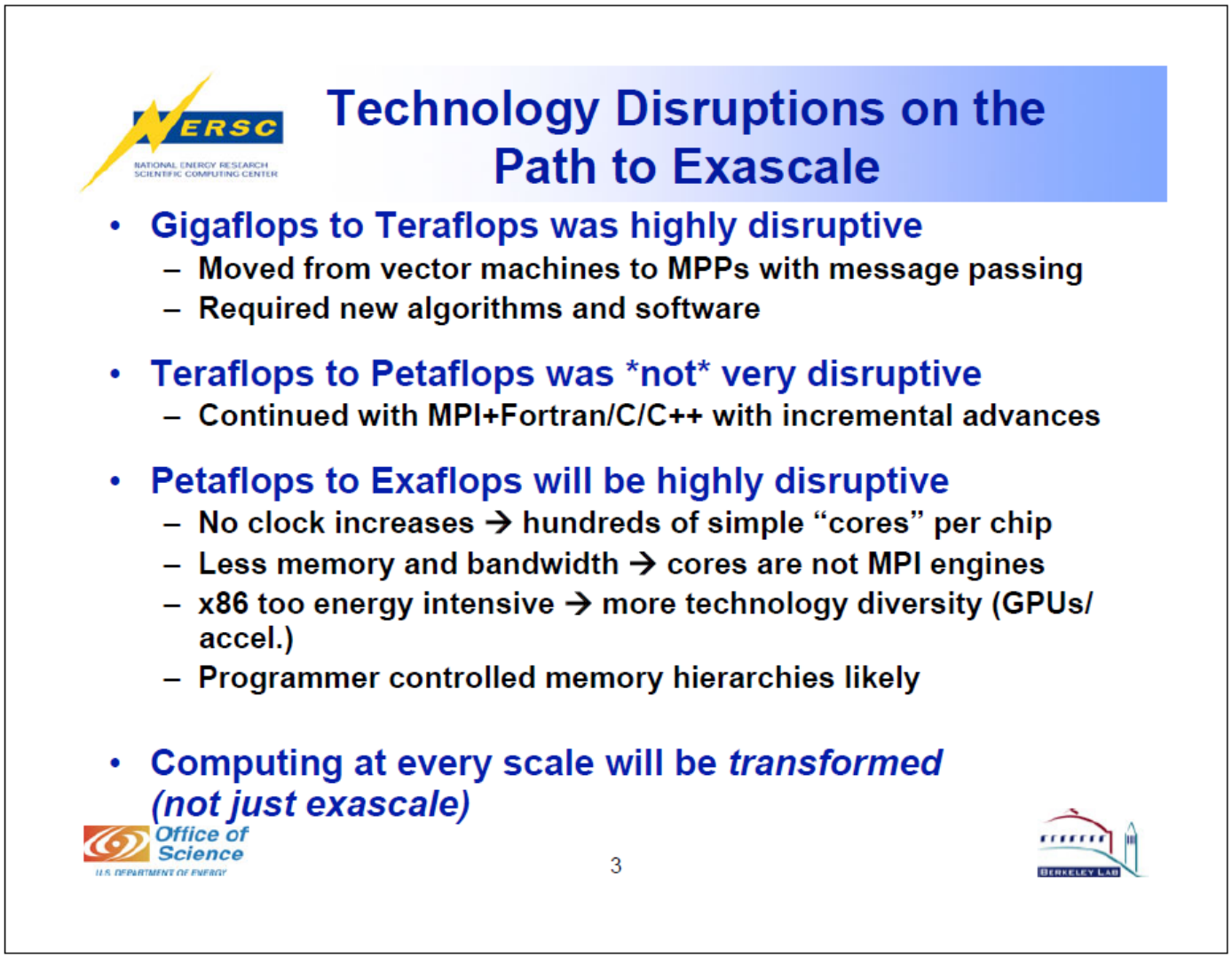
The influence of Moore's Law and friends on our computing environment!

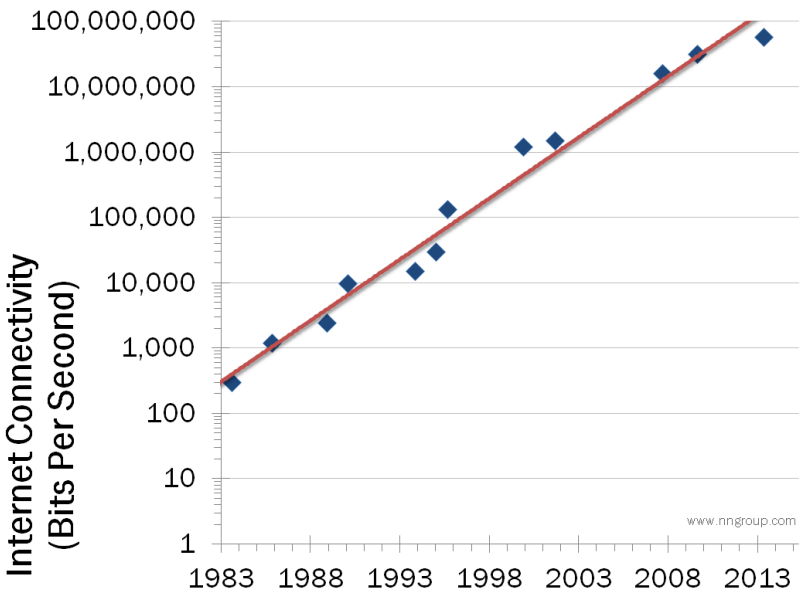
Bryan Lawrence
NCAS, STFC & The University of Reading

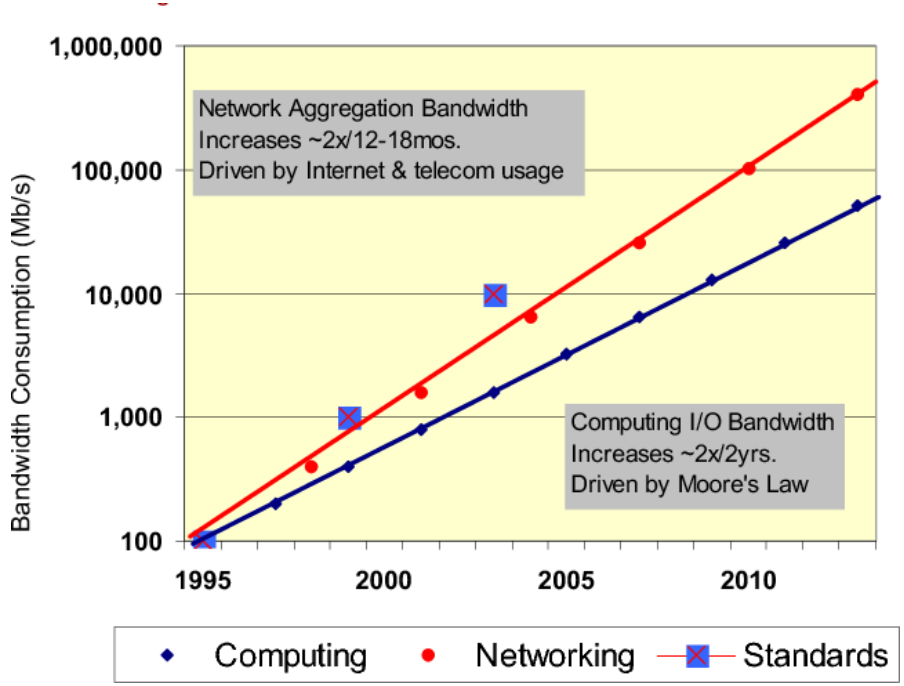
Outline

Moore's Law


NERSC
NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER
Technology Disruptions on the Path to Exascale
• Gigaflops to Teraflops was highly disruptive
– Moved from vector machines to MPPs with message passing
– Required new algorithms and software
• Teraflops to Petaflops was *not* very disruptive
– Continued with MPI+Fortran/C/C++ with incremental advances
• Petaflops to Exaflops will be highly disruptive
– No clock increases → hundreds of simple "cores" per chip
– Less memory and bandwidth → cores are not MPI engines
– x86 too energy intensive → more technology diversity (GPUs/ accel.)
– Programmer controlled memory hierarchies likely
• Computing at every scale will be *transformed (not just exascale)*
Office of Science
U.S. DEPARTMENT OF ENERGY
3
BERKELEY LAB

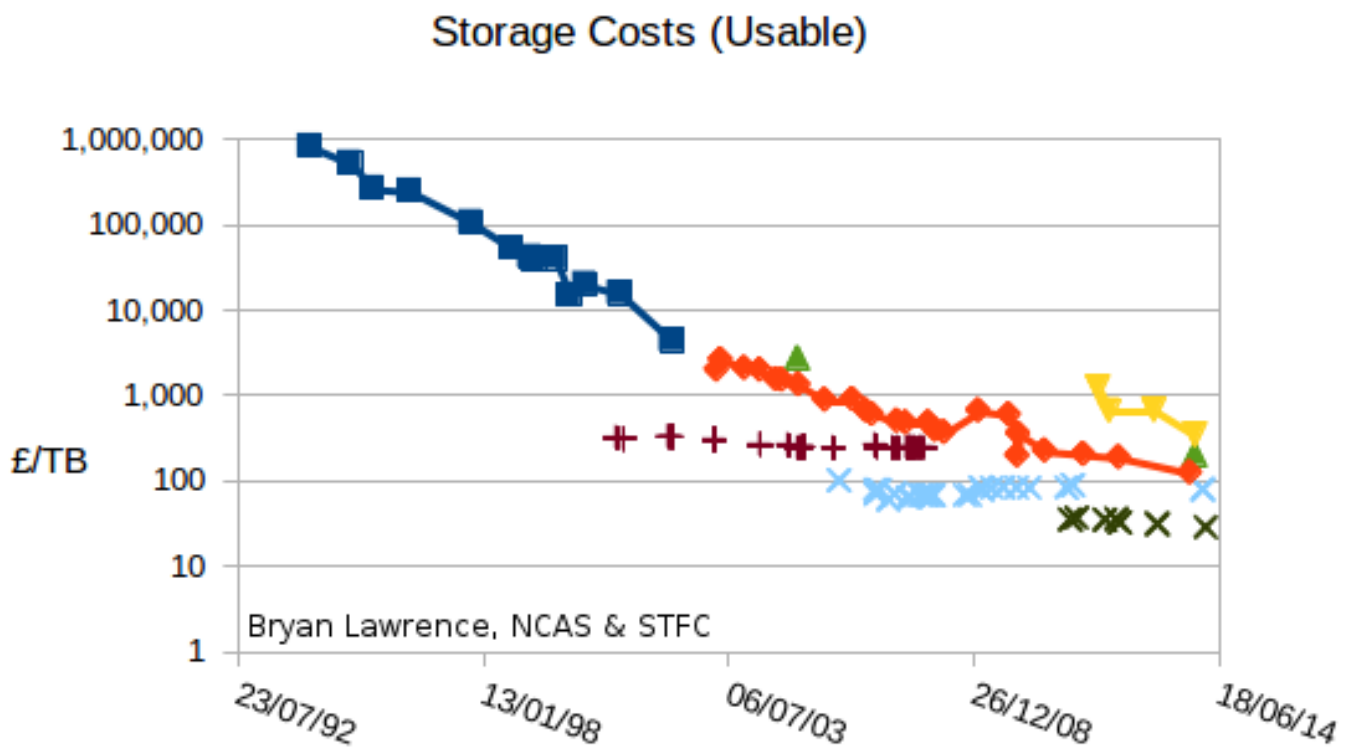
Nielsen's Law and Gilder's Law





End user bandwidth (Nielsen) slightly slower than Moore's law. Wide Area Network bandwidth (Gilder), faster than Moore's Law ... but I/O bandwidth (no name) isn't keeping up ...

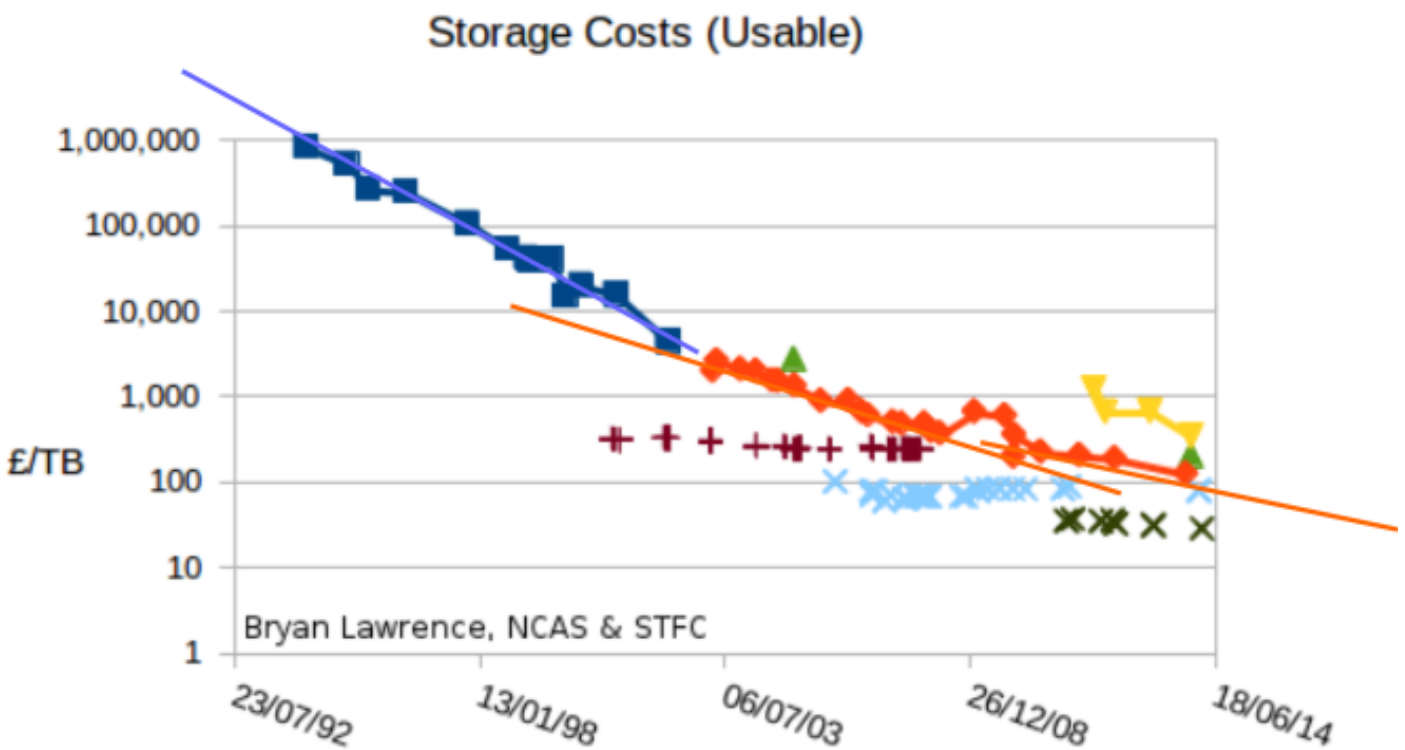
Kryder's Law



Solid objects: colours are different generations of disk. Crosses: different generations of tape.

(Data from Peter Chiu, Jonathan Churchill and Tim Folkes, STFC)

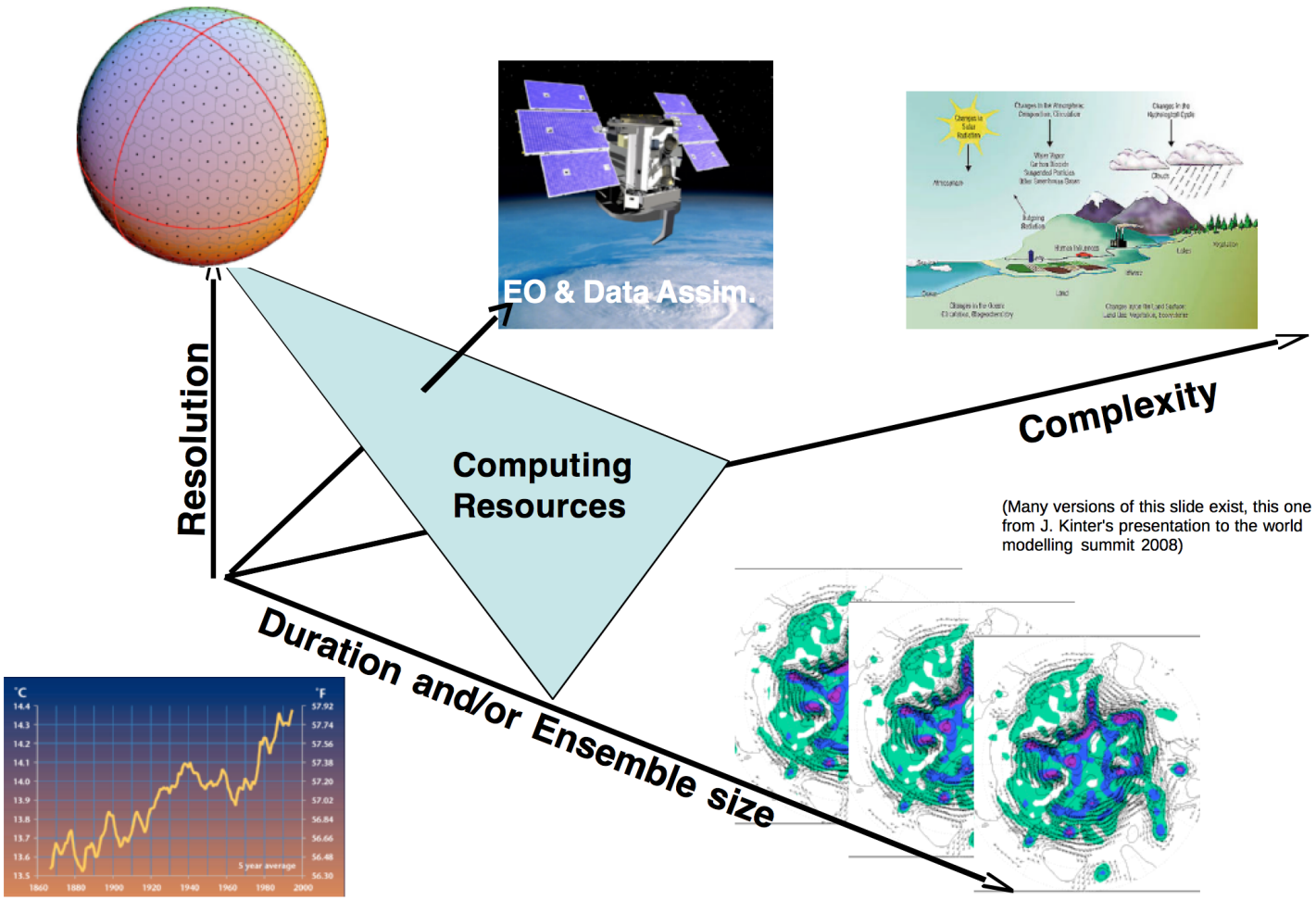
Kryder's Law



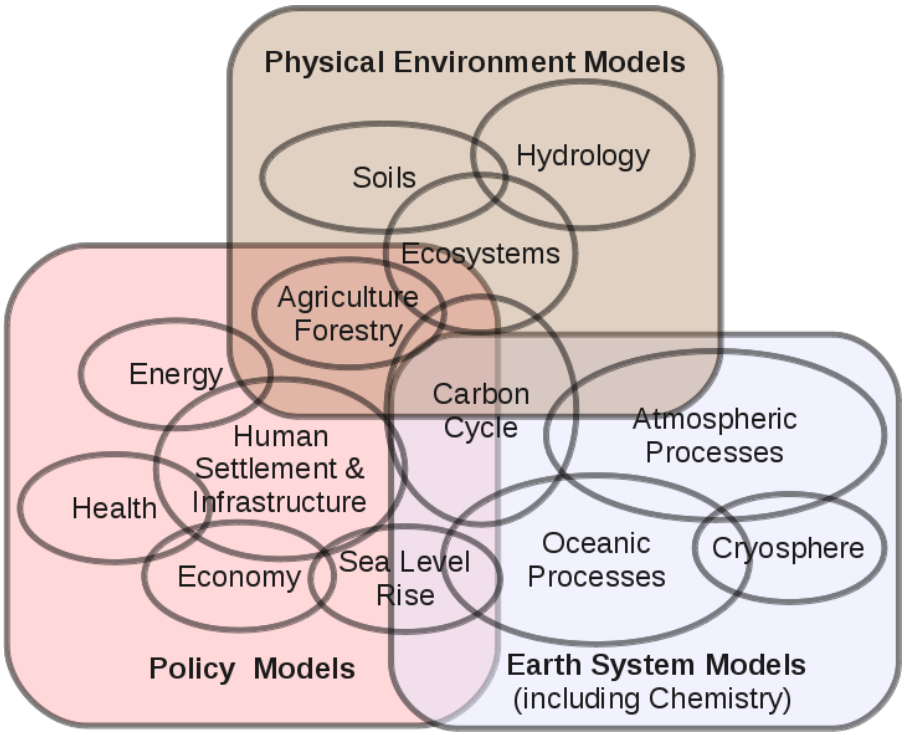
Solid objects: colours are different generations of disk. Crosses: different generations of tape.

Kryder's Law definitely slowing down! Plenty of mileage still in tape though!

Give me more computing?



Communities



Many interacting communities, each with their own software, compute environments etc.

Figure adapted from Moss et al, 2010

Archer

Systems | History

System	Year	Vendor	Cores	Rmax (GFlop/s)	Rpeak (GFlop/s)
eServer pSeries p5 575 1.5 GHz	2006	IBM	2,560	12,940	15,360
eServer pSeries p5 575 1.5 GHz	2005	IBM	1,536	7,395	9,216
eServer pSeries 690 (1.7 GHz Power4+)	2004	IBM	1,600	6,188	10,880
pSeries 690 Turbo 1.3GHz	2002	IBM	1,280	3,406	6,656

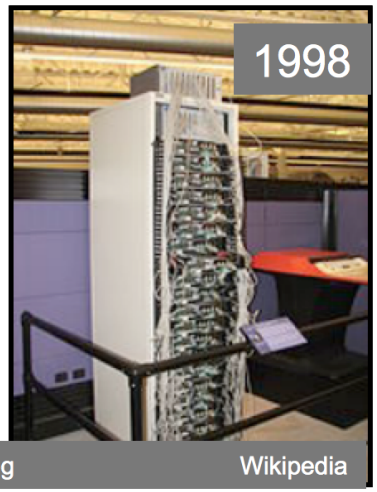
Ranking

List	Rank	System	Vendor	Total Cores	Rmax (TFlops)	Rpeak (TFlops)	Power (kW)
06/2014	25	Cray XC30, Intel Xeon E5 v2 12C 2.700GHz, Aries interconnect	Cray Inc.	76,192	1,367.5	1,645.7	
11/2013	19	Cray XC30, Intel Xeon E5 v2 12C 2.700GHz, Aries interconnect	Cray Inc.	76,192	1,367.5	1,645.7	

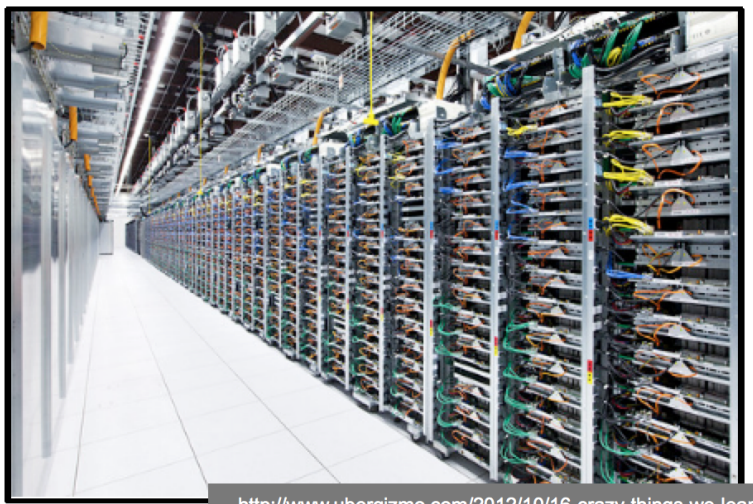
HPCx and Archer are 12 years apart!
We were using the UM then, and we are now!
But we couldn't claim that our workhorse simulations have scaled proportionally!

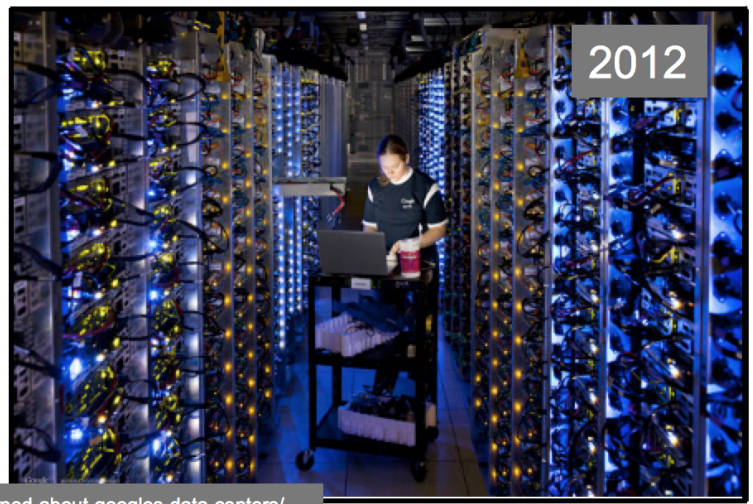
CEDA Evolution

Eerily similar to Google



http://infolab.stanford.edu/pub/voy/museum/pictures/display/GoogleBG.jpg Wikipedia



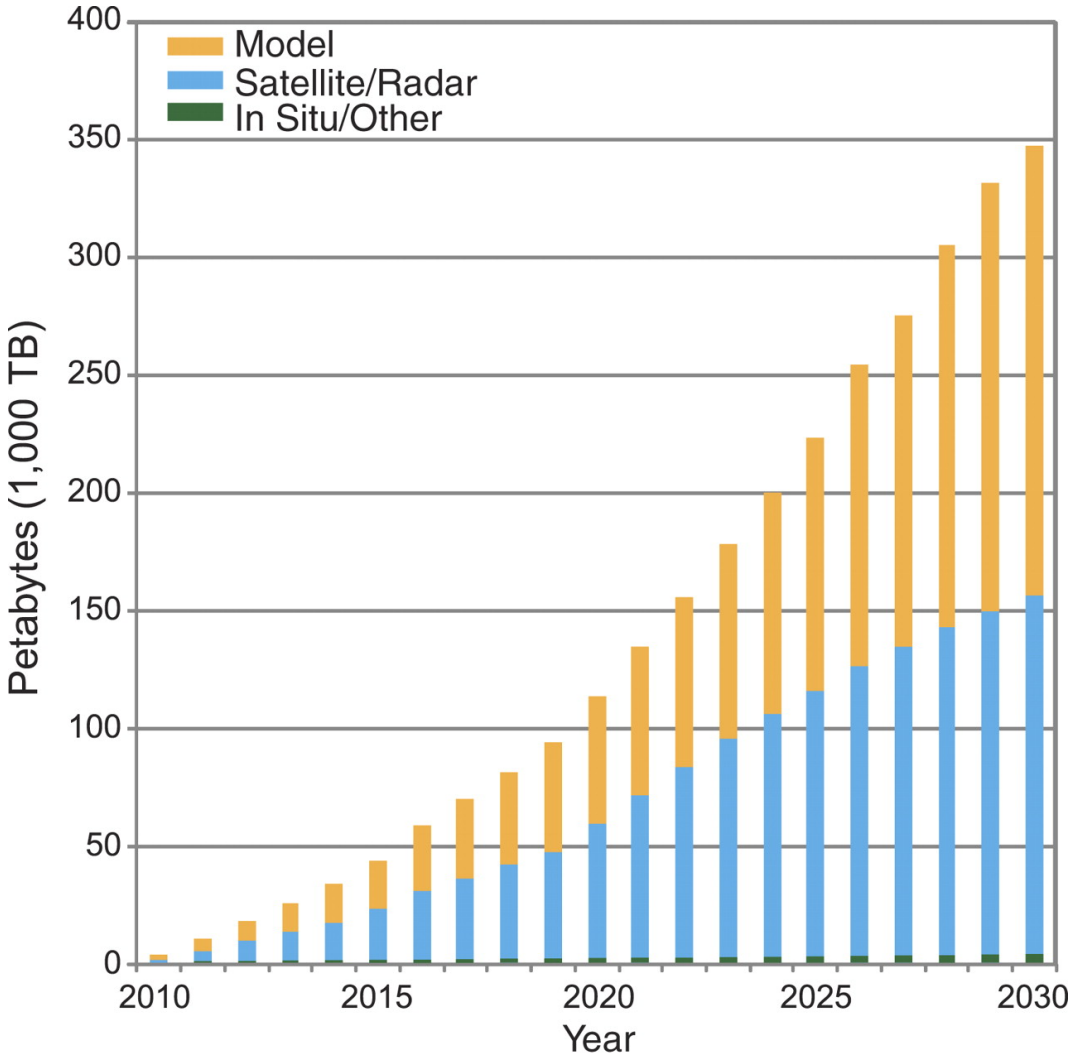


http://www.ubergizmo.com/2012/10/16-crazy-things-we-learned-about-googles-data-centers/,
http://blogs.wsj.com/digits/2012/10/17/google-servers-photos/

Global Data Archival

Fig. 2 The volume of worldwide climate data is expanding rapidly, creating challenges for both physical archiving and sharing, as well as for ease of access and finding what's needed, particularly if you're not a climate scientist.

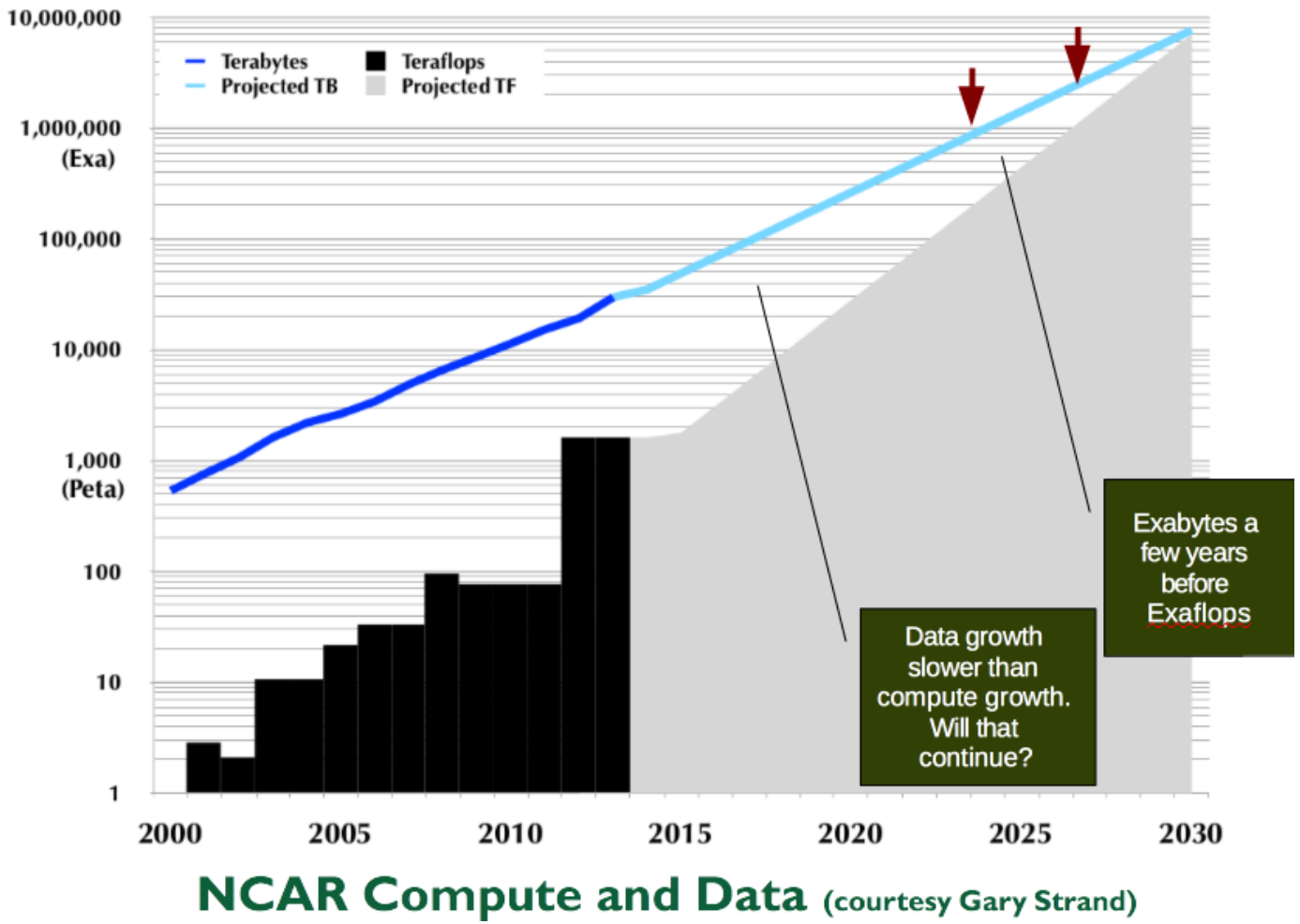
(BNL: Even if you are?)



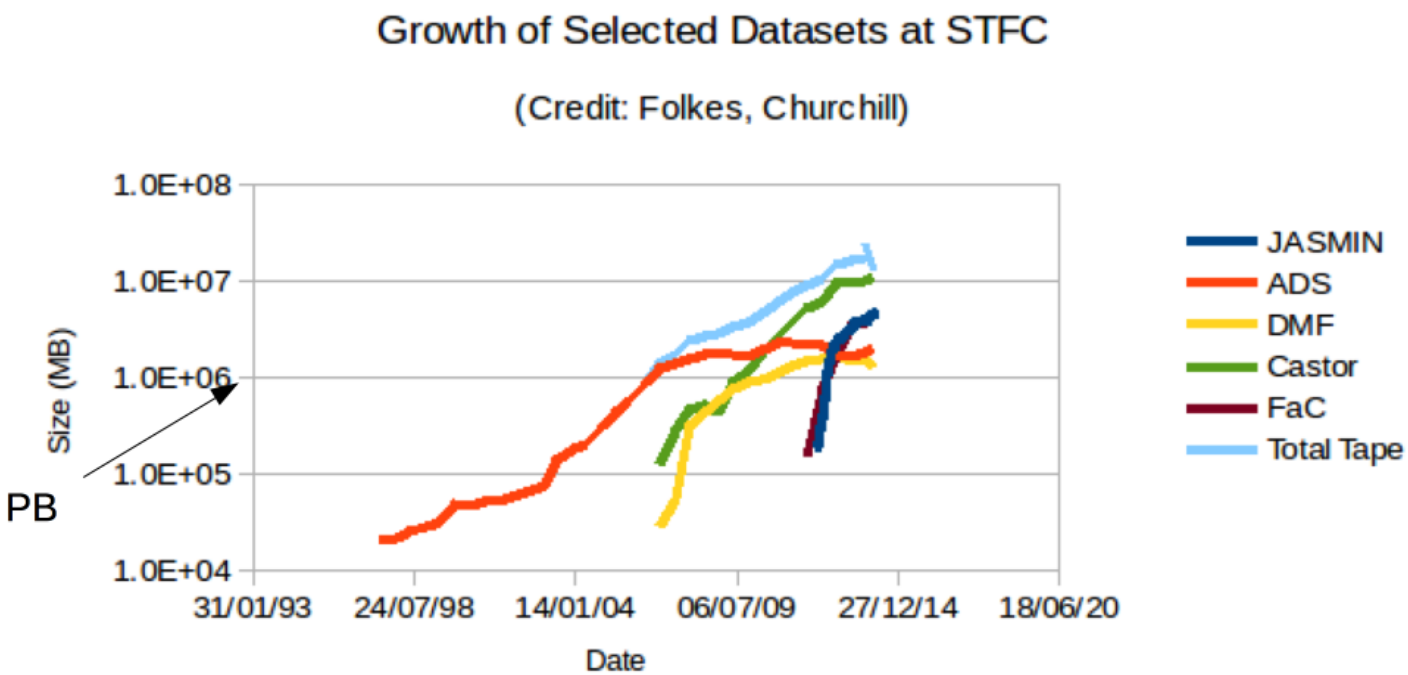
J T Overpeck et al. Science 2011;331:700-702

NCAR

Storage, and power for storage, will dominate NCAR's compute budget within a few years! (Rich Loft, 2014).



STFC and CEDA



Predictions for JASMIN in 2020? 30 - 85 PB of unique data[1]!
But we think we could only fit only 30 PB in the physical space available!

([1]Not including CMIP6, which might be anything from 30-300 PB, but we hope at the lower end!)

U.S. National Academy

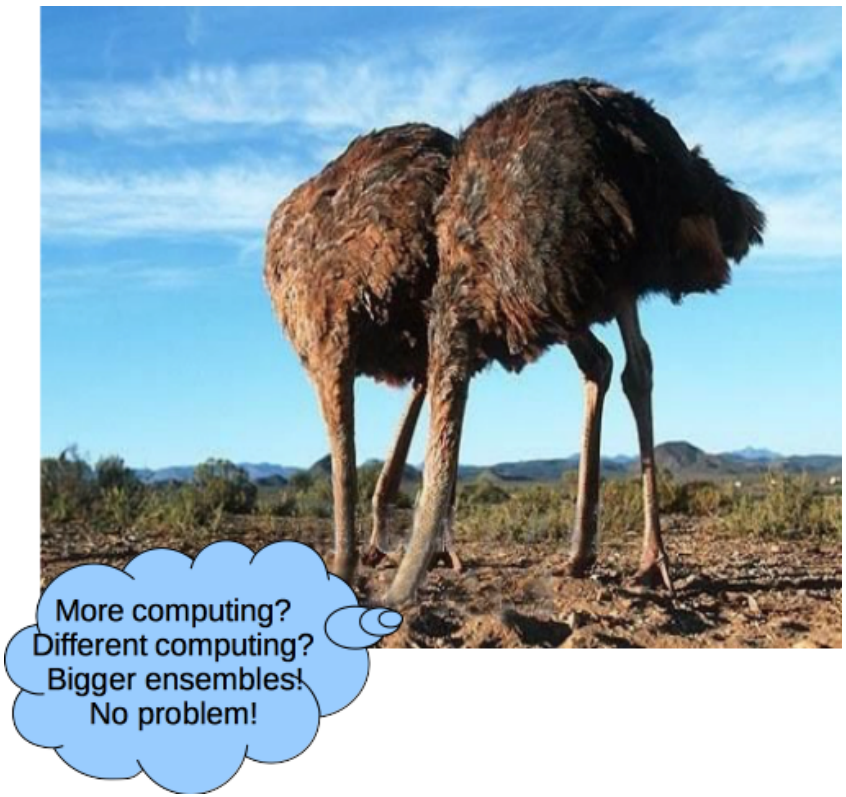
"Without substantial research effort into new methods of storage, data dissemination, data semantics, and visualization, all aimed at bringing analysis and computation to the data, rather than trying to download the data and perform analysis locally, it is likely that the data might become frustratingly inaccessible to users"

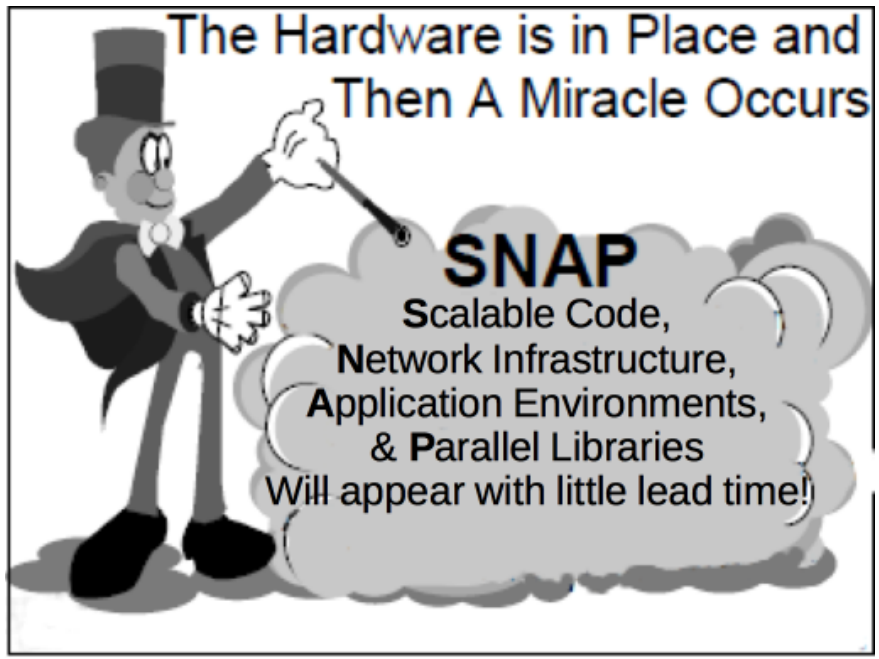
A National Strategy for Advancing Climate Modeling, 2012

Semantic Analysis: "substantial research effort" "new methods" "computation to data" "rather than trying to download" "frustratingly inaccessible" (to whom?)

What about software?

From some:





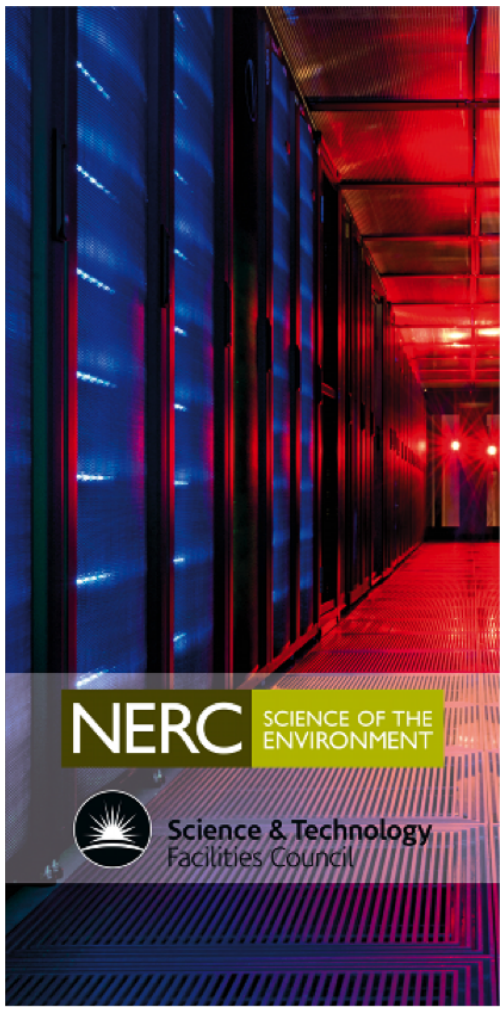
Summary so far

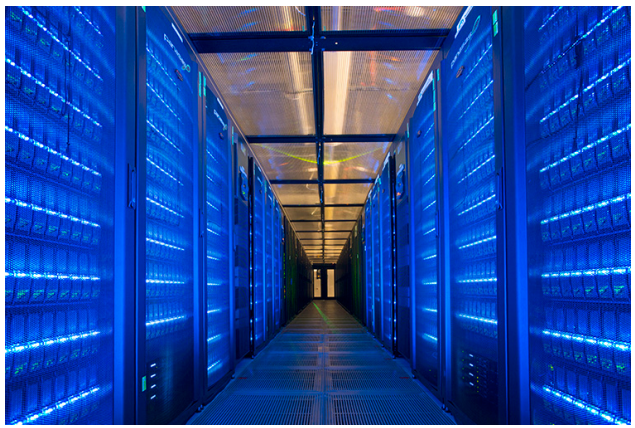
The technology drivers are tending towards infinitely cheap computing and infinitely expensive data systems!

(?tending?: tending, I just said tending, nothing ever asymptotes ok!)

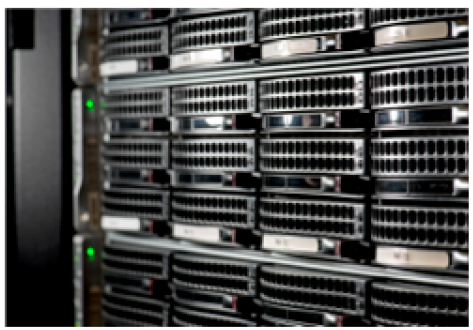
However, while the computing might be (relatively) cheap, exploiting it is likely to become harder and harder

Gratuitous Photos

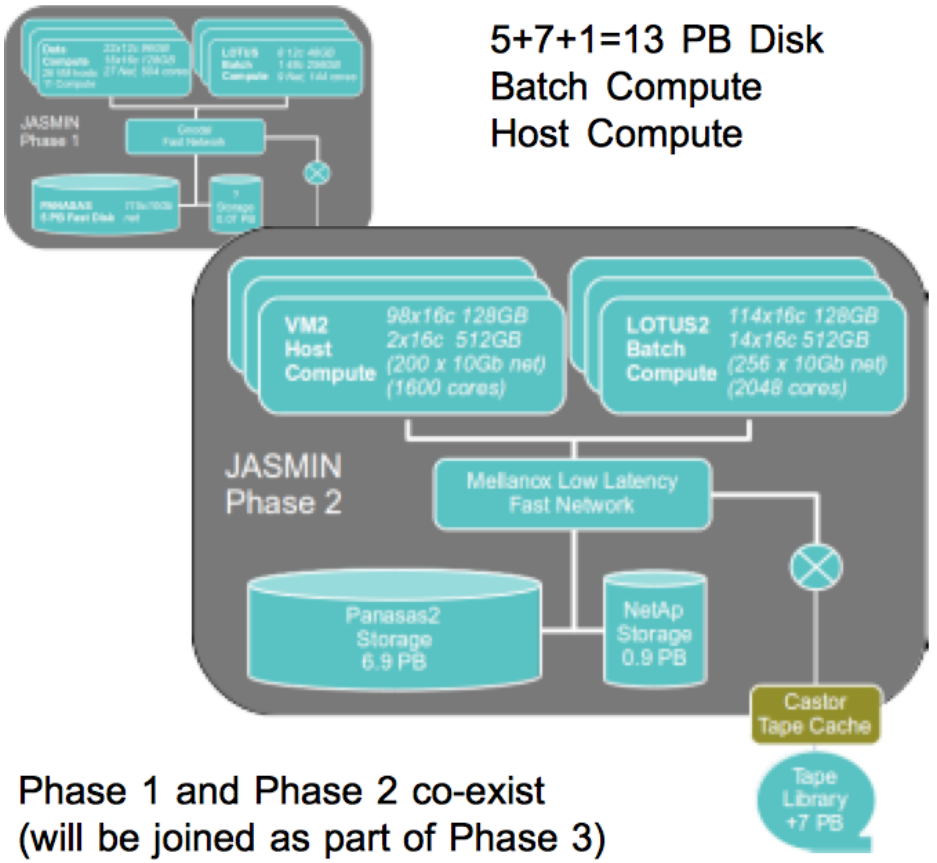


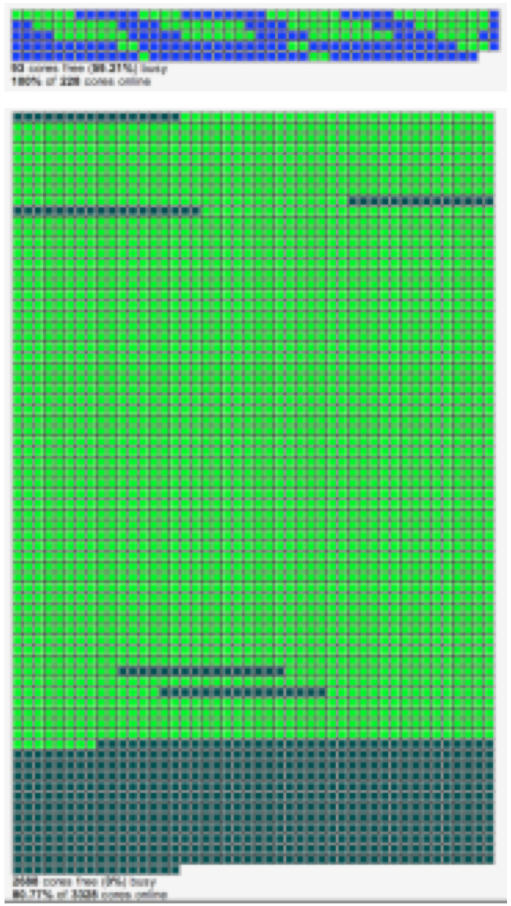


- 12 PB Fast Storage
- 1 PB Bulk Storage
- Elastic Tape
- 4000 cores: half deployed as hypervisors, half as the "Lotus" batch cluster.



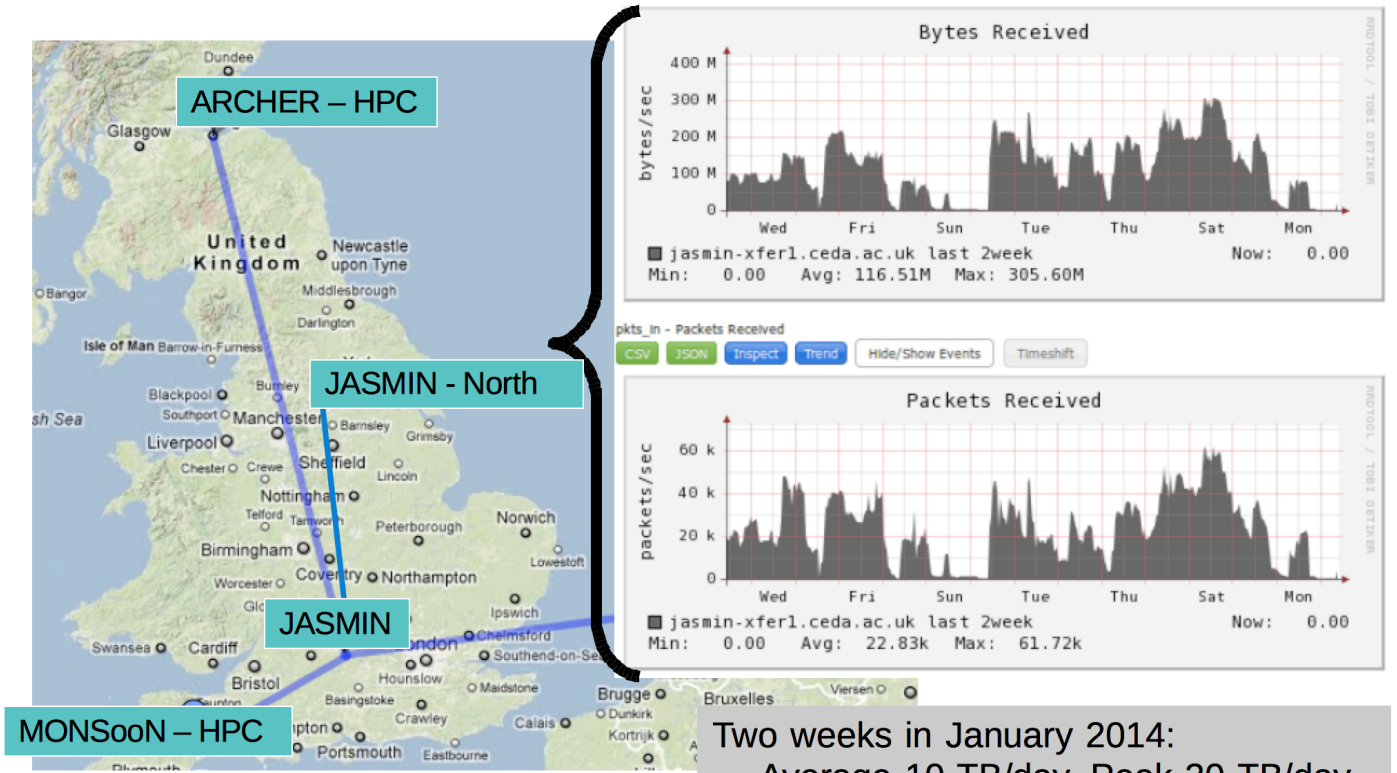
JASMIN - The upgrade





Doubled storage, but increased compute by a factor of six!

Making use of the bandwidth



We've had some network upgrades since then. The bottom line is that you should be able to move TBs per day - to JASMIN at least.

But while hardware is a necessary condition for making progress, it's far from sufficient!

It's really all about the software: the environment and the code!

Rest of talk:

- Portability, Scalability and Performance
- Cloud Environments (including the JASMIN Analysis Platform)
- The G8 ExArch Project
- Training

CMS support for Portability, Scalability and Performance

Simon Wilson

Model Support and Performance Specialist

Simon Read

UKESM1 core team – diagnostics

Karthee Sivalingam

JWCRP Porting, new architectures, high-res support

Jeff Cole

Model Support and Software Tools Specialist

Andy Heaps

System Administration and Visualisation Specialist

- Infrastructure Support (MONSOON, ARCHER, PUMA)
- Portability, Performance
- Mesoscale Support
- Parallel Software Engineering
- Input/Output, Diagnostics
- HPC Futures

Charles Roberts

CF-Python/HighRes Support

Rosalyn Hatcher

Software Management Specialist MMG

William McGinty

Modelling Support for the mesoscale community, Data

David Hassell

Software Tools specialist – CF-Python - IS-ENES2

Grenville Lister

Head of CMS, modelling and HPC Support

Luke Abraham

UKCA development and support

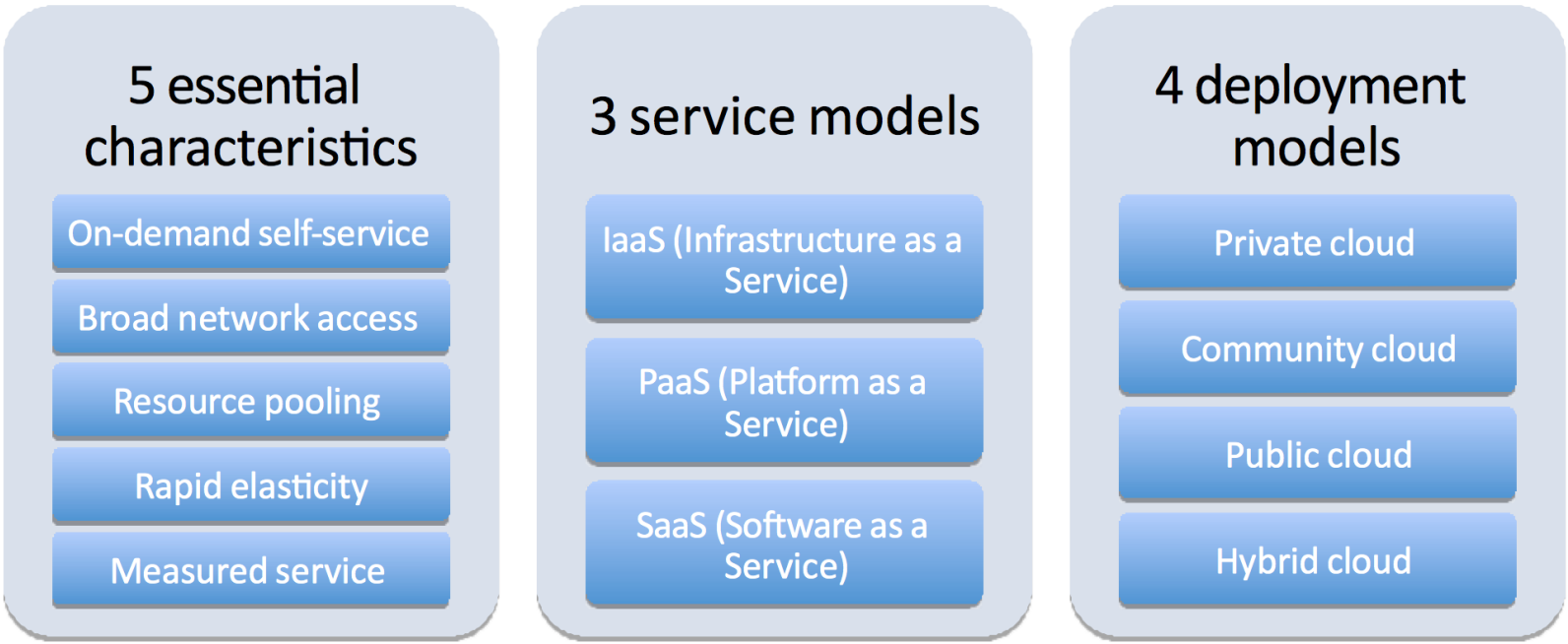
Annette Osprey

Modelling Support for Earth System Models, NEMO HPC futures

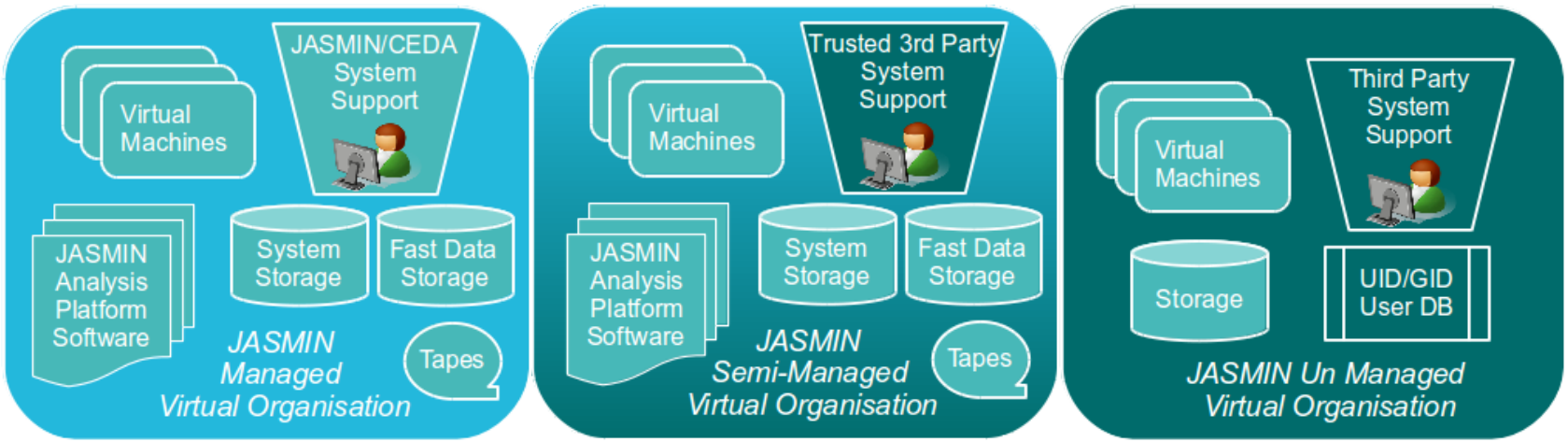
An introduction to the cloud

Why cloud? Remember all this communities, with their own software environments?

> *"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction."* — *NIST SP800-145*



Virtual Organisations



Platform as a Service $\longrightarrow$ Infrastructure as a Service

NCAS itself will run a semi-managed virtual organisation (with multiple group work spaces), but large groups within NCAS can themselves also run virtual organisations.

Some Special Virtual Organisations

CEDA: Centre for Environmental Data Archival

- Will provide archival services for the community.
- Data held in the archive will be managed, and made available to all the managed and semi-managed V.O.s directly (and indirectly to the un-managed V.O.s).
- Will provide "generic" access platforms for virtual organisations that do not wish to manage their own platforms and users who do not belong to specific virtual organisations.

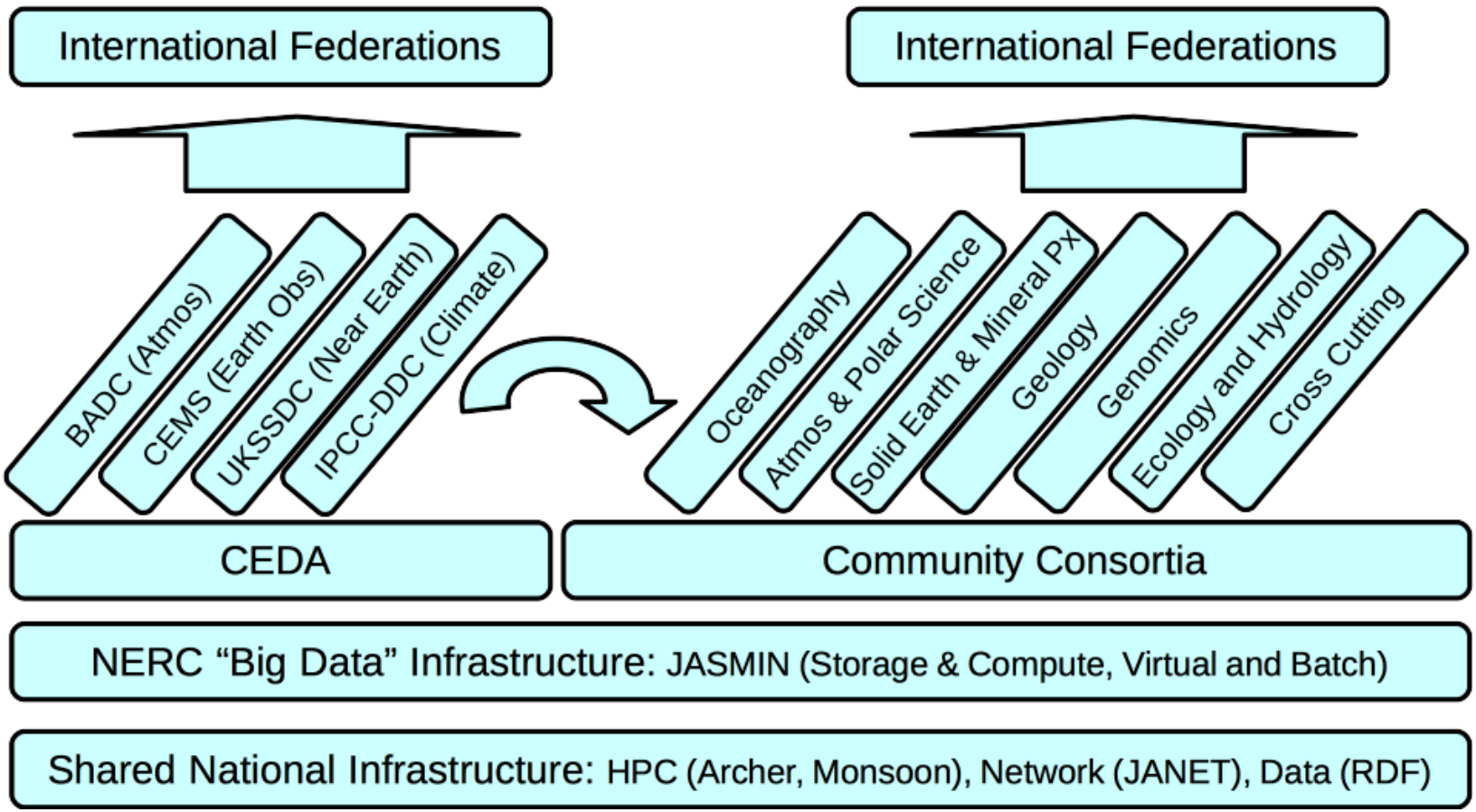
EOS Cloud

- Cloud services for the environmental 'omics community
- Delivered by JASMIN on behalf of the Centre for Ecology and Hydrology

CEMS: The facility for Climate, Environment and Monitoring from Space

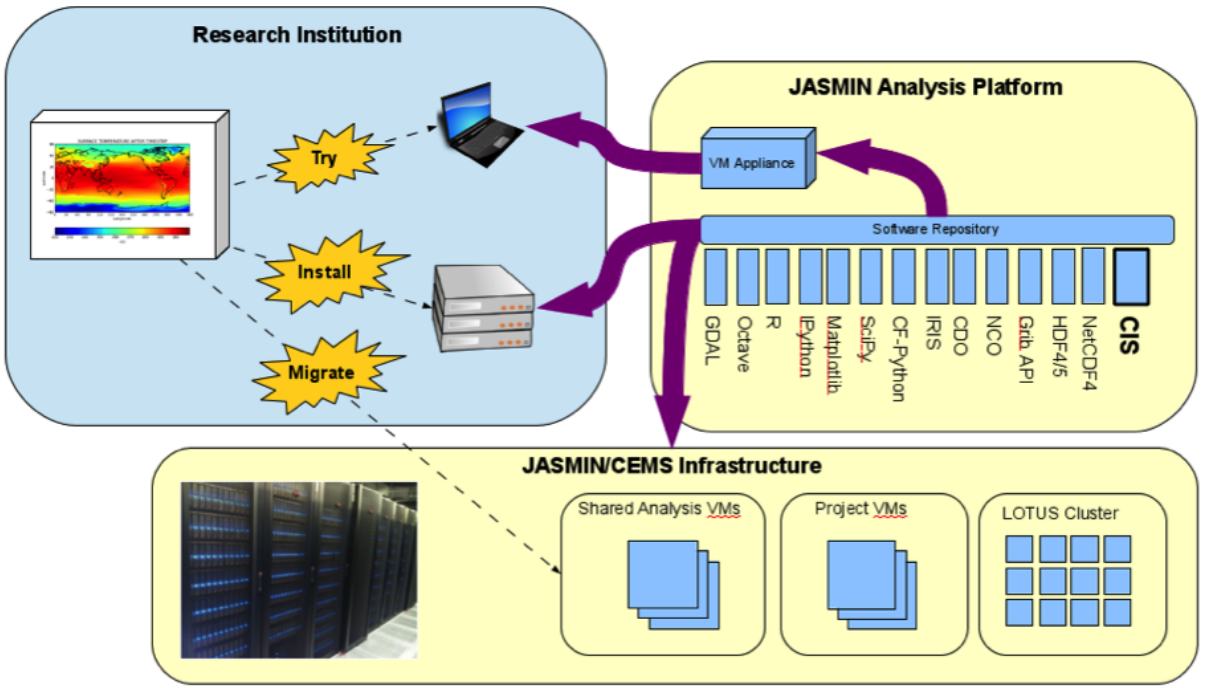
- Will acquire and archive (via CEDA) key third party datasets needed by the NERC science community.
- Will provide services for the Earth Observation Community, in particular, in partnership with Satellite Applications catapult (SAC), the UK and European space industry.
- The academic component will run on JASMIN, the bulk of the industrial component, in the SAC, with access to CEDA data.

The "headline" virtual organisations



Platform as a Service: The JASMIN Analysis Platform

- Multi-node infrastructure requires a way to install tools quickly and consistently
- The community needs a consistent platform where ever they need them.
- Users need help migrating analysis to JASMIN.



http://proj.badc.rl.ac.uk/cedaservices/wiki/JASMIN/AnalysisPlatform

What JAP Provides

Standard Analysis Tools

- NetCDF4, HDF5, Grib
- Operators: NCO, CDO
- Python Stack: Numpy, SciPy, Matplotlib, IRIS, cf-python, cdat_lite, IPython
- GDAL, GEOS
- NCAR Graphics, NCL
- R, octave
- IDL (. . . but)
- . . .

Parallelisation and Workflow

- Python MPI bindings
- Jug (simple python task scheduling)
- IPython notebook
- IPython-parallel
- JASMIN Community Intercomparrison Suite

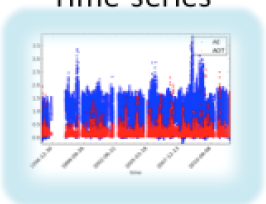
Science Codes

- JASMIN Community Intercomparrison Suite
- . . . soon: validation tooling (e.g ESMVal)

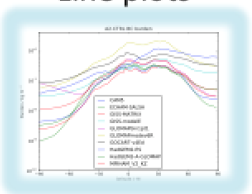
Community Intercomparison Suite

(Philip Stier, University of Oxford)

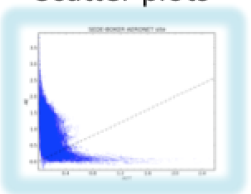
Time-series



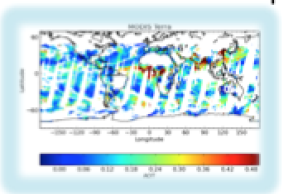
Line plots

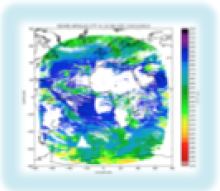


Scatter plots

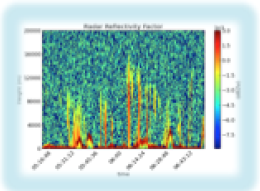


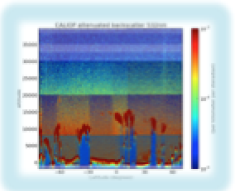
Global plots



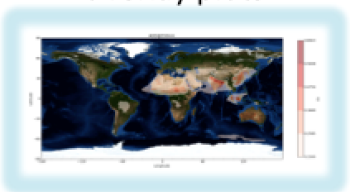


Curtain plots

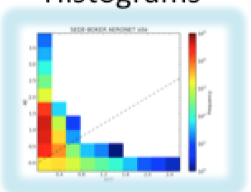




Overlay plots



Histograms



Dataset	Format
AERONET	Text
MODIS	HDF
CALIOP	HDF
CloudSAT	HDF
AMSRE	HDF
TRMM	HDF
CCI aerosol & cloud	NetCDF
SEVIRI	NetCDF
Flight campaign data	RAF
Models	NetCDF

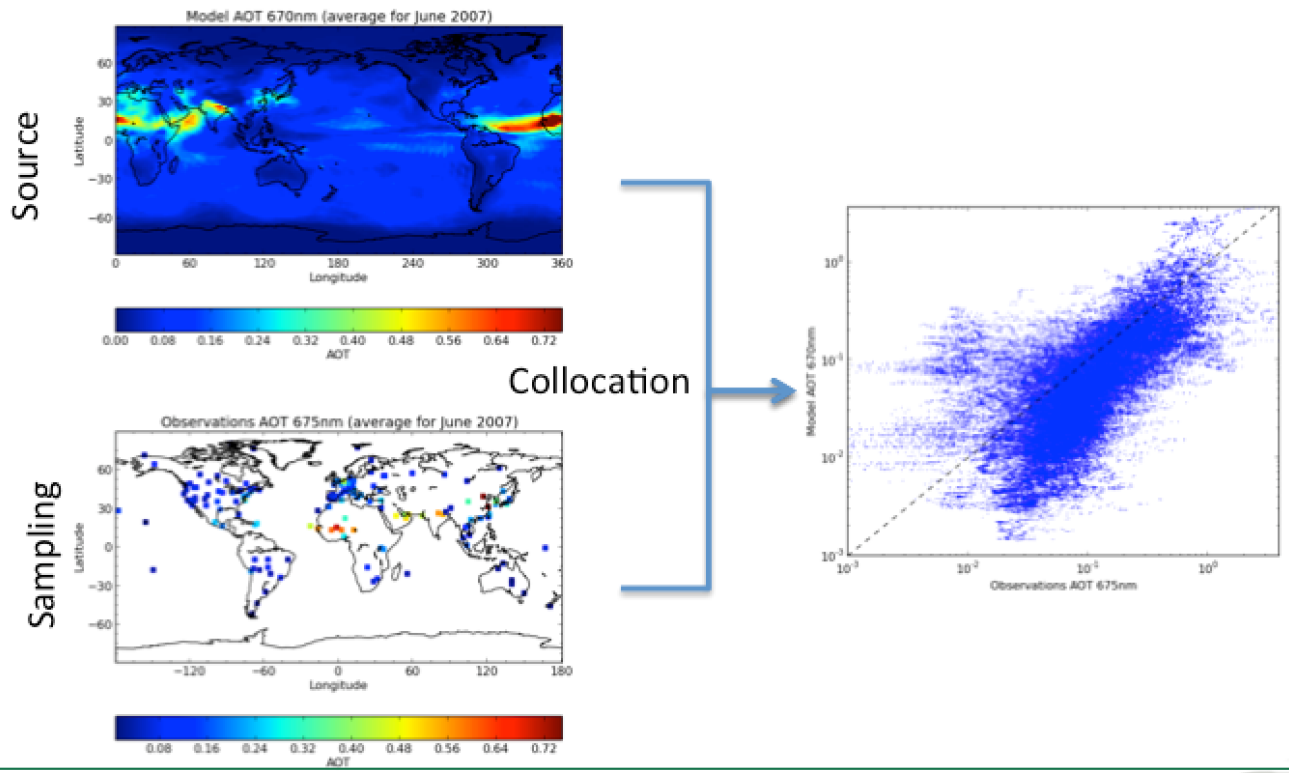
CIS Command line tools

(Philip Stier, University of Oxford)

```
cis  col  <variable>:<source file>  <sampling file>:colocator=lin  -o <new file>
cis  plot <variable>:<new file> <variable>:<sampling file> --type comparativescatter \
             --logx --xlabel 'Observations AOT 675nm' --xmin 1.e-3 --xmax 10   \
             --logy --ylabel 'Model AOT 670nm' --ymin 1.e-3 --ymax 10
```

Model gives global output every 3 hours for a full month

Observations are day-time site measurements, every 15 min for a full month

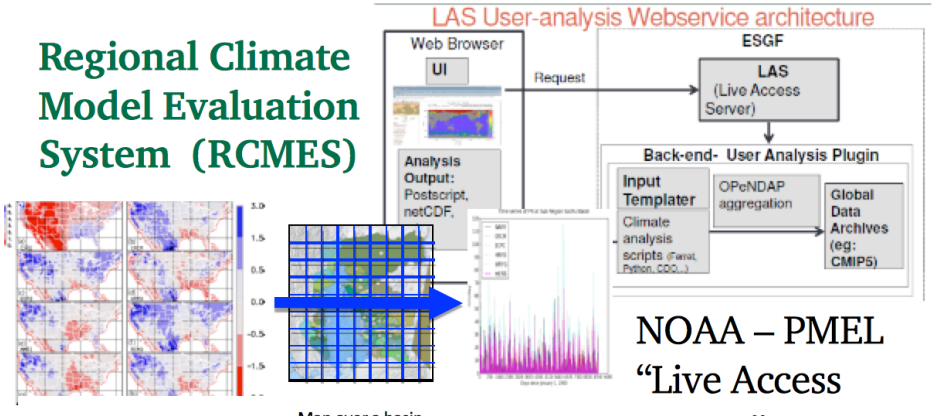


The ExArch Project - Taking compute to the data!

ExArch: Climate analytics on distributed exascale data archives (Juckes PI, G8 funded)

Martin Juckes, V. Balaji, B.N. Lawrence, M. Lautenschlager, S. Denvil, G. Aloisio, P. Kushner, D. Waliser, S. Pascoe, A. Stephens, P. Kershaw, F. Laliberte, J. Kim, S. Fiore

Regional Climate Model Evaluation System (RCMES)

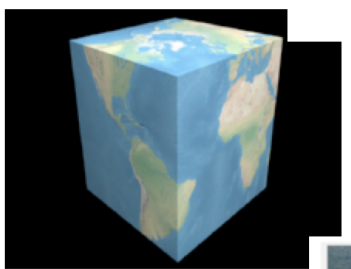


Observation/Model rainfall

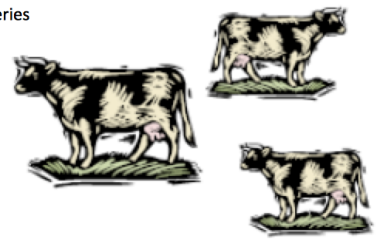
Map over a basin using an area-matching method

Get basin-mean time series

NOAA – PMEL "Live Access Server"



CMCC parallel data analytics framework



CEDA OGC Web Services

http://climate4impact.eu/

Training

- It's getting harder at the fundamental level to develop/improve models, even as we provide tools to more easily run models (yet even with them, it's hardly easy).
- It's getting harder at the fundamental level to develop new analysis routines for bigger (and/or more complex) data, even as we provide better and better toolsets.
  $\longrightarrow$ so we all need to "up-skill".
- Whatever you do, you need better control over how you record what you've done!
  $\longrightarrow$ so we need better version control.

All of which leads to increased requirement for generic courses like "software carpentry", the "NCAS Introduction to Scientific Computing", the "NCAS Training for the Unified Model" (and UKCA top up courses), and the WRF summer school.

Summary

- Moore's Law and friends are both providing opportunities and causing problems: difficulty of parallelisation (for both compute and data analysis), and cost of storage!
- NCAS CMS and CEDA providing software support for a range of relevant problems.
- Large NERC investment in JASMIN hardware is now being followed by (not so large) investments in software, but to make the best use of that, you need to:
  - Learn some new tricks (how to exploit cloud, and the JASMIN analysis platform), and
  - Possibly avail yourself of more training, especially in good digital hygiene (i.e. version control and provenance keeping).

P.S. and provide feedback, making progress on all of this requires feedback, not knock back