

Experiences Gained from Managing a Multi-Year, Multi-NWP Center Ensemble Archive

Doug Schuster – NCAR

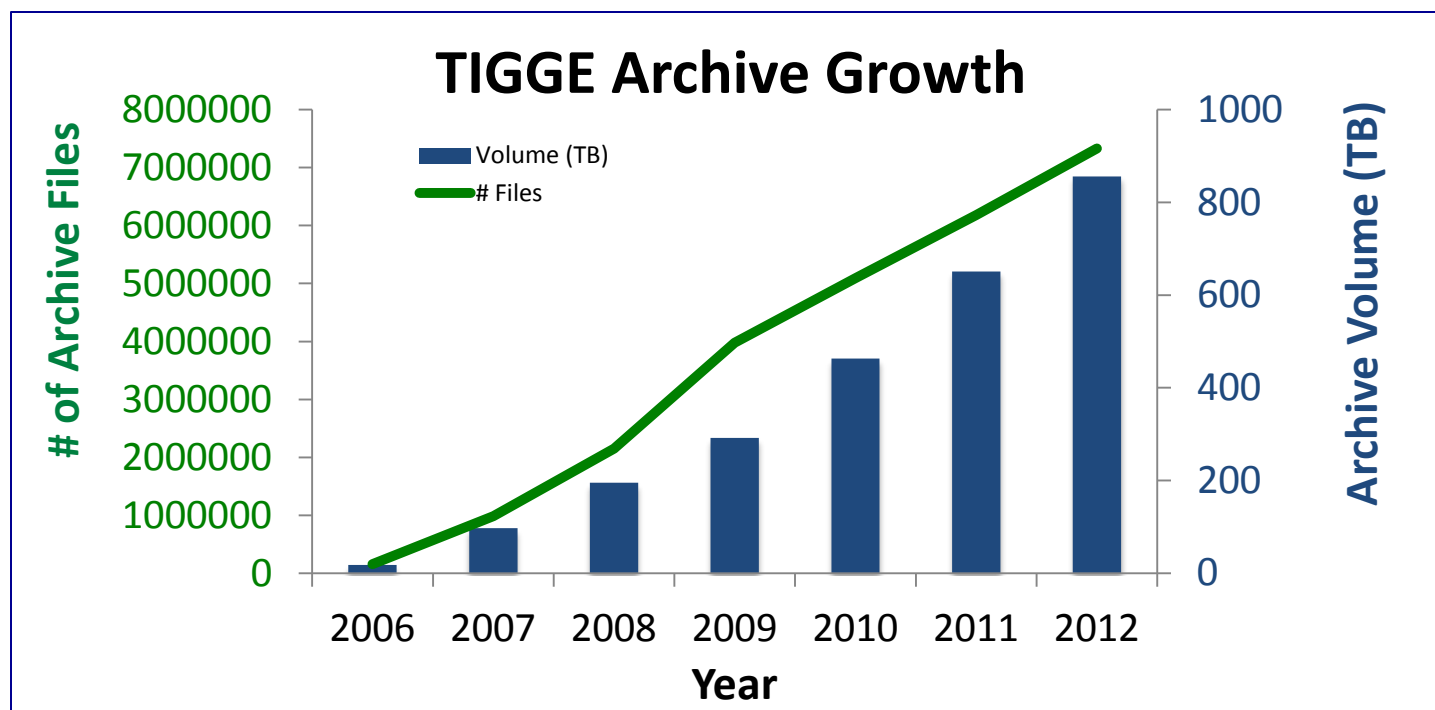
Many Partners at NCAR, ECMWF, and CMA

Outline

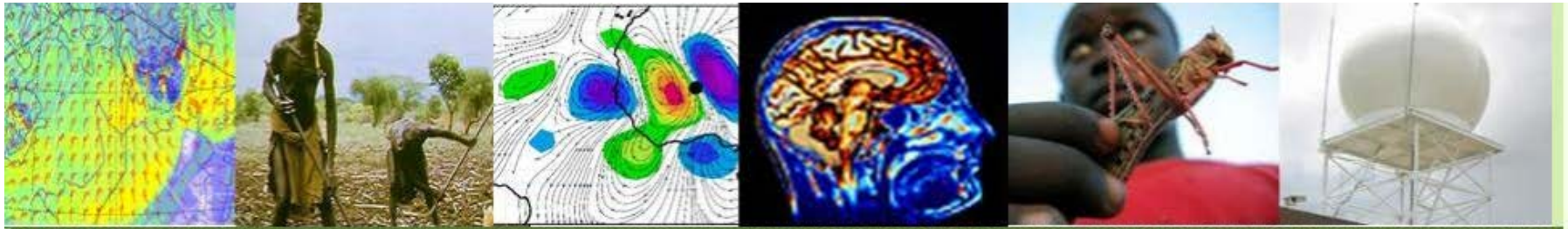
- Background
- TIGGE Project Design
- Resulting Operational Implementation
- Access Options
- Reflection

Background: TIGGE Project

- THORPEX Interactive Grand Global Ensemble
- Accelerate improvements in the accuracy of 1-day to 2 week high impact weather forecasts
- Global ensemble forecasts to around 15 days generated routinely at different NWP centres around the world
- 850 TB, growing by > 3.5 TB/week

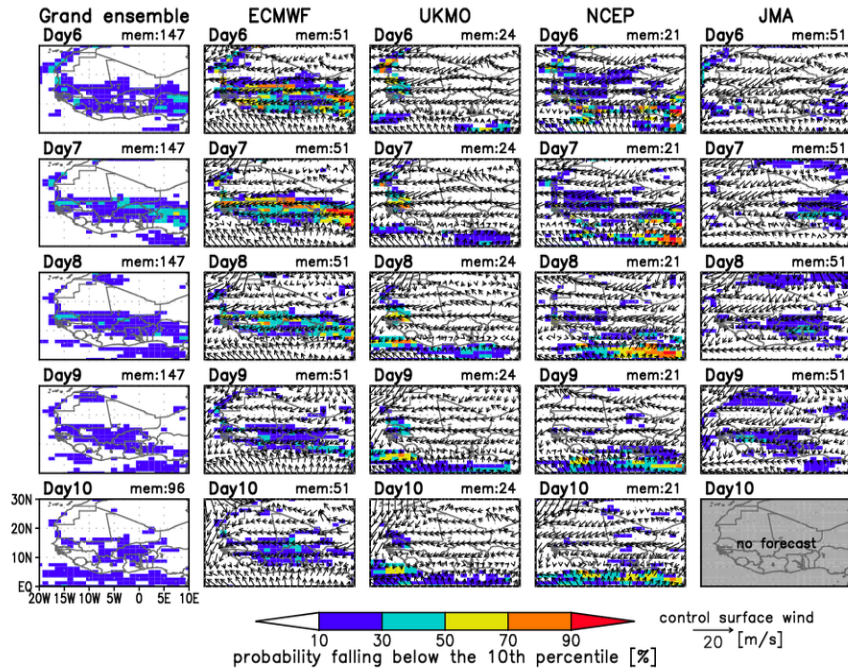


Background: TIGGE Project –Field Project Support

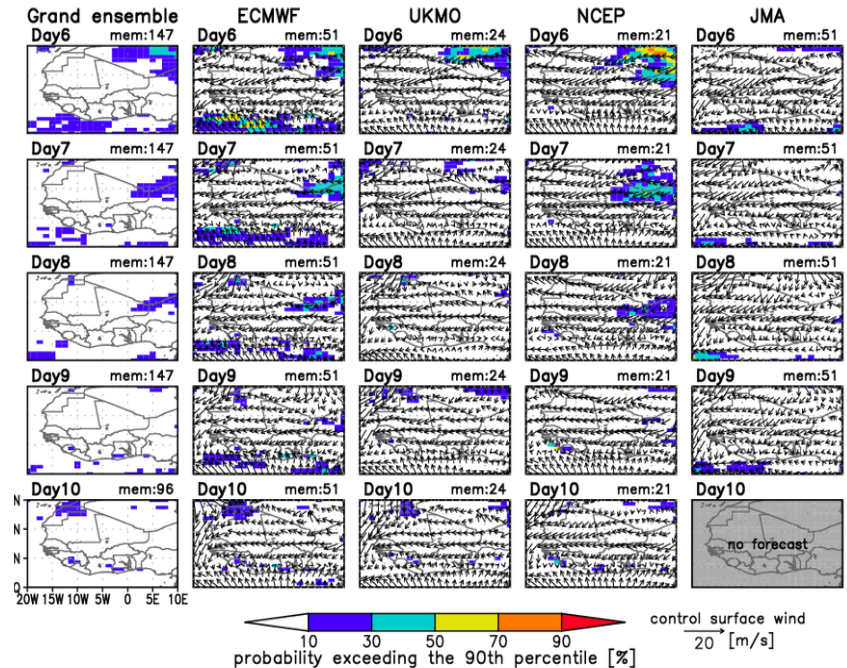


UCAR Africa Initiative

Probability of low humidity (10th percentile)
West Sahel Initial:12UTC 14DEC2012 Day6–10



Probability of high humidity (90th percentile)
West Sahel Initial:12UTC 14DEC2012 Day6–10

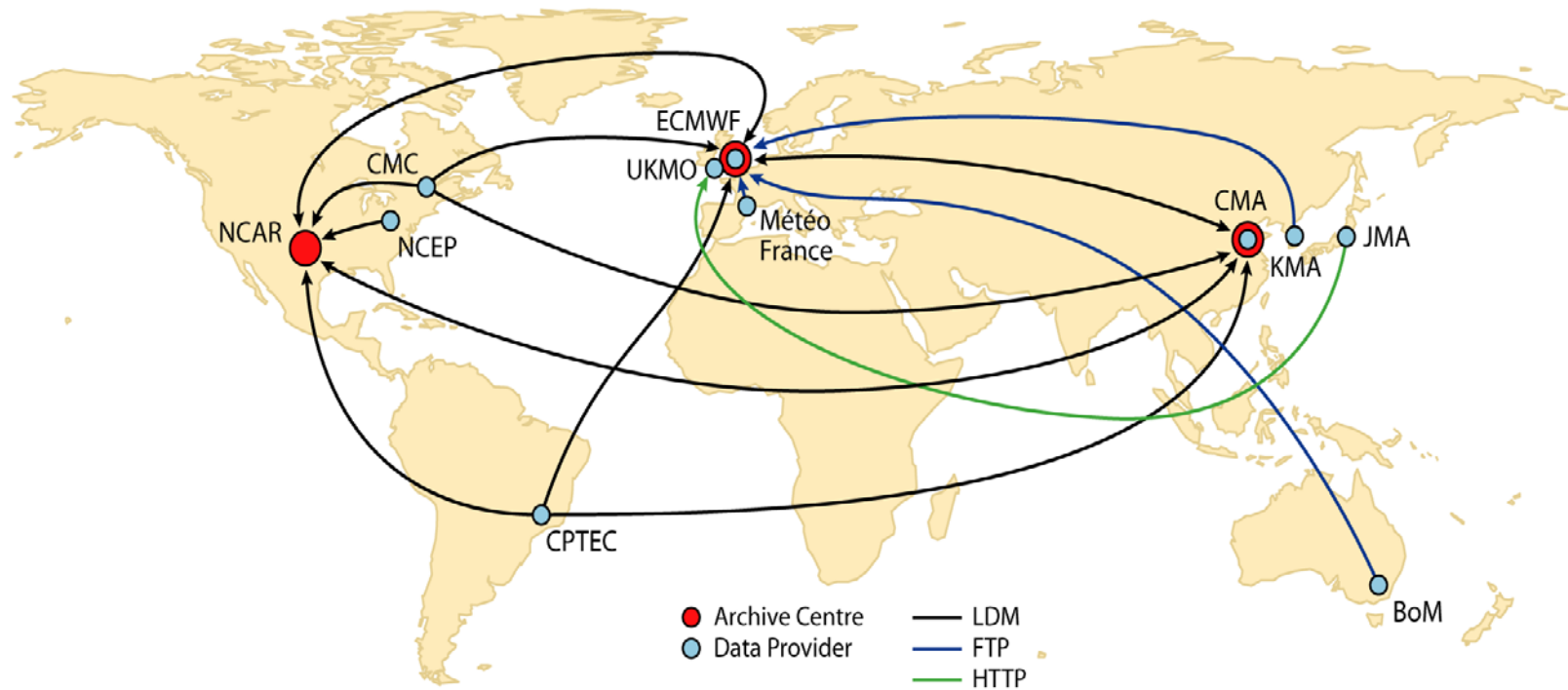


TIGGE Project Design: Common Standards -Homogeneity

- Homogeneity is paramount for TIGGE to succeed
 - The more consistent the archive, the easier it is to manage and to develop applications
- There are many aspects to homogeneity:
 - Common terminology (parameter names, file names,...)
 - Common data structures (format –WMO GRIB-2, units, ...)
 - Definition of an agreed list of products (parameters, steps, levels, ...)
 - Data transmission protocol for providers – Unidata
IDD/LDM
 - Software tools to facilitate and enforce homogeneity

TIGGE Project Design: Resulting Operational Implementation

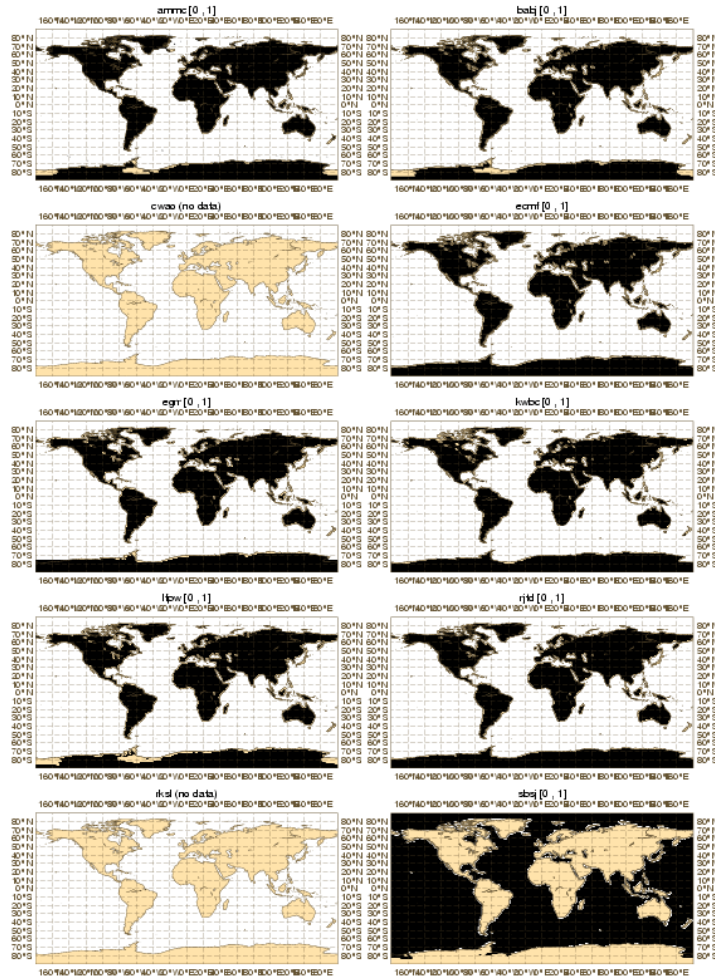
- Three archive centres: CMA, NCAR and ECMWF
- Ten data providers:
 - ECMWF, JMA (Japan), UK Met Office (UK), CMA (China), NCEP (USA), MSC (Canada), Météo-France (France), BOM (Australia), KMA (Korea), CPTEC (Brazil)
- Multiple data transmission methods
 - Problems implementing resend protocols



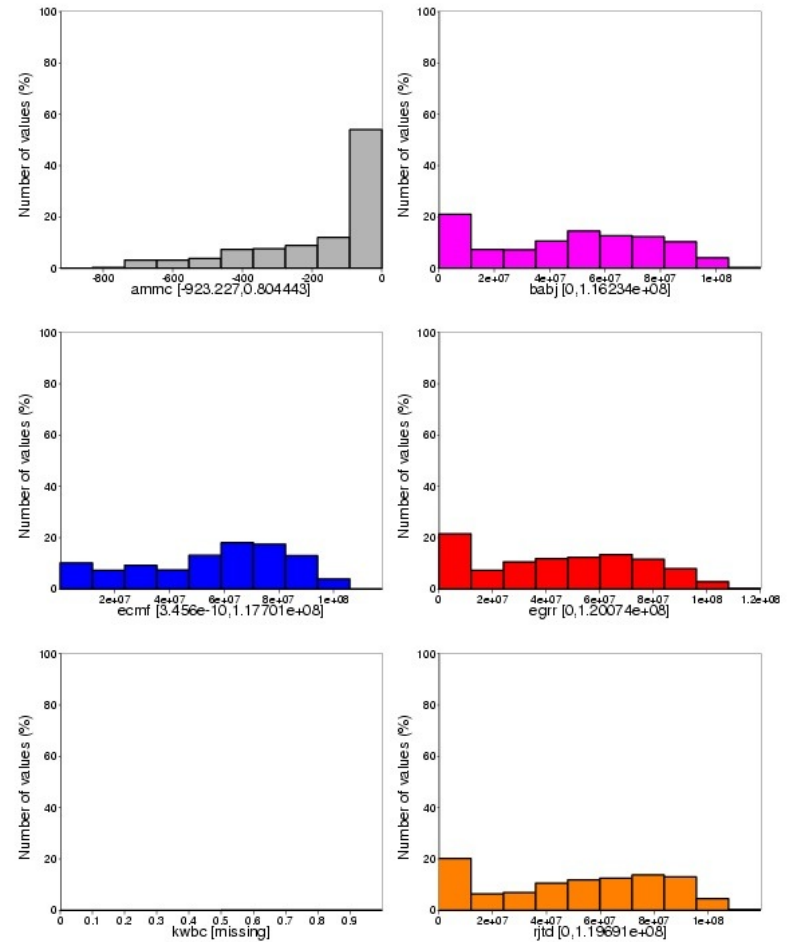
TIGGE Project Design: Resulting Operational Implementation

- Inconsistent product definitions –instantaneous and 6-hourly

Land sea mask (sfc), step 48, 20071107

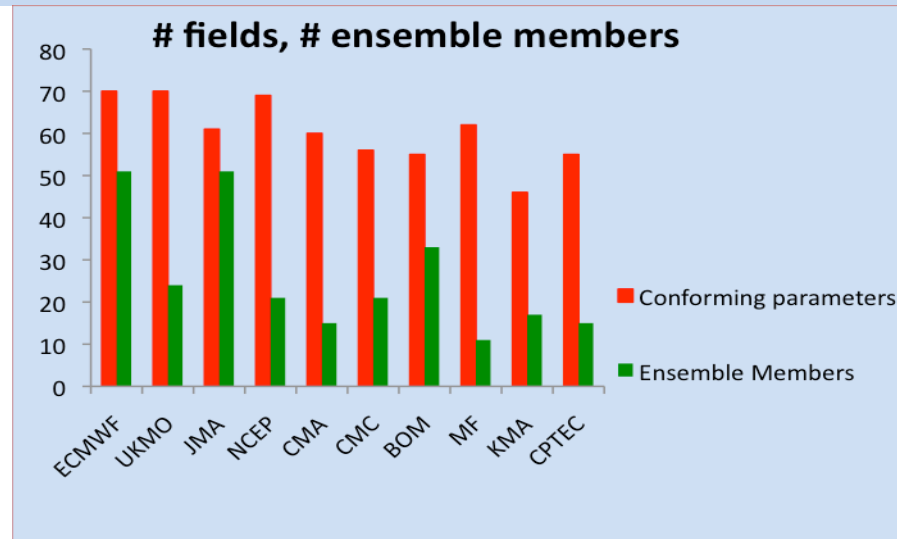
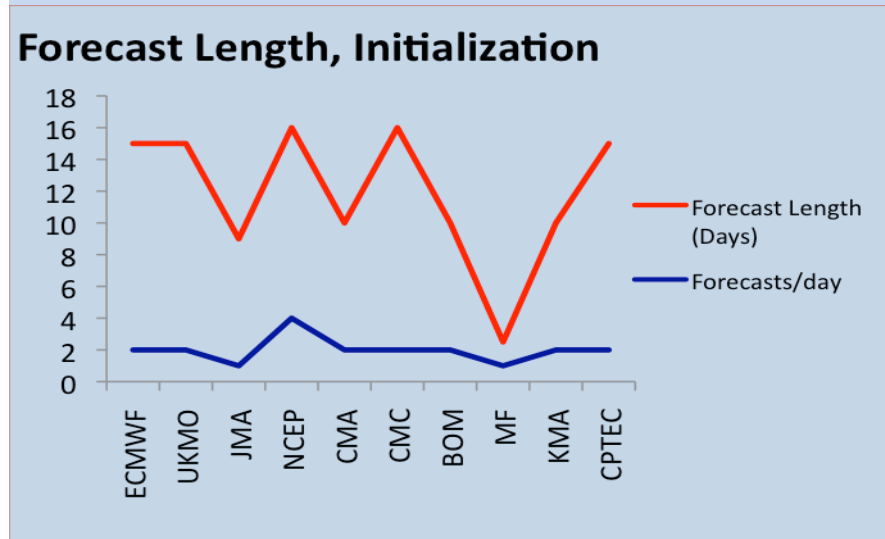
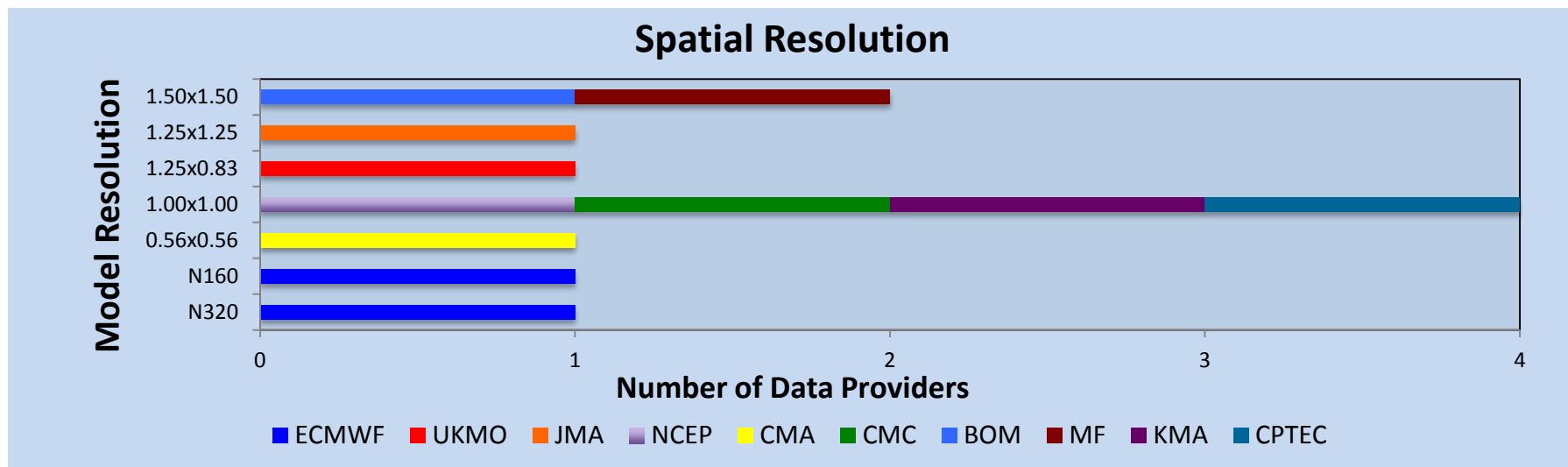


Time integrated surface net solar radiation (sfc), step 96, 20070608



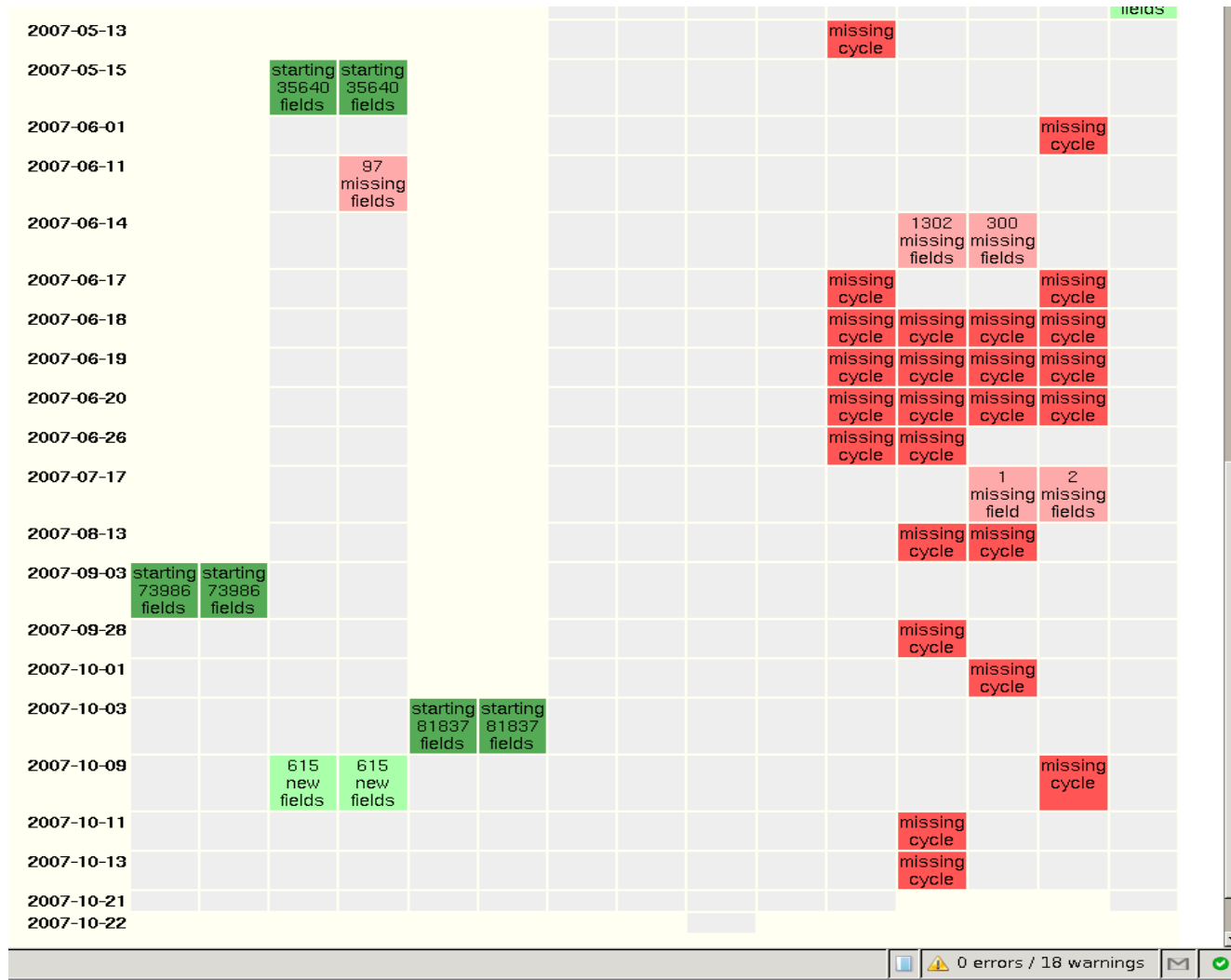
TIGGE Project Design: Resulting Operational Implementation

- No NWP center produces all agreed upon parameters
- Models typically interpolated to lower resolution grids



TIGGE Project Design: Resulting Operational Implementation

- Incomplete time-series issues



Access Options: User Data Access @ NCAR

THORPEX Interactive Grand Global Ensemble
TIGGE Data Archive Portal
 National Center for Atmospheric Research
 Computational and Information Systems Laboratory



Home | Get Forecast Data | Get Model Validation Data | Tools | Help | Documentation | Logout

Data Subset Selection

TIGGE Portal for Sub-setting

Select Centers

Australia Bureau of Met.	China Met. Administration	Met. Service of Canada	ECMWF	United Kingdom Met. Office	USA NCEP	Met. France	Japan Met. Agency	Korea Met. Administration	Brazil CPTEC
<input type="checkbox"/> AMMC	<input type="checkbox"/> BABJ	<input type="checkbox"/> CWA0	<input checked="" type="checkbox"/> ECMF	<input type="checkbox"/> EGRR	<input checked="" type="checkbox"/> KWBC	<input type="checkbox"/> LFPW	<input checked="" type="checkbox"/> RJTD	<input type="checkbox"/> RKSL	<input type="checkbox"/> SBSJ

Output File Format

NetCDF

Retrieve Matching Model Validation Data:

Check the box above to retrieve observational data for model validation. This will be processed as a separate data request from the NCAR RDA data portal and will match the available parameters chosen in your TIGGE data request.

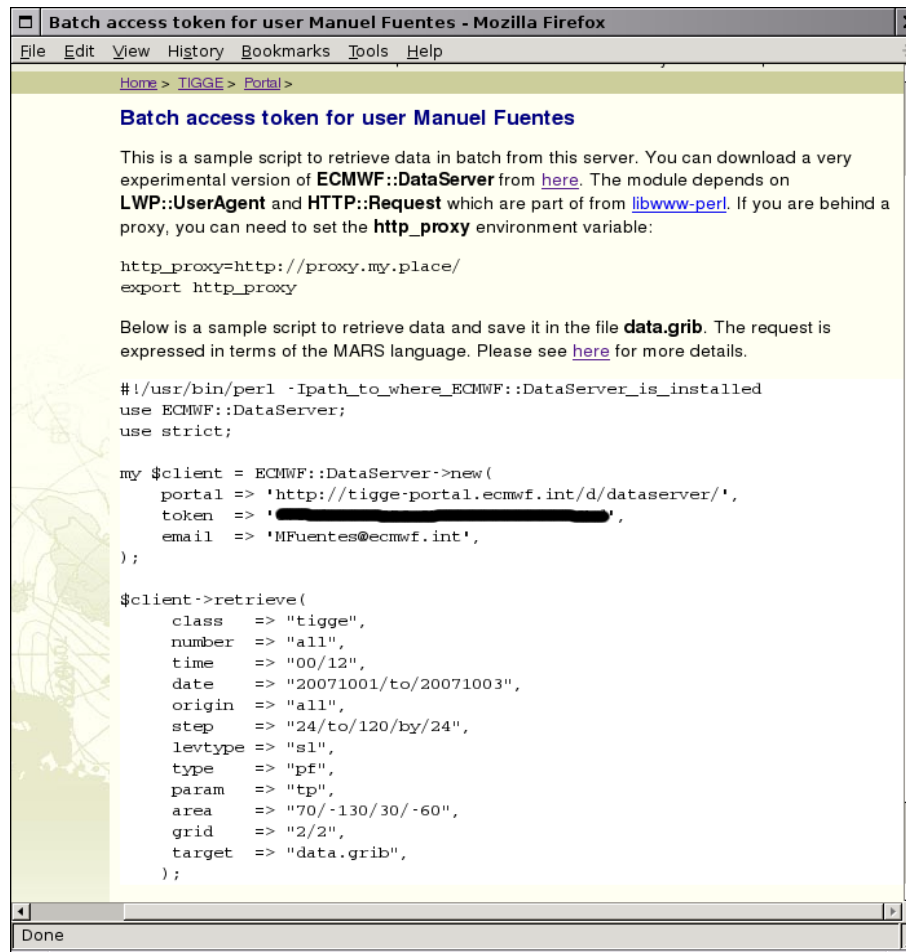
To make changes, use your browser back button to return to previous page. OR,

Code	Parameter Name	Units	SL	Value	Frequency	Time
<input type="checkbox"/>	10v	10_meter_v_velocity	m s ⁻¹	SL		
<input type="checkbox"/>	2d	surface_air_dew_point_temperature	K	SL		
<input type="checkbox"/>	2t	surface_air_temperature	K	SL		
<input type="checkbox"/>	ism	land_sea_mask	0 or 1	SL		
<input type="checkbox"/>	mn2t6	surface_air_minimum_temperature	K	SL		
<input type="checkbox"/>	msl	mean_sea_level_pressure	Pa	SL		
<input type="checkbox"/>	mx2t6	surface_air_maximum_temperature	K	SL		
<input type="checkbox"/>	orog	orography	gpm	SL		
<input type="checkbox"/>	sd	snow_depth_water_equivalent	kg m ⁻²	SL		
<input type="checkbox"/>	skt	skin_temperature	K	SL		
<input type="checkbox"/>	slhf	time_integrated_surface_net_solar_radiation	W m ⁻² s	SL		
<input type="checkbox"/>	sm	soil_moisture	kg m ⁻³	SL		
<input type="checkbox"/>	sp	surface_pressure	Pa	SL		
<input type="checkbox"/>	sshf	time_integrated_surface_sensible_heat_flux	W m ⁻² s	SL		
<input type="checkbox"/>	ssr	time_integrated_surface_net_solar_radiation	W m ⁻² s	SL		
<input type="checkbox"/>	str	time_integrated_surface_net_thermal_radiation	W m ⁻² s	SL		
<input type="checkbox"/>	tcw	total_column_water	kg m ⁻²	SL		
<input checked="" type="checkbox"/>	tp	total_precipitation	kg m ⁻²	SL	255	6 12 18 24 hrs
<input type="checkbox"/>	ttr	time_integrated_outgoing_long_wave_radiation	W m ⁻² s	SL		



Access Options: User Data Access @ ECMWF

- ECMWF provides similar access options to the complete archive + batch access capability.



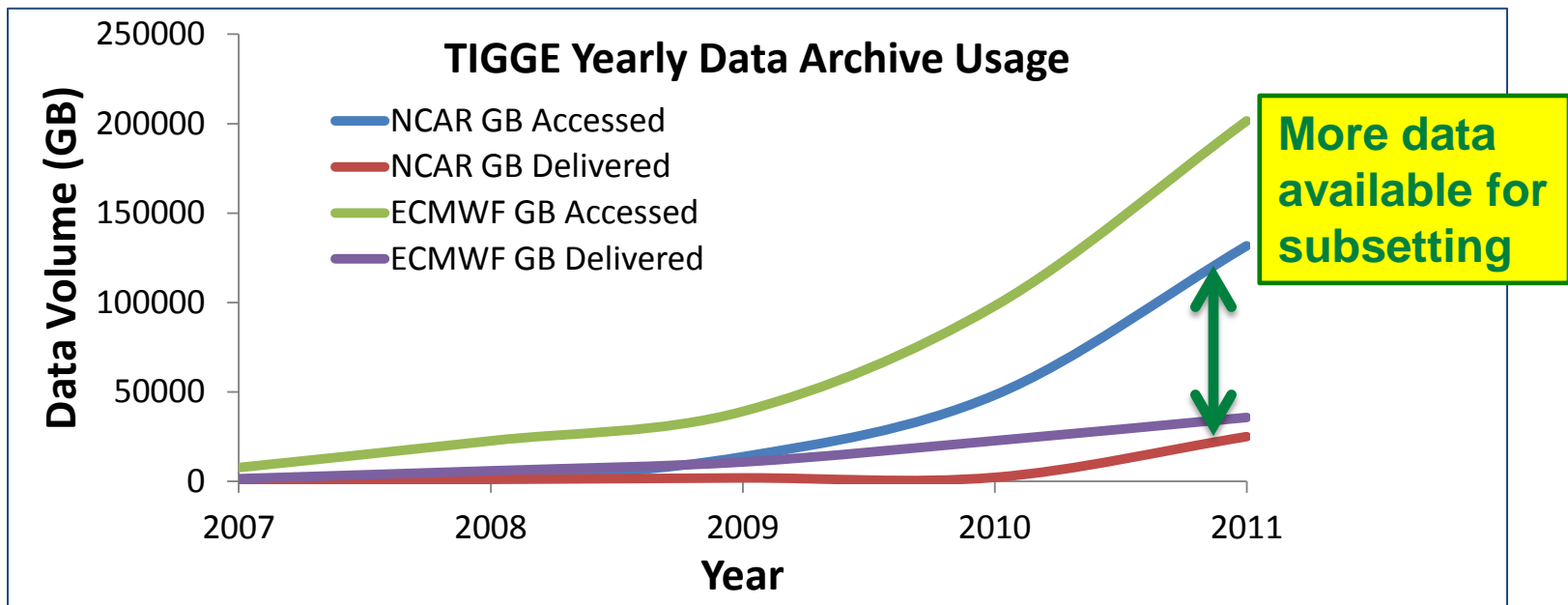
The screenshot shows a Mozilla Firefox browser window with the title "Batch access token for user Manuel Fuentes". The page content includes a heading "Batch access token for user Manuel Fuentes" and a paragraph explaining that it is a sample script to retrieve data in batch. It mentions the use of the `ECMWF::DataServer` module, which depends on `LWP::UserAgent` and `HTTP::Request`. It also provides instructions on how to set the `http_proxy` environment variable if behind a proxy. Below this, it shows a sample script to retrieve data and save it in a file named `data.grib`. The script is written in Perl and uses the `ECMWF::DataServer` module. The script includes the following code:

```
#!/usr/bin/perl -Ipath_to_where_ECMWF::DataServer_is_installed
use ECMWF::DataServer;
use strict;

my $client = ECMWF::DataServer->new(
    portal => 'http://tigge-portal.ecmwf.int/d/dataserver/',
    token => '██████████',
    email => 'MFuentes@ecmwf.int',
);

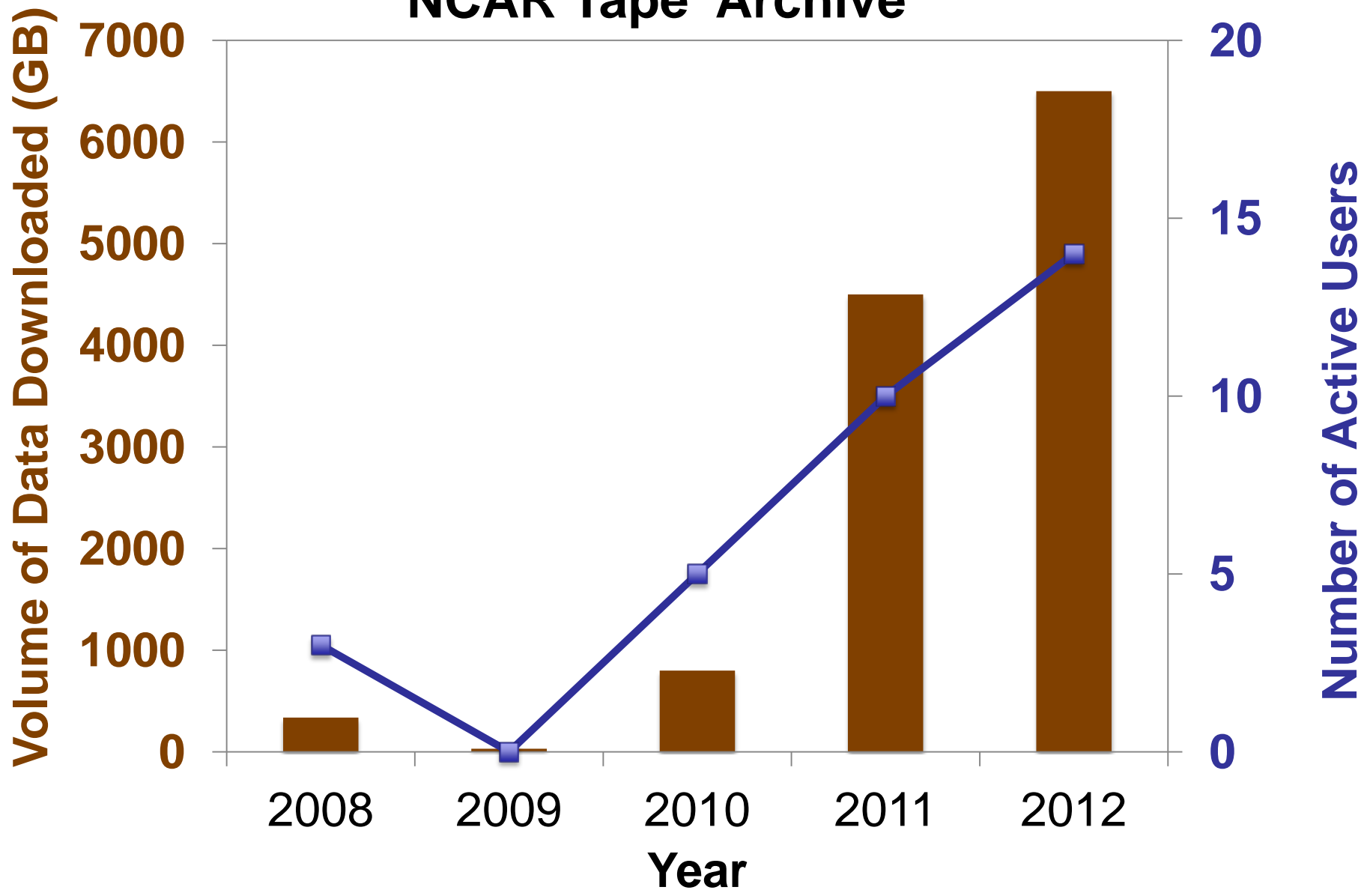
$client->retrieve(
    class => "tigge",
    number => "all",
    time => "00/12",
    date => "20071001/to/20071003",
    origin => "all",
    step => "24/to/120/by/24",
    levtype => "sl",
    type => "pf",
    param => "tp",
    area => "70/-130/30/-60",
    grid => "2/2",
    target => "data.grib",
);
```

Access Options: TIGGE Archive Usage



YEAR	NCAR GB Accessed Delivered		ECMWF GB Accessed Delivered		NCAR Active Users ECMWF Active Users	
	Accessed	Delivered	Accessed	Delivered	NCAR Active Users	ECMWF Active Users
2007	900	900	7600	1,500	10	10
2008	1,300	1,100	22,600	5,900	10	20
2009	13,700	1,900	39,000	10,800	10	35
2010	48,200	2,200	98,000	22,700	12	45
2011	131,700	25,000	201,500	35,800	15	70

Access Options: Data Downloaded from NCAR Tape Archive



Reflection: Scope

- What are the research objectives?
- What is **REALLY** needed to support those objective?
 - Is the project sustainable?
 - Can the project's support system scale as the project requires?
 - Is mission creep allowed?
 - Are all parties **completely** committed?
- Is the user community large enough, sufficiently aware, and well enough prepared to justify the storage and support costs?

Reflection: Science Partner Commitment

- Are all partners committed to implementing and supporting the agreed upon protocols, standards, etc..?
 - Data Transmission, Data Format, Format Conventions
 - Parameter definitions, Parameter Encoding
 - Science partners need to negotiate carefully the agreed set of parameter fields. There is a tendency to let this grow too large (many fields in TIGGE are infrequently or unused).
- Avoid 1-off solutions
 - **Very resource intensive!!!**



Reflection: Archive Structure

- What are the user/research requirements?
- What services will fulfill those requirements?
- What is the most efficient structure to support desired services?
 - TIGGE –Centralized Archive
 - Would distributed work better for this scale of archive (850 TB, 7 Million+ files? Use of common Service Oriented Architecture, and data brokering software?
 - TIGGE Archive @ NCAR organized in file groups by provider and forecast initialization time
 - Pressure Level, Θ Level, PV Level, Single/Surface Level
 - Contrast: TIGGE @ ECMWF is a one-off quasi-DB structure leveraging long-standing operational services

Conclusions

- Scope –Stay focused on achieving research objectives
 - Are support systems scalable to meet future needs?
- Project Design –**All partners must agree upon and support common standards and apply them**
 - Ensures consistent parameter definitions/encoding
 - Supports completeness and interoperability across tools
- Archive Structure –Find an archive structure that best supports user research needs and is scalable
 - Data access options and services are impacted by choice of archive structure

Questions?

schuster@ucar.edu

International TIGGE Archive Centers

ECMWF	http://tigge.ecmwf.int
NCAR	http://tigge.ucar.edu
CMA	http://wisportal.cma.gov.cn/tigge