

# Cyberinfrastructure

8<sup>th</sup> Team Meeting  
La Jolla, CA  
January, 2010

John Helly

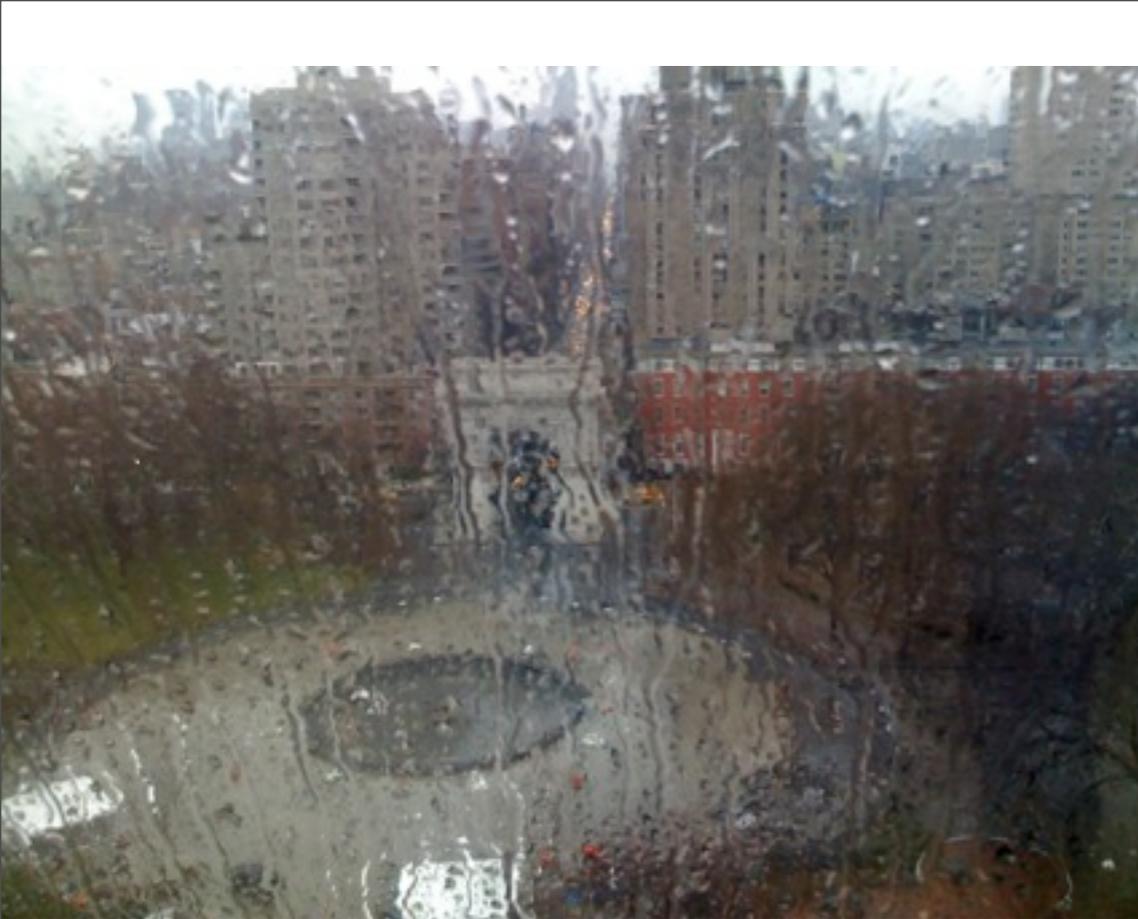


# Outline

- Management
  - CIWVG (Cyberinfrastructure Working Group)
  - Data Policy
  - Software Policy
- Resources
  - Computing
  - Data
- Community Support
  - Data interoperability
  - Model code portability

# Objective 6

- Management, analysis and visualization of very large model output datasets - creation of infrastructure
- Resource procurement & distribution
- Data management
- Analysis and visualization of model results



# Management CIWG

# CIWG Objectives

- Make efficient use of computing and data resources
  - acquire resources
  - coordinate resource utilization
  - collaborate to leverage joint efforts
- Provide technology look-ahead
- Validate goals and provide advice and consent to Executive Committee

http://cmmmap.sdsc.edu/

# CMMAP Digital Library

Active forum topics

- Allocation Proposal Schedule (as of: 2008-07)
- Computing Time
- Proposals
- Fortran90
- Data

hellyj

- Create content
- Login
- Search Data Catalogue
- My account
- Recent posts
- Administer
- Log out

Allocation Proposal Schedule (as of: 2008-07)

Tue, 07/29/2008 - 2:46pm — hellyj

Latest summary schedule

Add new comment 1 attachment CIWG

2008 Annual Meeting

Tue, 07/29/2008 - 1:02pm — hellyj

First blog entry

hellyj's blog 2 comments

Developer Tools

- Migrating code to Future NSF resources (<http://www.sdsc.edu/us/consulting/migration.html>)
- Script automation (<http://www.ibm.com/developerworks/aix/library/au-satbuildscript.html>)
- Compiler quirks (<http://en.wikipedia.org/wiki/Compiler>)

Read more

Recent blog posts

- 2008 Annual Meeting

Recent comments

- Model Taxonomy has been updated 2 min 7 sec ago
- ECMWF Multi-grid modeling idea 2 hours 44 min ago

Who's new

- mikepritchard
- kelly
- mbranson
- leoners56
- jasonc

Who's online

RSS

CMMAP Digital Library - Reach for the sky - RSS (8 messages, 3 unread)

Date Received	From	Subject
Today 2:46 PM	CMMAP Digital Library - Reach for the sky	Allocation Proposal Schedule (as of: 2008-07)
Today 1:02 PM	CMMAP Digital Library - Reach for the sky	2008 Annual Meeting
Today 10:18 AM	CMMAP Digital Library - Reach for the sky	Developer Tools
Today 10:01 AM	CMMAP Digital Library - Reach for the sky	Visualization Resources
Today 9:54 AM	CMMAP Digital Library - Reach for the sky	Computing Allocation Resources
July 26, 2008 8:57 PM	CMMAP Digital Library - Reach for the sky	Example of Creating HTML from External Editor
July 26, 2008 7:52 PM	CMMAP Digital Library - Reach for the sky	How to Create Drupal Content
February 11, 2008 6:51 PM	CMMAP Digital Library - Reach for the sky	Preparing for Track 2 NSF Systems

## CMMAP Digital Library - Reach for the sky

### Developer Tools

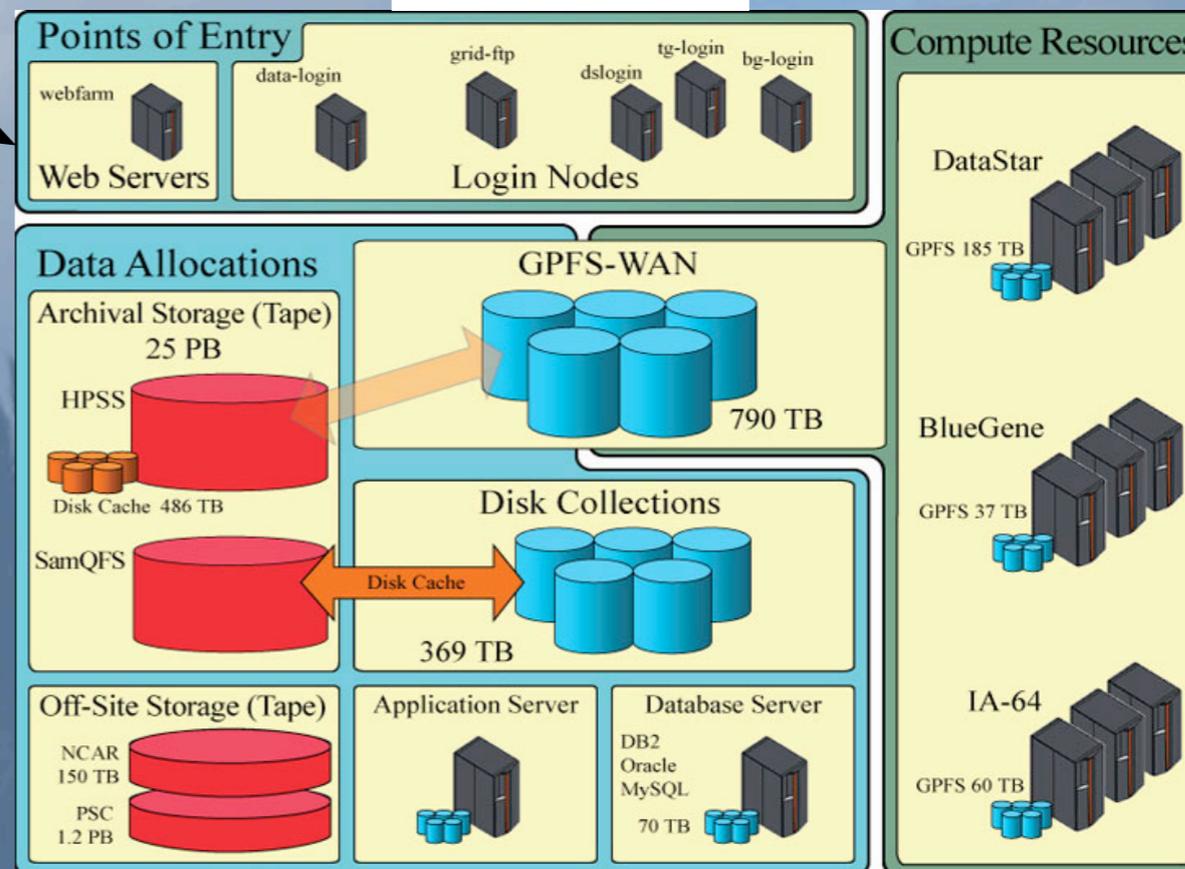
July 29, 2008 10:18:40 AM PDT

- Migrating code to Future NSF resources (<http://www.sdsc.edu/us/consulting/migration.html>)
- Script automation (<http://www.ibm.com/developerworks/aix/library/au-satbuildscript.html>)
- Compiler quirks (<http://en.wikipedia.org/wiki/Compiler>)

read more

Read more...

**SDSC**  
SAN DIEGO SUPERCOMPUTER CENTER



URL data retrieval

OAI metadata

200 GB/snapshot = 1/2 hr

# MMF-LES Model Data Output Budget

1 RUN = 24 HRS

200 GB  
1/2 HR  
3D

slightly avg  
3D flow  
fields

static  
history Q<sub>2</sub>  
Q<sub>3</sub>

2D  
slice

16 · 10<sup>9</sup>  
≈ 2 · 10<sup>10</sup>

4 × 4 × 2 = 32

4500 × 4500  
× 256

LES  
DEEP

CONVERT  
3D

400 GB/hour - 3D

400 GB/hour - degraded 3D

10 GB/hour - slices

810 GB/hour - all

x 24  
324  
162  
19,440

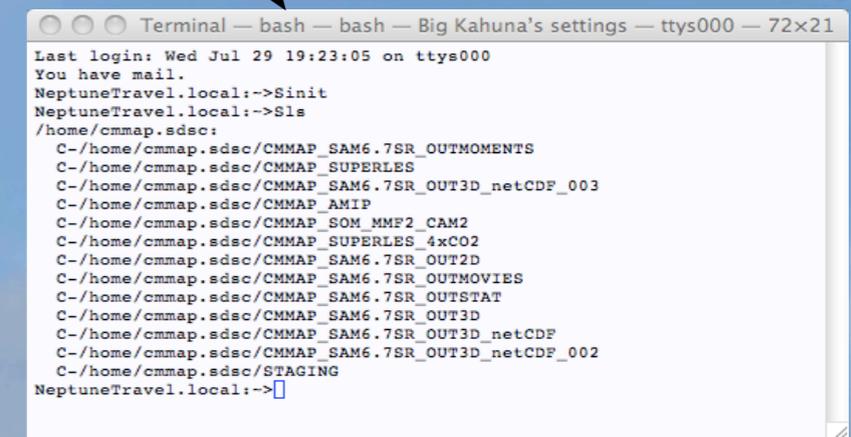
20 TB/day

# Simple Methods of Data Download (2) and Upload (1)



Selective Data Transfer

Bulk Data Transfer SRB Client



Storage Resource Broker (SRB)

- SAM-QFS has 16 tape drives, an 304-TB disk cache
- Meta Data Servers: 2 x Sun x4600
- Data Login Servers: 2 x Sun v40z /
- 6 x 1GbE connections to the TeraGrid/HPC network
- Tape Drives
- 16 STK 9940-B tape drives
- 24 IBM J2 3592 tape drives

# The GCRM Tsunami

## 2 km, 100 levels, hourly data

- ~4 TB / simulated hour
- ~100 TB / simulated day
- ~35 PB / simulated year

## 4 km, 100 levels, hourly data

- ~1 TB / simulated hour
- ~24 TB / simulated day
- ~9 PB / simulated year

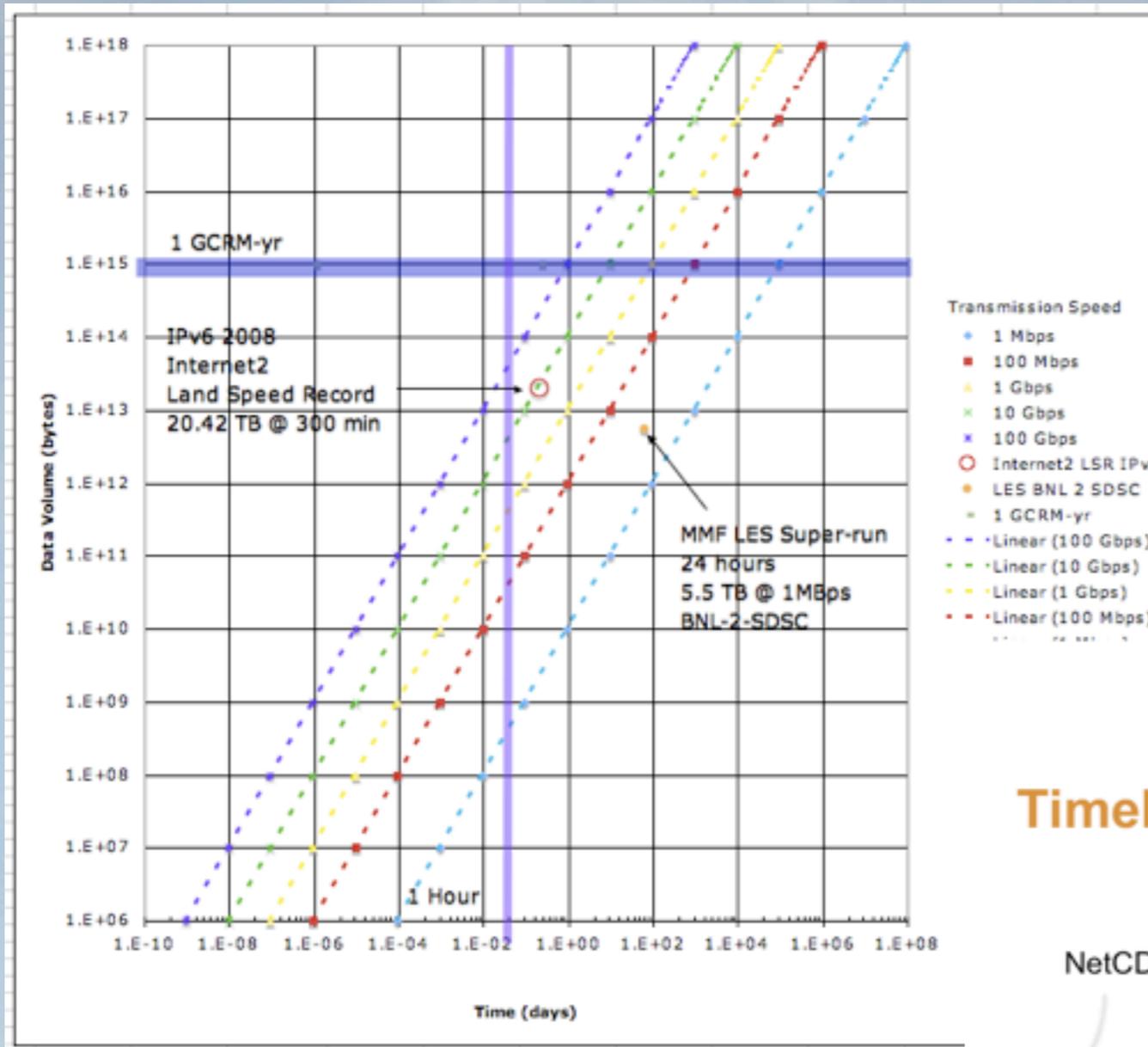
## Other Data Tsunamis

- 30 TB/night: Large Synoptic Survey (LSS) Telescope (2014)
- 15 PB/year: CERN's Large Haydron Collider ( May 2008)
- 1 PB over 3 years: EOS (Earth Observing System) data (2001)



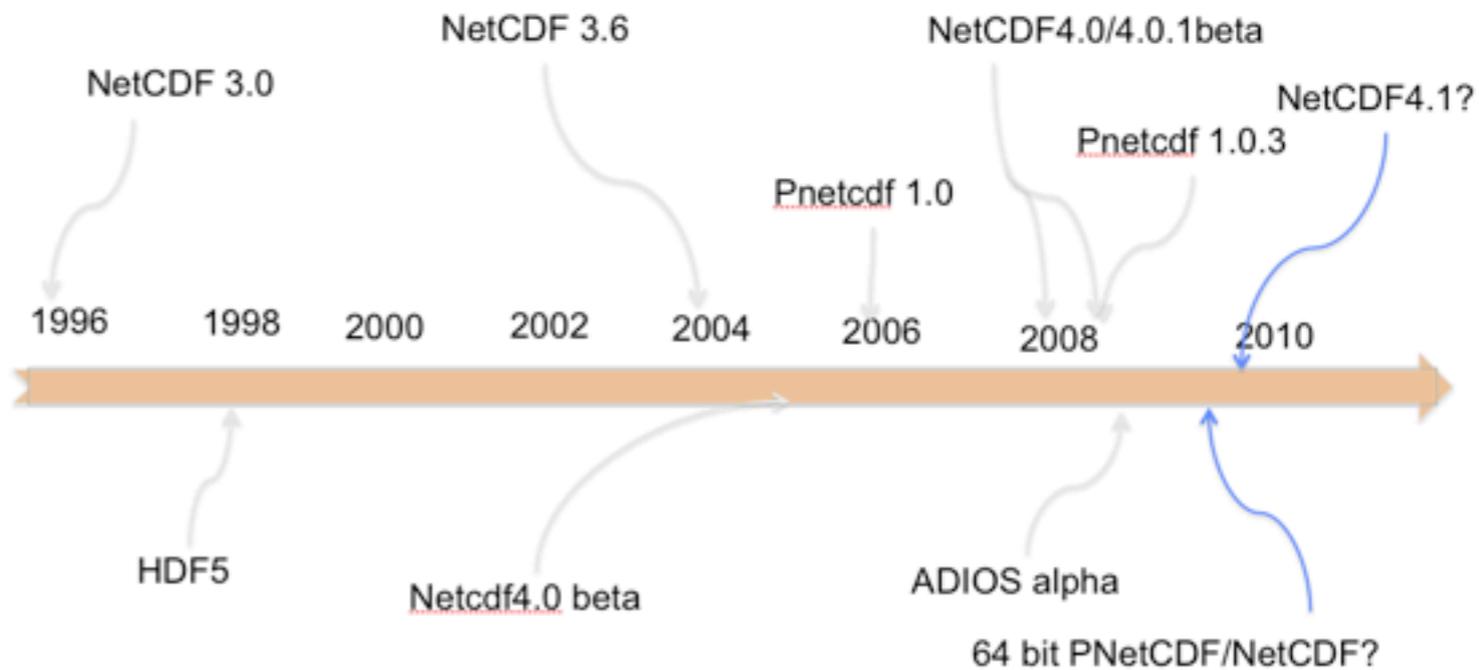
(from K. Schuchart presentation January, 2008)

# Internet Data Transfer Capacity

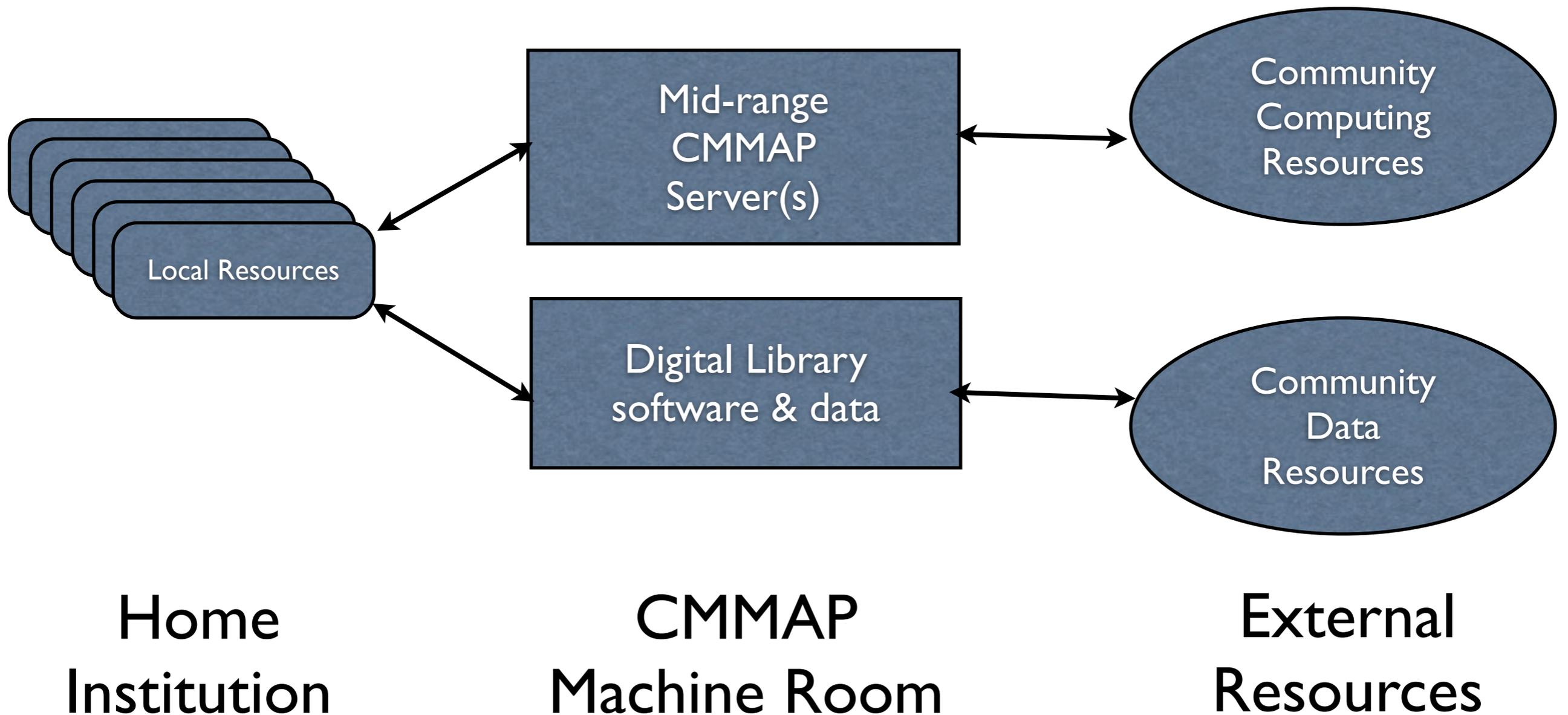


# Parallel Input/Output Technology Progress

## Timeline



# Architectural Goals





# Management Data Policy



# Data Policy

- Data published on the CMMAP Digital Library is in the public domain but registration and authorization required to access it
  - this is to prevent hacking and bot-crawling and
  - provide tracking of who is accessing the data
- All metadata is public
  - will be published via a new OAI (Open Archive Initiative) service to be instituted this year
  - CF metadata conventions are followed
- We are investigating the use of DOIs (digital object identifiers) for data consistent with the scholarly publication process



# CMMAP Digital Library



## Computing Resources

- [NSF Teragrid](#)
- [DOE INCITE](#)

## Data Resources

- [Search for Data](#)

## Software

- [Bulk Data Transfer Client](#)
- [Multi-scale Modeling Framework \(MMF\)](#)

## hellyj

- [My blog](#)
- ▷ [Create content](#)
- [Search for Data](#)
- [My account](#)

## INCITE Resources & Allocations

Mon, 05/18/2009 - 23:25 — hellyj

### 2010 INCITE Call for Proposals

[Add new comment](#) [Read more](#)

[Computational Resources](#)

## Teragrid Resources & Allocations

Mon, 05/18/2009 - 23:14 — hellyj

[Add new comment](#) [Read more](#)

[Computational Resources](#)

## Sample Fortran program to read GIGALES netcdf 3D snapshot data

This program will read one variables for a subset of the horizontal domain for all vertical levels of a given 3D snapshot time.

1 attachment

## Data Conversion Tools

Thu, 11/27/2008 - 20:54 — hellyj



# Management Software Policy



# Software Policy

- Software developed with CMMAP funding is covered by an intellectual property agreement between CSU and CMMAP partners
- MMF code is controlled by the author of the code
- GCRM code is very new and developed under SCIDAC funding
- Subversion source control system has been provisioned but has not yet been required
- CCSM coding conventions are encouraged

## Community Sea Ice Model (CSIM) Developer's Guide Code Reference for Version 5.0

Released with CCSM3.0

Julie Schramm  
Cecilia Bitz  
Bruce Briegleb  
Marika Holland  
Elizabeth Hunke  
Bill Lipscomb  
Dick Moritz

ice\_refdoc

Next Up Previous Contents

Next: [Contents](#) [Contents](#)

---

5.1 Style Guidelines

5.1 Style Guidelines

Next Up Previous Contents

Next: [5.2 Content Guidelines](#) Up: [5 Coding Standard](#) Previous: [5 Coding Standard](#) [Contents](#)

Subsections

- [General Guidelines](#)
- [Modules](#)
- [Subroutines](#)
- [Loops](#)
- [Array Syntax](#)
- [Allocatable Arrays](#)
- [Variable Names](#)
- [Variable Declarations](#)
- [Code indentation](#)
- [Commenting of code](#)
- [Portability](#)
- [Incomplete and dead code](#)
- [Miscellaneous](#)

---

### 5.1 Style Guidelines

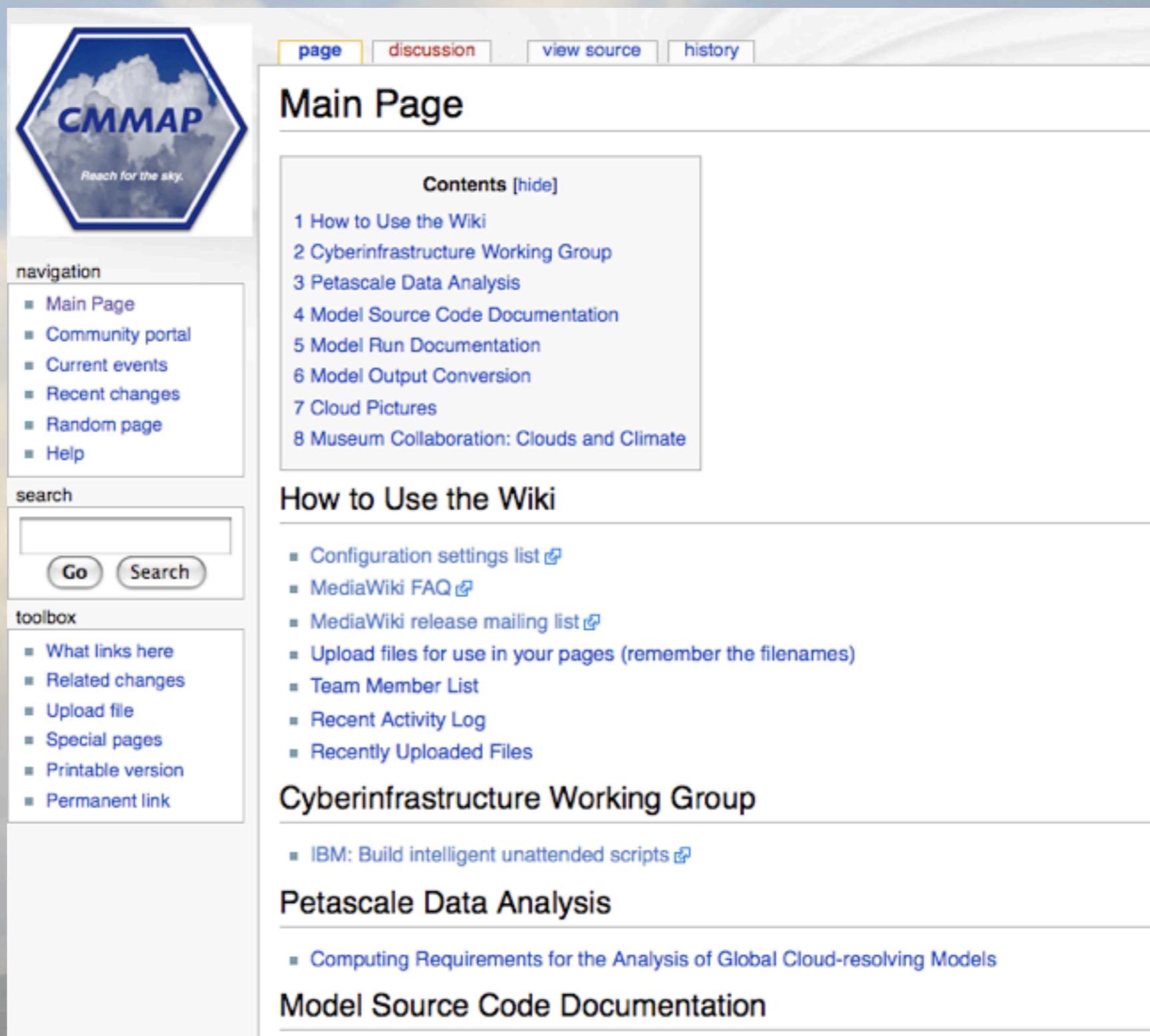
#### General Guidelines

- **Fortran 90 Standard** CSIM uses the F90 standard.
- **Preprocessor** The C preprocessor is required as the macro processor (cpp). All tokens should be uppercase to distinguish them from fortran code. The use of ifdef constructs should be avoided wherever possible. The option should be set in the namelist instead.
- **Fixed-Form Source** Fixed-form source will be used for readability. Columns 7 to 72 of a line may contain a Fortran statement, while columns 1 to 6 are reserved for special purposes. Blanks may be used freely anywhere. Exclamation marks should be used to denote comments. A statement may include a maximum of 19 continuation lines.
- **Bounds Checking** All code should be able to run when the following compiler options are set:
  - Array bounds checking
  - Automatic variables are initialized to NaN
  - Floating point exception conditions: overflow, division by zero, and invalid operations

#### Modules

- The use of modules is strongly encouraged since they provide global access to variables.
- Modules must have the same name as the file in which they reside, due to dependency rules used by "make" programs. This implies that multiple modules are not allowed in a single file.
- Module names must begin with "ice\_". This will provide unique module names when CCSM is released as a single executable version.
- Related subroutines and variables should be placed in a single module/file rather than keeping a single routine per file.

# Documentation



The screenshot shows the main page of the CMMAP Wiki. At the top left is the CMMAP logo, a blue hexagon with a cloud image and the text 'CMMAP' and 'Reach for the sky.'. Below the logo is a navigation menu with links to Main Page, Community portal, Current events, Recent changes, Random page, and Help. To the right of the navigation menu is a search box with 'Go' and 'Search' buttons. Below the search box is a toolbox with links to What links here, Related changes, Upload file, Special pages, Printable version, and Permanent link. The main content area has tabs for 'page', 'discussion', 'view source', and 'history'. The main heading is 'Main Page'. Below it is a 'Contents [hide]' section with a numbered list of 8 items: 1 How to Use the Wiki, 2 Cyberinfrastructure Working Group, 3 Petascale Data Analysis, 4 Model Source Code Documentation, 5 Model Run Documentation, 6 Model Output Conversion, 7 Cloud Pictures, and 8 Museum Collaboration: Clouds and Climate. Below the contents is a section titled 'How to Use the Wiki' with a list of links: Configuration settings list, MediaWiki FAQ, MediaWiki release mailing list, Upload files for use in your pages (remember the filenames), Team Member List, Recent Activity Log, and Recently Uploaded Files. Below that is a section titled 'Cyberinfrastructure Working Group' with a link to IBM: Build intelligent unattended scripts. Below that is a section titled 'Petascale Data Analysis' with a link to Computing Requirements for the Analysis of Global Cloud-resolving Models. The final section is titled 'Model Source Code Documentation'.



# Resources

# Computing



# NSF Teragrid

	Startup/Educational Allocation	Research Allocation (TRAC)	
Units Requested <i>Service Units (SUs) on Compute Resources</i> <i>Terabytes (TBs) on Data Resources</i>	Maximum compute request <ul style="list-style-type: none"> <li>Most systems have a limit of 30,000 SUs, though some may be as high as 200,000 SUs. Please see the Resource Catalog for specific resource limits.</li> <li>Maximum Startup request for TeraGrid Roaming: 50,000 SUs</li> <li>Aggregate request for multiple compute resources cannot exceed 200,000 SUs.</li> </ul>	30,000 – Unlimited	
	<ul style="list-style-type: none"> <li>Storage on disk: 5 TB</li> <li>Storage on tape: 25 TB</li> </ul>		
Deadlines	N/A	Open Submissions	Close Submissions
		Dec. 15 Mar. 15 <sup>1</sup> Jun. 15 Sept. 15	Jan. 15 <sup>1</sup> Apr. 15 Jul. 15 Oct. 15
Allocations Begin	Two weeks after submission	April 1 July 1 October 1 January 1	
Review Cycle	Within one week	Quarterly <sup>2</sup>	
Typical Use	Classroom or training accounts and startup accounts requiring small amounts of time	Experienced users with research projects	

<sup>1</sup> The TRAC Meeting schedule is available for reviewers on the [POPS Review page](#).

<sup>2</sup> See the [Past Allocations](#) for a list of previous awards.

# DOE Earth System Grid

**NCCS.GOV**  
NATIONAL CENTER FOR COMPUTATIONAL SCIENCES

Home | About | Leadership Science | Computing Resources | User Support | Media Center

Home » User Support » Access & Accounts

### Access & Accounts

Resources of the NCCS are dedicated to a limited number of computationally intensive research projects. Ninety-five percent of the Oak Ridge Leadership Computing Facility (OLCF) resources are dedicated to these projects.

#### Types of Projects

We currently offer two types of projects for which you can request a new account or join an existing project. INCITE projects and Director's Discretion projects.

Details	INCITE	Director's Discretion
<b>Duration</b>	1 year	1 year
<b>Allocations</b>	Large	Small
<b>Call for Proposals</b>	Once per year	At any time
<b>Priority</b>	High	Below INCITE
<b>Quarterly Reports</b>	✓	
<b>Closeout Report</b>	✓	✓

Apply      Apply

# Leveraging National & Partner Resources

	Organization	Resource	2007	2008	2009	2010
Data Allocations	San Diego Supercomputer Center (SDSC)	Disk	15 Terabytes	15 Terabytes	30 Terabytes	45 Terabytes
		BlueGene			30,000 SUs*	
Computing Allocations	Teragrid (multi-institution)	SDSC DataStar (IBM SP4)	600,000 SUs	1,200,000 SUs		
		Grid Roaming			600,000 SUs	
		LSU Steele			900,000 SUs	
	Lawrence Berkeley National Laboratory (LBNL)	National Energy Research Scientific Computing Center (NERSC)			700,000 SUs	
	Oak Ridge National Laboratory (ORNL)	Cray XT			2,000,000 hrs	3,000,000 hrs
	National Center for Atmospheric research (NCAR)	Bluelce IBM Power5			500,000 SUs	
	IBM Watson Research Center	BGW - eServer Blue Gene Solution			TBD	
	Stonybrook				TBD	



# Resources

# Data





# CMMAP Digital Library

Edit primary

Home

## Search for Data

Search

### Computing Resources

- NSF Teragrid
- DOE INCITE

### Data Resources

- Search for Data

### Software

- Bulk Data Transfer Client
- Multi-scale Modeling Framework (MMF)

### hellyj

- My blog
- Create content
- Search for Data
- My account
- Recent posts
- News aggregator
- Administer
- Log out

#### Search Parameters

- CMMAP\_AMIP
- CMMAP\_SAM6.7SR\_OUTMOMENTS
- CMMAP\_SAM6.7SR\_OUTMOVIES
- CMMAP\_SAM6.7SR\_OUTSTATa
- CMMAP\_SAM6.7SR\_OUT2D
- CMMAP\_SAM6.7SR\_OUT3D
- CMMAP\_SAM6.7SR\_OUT3D\_netCDF
- CMMAP\_SUPERLES
- CMMAP\_SUPERLES\_4xCO2

#### Navigation

North:

South:

East:

Wast:

Submit Query

Filename	Collection	Subject	Keywords	Description	Type
<input type="checkbox"/> GATE_IDEAL_S_2048x2048x256_100m_2s_cldtop.raw	metadata		CMMAP_SAM6.7SR_OUTMOVIES	MMF Output for the SUPERLES project	
<input type="checkbox"/> GATE_IDEAL_S_2048x2048x256_100m_2s_cwp.raw	metadata		CMMAP_SAM6.7SR_OUTMOVIES	MMF Output for the SUPERLES project	
<input type="checkbox"/> GATE_IDEAL_S_2048x2048x256_100m_2s_lwp.raw	metadata		CMMAP_SAM6.7SR_OUTMOVIES	MMF Output for the SUPERLES project	
<input type="checkbox"/> GATE_IDEAL_S_2048x2048x256_100m_2s_mse.raw	metadata		CMMAP_SAM6.7SR_OUTMOVIES	MMF Output for the SUPERLES project	
<input type="checkbox"/> GATE_IDEAL_S_2048x2048x256_100m_2s_qvsfc.raw	metadata		CMMAP_SAM6.7SR_OUTMOVIES	MMF Output for the SUPERLES project	
<input type="checkbox"/> GATE_IDEAL_S_2048x2048x256_100m_2s_sfcprec.raw	metadata		CMMAP_SAM6.7SR_OUTMOVIES	MMF Output for the SUPERLES project	

# Conversion of SAM6.7SR com3D to netCDF

```
jitterbug:/archive/science/cmmmap-staging>du -kh
```

```
114G ./staging/CMMAP_AMIP
91G  ./staging/CMMAP_SUPERLES_4xCO2
92G  ./staging/CMMAP_SUPERLES
14G  ./staging/CMMAP_SAM6.7SR_OUT2D
3.7T ./staging/CMMAP_SAM6.7SR_OUT3D
113G ./staging/CMMAP_SAM6.7SR_OUTMOMENTS
51G  ./staging/CMMAP_SAM6.7SR_OUTMOVIES
137M ./staging/CMMAP_SAM6.7SR_OUTSTAT
4.2T ./staging
```

```
781G ./source/PRITCHARD
114G ./source/CMMAP_AMIP
92G  ./source/CMMAP_SUPERLES
91G  ./source/CMMAP_SUPERLES_4xCO2
14G  ./source/CMMAP_SAM6.7SR_OUT2D
113G ./source/CMMAP_SAM6.7SR_OUTMOMENTS
51G  ./source/CMMAP_SAM6.7SR_OUTMOVIES
137M ./source/CMMAP_SAM6.7SR_OUTSTAT
2.1T ./source/CMMAP_SAM6.7SR_OUT3D_com3D
3.1T ./source/CMMAP_SAM6.7SR_OUT3D_netCDF
6.3T ./source
11T  .
```

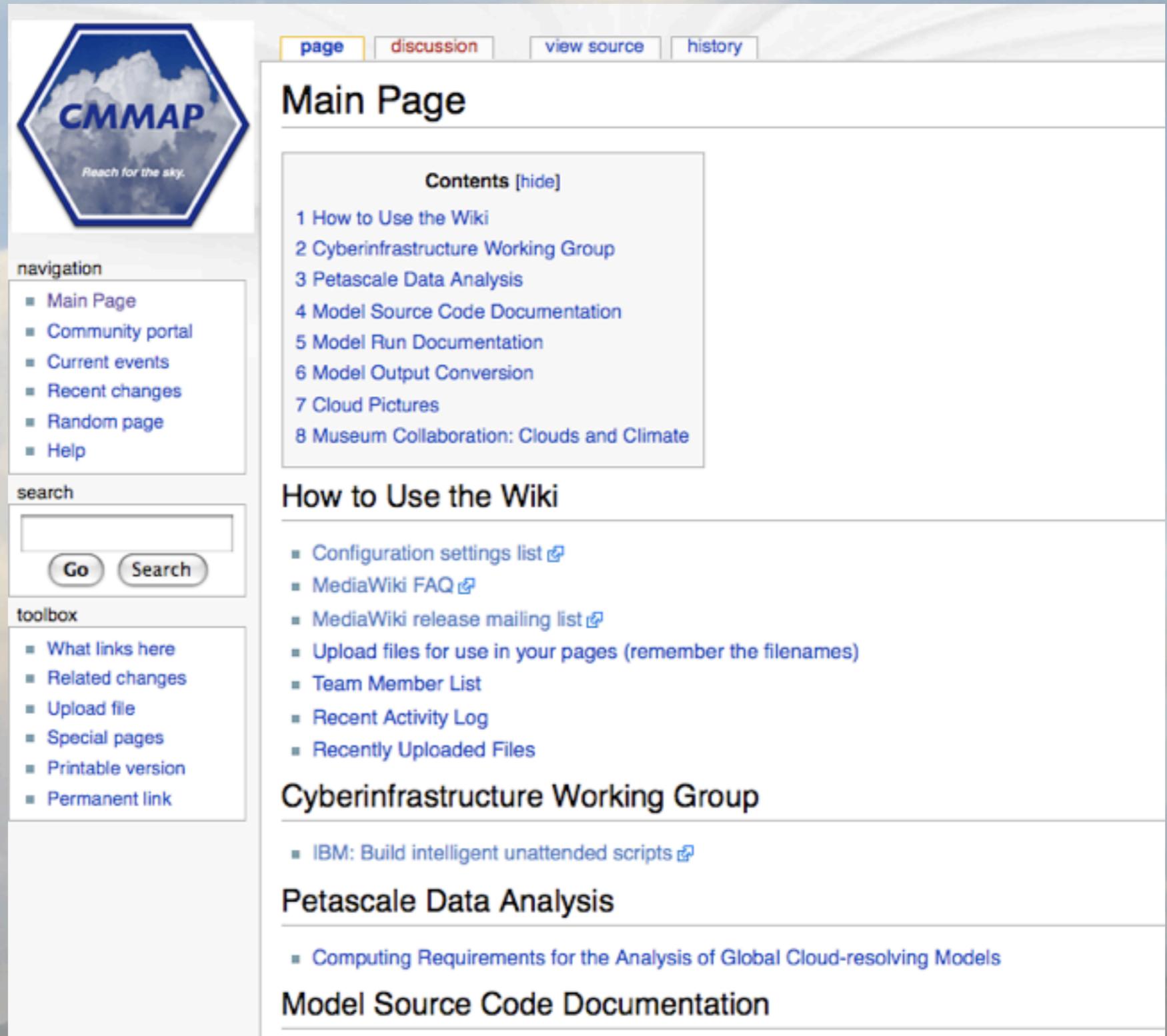
Published in  
Digital Library

Stored in  
Archival File  
System

\*AI: Rename SUPERLES to SPCAM

<http://cmmap.sdsc.edu/wiki/>

Same as  
Wikipedia



The screenshot shows the main page of the CMMAP Wiki. At the top left is the CMMAP logo, a blue hexagon with a cloud image and the text "CMMAP" and "Reach for the sky." Below the logo is a navigation menu with links to Main Page, Community portal, Current events, Recent changes, Random page, and Help. To the right of the navigation menu is a search box with "Go" and "Search" buttons. Below the search box is a toolbox with links to What links here, Related changes, Upload file, Special pages, Printable version, and Permanent link. The main content area has tabs for "page", "discussion", "view source", and "history". The title "Main Page" is followed by a "Contents [hide]" section with a numbered list of eight items: 1 How to Use the Wiki, 2 Cyberinfrastructure Working Group, 3 Petascale Data Analysis, 4 Model Source Code Documentation, 5 Model Run Documentation, 6 Model Output Conversion, 7 Cloud Pictures, and 8 Museum Collaboration: Clouds and Climate. Below this is a section titled "How to Use the Wiki" with a list of links: Configuration settings list, MediaWiki FAQ, MediaWiki release mailing list, Upload files for use in your pages (remember the filenames), Team Member List, Recent Activity Log, and Recently Uploaded Files. The next section is "Cyberinfrastructure Working Group" with a link to IBM: Build intelligent unattended scripts. This is followed by "Petascale Data Analysis" with a link to Computing Requirements for the Analysis of Global Cloud-resolving Models. The final section is "Model Source Code Documentation".

Community Support

# Data Interoperability





# CMMAP Digital Library



## Computing Resources

- [NSF Teragrid](#)
- [DOE INCITE](#)

## Data Resources

- [Search for Data](#)

## Software

- [Bulk Data Transfer Client](#)
- [Multi-scale Modeling Framework \(MMF\)](#)

## hellyj

- [My blog](#)
- ▷ [Create content](#)
- [Search for Data](#)
- [My account](#)

## INCITE Resources & Allocations

Mon, 05/18/2009 - 23:25 — hellyj

### 2010 INCITE Call for Proposals

[Add new comment](#) [Read more](#)

[Computational Resources](#)

## Teragrid Resources & Allocations

Mon, 05/18/2009 - 23:14 — hellyj

[Add new comment](#) [Read more](#)

[Computational Resources](#)

## Sample Fortran program to read GIGALES netcdf 3D snapshot data

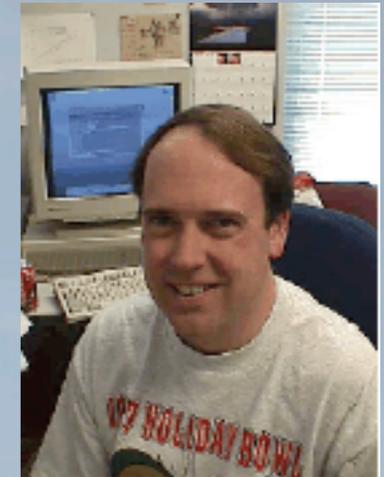
This program will read one variables for a subset of the horizontal domain for all vertical levels of a given 3D snapshot time.

1 attachment

## Data Conversion Tools

Thu, 11/27/2008 - 20:54 — hellyj

# Data Conversion (com3D->netCDF)



by Mark  
Branson/CSU

**CMMAP Digital Library**

Home

### SAM output conversion programs

[View](#) [Edit](#) [Outline](#)

Two of the programs that convert the 3D model output which is made up of compressed 2-byte integers to either binary floating point values (com3D2bin.f) or netcdf files (com3D2nc.f).

Attachment	Size
<a href="#">com3D2bin.f</a>	3.5 KB
<a href="#">com3D2nc.f</a>	5.5 KB

[Add new comment](#)

*powered by the San Diego Supercomputer Center and sponsored by the National Science Foundation*

**Active forum topics**

- What software should we use to produce the SUPERLES derivatives?
- We have identified netCDF as the default format for data products. Any discussion needed?
- data format conversion

[more](#)

**hellyj**

- My blog
- Create content
- Search for Data
- My account
- Recent posts
- News aggregator

**CMMAP Digital Library**

Home

### Sample Fortran program to read GIGALES netcdf 3D snapshot data

[View](#) [Edit](#) [Outline](#)

This program will read one variables for a subset of the horizontal domain for all vertical levels of a given 3D snapshot time.

Attachment	Size
<a href="#">read_gigales3d.f90</a>	2.01 KB

[Add new comment](#)

*powered by the San Diego Supercomputer Center and sponsored by the National Science Foundation*

**Active forum topics**

- What software should we use to produce the SUPERLES derivatives?
- We have identified netCDF as the default format for data products. Any discussion needed?
- data format conversion

[more](#)

**hellyj**

- My blog
- Create content
- Search for Data
- My account
- Recent posts
- News aggregator

**Recent blog posts**

- Datastar Alternatives
- CMMAP Digital Library URL updated to <http://cmmmap.sdsc.edu> (don't need /drupal)
- CMMAP Digital Library upgrade completed
- Collections Status

[more](#)

**Recent comments**

- netcdf on Datastar 17 weeks 4 days ago
- IEEE format means unformatted binary from FORTRAN writes 17 weeks 4 days ago
- netCDF or ieee 17 weeks 4 days ago
- You can use the reply

# Circulating External Disk Drives



- FAT32 volume format
- USB 2.0 connector

- SRB-client can also be used for non-web data transfer ala rsync
- List-base scripts are also possible if they are desired



Resources

# Visualization of Very Large Datasets



# Testing (preliminary data point)

- Remote viz on Teragrid
  - TACC (UT Austin) / SPUR
  - Super-LES data
  - ParaView (client-server, remote X-session over ssh)
- TACC<-->Cox Cable (residential ~10Mbps)
  - ‘works’ but not usable
- Need better bandwidth for both data transfer and interactivity

Time: 0

Result
 
 Surface

Pipeline Browser

- cs://mpc1074:11111
- comet\_p4.pvtu
- Calculator2
- Contour2

Object Inspector

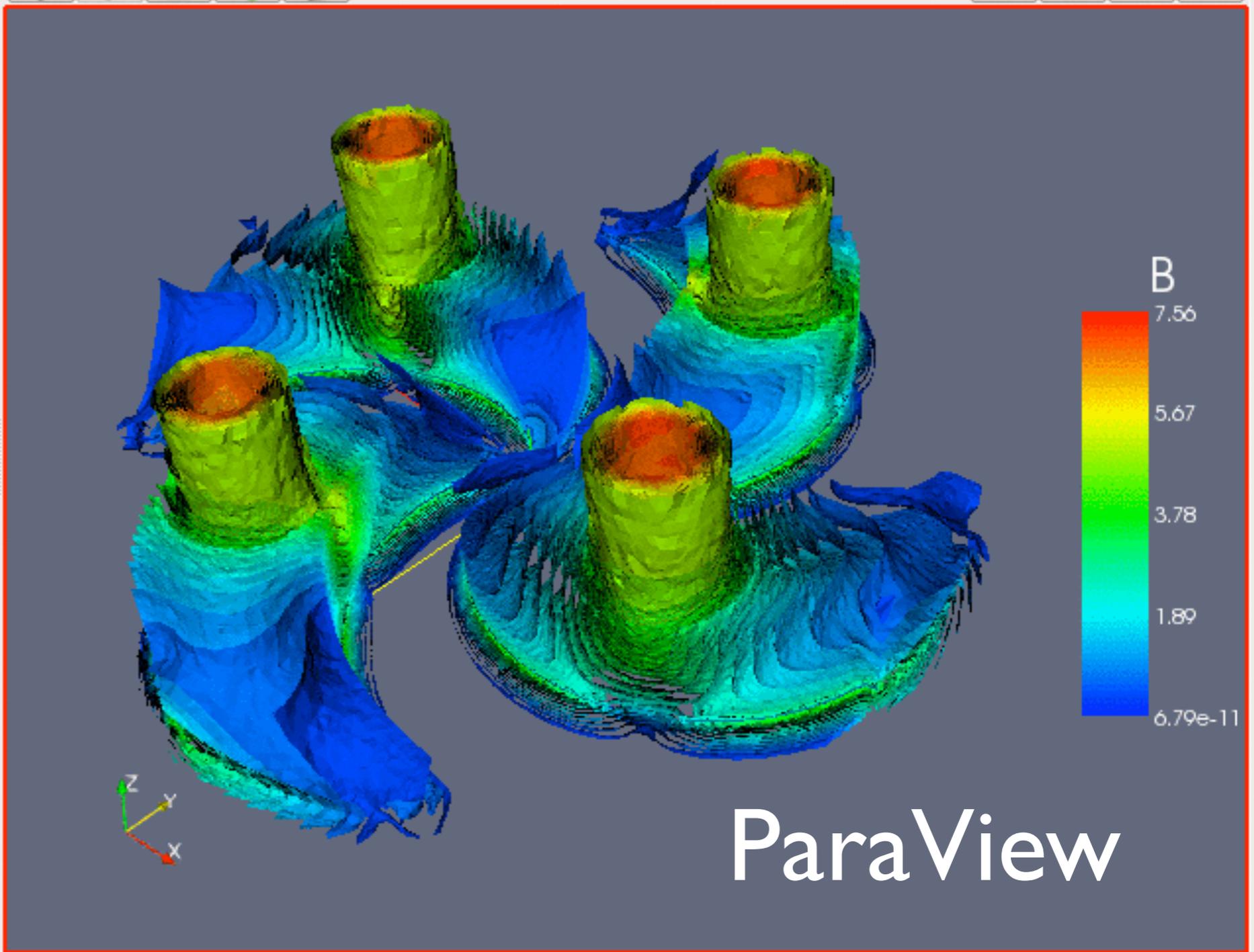
Properties | Display | Information

Apply | Reset | Delete

Isosurfaces

Value Range: [6.79195e-11, 7.55638]

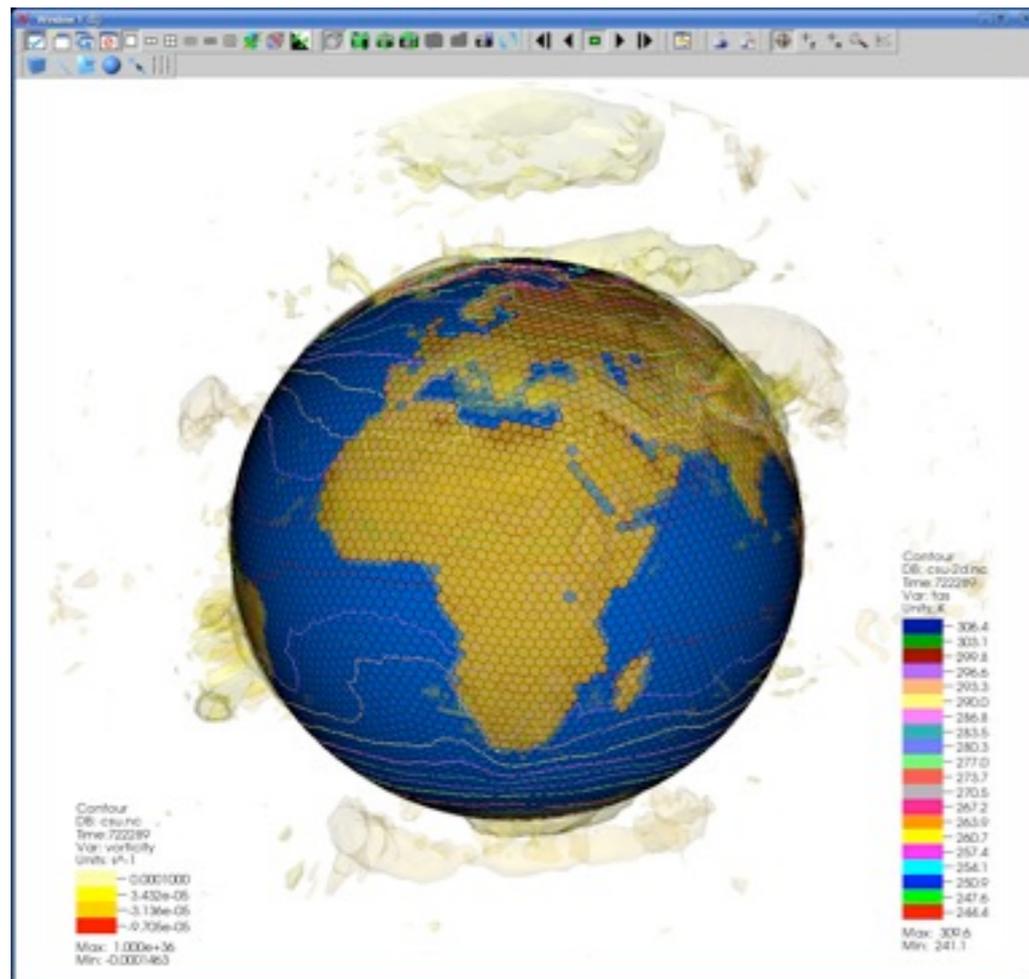
6.79e-11	Delete
0.329	Delete All
0.657	New Value
0.986	New Range
1.31	<input type="checkbox"/> Scientific
1.64	
1.97	
2.3	
2.63	



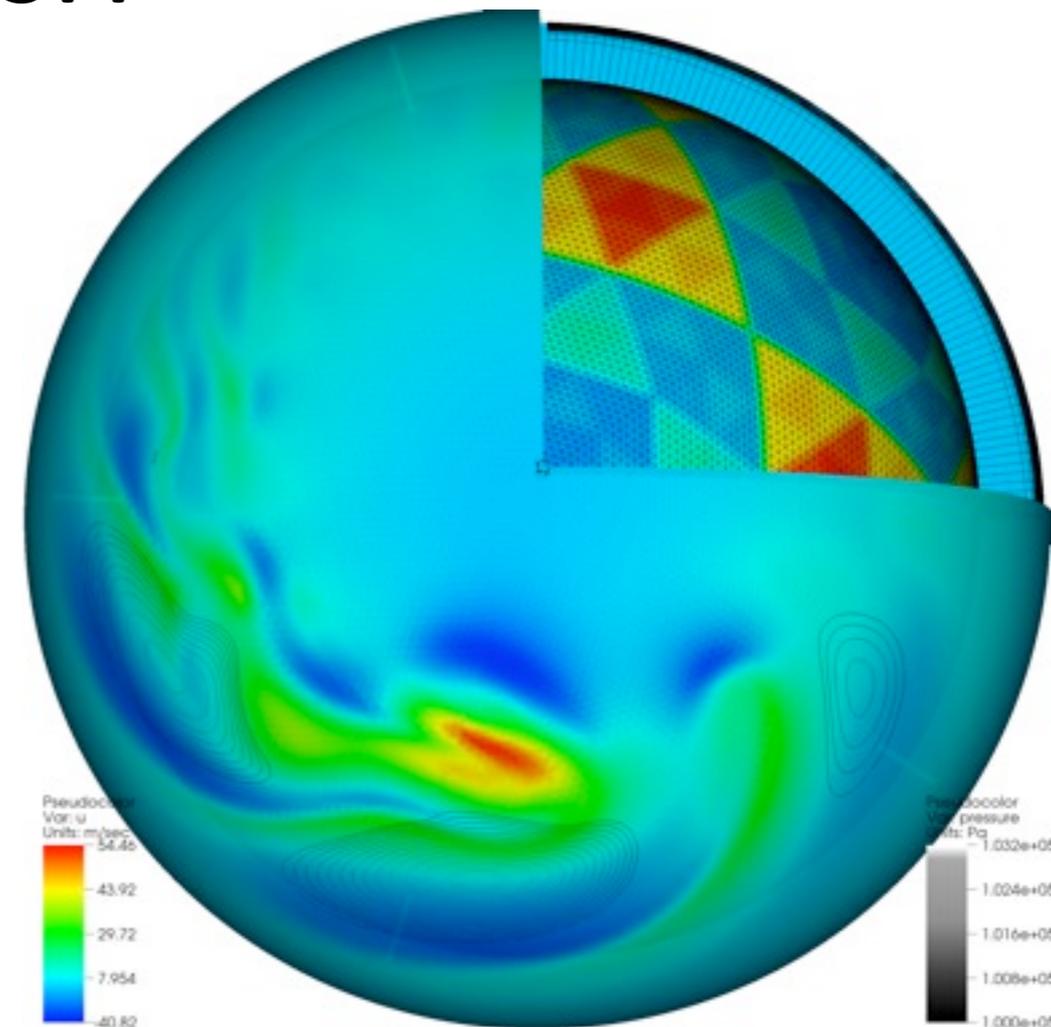
ParaView

# 3D visualization of geodesic data

## VisIT



3D isocontours of vorticity.



Composite plot of multiple mesh types and variables in the geodesic grid. Cell area (2D cell-centered data) and wind velocity (3D corner-centered on layers) data is shown by pseudocolor plots. Pressure (3D cell-centered on layers) is shown by contour lines.

# Future Testing

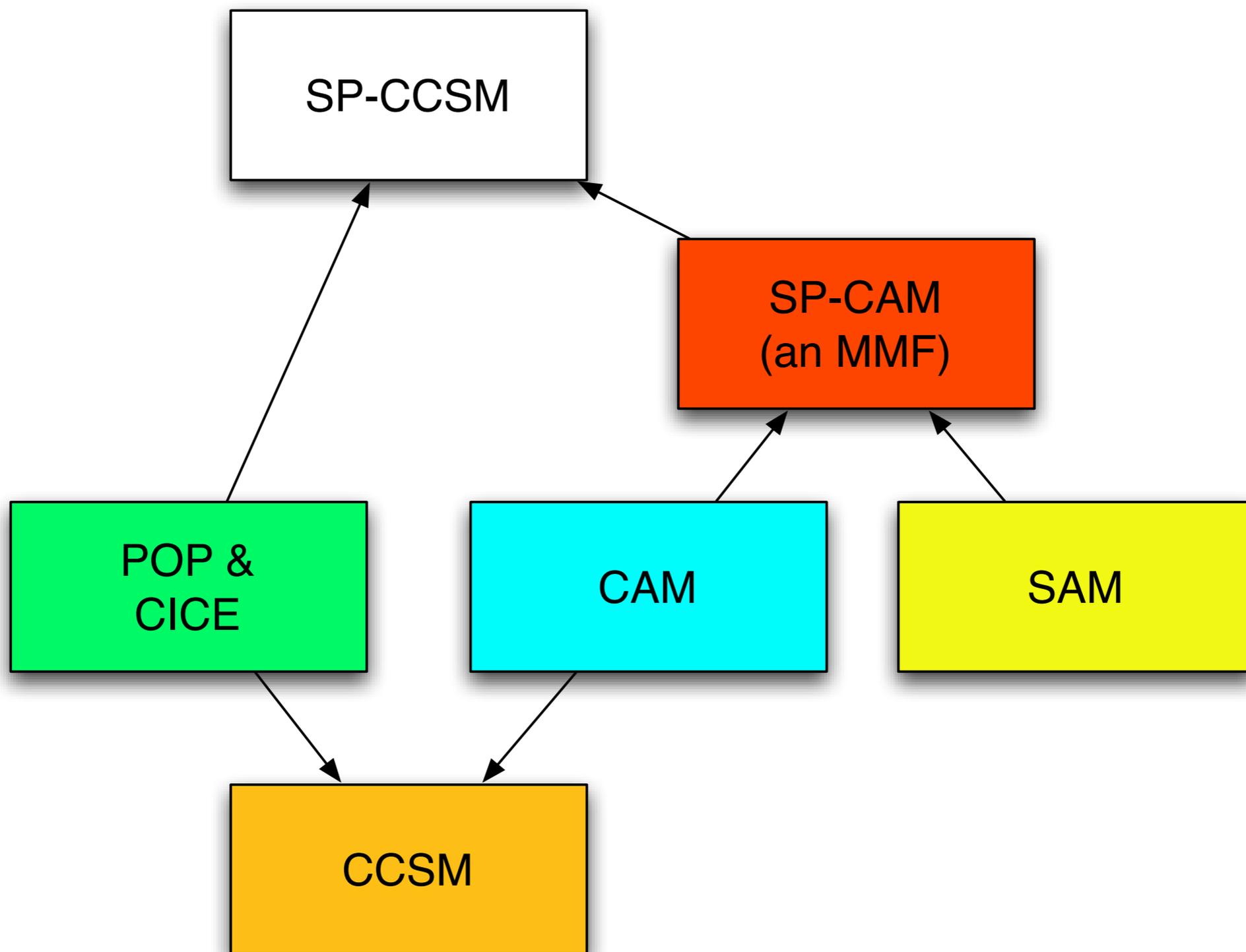
- Remote viz on Teragrid
  - TACC (UT Austin) / SPUR
  - Super-LES data
  - ParaView (client-server, remote X-session over ssh)
- UCSD Cave
- SDSC high-capacity network connections
- CSU network connections
- Other interested parties?



# Community Support

# Model Code Portability





# Model Taxonomy & Software Configuration Management

- Model codes
- Utility codes



# Mark Branson CMMAP User Support

---

## Connecting to TeraGrid

---

### Community Accounts

To address scalability issues, many gateways provide access to TeraGrid resources through a community account rather than setting up unique TeraGrid accounts for each gateway user.

A community account has the following characteristics:

- Only a single community user account (i.e., a TeraGrid username/password) is created.
- The Science Gateway uses the single TeraGrid community user account to launch jobs on the TeraGrid.
- The gateway user running under the community account typically has privileges to run only a limited set of applications.

The chief difference between an individual and a community account is that a community account is essentially a single username on the TeraGrid shared by many (human) users. While this eliminates the need for individual gateway users to request their own TeraGrid accounts, it places additional accounting and security burdens on the gateway developers. To distinguish one gateway user from another, the gateway developer has to institute a user registry and gateway authentication mechanism.

## User Support & Documentation

### Advanced Support for TeraGrid Applications (ASTA)

[Home](#) > [User Support](#) > ASTA

Advanced Support for TeraGrid Applications (ASTA) provides collaboration between Advanced User Support (AUS) staff and users of TeraGrid resources. The objective of the program is to enhance the effectiveness and productivity of scientists and engineers. As a part of the ASTA program, guided by the allocation process, one or multiple AUS staff will join the principle investigator's (PI's) team to collaborate for up to a year, working with users' applications.

AUS staff from TeraGrid resource provider sites have expertise and experience in many areas of high performance computing, domain sciences, data management, and grid computing. Collaborative work can include any of the following:

- Porting applications to new resources
- Implementing algorithmic enhancements
- Implementing parallel math libraries
- Improving the scalability of codes to higher processor counts
- Optimizing codes to efficiently utilize specific resources
- Providing help for portal and gateway development
- Assisting with visualization, workflow, and data analysis/transfer

In addition to requesting ASTA at the four quarterly allocations meetings, ASTA can also be requested throughout the year, as a Startup or Supplemental request, via [POPS](#). Please visit the [ASTA Project List](#) page to view a list of current ASTA projects.

#### On this page

- [How to Apply](#)
- [ASTA Selection Process](#)

#### Related Links

- [ASTA Project List](#)
- [Allocations Information](#)

#### Need Help?

- Phone Toll-free  
1.866.907.2383
- [Submit a Ticket \(online form -- fastest\)](#)
- [Submit a Ticket via email](#)
- [TeraGrid Knowledge Base](#)



# Backup



# TeraGrid™ User Portal

[Logout](#)

Welcome, John Helly

- Home
- My TeraGrid
- Resources
- Documentation
- Training
- Consulting
- Allocations

- Allocations/Usage**
- Accounts
- Profile
- Registered DNs
- Change Portal Password
- Add/Remove User
- Community Account
- SSH Terminal

## Allocation Usage

### Projects

[Show Inactive Projects](#) | [Show Expired Allocations](#)

#### Remote Visualization of Clouds

[Show Project Details](#)

##### Allocations

Start Date	End Date	Resource	SUs Remaining	SUs Awarded	My Usage (SU)	% Remaining	Alloc. Type	State
2009-11-07	2010-11-07	Spur	30,000	30,000	0.0	100 %	new	active

[Show Users on Spur](#)

#### Regionalization of Anthropogenic Climate Change Simulations

[Show Project Details](#)

##### Allocations

Start Date	End Date	Resource	SUs Remaining	SUs Awarded	My Usage (SU)	% Remaining	Alloc. Type	State
2009-04-01	2010-03-31	Ranger	1,989,328	3,950,000	0.0	50 %	new	active
2009-04-01	2010-03-31	Spur	500	500	0.0	100 %	new	active

#### Modeling Global Climate Variability with the Multi-scale Modeling Framework: The Boundary-layer Cloud Problem

[Show Project Details](#)

##### Allocations

Start Date	End Date	Resource	SUs Remaining	SUs Awarded	My Usage (SU)	% Remaining	Alloc. Type	State
2009-04-01	2010-03-31	Steele	812,564	950,000	0.0	86 %	new	active

[Show Users on Steele](#)

# Data Challenges of GCRM

- Extremely high volumes of data
  - 10 GB/ variable / step
  - 1-10 petabytes / simulated year
  - Can't just move data to local systems
  - Data will have to be on-offline
  - 4 byte offsets exceeded
  - Huge number of files per simulation
  - Model for running analysis on the entire data set needed

# Data Challenges of the GCRM (cont)

- Geodesic Grid
  - Preliminary (but not sufficient) support in some analysis tools
  - Standards for complete description not defined
  - Hyperslab-ing on coordinate values not supported and very costly
  - Grid itself is large (~ 2GB)

# CMMAP's Stable of Models

● **Current**

▲ **Super-CAM**

▲ **SAM**

▲ **VVM**

● **Under development**

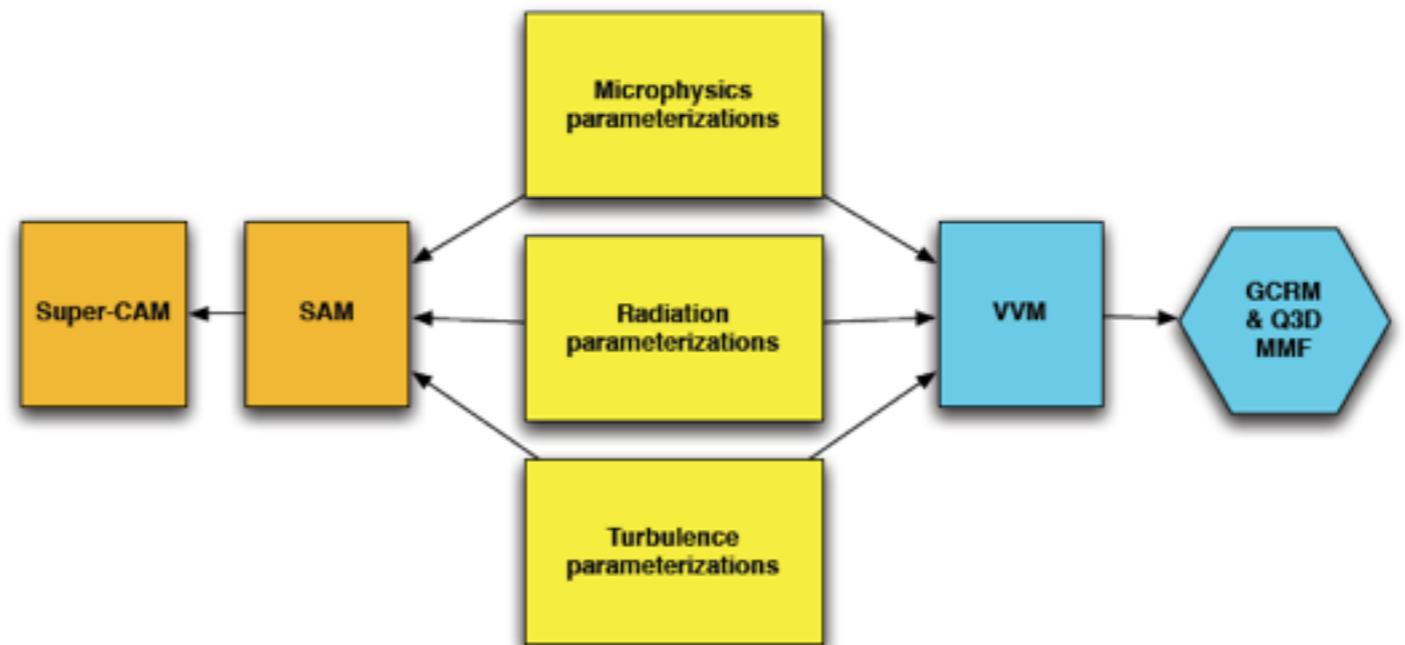
▲ **GCRM**

▲ **Q3D MMF**

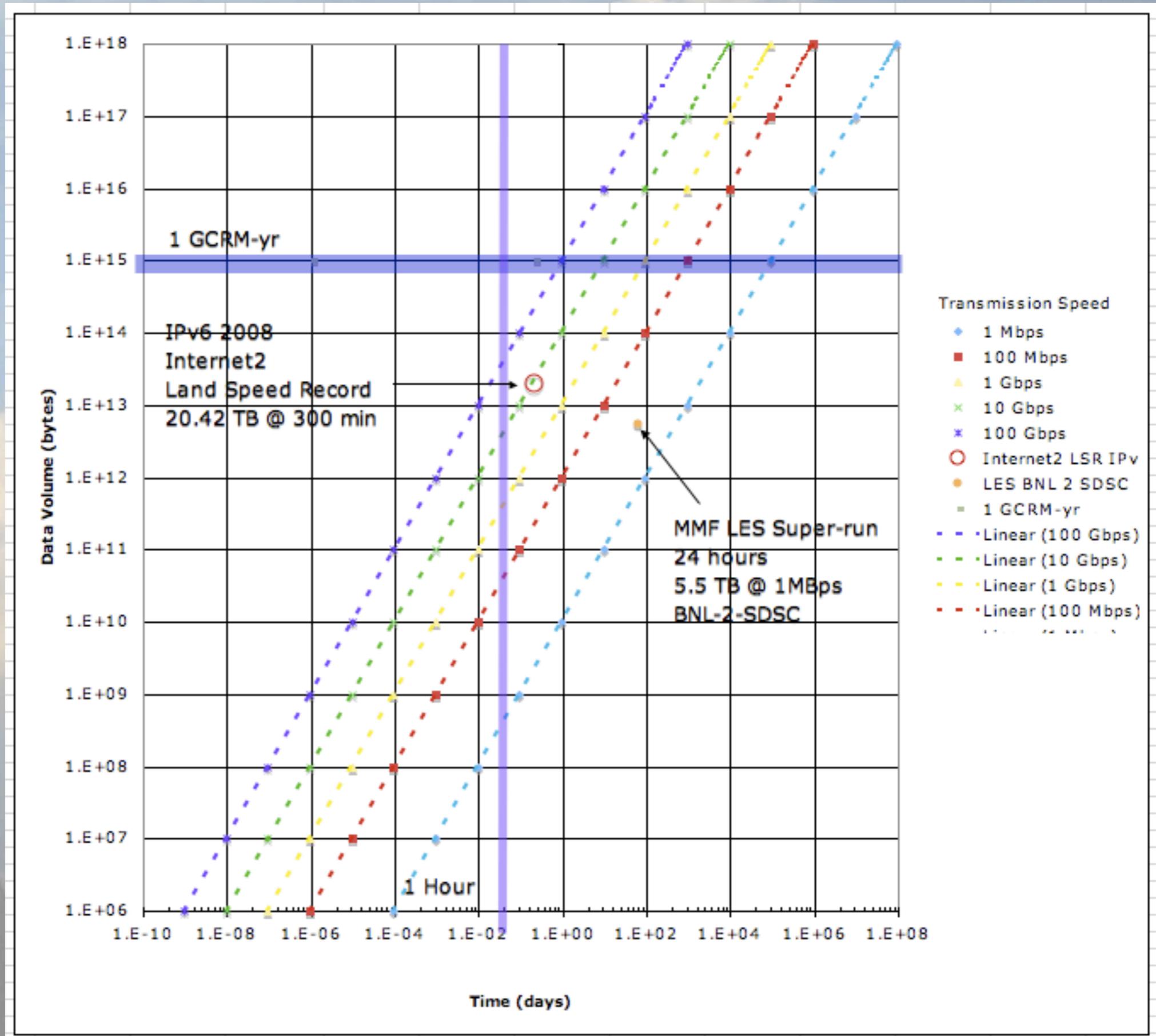
Who's using what and what modifications are being made?

What do we mean when we say we '...are running the MMF'

## Testing parameterizations



# Internet Data Transfer Capacity



# Conventions for Naming Datasets

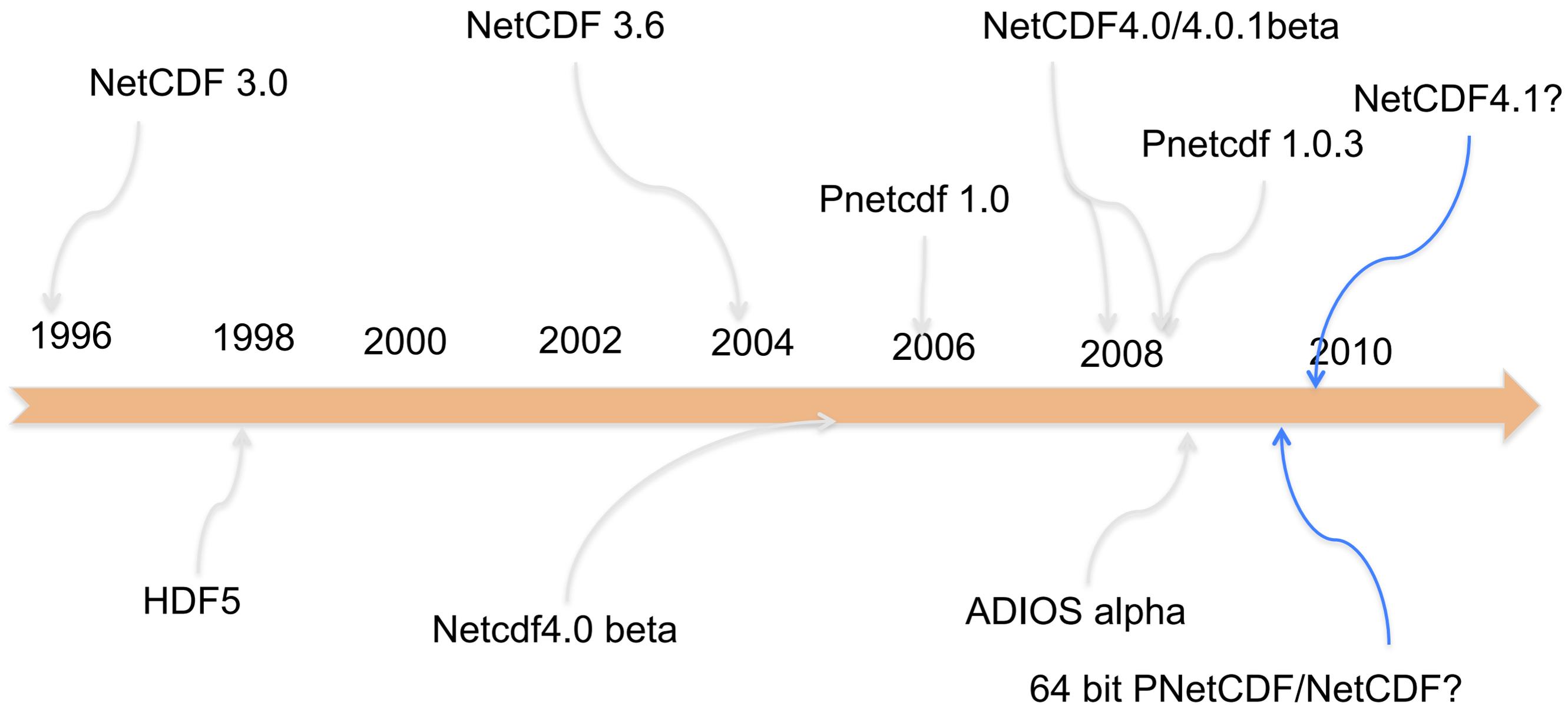
- `spcam_climo_-3C-12C_1e-5_re7-10_vt01---` → `CMMAP_SUPERLES`
- Standardized names aid communication and data management
  - current convention is adapted from oceanographic collections done for SIO and IODP
  - none apparent for climate modeling (e.g., AMIP, CF)
  - DOI (digital object identifiers) are used in the publishing world)
  - this is similar to a DOI and we may add a DOI if a DOI authority emerges for atmospheric data
  - In the meantime we should have something equivalent

`CMMAP_SUPERLES_20080725014445402_20080725014445402_spcam3.clm2.r.0003-01-01-00000.mif`

Standardized prefix is added  
to facilitate search and identification

Original filename is untouched

# Timeline

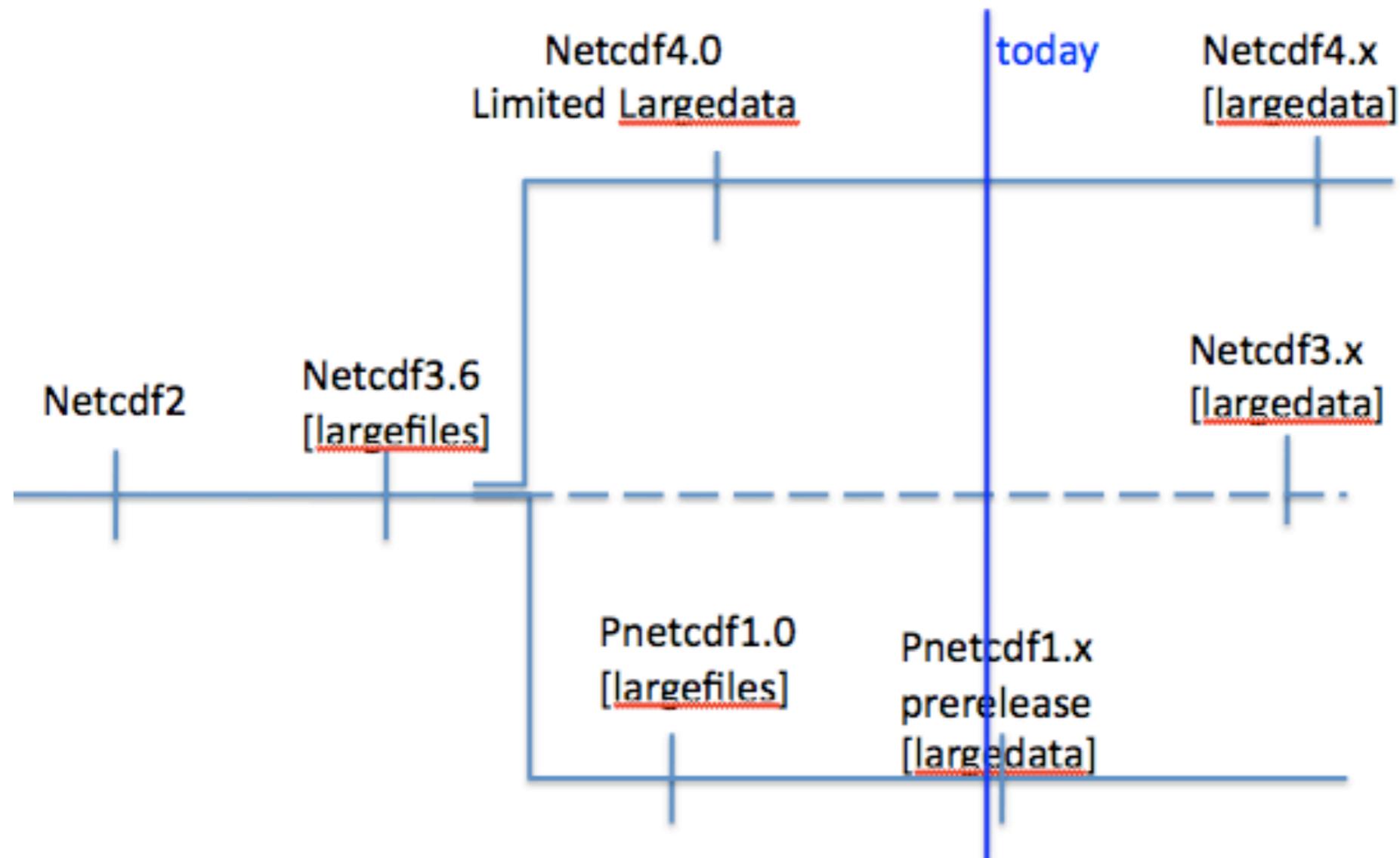


# Action Items

- Data conversion code: com3d-netCDF (Helly, Branson)
  - add sample dataset
- Data Distribution of LES dataset (Helly)
- NSF LRAC Allocation Proposal w/technical support (Helly, Randall, Ackerman/Bretherton, Krueger)
- Digital Library search interface extension (Helly)
- OAI metadata server for external interoperability (Helly)
- Intercalibration observations for digital library (Helly, Ackerman)
- Prepare version control for MMF 2.0 and GCRM

# NetCDF Status

Real large data support coming soon!



Caution:  
Pnetcdf output not  
currently compatible  
with other formats

# Model Taxonomy

