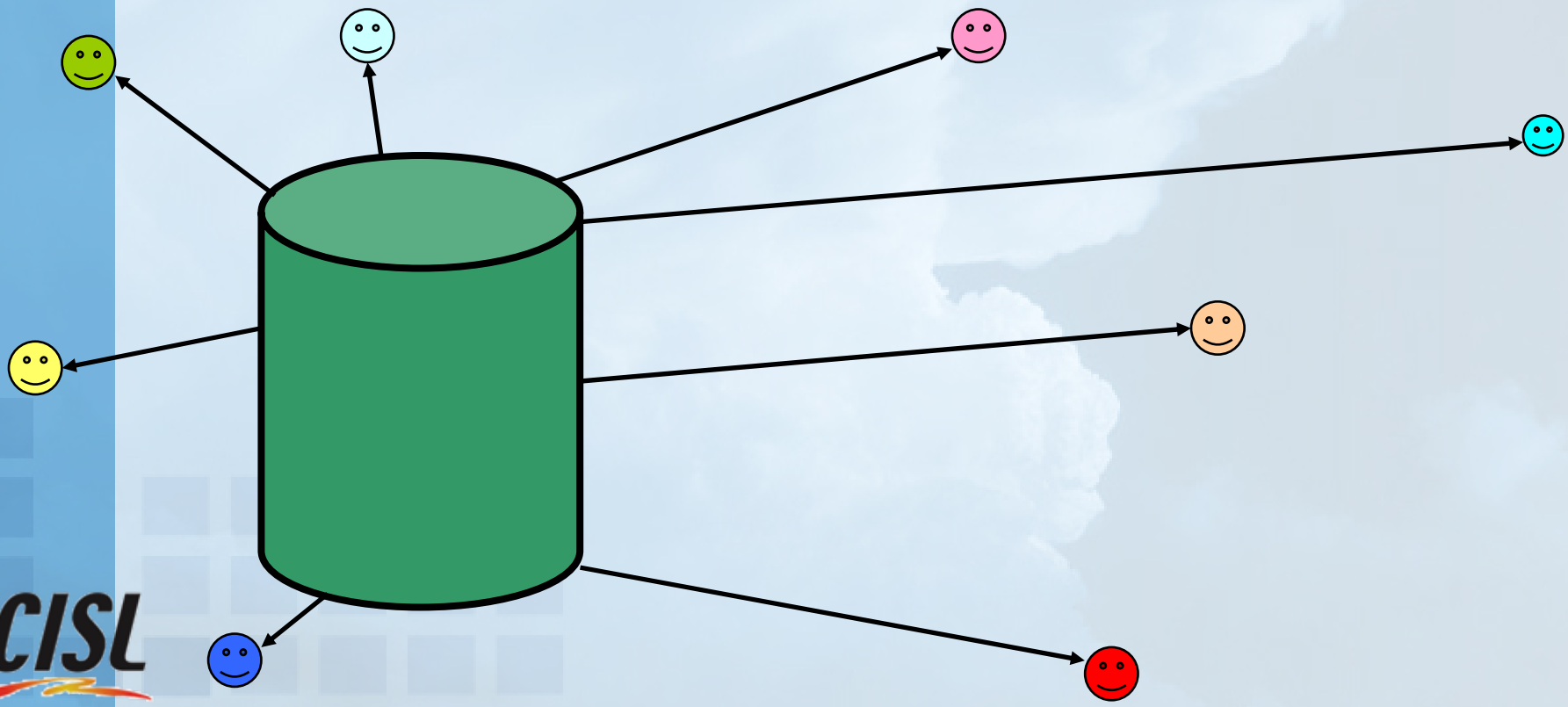


CISL Computational and Systems Laboratory

NCAR

NCAR TIGGE archive usage

- 1.0 TB, 65 K file, downloaded (11/11/07)
- 61 Unique data users



CISL
Computational and Information Systems Laboratory

Comparisons with partners; ECMWF

- NCAR and ECMWF have fully mirrored archives
- ECMWF uses a storage and access model based on individual fields (MARS)
  - *Quite different than NCAR files based system*
- ECMWF and NCAR have interfaces with the **same** look and feel
- ECMWF is a data provider and an archive center
  - *Has 160+ GB/day data produced locally (EC and UKMO)*
  - *Does significant data processing to prepare TIGGE fields from operational output*
  - *Assists UKMO, JMA, BoM, KMA, and MeteoFrance in building the TIGGE archive*

Comparisons with partners; CMA

- Uses file-based system to save all data at present
  - *Plan to deploy MARS in the near future*
- Designing a portal similar to NCAR and ECMWF
  - *Same look and feel*
  - *Same access options and development plan*
- Data provider and an archive center
  - *Receives data via IDD/LDM, same data as ECMWF and NCAR*
  - *Provide TIGGE data to support internal research program*
- Future plan at CMA
  - *Integrate data access portal interface with MARS*
  - *Enhance portal and open for wide data distribution*

Future at NCAR

- Complete advanced subsetting features
  - *Spatial, grid interpolation, and user selected output format (GRIB2 and NetCDF)*
- Add new contributors into the archive
  - *All have committed to doing so in 2007*
- Continue data analysis tool development
- Develop web service protocols for uniform direct access at distributed centers
  - *Termed as Phase II in TIGGE documentation*
  - *Could enable data provider host their data directly*
- Automated user access to long-term TIGGE holdings from the NCAR MSS

TIGGE Portals

Portals

- http://tigge.ucar.edu
- http://tigge.ecmwf.int
- http://tigge-portal.ecmwf.int

Doug Schuster, schuster@ucar.edu

NCAR

CISL
Computational and Information Systems Laboratory

Comparisons with partners; ECMWF

- Website/Portal (http://tigge.ecmwf.int)
- Primary Information
  - *Meeting Reports and Documentation*
  - *Technical information for Data Providers*
  - *Downloadable scripts to implement TIGGE IDD/LDM protocol*
  - *Detailed description of agreed GRIB 2 encoding*
- ECMWF Archive Status
  - *Monitoring plots showing each parameter from each Data Provider, use for quality assurance (e.g. correct units):*
  - *History web page: record of events, such as addition of new fields or missing cycles*

    (http://tigge.ecmwf.int/tigge/d/tigge_history/)

Summary Lessons

- Formal agreements on formats and variables are essential
  - *Small loop holes, anomalies, are problematic*
- Work sharing ethics between skilled partners allows rapid progress - TIGGE Archive partners are excellent
- Pushing the technical and experience limits forces leading edge developments, preparation for the future

Comparisons with partners; ECMWF

- Data Retrieval Interface
  - *User Registration*
  - *Access to **all** available data, including data off-line (on tape)*
    - Integrated with MARS
  - *Smallest accessible item: one 2D field*
  - *Subset by space, time, variable, level, etc.*
  - *Interpolation capabilities (re-gridding)*

Comparisons with partners; ECMWF

- Usage:
  - *45 registered users*
  - *2.5 TB extracted from the archive*
  - *After interpolation, 353 GB delivered to users*

- Future
  - *Add new data providers*
  - *Offer netCDF format output*
  - *Enable web service access*

Agreement Process

- Chronology of major workshops and outcomes
  - *First Workshop on TIGGE, March 2005, Reading UK*
  - *TIGGE - Archive Working Group, September 2005, Reading UK*
  - *2nd GIFS-TIGGE Working Group, March 2006, Reading UK*
  - *3rd GIFS-TIGGE Working Group, December 2006, Landshut Germany*
  - *4th GIFS-TIGGE Working Group, March 2007, Beijing China*
    - Establish data policy and requirements
    - Get agreement to participate from 10 NWP centers
    - Target support for IPY and Beijing Olympics '08
- Archive relevance
  - *Standardized data products, formats, distribution policy*

TIGGE Objectives

- Enhance collaboration on ensemble prediction, internationally and between operational centers and universities
- Develop new methods to combine ensembles from different sources and to correct for systematic errors (e.g. biases, etc)
- Achieve a deeper understanding forecast errors contributed by the observation, and initial and model uncertainties
- Enable evolution towards an operational “Global Interactive Forecast System”.

From Philippe Bougeault, ECMWF

Technical Challenges

- Why use IDD/LDM?
  - *Advantages*
    - **Application coordinates data transfer between sending and receiving queues - very automated**
    - **Queue size and TCP/IP packet size are configurable to optimize transfer rate and success**
    - **Developed and supported by Unidata, a UCAR program**
    - **Used in many other real-time data transport scenarios, e.g. education, field projects, US National Weather Service**
    - **Easy to coordinate multi-center exchanges, *one can feed many, CPTEC***
  - *Disadvantages*
    - **Somewhat complex to configure and tune**
    - **Monitoring software must be developed to assure archive completeness**
      - **Verify receipt against a manifest list, request data resend**

Technical Challenges

- Alternate Approach
  - *Use on 'old reliable' HTTP/FTP*
    - Exclusively a two-way exchange
    - Must arrange agreements and processes independently at both ends
    - Not complex
    - Works best for small to moderate data volume, e.g. JMA, KMA, BoM, MeteoFrance feeds to ECMWF