Distributing climate model data to the world

A 10-year retrospective

Gary Strand

1998

Parallel Climate Model (PCM) experiments begin

Initial interest in distributing output data outside the “anointed” few

1999

PCM data sent to PCMDI for distribution via ftp

(2,000 files, ~860 GB available)

ESG-I project

2000

@PCMDI, additional
20,000 files, ~3.7 TB available

ESG I wraps up

2001

@PCMDI, total
30,000 files, ~6.8 TB available

ESG II begins

IPCC TAR released

2002

@PCMDI, total
53,000 files, ~13 TB available

Initial ESG II meetings

2002 Data Users and Collaborators

Aiguo Dai, NCAR
Alan Ziegler, Princeton University
Alex Sim, LBNL
Andrew Wood, U. Washington
B. Govindasamy, LLNL
Bill Anderson, NCAR
Bob Drach, LLNL
Bob Malone, LANL
Cecilia Bitz, U. Washington
Charles Doutriaux, LLNL
Charles Hakkarinen, EPRI
Chick Keller, LANL
Clara Deser, NCAR
Curtis Covey, LLNL
David Pierce, Scripps
Dean Williams, LLNL
Dennis Shea, NCAR
Dmitry Vjushin, Bar-llan U. Israel
Eric Leuliette, CU
Gerald Meehl, NCAR
Gerald Potter, LLNL

Greg Ostermeier, U. Washington
Hans Luthardt, DKRZ
Helen Kettle, Edinburgh U.
Irene Fischer-Burn, DKRZ
James Boyle, LLNL
James Hack, NCAR
Jeff Kiehl, NCAR
Jeffrey Annis, Scripps
John Gregory, IPCC
John Taylor, Argonne
John Weatherly, CRREL
Jouni Raisane, SMHI
Julie Arblaster, NCAR
Justin Hnilo, LLNL
Kao J. Chin Yue, LANL
Ken Sperber, LLNL
Kevin Keay, U. Melbourne
Kwang Yul Kim, Florida State
M. Kanamitsu, U. California
Marcia Branstetter, U. Texas
Michael Wehner, LLNL

Pat Behling, U. Wisconsin
Peter Gleckler, LLNL
Raymond Arritt, Iowa State
Regine Hock, Swedish Royal Inst. Tech.
Robert Oglesby, NASA Ames
Ronald Stouffer, NOAA
Roy Jenne, NCAR
Scott Smith, LANL
Sylvia Murphy, NCAR
Thomas Reichler, U. California
Tim Barnett, Scripps
Tom Knutson, NOAA
Tony Hirst, CSIRO
Tony Tubbs, Scripps
Ulrich Cubasch, DKRZ
Vadim Guzey, U. Adelaide, Australia
Vadim Matyugn, RAS
Vladimir Kattsov, RAS
Wesley Ebisuzaki, NOAA
Wilfred Post, ORNL
Youichi Tanimoto, Japan NIC

2003

@PCMDI, total
100,000 files, ~16 TB available

ESG II still meeting

1st GO-ESSP, Jan
2nd GO-ESSP, Sep

2004

@PCMDI, total
184,000 files, ~30 TB available

ESG II opened to friendly users

3rd GO-ESSP

2005

@PCMDI, total
230,000 files, ~33 TB available

ESG II gains traction
Replicated at PCMDI
for CMIP3
4th GO-ESSP

2006

@PCMDI, total
260,000 files, ~36 TB available

ESG II continues to grow
CMIP3 archive opened

5th GO-ESSP

EGU 2006

Climate Model Data Management

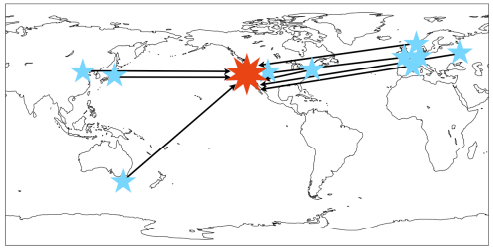
Gary Strand
strandwg@ucar.edu
National Center for Atmospheric Research, USA

Introduction

The IPCC Fourth Assessment Report (IPCC AR4) required the largest collection of climate model data assembled and distributed to analysts to date, and was unprecedented in scale and scope:

- ~500 TB (10^{12} bytes) of data generated
- 35+ TB of data available for access
- 80+ TB of data downloaded
- 11 experiments, 23 models from 10 nations
- 500+ registered users
- More than 200 refereed papers

IPCC AR4 (2004-2005)

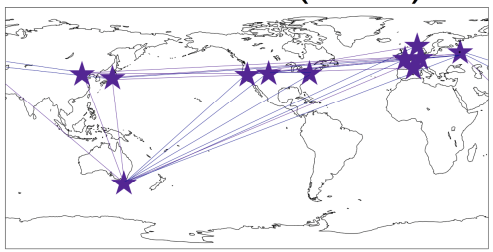


A **single central** online repository for all data. Access via web and ftp. Simple registration process.

IPCC AR4 contributions

Model	Country	Data created (GB)	Submitted for AR4 (GB)
BCC-CM1	China	12,500	530
BCCR-BCM2.0	Norway	22,180	1,080
CCSM3	USA	116,340	9,610
CGCM3.1(T63)	Canada	12,080	880
CGCM3.1(T47)		22,320	1,320
CNRM-CM3	France	16,500	945
CSIRO-Mk3.0	Australia	3,700	1,500
ECHAM5/MPI-OM	Germany	115,000	1,230
ECHO-G	Germany S. Korea	5,100	480
FGOALS-g1.0	China	15,000	2,870
GFDL-CM2.0	USA	155,000	1,660
GFDL-CM2.1			990
GISS-EH	USA	3,250	830
GISS-ER			340
GISS-AOM		-	210
INM-CM3.0	Russia	500	370
IPSL-CM4	France	50,000	1,120
MIROC3.2(hires)	Japan	60,000	2,180
MIROC3.2(medres)		40,000	2,500
MRI-CGCM2.3.2	Japan	-	1,025
PCM	USA	6,300	2,030
UKMO-HadCM3	UK	-	930
UKMO-HadGEM1		-	850
Totals		**655,770**	**35,480**

IPCC AR5 (2012)



Federated data centers, replicas at multiple sites. Certificates for secure registration and access, via web and aware clients.

Recommendations for AR5

Based on the AR4 experience, by the time of AR5 (around 2012) the practices of climate model data management must change. The IPCC AR5 will require hundreds of TB to tens of Petabytes (10^{15}) of data to be made available for analysis.

Distribution

- Multiple federated data centers
- Replicas across some/all sites
- Data retrieval via web & clients
- Server-side processing
- Visualisation

Standards recommendations

- CF-2.0 for metadata
- netCDF-4/HDF-5 for file format
- Must be set well in advance (~2008)

Security recommendations

- More formal and strict procedures
- Certificate authority (CA) likely
- Allow one-time password (OTP) tokens

Next steps

- **Data management must become a formal part of AR5's process, not left as an afterthought**
- Planning needs to be integrated from the beginning of the process
- Formation of a data management group
- Data managers **must** be more involved
- More interaction with analysis community
- Resources for proper and efficient data management must be made available
- Changes in data center policies examined

Acknowledgements

Thanks to the following individuals who provided information and advice: Julie Arblaster, Lawrence Buja, Michel Crucifix, Kenneth Lo, Keith Dixon, Xu Ying, Jean-Louis Dufresne, Daniel Robitaille, Erich Roeckner, Ingo Bethke, Martin Dix, Jean-François Royer, Toru Nozawa, Yongqiang Yu, Evgeny Volodin, Seung-Ki Min.

2007

@PCMDI, PCM total
280,000 files, ~37 TB available

IPCC 4th Assessment Report released

6th GO-ESSP

2008

ESG II to continue as ESG-CET

7th GO-ESSP

2009 data users

Olemorten Aamo
Toomas Aas
Jorge Abad
Abdelouahed Abahamid
Ricardo Abarca
Sergio Abarca
Rodrigo Abarcadelrio
John Abatzoglou
Khalid Abbas
Mark Abbott
Rachid Abboubi
Sameh Abdelazim
Tilioua Abdelaziz
Razagui Abdelhak
Mohamed Abdelkader
Fekih Abdellali
Sayed Abdel_rahman
Moujane Abderrahim
Abdellatif Abdou
Bannu Abdulsamad
Ryo Abe
Nibret Abebe
Xavi Abellan
Tekeste Abezghi
Sukumarapillai Abhilash
Inma Abia
Muhammadadnan Abid
Zainal Abidin
Babatunde Abiodun
Neil Ablang
Maria Abrahamowicz
Alicia Acero
Krishna Achutarao
Melissa Ackerman
Drgeeta Agnihotri
Geeta Agnihotri
Lucia Agnoletto
Vikram Agone
Emiliano Agosta
Monica Agostinone
Giuseppe Agrillo
Agrsianp Agrsianp

Oscar Aguado
Alejandro Aguilarsierra
Jose Aguilera
Adolfo Aguirre
Nazmul Ahasan
David Ahijeyvch
Hugo Ahlenius
David Ahlfeld
Jon Ahlquist
Azadeh Ahmadi
Ahmed Ahmed
Farzan Ahmed
Sultan Ahmed
Gokhan Ahmet
Hyun-ju A
Kaiden A
Kwang- Ahn
Leena Aho-mantila
Mehran Ahsant
Hsuehfang Ai
Nian Aibing
Suchandra Aichbhowmick
Giuseppe Aiello
John Aiello
Lv Aifeng
Zhu Aijun
Christopher Aiken
Hidenori Aiki
Luis Aimola
Bruce Ainslie
Air Air
Filipe Aires
Anantha Aiyyer
Litta A.j
Steven Alderson
Corina Alecu
Maryana Alegro
Omid Alemi
Andrea Alessandri
Nicola Alessi
Alex Alex
Curtis Alexander

Gavrikov Alexander
Joshua Alexander
Lisa Alexander
Sadretdinov Alexandr
Vassil Alexandrov
Dumitrescu Alexandru
Vladimir Alexeev
Ioannis Alexiou
Reda Algendy
Andrea Algieri
Chris Algieri
Juana. Algodon
Juan Algodon
Hussain Ali
Izhar Ali
er
Ali
Mohamed Alioua
Kais Al-jumily
Cem Alkan
Opesh Alkara
Patrick Alken
Mike Allard
Pierre Allemand
David Allen
John Allen
Maegan Allen
Matthew Allen
Melissa Allen
Michael Allen
Pat Allen
Robert Allen
Teddy Allen
Tim Allen
Carmen Alvarezcastro
Oscarandres Alvarezmontero
Fernando Alves
Marcelo Alves
Sultan Al-yahyai
Toshiyuki Amagasa
Amalia Amalia
Anders Aman

Brard Amand
Lior Amar
Jaisonthomas Ambadan
Bas Amelug
Alexander Ames
Farid Amid
Eddie Amir
Muhammad Amjad
Caspar Ammann
Teerachai Amnauylawjarurn
Ricardo Amorim
Jeff Amthor
Jeff Amthor
Jaeho An
Xiaocun An
Harsh Anand
Nikil Anand
Aayudai Anandhi
Anantharaj
John And
Not And
Søren
Theresa Andersen
Aaron Anderson
Bret Anderson
Brian Anderson
Brucet. Anderson
Charles Anderson
Chris Anderson
David Anderson
Sergey Anikin
Man Anil
Mihai Anitescu
Anila Anitha
Stoens Anne
David Annetts
Jeffrey Annis
Sandrine Anquetin

Carlos Anselmi
Faron Anslow
Steve Anthonio
Froissart Antoine
Marta Antonelli
Chiara Antoniazzi
Joseph Antony
Edward Antrobus
Lai Anwei
Richard Anyah
Shigeaki Aoki
Bonnie Aona
Phornnarong Aonchart
Josep Aparicio
Duane
Amy Apon
Andreas Appelhagen
Scott Applegquist
Todd
Joh
Dem
Amol Apte
Ragi A.r
Osamu Arakawa
Nilesh Araligidad
Afonso Araujo
Aroldo Araujo
Sebastián Araujo
Vibhas Aravamuthan
Jose Aravequia
Brian Arbic
Jon Aronsson
Jonatan Aronsson
David Aronstein
Mary Arras
Jose Arrecio
Jose Arreola
Joseluis Arreola
Chris Arrington

Raymond Arritt
Alfredo Arroyo
Rodrigo Arroyo
Kristi Arsenault
Joaquim Arteta
Craig Arthur
Dan Arthur
Artixit Artixit
Saravanan Arunachalam
Arun Aryasomayajula
Ali Asadi
Hirofumi Asahi
Ichio Asanuma
Jun Asanuma
Yoshi Asaoka
Ernes
Andy Aschwanden
Asd Asd
Asd Asda
Asd Asdasdas
Asdf
Asdfg Asdfg
Hermann Asensio
Shakeel Ashara
Vinit Asher
Moet Ashfaq
Moetasim Ashfaq
Yosef Ashkenazy
Parisa Ashofteh
Bahman Ashrafijoo
Raghavendra Ashrit
Emmanuelle Autret
Michael Auyeung
James Avelar
Clayton Avila
Francia Avila
Norman Avila
Ruben Avila
Serena Avolio

Yusuf Awaluddin
Nauman Awan
Philip Axe
David Axtman
Orlando Ayala
Javier Ayllon
Jouon Aymeric
Majid Azadi
Muhammadirfan Azam
Akira Azami
Arash Azaranfar
Mina Azer
Al Azhar
Abdou Aziz
Subbar B
Vladimir B
Nils Babel
Larry Baber
Andrey Baboshin
Praveen Babu
Alfred Bachmann
Larissa Back
Daniel Bader
David Bader
Juergen Bader
Eunhyuk Baek
Yulim Baek
Michael Baldauf
Michael Baldwin
Dimitra Bali
Nebojsa Balic
Srabanti Ballaav
Joaquim Ballabrera
Mario Ballario
Srabanti Ballav

Robert Ballentine
Joan Ballester
Guillermo Ballestervalor
Szilveszter Balogh
Angela Balou
Michael Balshi
Haythem Balti
Magnus Baltscheffsky
Muthukumar Balu
Wolfgang Balzer
Ray Bambha
Jun Ban
Junmei Ban
Lawrence Band
Ste
Sudipta Banerjee
Anantha Bangalore
Maria Banks
Sandra Banson
Eddie Bao
Bao
Shaowu Bao
Xu Bao
Yan Bao
Pan Baobao
Li Baofeng
Xie Baoguo
Ren Baohua
Pan Baojun
Wan Baolu
Malcolm Baptie
Sed Bar
Chip Barrere
Nilzair Barreto
Andrew Barrett
Brad Barrett
Daniel Barrie
Nestorjose Barrionuevo
Eduardo Barrios
Guillermo Barrios

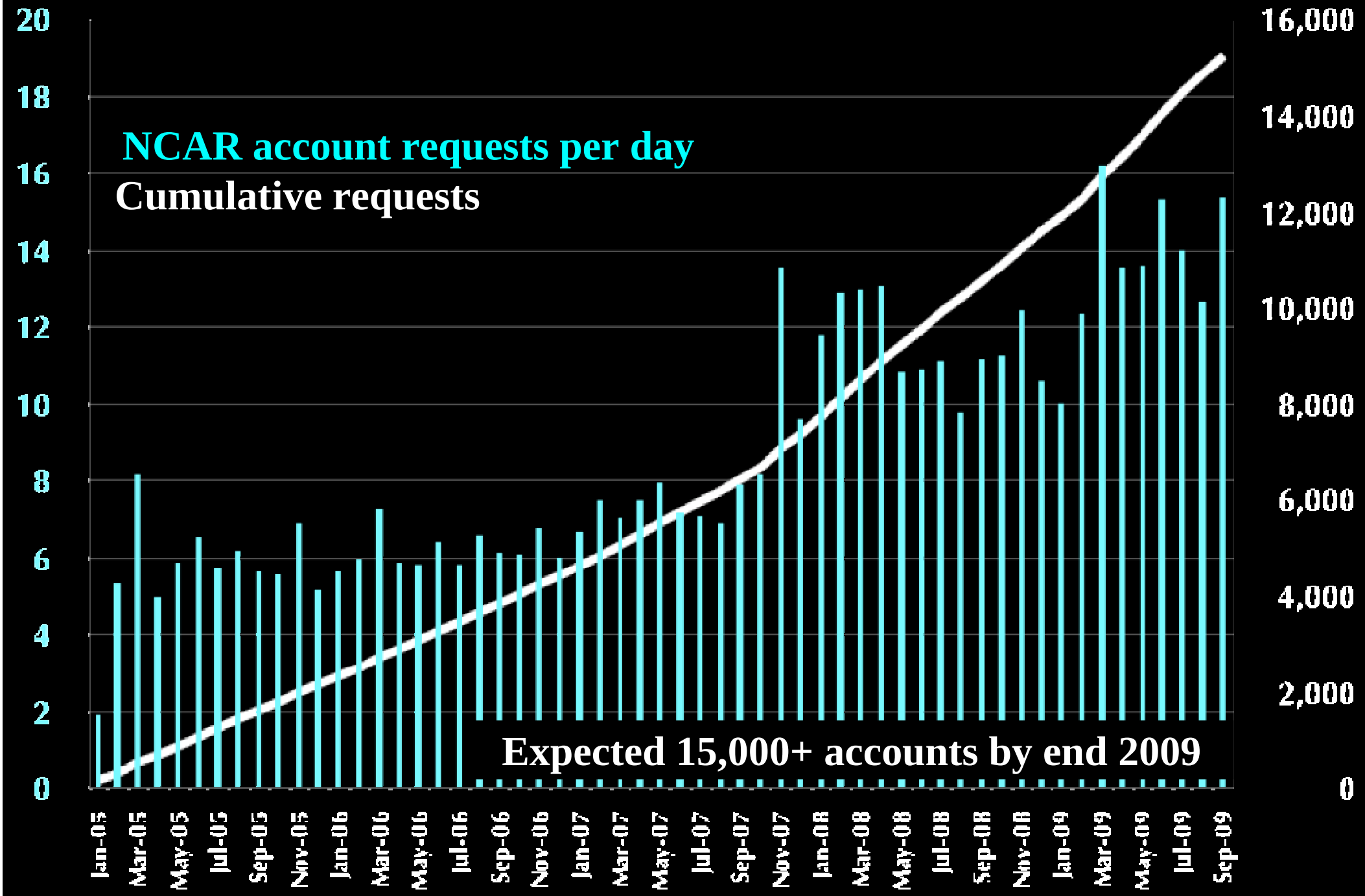
Felipe Barro
Tom Barron
Ana Barros
John Barry
Idar Barstad
Joseph Barsugli
Laio Bart
Timothy Barth
Christelle Barthe
Patrick Bartlein
Mike Barton
Saurabh Barve
Jean-christophe Basaille
Umutgul Basar
Enils Bashi
Preethi Basa
Latha Baskaran
Senaka Basnayake
Jim Basney
Jose Bassan
Gene Bassett
Radovan Bast
Ivan Bastak
Sophie Bastin
Hugo Bastosdeoliveira
Chandan Basu
Sourish Basu
Sukanta Basu
Ayon Basumallik
John Bateman
Gary Bates
Kevin Bates
Michael Bates
Brad Beechler
Naseema Beegum
Winfried Beer
Tony Beesley
Jonathan Beezley
Fred Begay
Andreas Béguin
P Begwinter

Ambica Behara
Pat Behling
Ruben Behnke
Rainer Behrendt
Erik Behrens
Xu Bei
Jens-olaf Beismann
Tor Bekkvik
Megan Bela
James Belanger
Sylvain Belanger
Gennady Belchansky
Larry Belcher
Michal Belda
Mark Belding
Kader Beldjilali
Andre Belem
Abderrahim Belhajji
Dmitry Belikov
Jason Bell
Jeanpierre Bell
Ken Bell
Marty Bell
Michael Bell
Shaun Bell
Stephen Bell
Elizabeth Bella
Richard Belles
Gilles Bellon
Omar Bellprat
Ali Belmadani
Alexei Belochitski
Josivan Beltrão
Danijel Belusic
J.scott Berg
Larry Berg
Andres Berger
Howard Berger
Sascha Berger
Tommi Bergman
Steffen Beringer
Mike Berkley

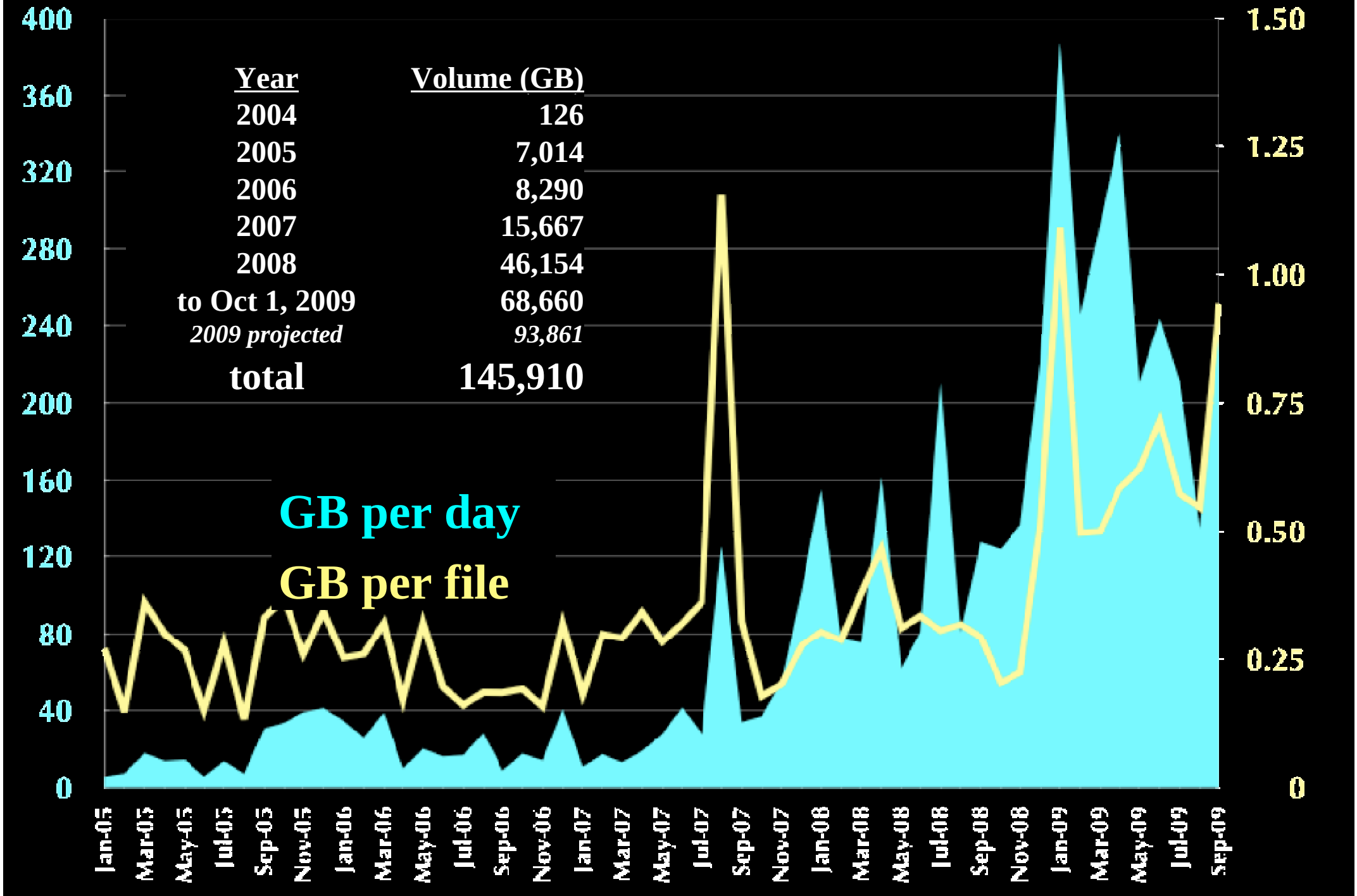
Plus another 38 slides like this one…

(which I will skip today)

Metrics

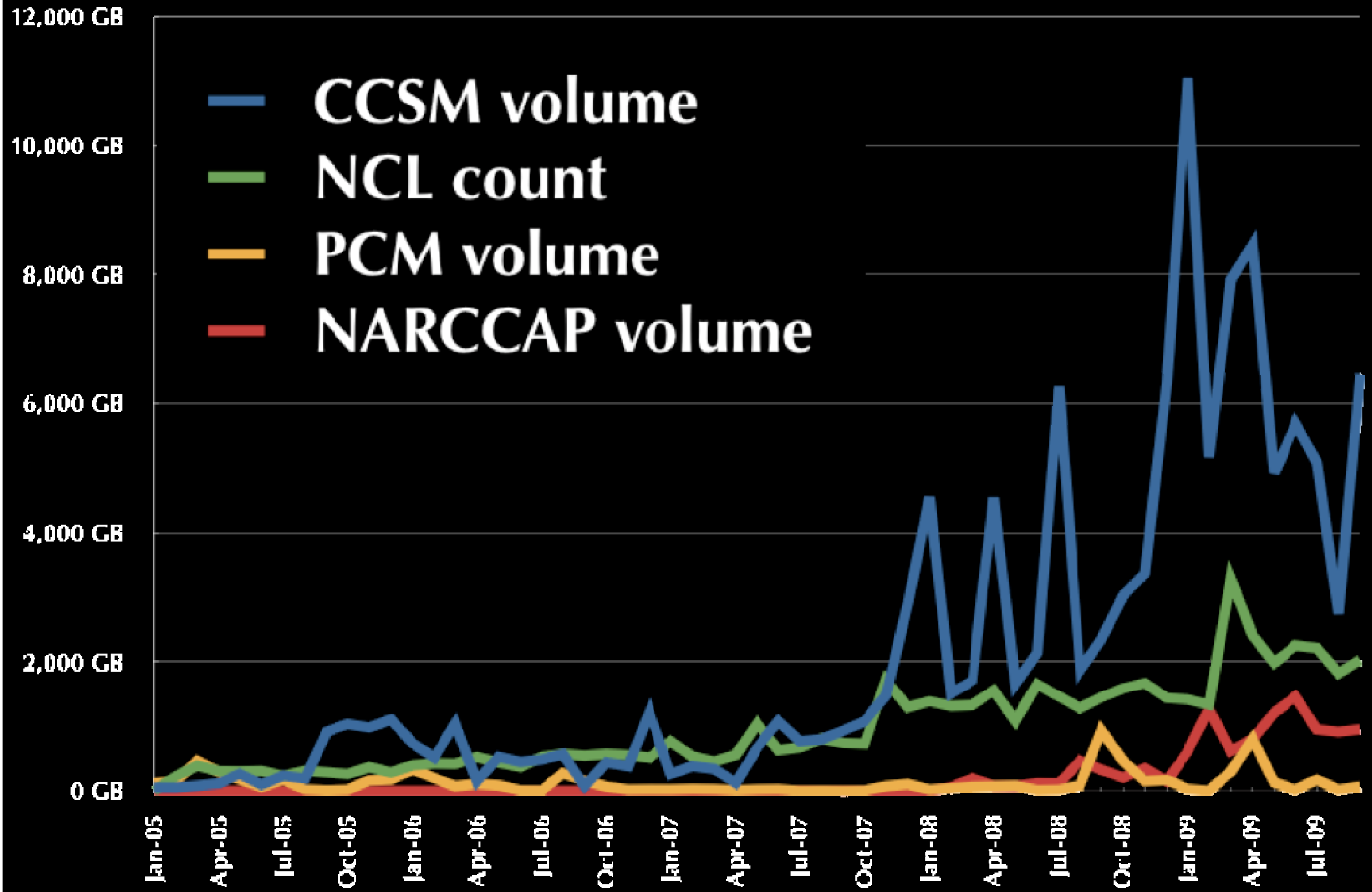

NCAR account requests per day
Cumulative requests
Expected 15,000+ accounts by end 2009
0
2
4
6
8
10
12
14
16
18
20
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
Jan-05
Mar-05
May-05
Jul-05
Sep-05
Nov-05
Jan-06
Mar-06
May-06
Jul-06
Sep-06
Nov-06
Jan-07
Mar-07
May-07
Jul-07
Sep-07
Nov-07
Jan-08
Mar-08
May-08
Jul-08
Sep-08
Nov-08
Jan-09
Mar-09
May-09
Jul-09
Sep-09

Metrics

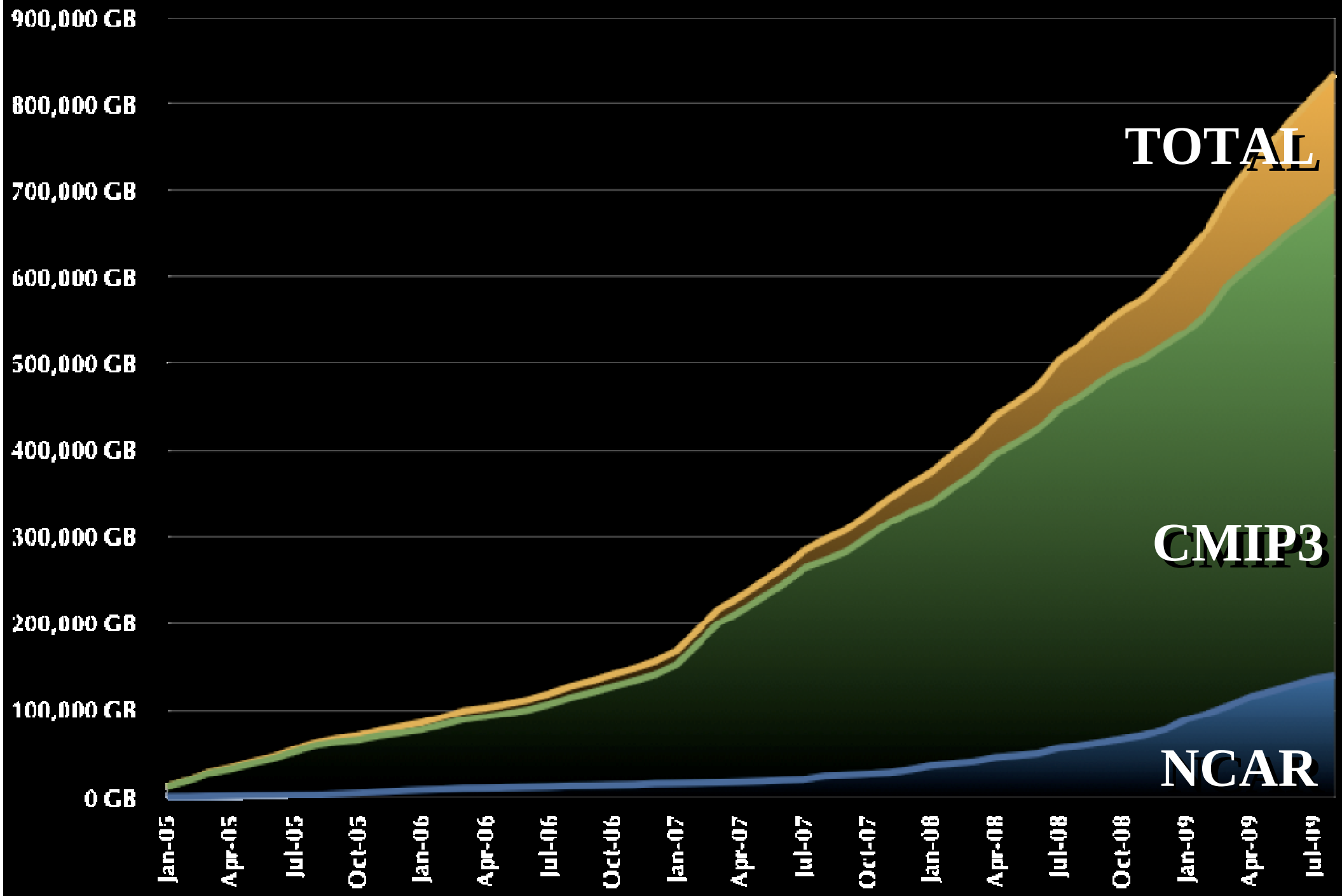


Year	Volume (GB)
2004	126
2005	7,014
2006	8,290
2007	15,667
2008	46,154
to Oct 1, 2009	68,660
2009 projected	*93,861*
total	**145,910**

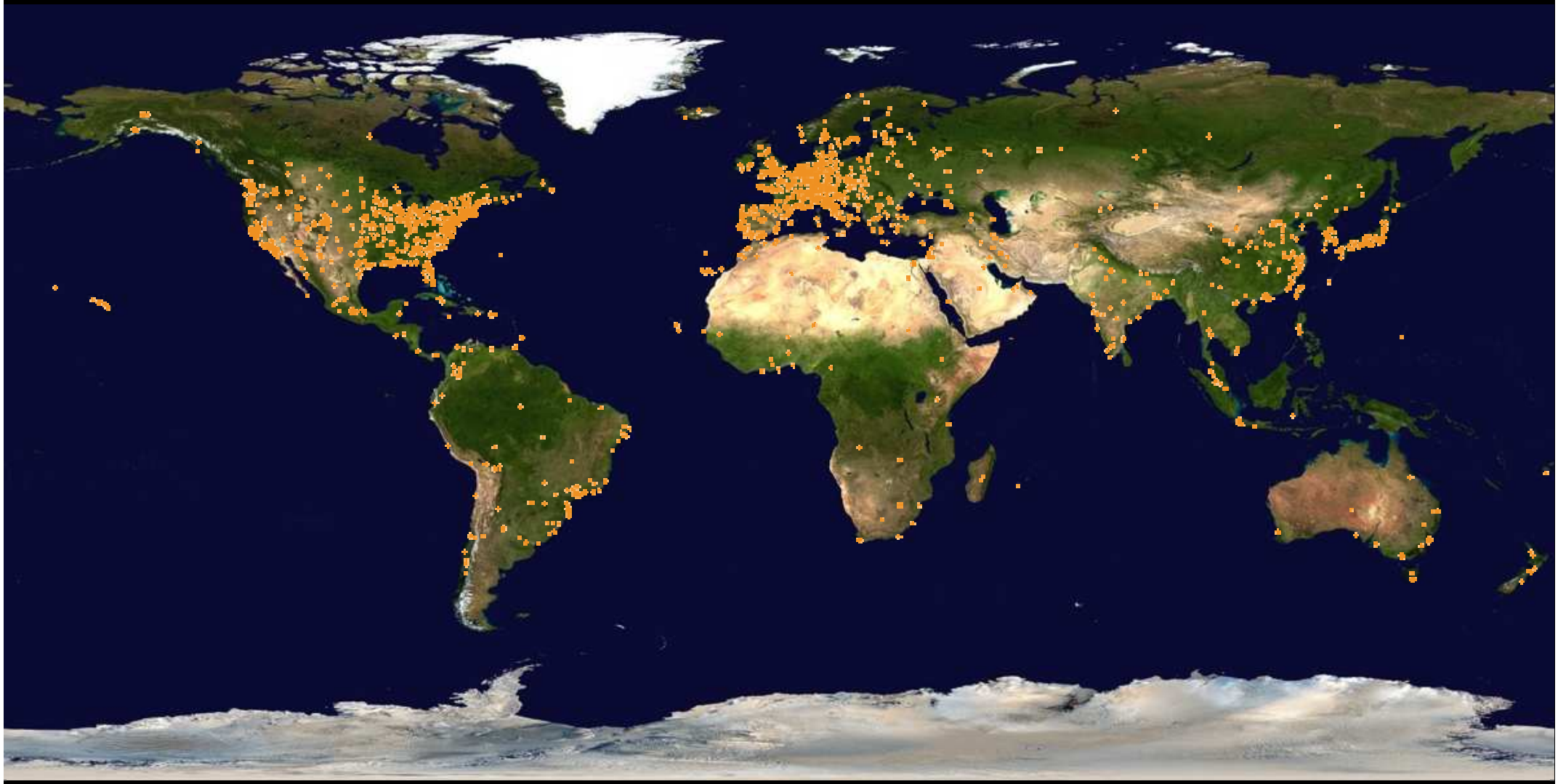
Metrics


12,000 GB
10,000 GB
8,000 GB
6,000 GB
4,000 GB
2,000 GB
0 GB
CCSM volume
NCL count
PCM volume
NARCCAP volume
Jan-05
Apr-05
Jul-05
Oct-05
Jan-06
Apr-06
Jul-06
Oct-06
Jan-07
Apr-07
Jul-07
Oct-07
Jan-08
Apr-08
Jul-08
Oct-08
Jan-09
Apr-09
Jul-09

Metrics


900,000 GB
800,000 GB
700,000 GB
600,000 GB
500,000 GB
400,000 GB
300,000 GB
200,000 GB
100,000 GB
0 GB
TOTAL
CMIP3
NCAR
Jan-05
Apr-05
Jul-05
Oct-05
Jan-06
Apr-06
Jul-06
Oct-06
Jan-07
Apr-07
Jul-07
Oct-07
Jan-08
Apr-08
Jul-08
Oct-08
Jan-09
Apr-09
Jul-09

Metrics



Metrics

5779 United States
1239 China
537 Germany
498 Japan
457 South Korea
397 France
388 Canada
343 Brazil
338 Taiwan
301 United Kingdom
281 Spain
259 Italy
252 Australia
237 India
142 Switzerland
139 Mexico
138 Russian Federation
118 Vietnam
115 Netherlands
111 Turkey
110 Hong Kong
94 Greece
93 Norway
89 Portugal
85 Sweden

81 Thailand
67 Chile
58 Singapore
46 Iran
45 Poland
45 Austria
43 New Zealand
40 Ireland
39 Indonesia
38 Colombia
36 Egypt
35 Denmark
32 Argentina
29 South Africa
29 Czech Republic
27 Ukraine
26 Finland
23 Israel
22 Hungary
20 Senegal
20 Morocco
19 Malaysia
19 Belgium
19 Algeria
18 Puerto Rico

18 Bulgaria
17 Luxembourg
16 Croatia
14 Saudi Arabia
14 Romania
14 Barbados
13 Peru
12 United Arab Emirates
11 Venezuela
11 Iceland
11 Costa Rica
10 Slovakia
10 Cuba
10 Bolivia
9 Slovenia
9 Nepal
9 Estonia
9 Bangladesh
8 Serbia
8 Philippines
8 Lithuania
8 Europe
8 Cameroon
7 Macau
7 Ecuador

7 Cape Verde
6 Uruguay
6 Kazakhstan
6 Belarus
5 Paraguay
5 Pakistan
5 Cyprus
5 Armenia
4 Niger
4 Mozambique
4 Kyrgyzstan
4 Ghana
4 Bosnia and Herzegovina
3 Trinidad and Tobago
3 Reunion
3 Panama
3 Mongolia
3 Madagascar
3 Jordan
2 Tunisia
2 Tanzania
2 Guam
2 Fiji
2 El Salvador
2 Bermuda

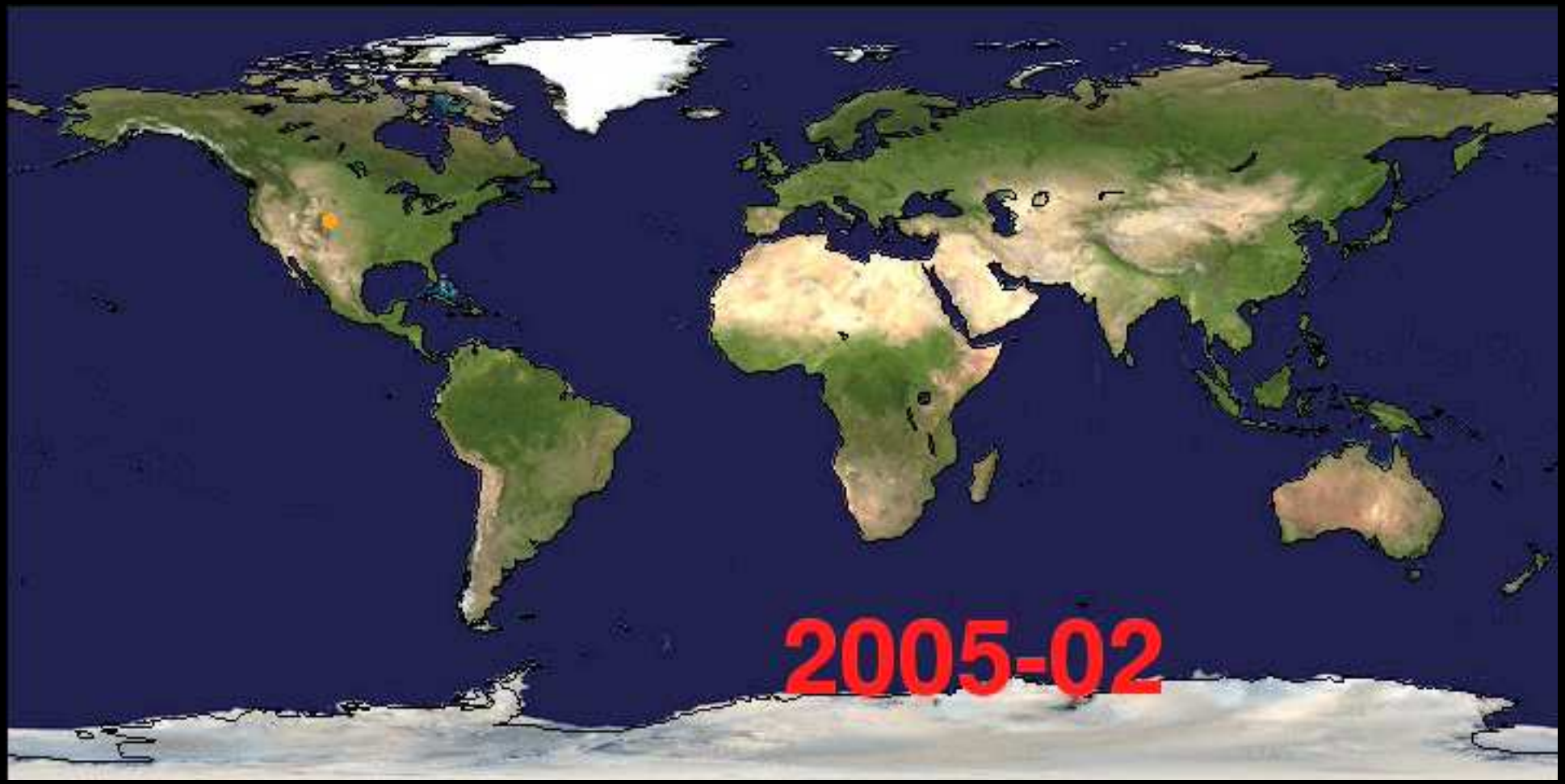
2 Asia/Pacific Region
1 Zambia
1 Syria
1 Sudan
1 Oman
1 Nigeria
1 Lebanon
1 Latvia
1 Kuwait
1 Kenya
1 Jamaica
1 Honduras
1 Georgia
1 Ethiopia
1 Dominican Republic
1 Cote D'Ivoire
1 Benin
1 Azerbaijan
1 Angola
1 Afghanistan

120+ countries
2750+ unique sites

Metrics

5436 com	21 sg	97 mx	7 ve	2 sn	1 sd
2869 edu	21 il	91 ch	7 si	2 om	1 rs
777 cn	21 id	90 mil	7 ph	2 np	1 pf
717 gov	21 be	79 no	7 my	2 nc	1 ne
464 jp	15 hu	68 hk	7 lt	2 mk	1 na
427 fr	14 th	62 se	7 is	2 lv	1 mz
416 de	14 hr	62 gr	7 fm	2 kz	1 gh
334 net	13 pe	55 tr	7 cat	2 kg	1 er
297 ca	11 ua	51 pl	6 lu	2 ke	1 eg
283 uk	11 ro	50 nz	5 cu	2 fj	1 dz
277 it	11 int	45 pt	4 info	2 cy	1 du
237 kr	11 co	39 cl	4 ec	2 bt	1 cx
213 org	9 vn	38 at	4 bm	2 biz	1 by
209 au	9 pk	37 za	3 tz	2 az	1 bw
201 tw	9 mo	28 cz	3 ma	2 am	1 bd
189 es	9 ir	27 us	3 eu	2 ae	1 ba
187 br	8 sk	26 dk	3 bo	1 www	1 aq
182 in	8 ee	24 ie	3 bb	1 ws	1 an
143 ru	8 cr	23 ar	2 yu	1 tn	
98 nl	8 bg	22 fi	2 uy	1 sv	

Metrics



The future

Fall 2009

Release of overhauled ESG-CET for CMIP5 and IPCC AR5

Winter 2009

Testing of ESG-CET gateways and node at > 5 sites worldwide

Spring-Summer 2010

Begin making CMIP5/IPCC AR5 data available to community

Thank you!

Questions?