The road to exascale for climate science: crossing borders or crossing disciplines, can one do both at the same time?

Brief abstract: The grand challenges of climate science have significant infrastructural implications, which lead to requirements for integrated e-infrastructure - integrated at national and international scales, but serving users from a variety of disciplines. We begin by introducing the challenges, then discuss the implications for computing, data, networks, software, and people, beginning from existing activities, ~~and looking out as far as we can see (spoiler alert: not far!)~~

Bryan Lawrence

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL
University of Reading
Science & Technology Facilities Council

Motivation – from the large

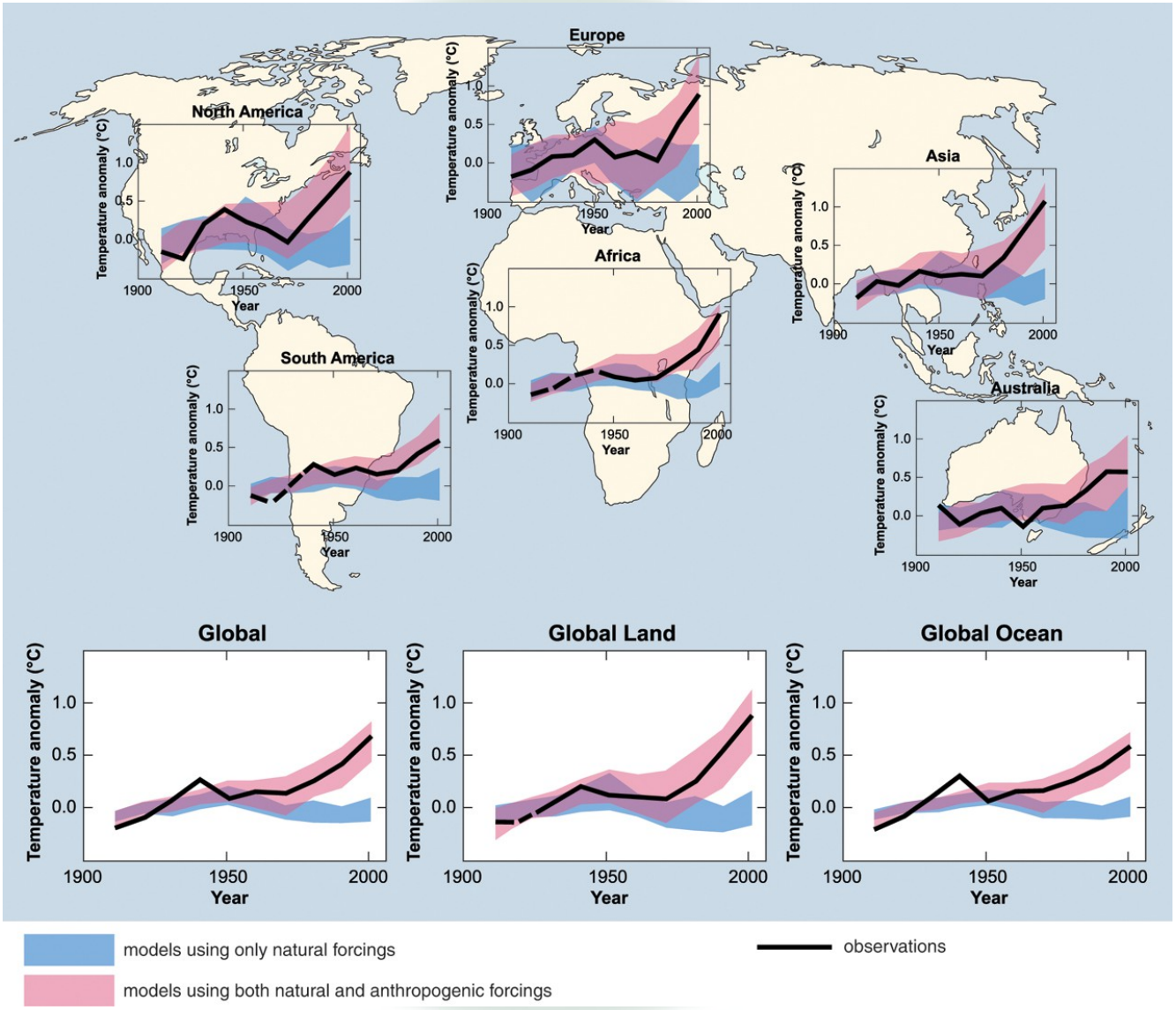


Fig 2.5 AR4 Synthesis Report

Motivation – to the small

How will climate change affect the global distribution of malaria?

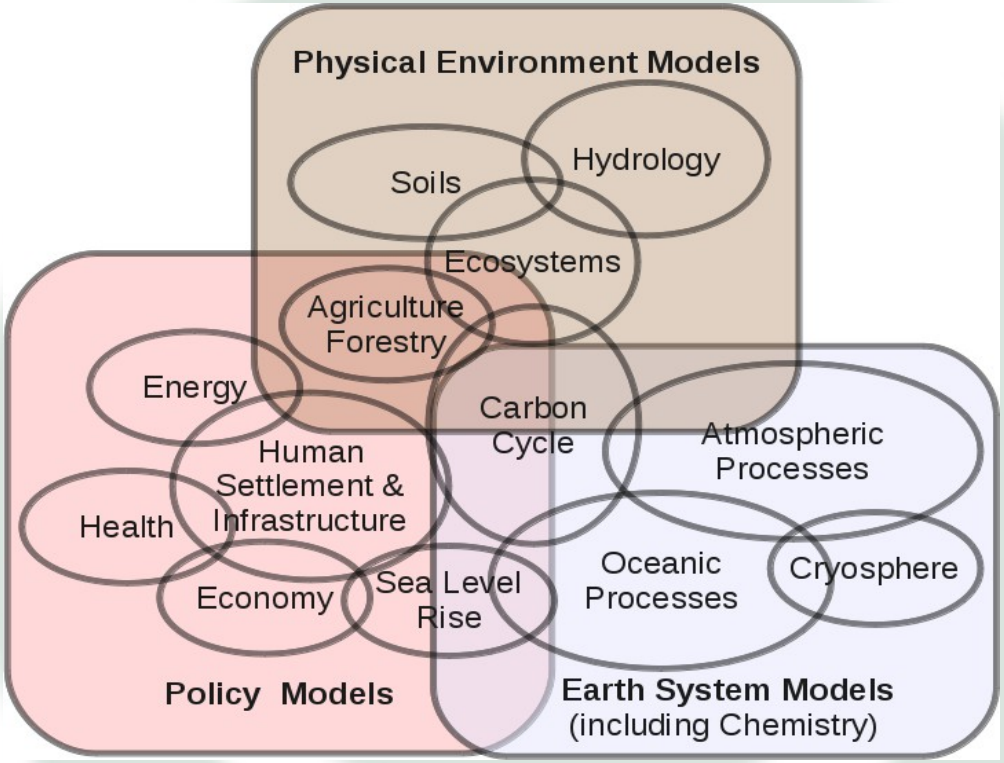
July 2007 Tewkesbury flood: 3B€ loss!

Can we predict risk into the future?

How will climate change affect the incidence of road and rail closures due to landslides?

What would be the impact of leakage from an oil and gas well in UK waters on the national economy, coastal and marine biodiversity and the well-being of the population affected?

Many, many processes, many, many communities!



Interconnected communities have problems which require coupling of models and sub-models between communities!

Not just a technical problem … language problems … scientific understanding problems … and …

(Figure adapted from Moss et al., 2010).

James Lovelock at the Geological Society, Burlington House, 5th May 2011

Science is still divided into co-existing disciplines each with its own language, journals and forceful defenders. We are tribal animals and such a trait is hard to resist.

What price sharing infrastructure then?

Can we share models?

(not enough time to talk about this, meet me in the bar!)

Can we share infrastructure (in particular, data infrastructure)?

Outline: The Big Trends

Science Drivers:

- More Direct Numerical Simulation
- More Interdisciplinarity
- More Data
- Smarter Algorithms

Infrastructure Drivers:

- More (not much faster) Processors
- Cheaper Disk
- Better Network (everywhere)
- Better Software Tools

Funding Driver:

- More efficiency
- (aka Spend Less)

(UK institutional landscape)

Collaboration Environments

- Climate as an example

Infrastructure

- Global (ESGF)
- National (JASMIN)

Putting the pieces together.

Direct Numerical Simulation

Primarily mathematical representation of a complex system of processes

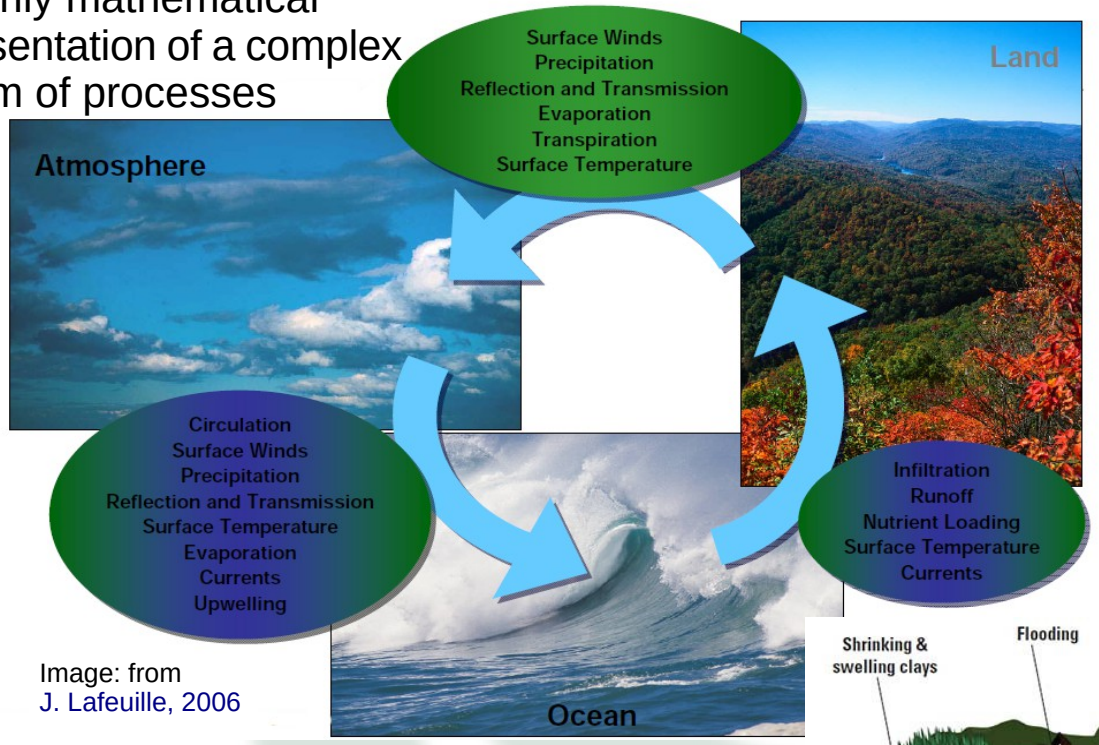
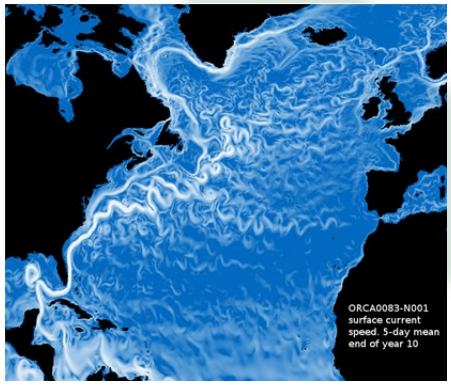
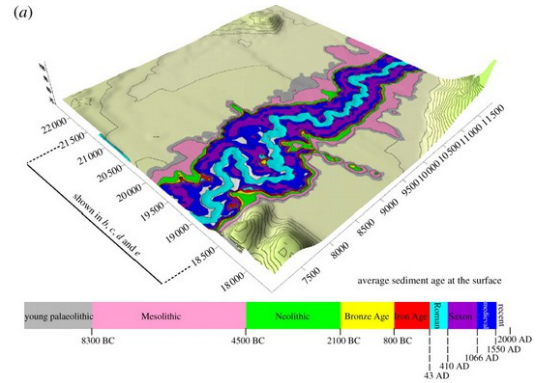
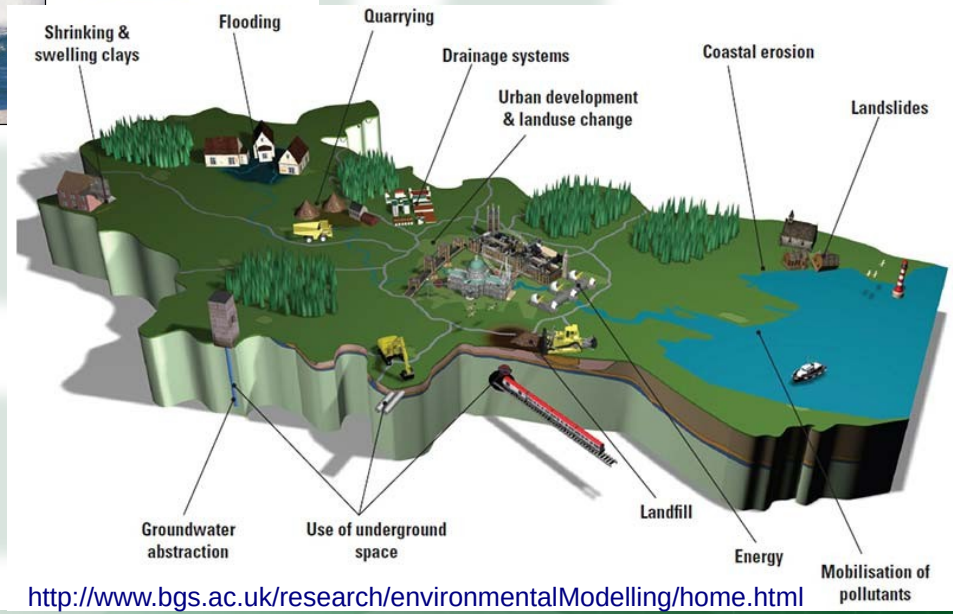


Image: from
J. Lafeuille, 2006





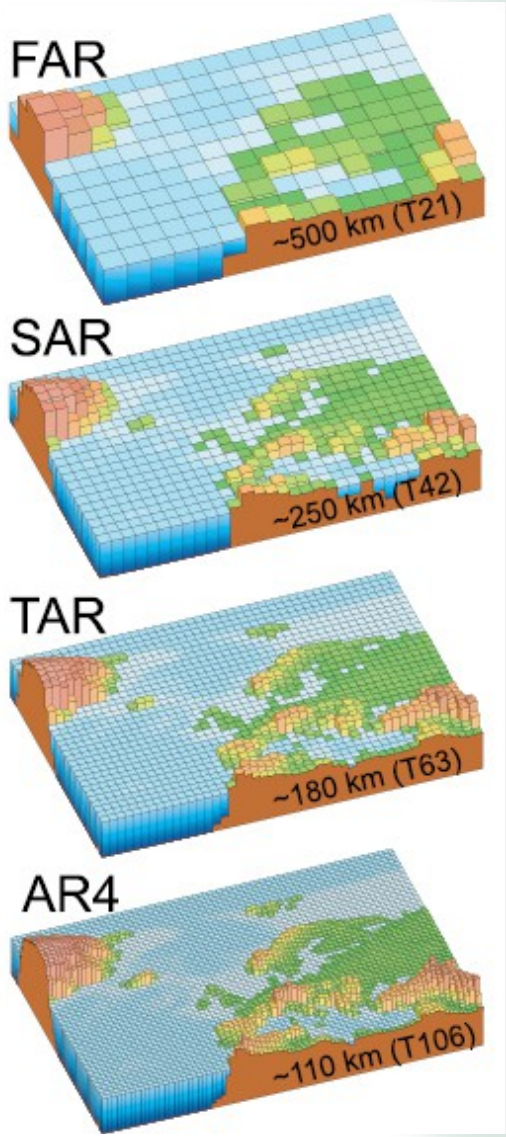
Coulthard and Van De Wiel lDoi:
10.1098/rsta.2011.0597

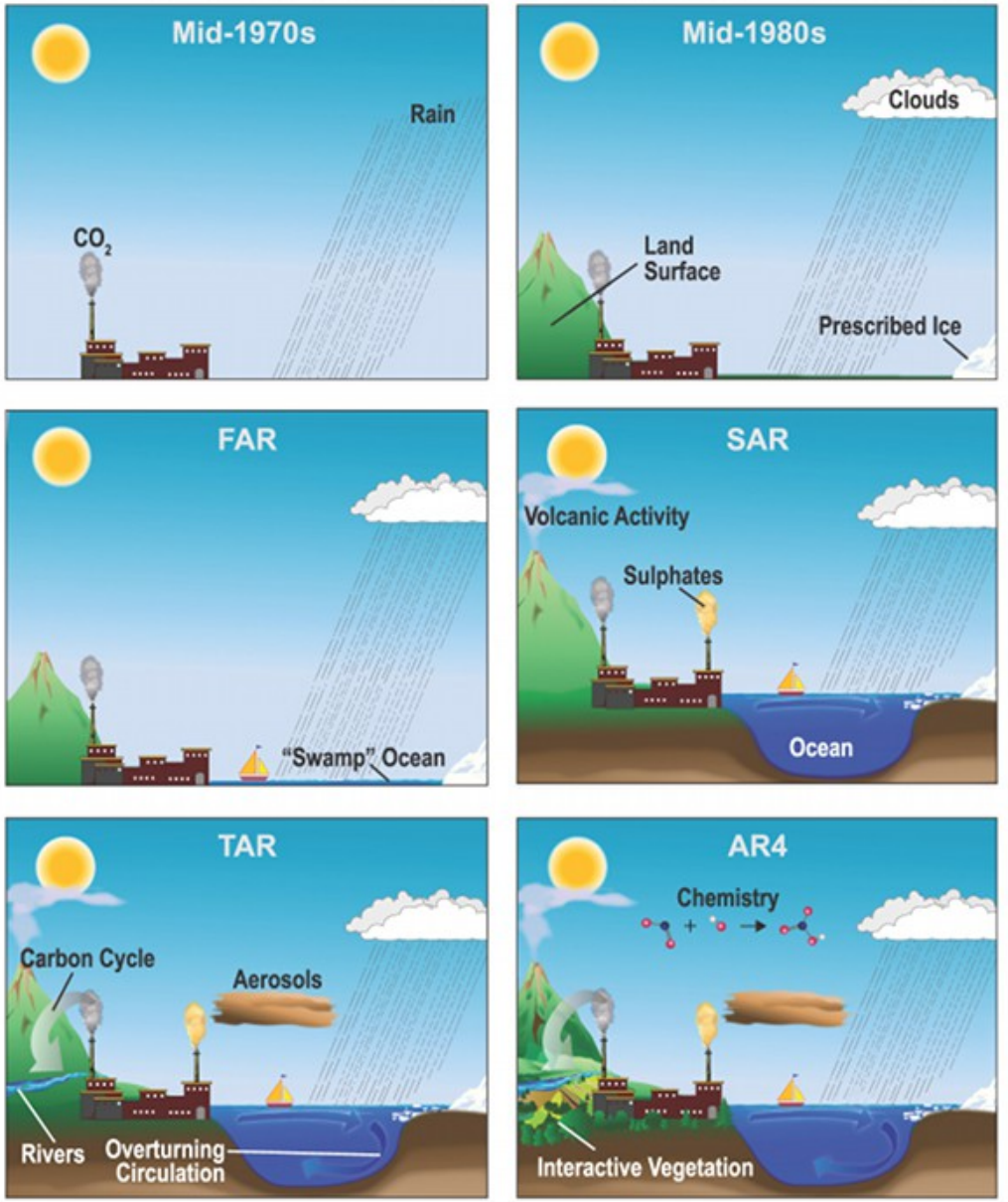


http://www.bgs.ac.uk/research/environmentalModelling/home.html

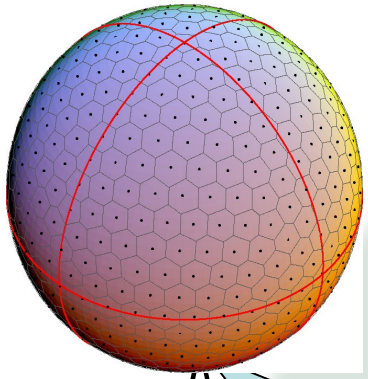
The World in Global Climate Models

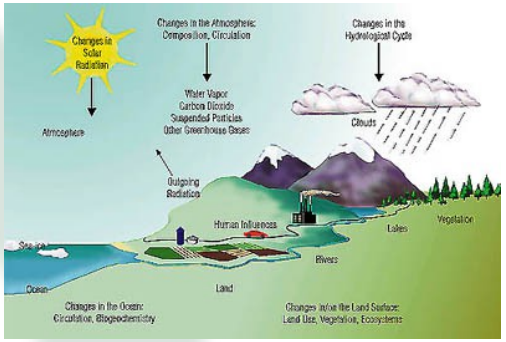
FAR:1990
SAR:1995
TAR:2001
AR4:2007
AR5:2013





Give me more computing: Whither Numerical Modelling?





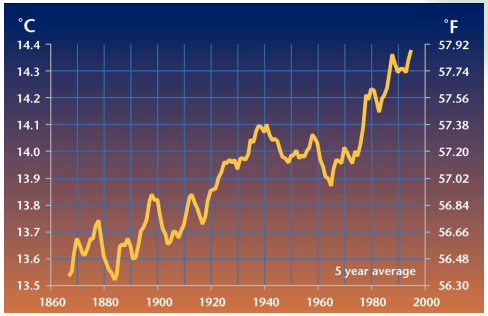
Resolution

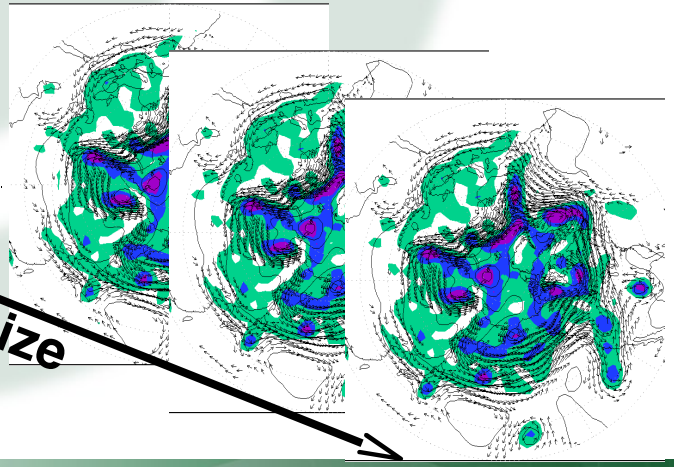
Enhanced computing resources produce MORE DATA

Complexity

Duration and/or Ensemble size

(Many versions of this slide exist, this one from J. Kinter's presentation to the world modelling summit 2008)

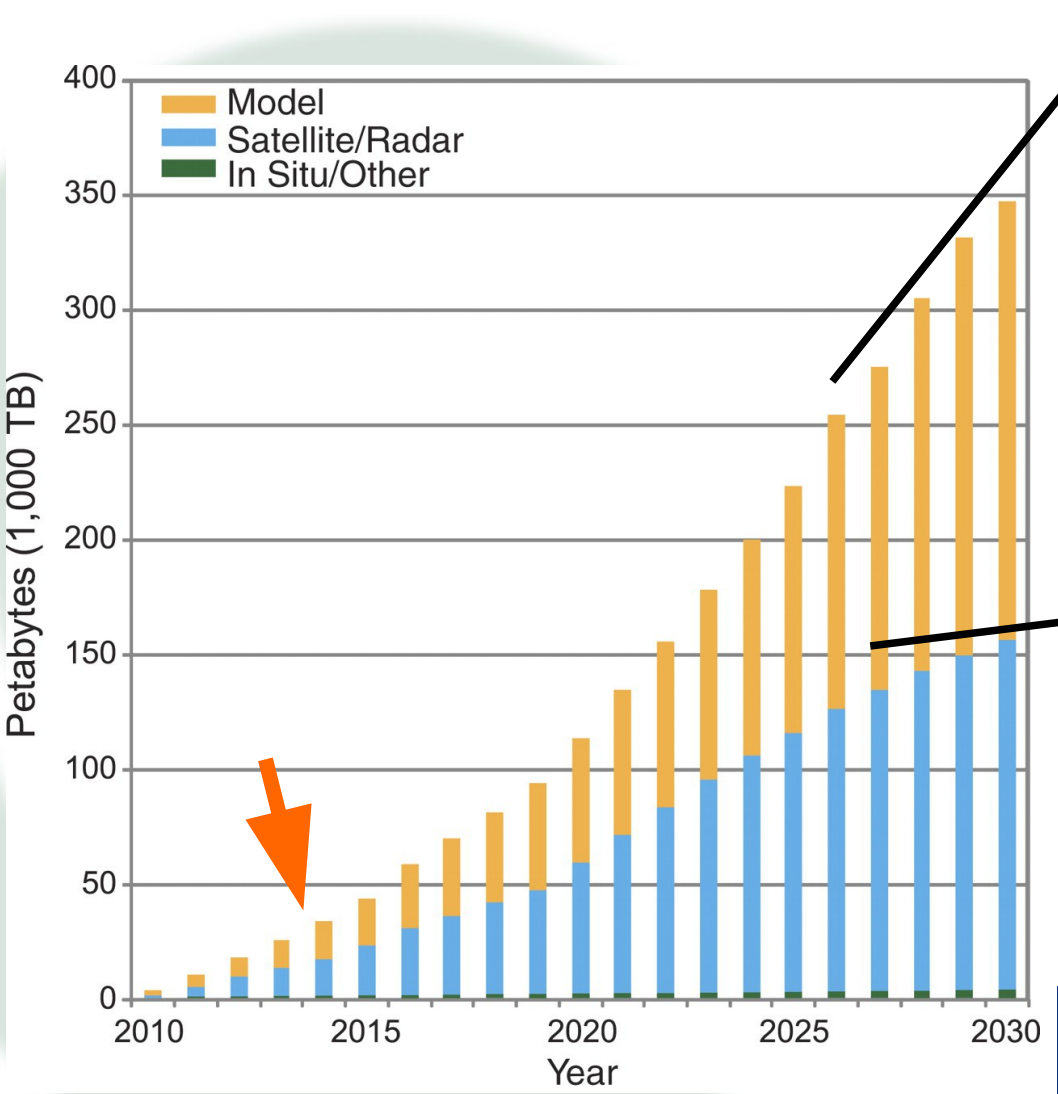




Gross underestimates ?!

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 "big data" specialist (who is?)

(Their words, not mine!)



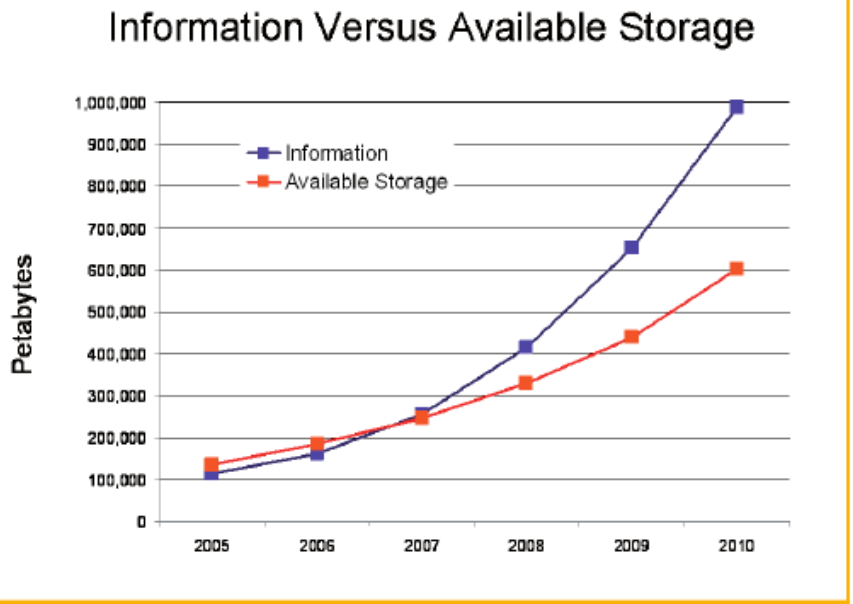
NB Climate
>
EO

Science
AAAS

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

Storage can't keep up!

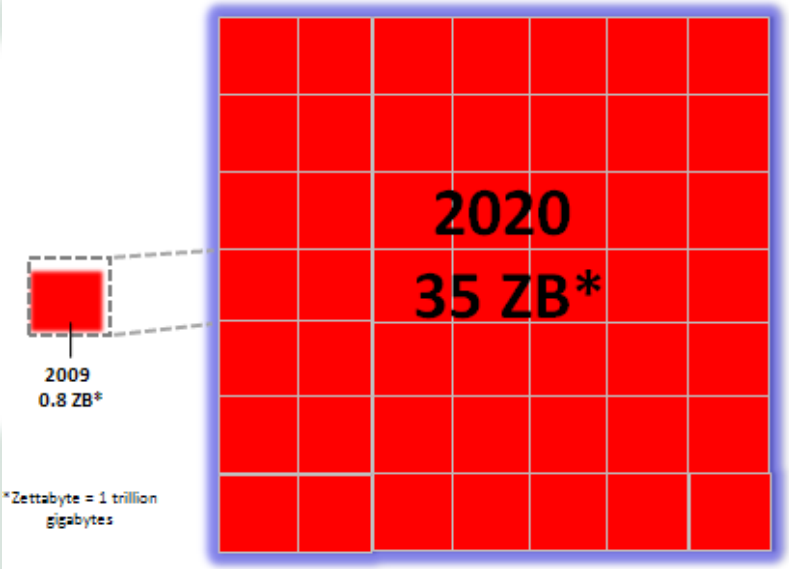
Figure 2



Source: IDC, 2007

Figure 1: The Digital Universe 2009 – 2020

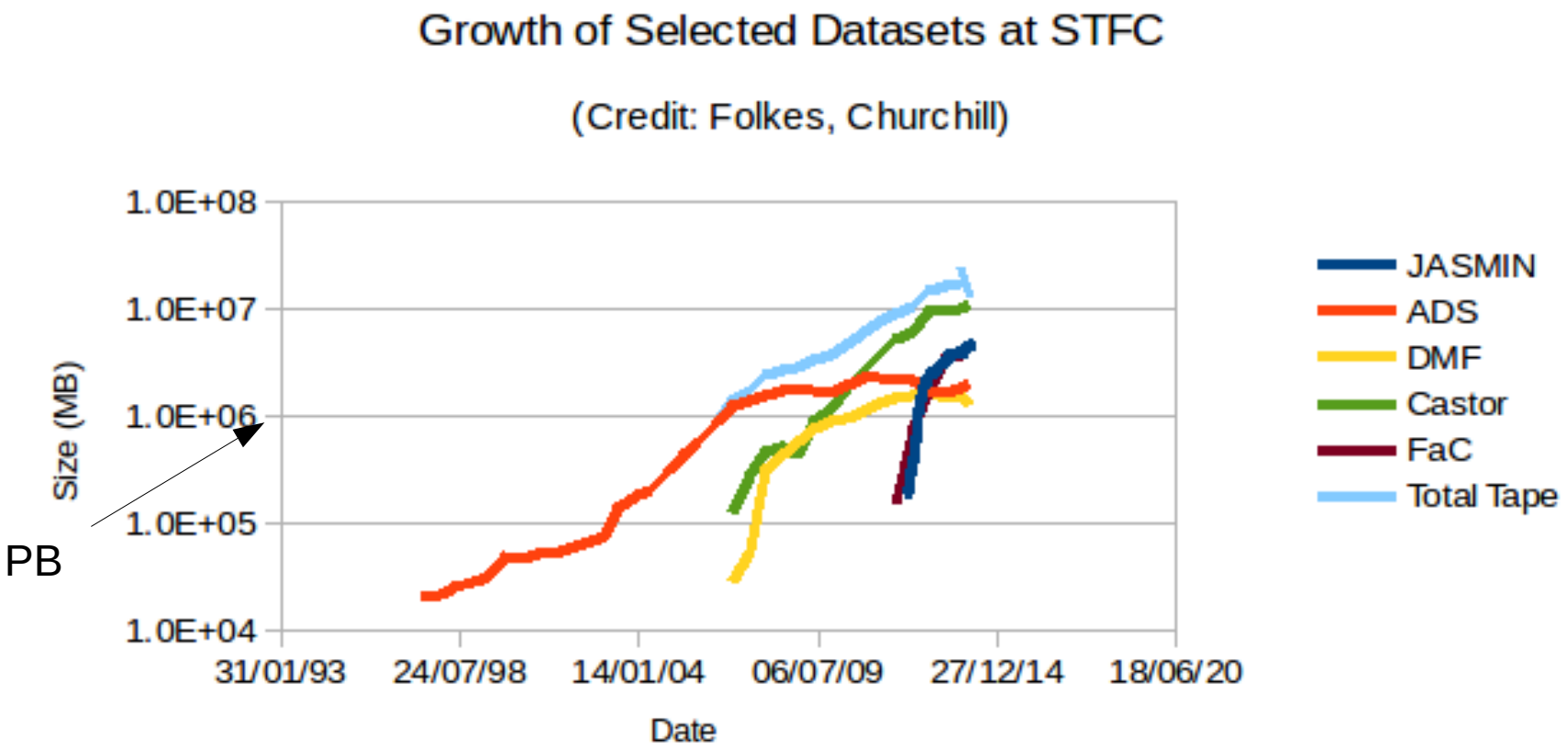
Growing by a Factor of 44



Source: IDC Digital Universe Study, sponsored by EMC, May 2010

(All data, not just scientific data)

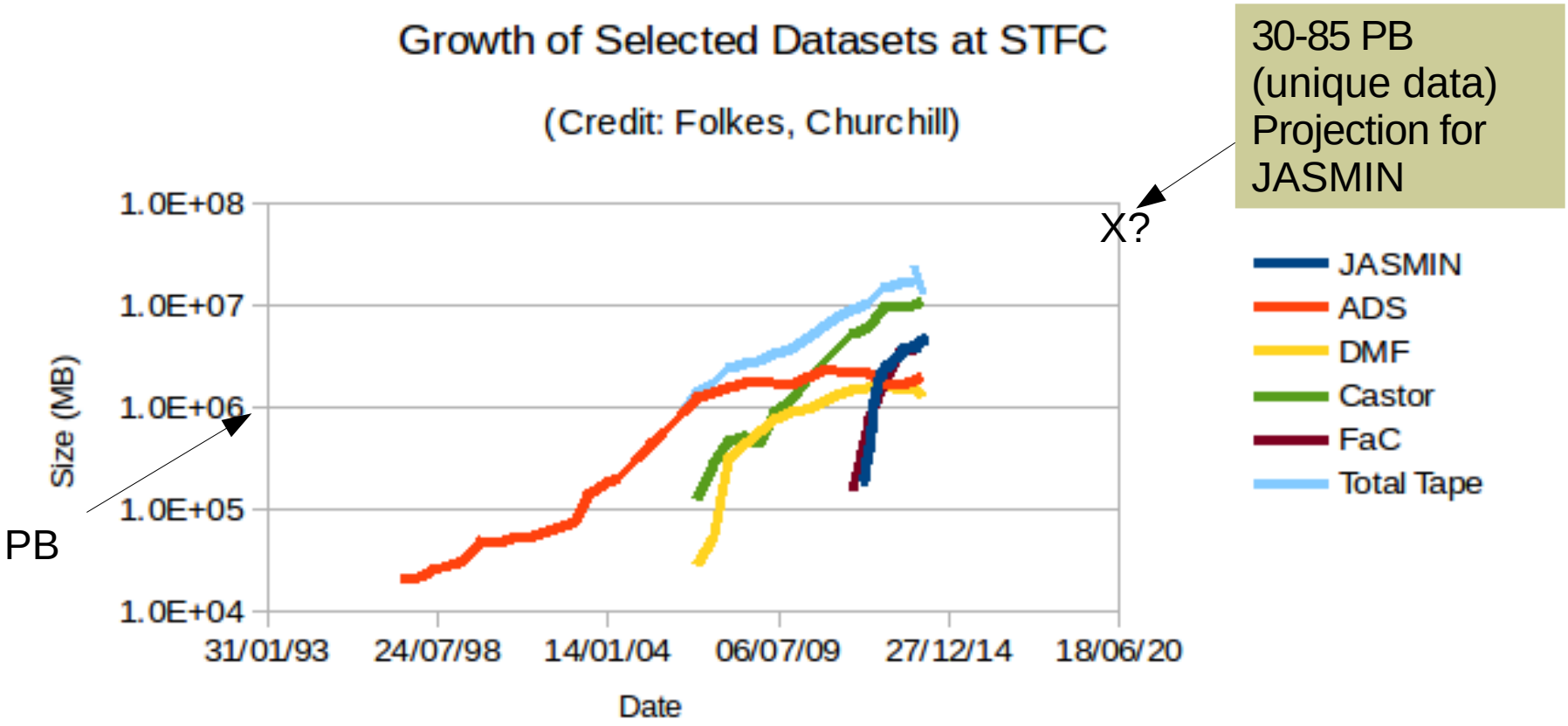
Data growth at STFC



The light blue line is the total of all the data stored on tape in the STFC Scientific Computing Department.

The green line is the LHC Tier 1 data on tape.
The dark blue line is the data stored on **disk** in JASMIN.

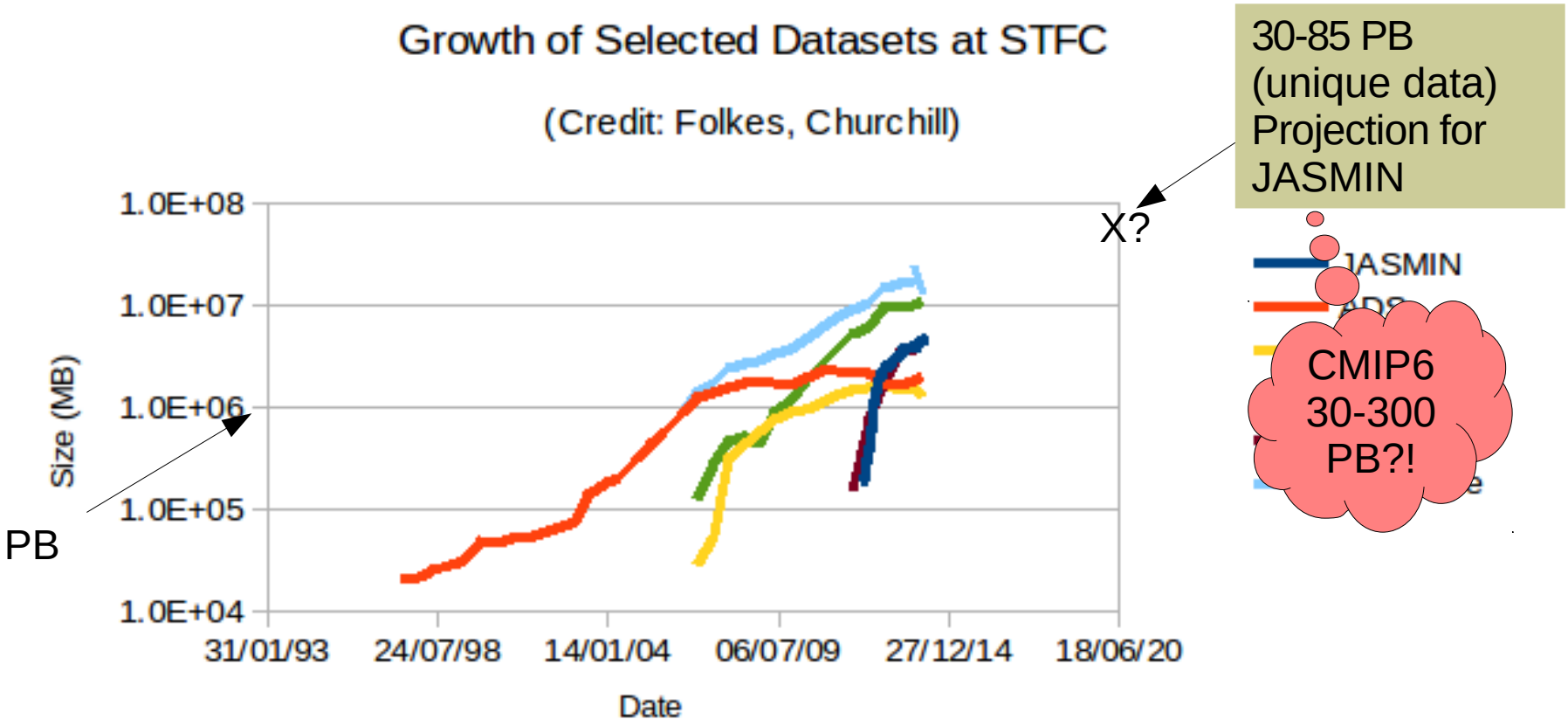
Data growth at STFC



The light blue line is the total of all the data stored on tape in the STFC Scientific Computing Department.

The green line is the LHC Tier 1 data on tape.
The dark blue line is the data stored on **disk** in JASMIN.

Data growth at STFC



The light blue line is the total of all the data stored on tape in the STFC Scientific Computing Department.

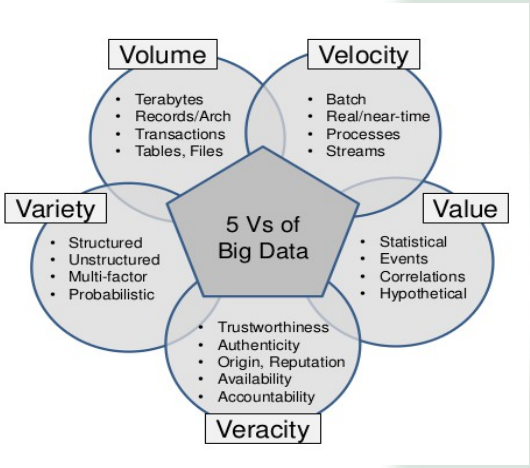
The green line is the LHC Tier 1 data on tape.
The dark blue line is the data stored on **disk** in JASMIN.

SI Prefixes

SI prefix	Name	Power of 10 or 2		Status
k kilo	thousand	10^{3}	2^{10}	Count on fingers
M mega	million	10^{6}	2^{20}	Trivial
G giga	billion	10^{9}	2^{30}	Small
T tera	trillion	10^{12}	2^{40}	Real
P peta	quadrillion	10^{15}	2^{50}	Challenging
E exa	quintillion	10^{18}	2^{60}	Aspirational
Z zetta	sextillion	10^{21}	2^{70}	Wacko
Y yotta	septillion	10^{24}	2^{80}	Science fiction

Stuart Feldman, Google

Which brings us to "big data" - as a driver!



Yuri Demchenko (meeting report, Amsterdam, July 2013)

Volume
- Bytes
- Records
- Tables

Variety
- Variables
- Models
- Experiments

Veracity

Which model? Can I trust it?

HYPE?

Velocity
- Bytes/second
- Soft-/Wet-/Hard-ware revolutions.

Voting

Multiple standards across multiple communities

Value

Decision relevant → decision critical

Martin Juckes

Humans and the Data Deluge

A person working full time for a year has about 1500 hours to do something.

(In the UK 220 working days a year is about standard. Let's remove about 20 days for courses, staff meetings etc ... so that leaves about 200 days or, for a working day of 7.5 hours, a working year of about 1500 hours.)

What does a 50 TB dataset mean?

- Take a set of climate predictions.
- A single lat/lon map might be of order 50 Kb … so we have of the order of 10 billion maps.
- Looking at each map for 10s, one individual could quality control those maps in approximately two thousand years of work!
- Bring on crowd sourcing … but there's only so many people in the world!

If it takes 2 minutes to find something, and have a quick look at it and extract a (e.g.) parameter name,

- You can process 45,000 items a year
- But no human could do that full time (repetitive boredom)!
- Maybe 30K in two years?

Your examples will differ, but your conclusions are unlikely to:

We can't manage big data relying on humans! We need automation!

And so to the Technology Drivers

John Shalf, LBNL: May 2010

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)***

I suspect we're in the phony war right now! "Playing" with GPU/MIC

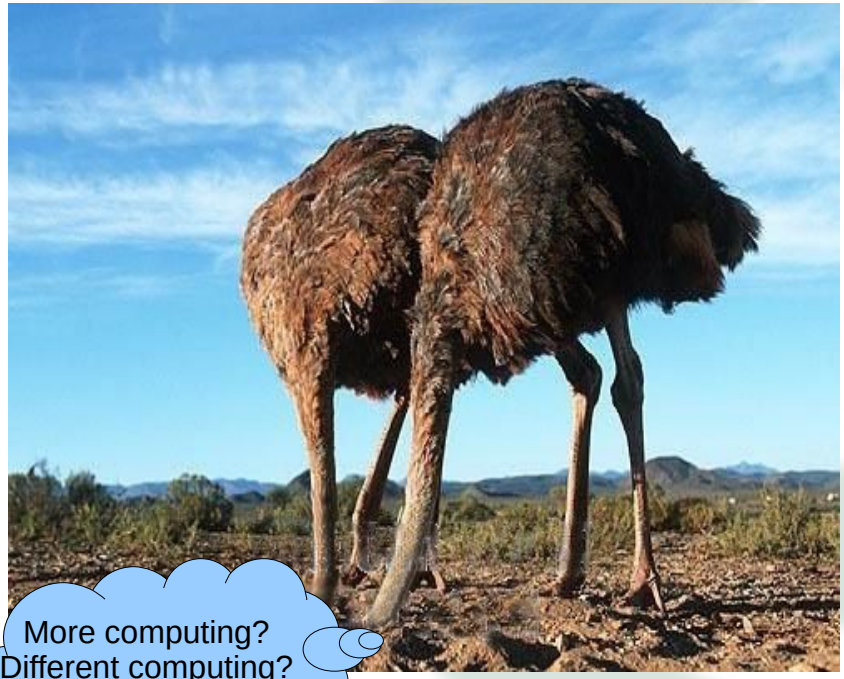
Office of Science U.S. DEPARTMENT OF ENERGY

3

www.spscicomp.org/ScicomP16/presentations/ExascaleChallenges.pdf
Accessed 24/01/2011

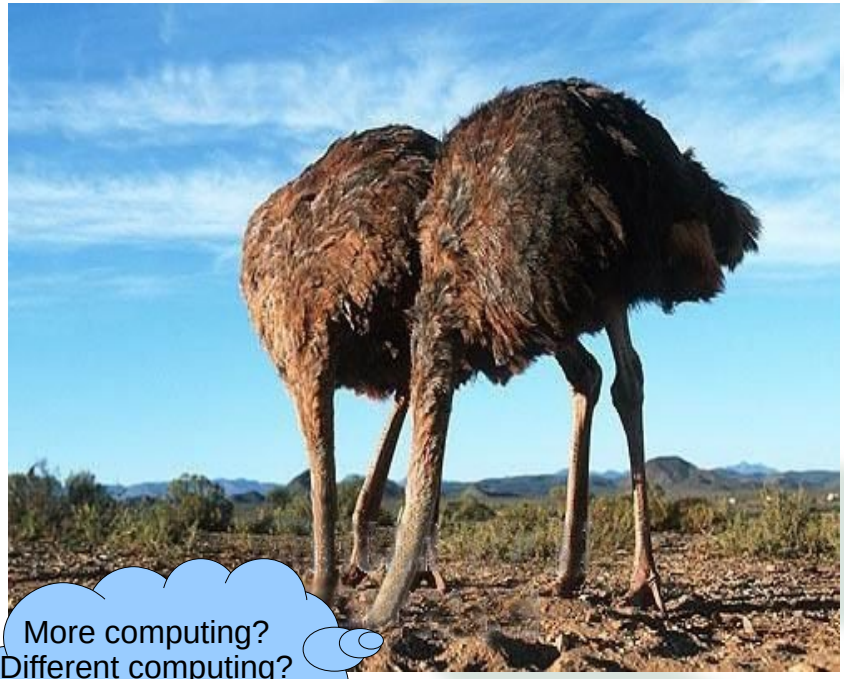
Software Progress Status
Some views of community readiness

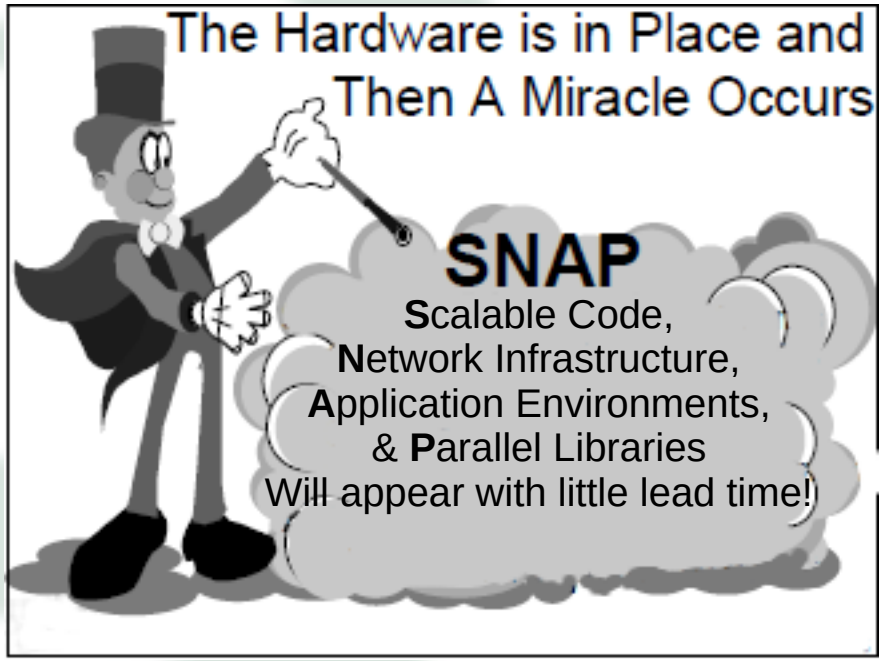
According to Ken Batcher, "A supercomputer is a device for turning compute-bound problems into I/O-bound problems."



Software Progress Status
Some views of community readiness

According to Ken Batcher, "A supercomputer is a device for turning compute-bound problems into I/O-bound problems."

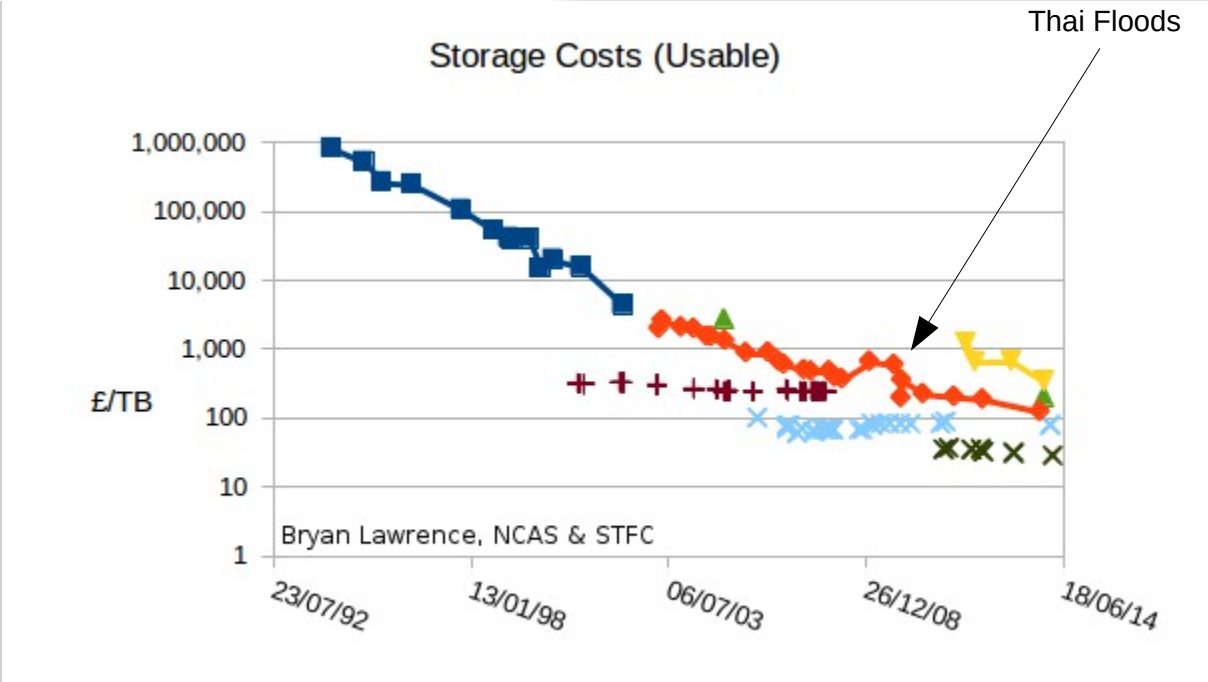




… which is a little unfair, but I think it is fair to say that (some of) the community underestimates the effort ahead!

Cost of storage likely to increase!

Actual costs from STFC:



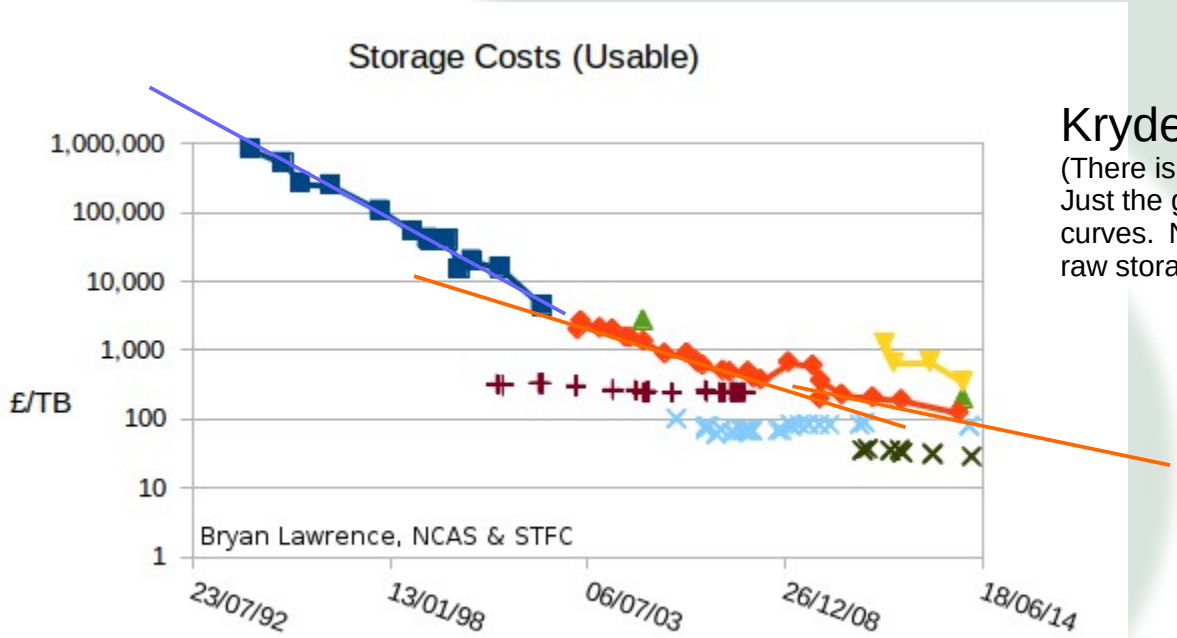
Filled characters and lines: different generations and disk technologies.
- Yellow is parallel disk: PanFS

Crosses: different tape technologies.

Data courtesy of Peter Chiu, Jonathan Churchill and Tim Folkes

Cost of storage likely to increase!

Actual costs from STFC:



Kryder's Law slowing down!

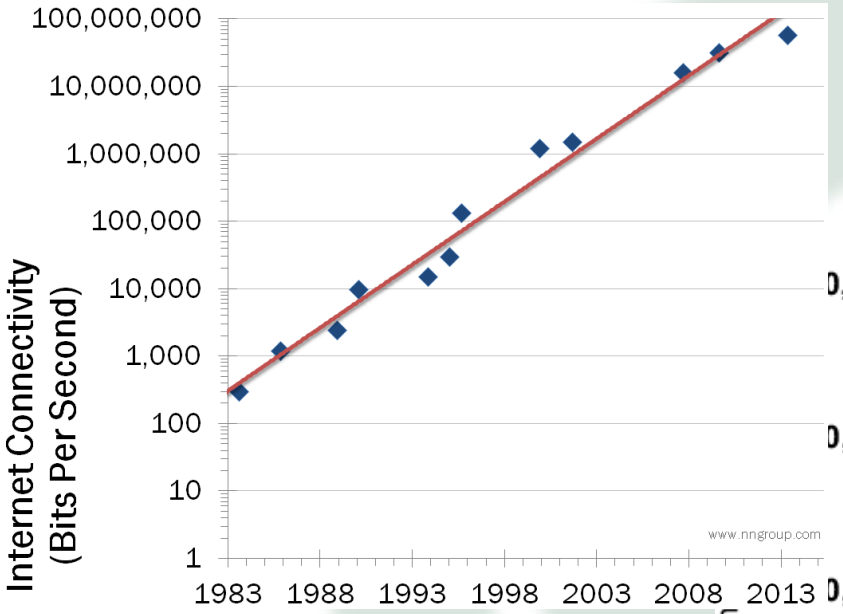
(There is no such thing as an exponential in real life, Just the growth part of an s-curve, or several s-curves. NB same three "eras" even when we use raw storage costs before RAID and friends.)

Tape technology looks like it has a lot to give us yet, while disk technology is struggling (for the moment a bit like Fusion, the next technology is "just over the horizon").

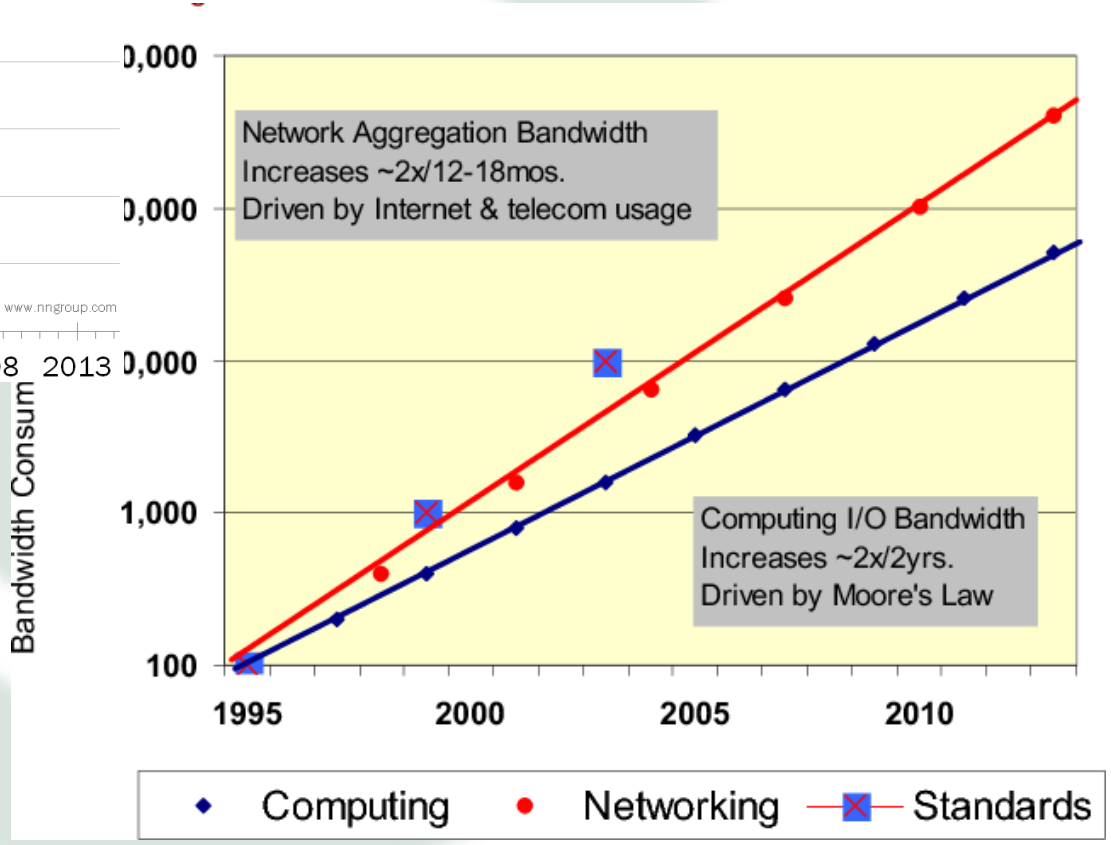
Whatever, cost of disk is increasing faster than the cost of compute! Especially the cost of "usable" disk.

Data courtesy of Peter Chiu, Jonathan Churchill and Tim Folkes

Better Networks? It's complicated!



Nielsen's Law;
Gilder's Law …

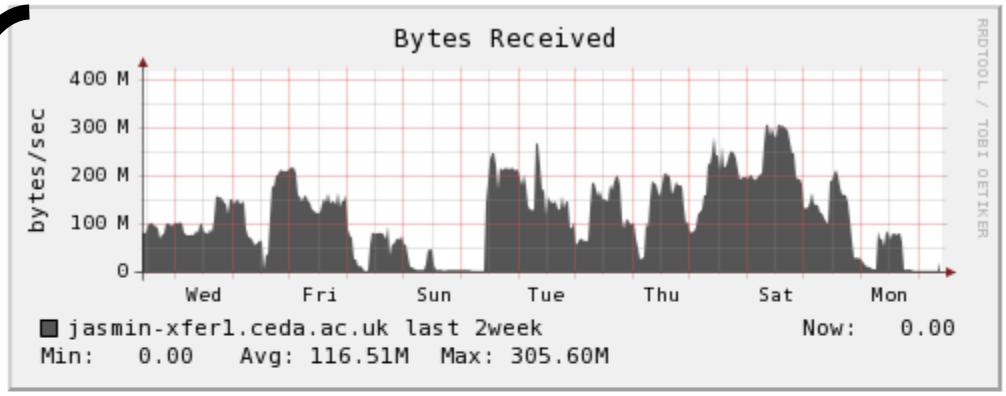


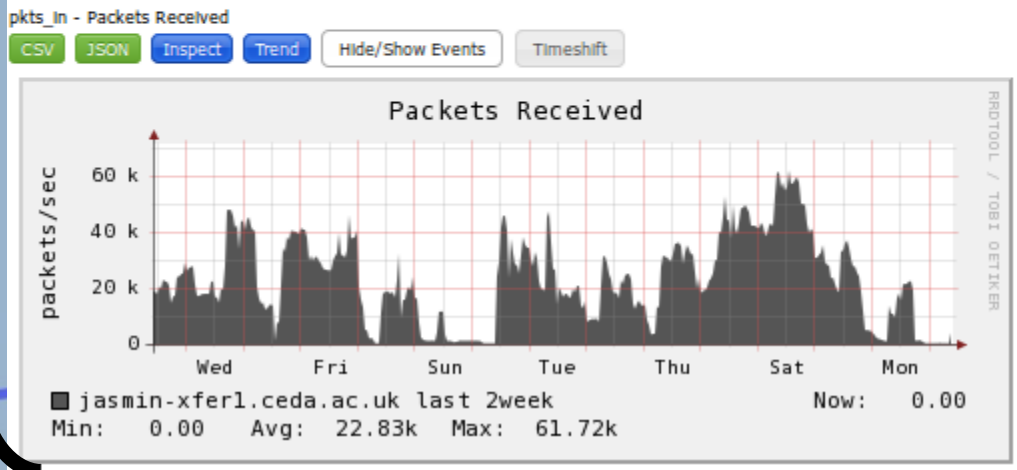
General Trends?

End-User and I/O connectivity growing slower than Moore's Law.

Network aggregation bandwdith, growing faster!

JASMIN Network





Dedicated Lightpath Network

Two weeks in January 2014:
→ Average 10 TB/day, Peak 30 TB/day
→ Inbound onto JASMIN Storage

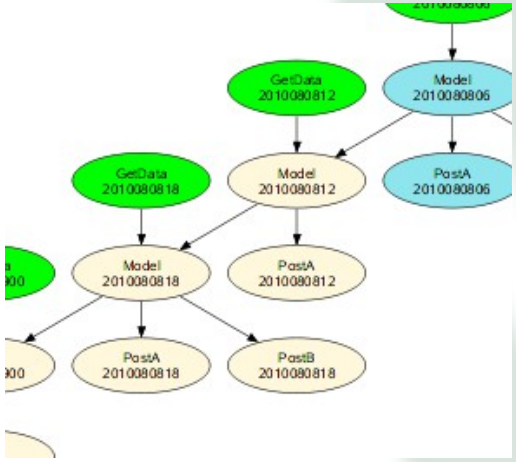
Better Software (1)

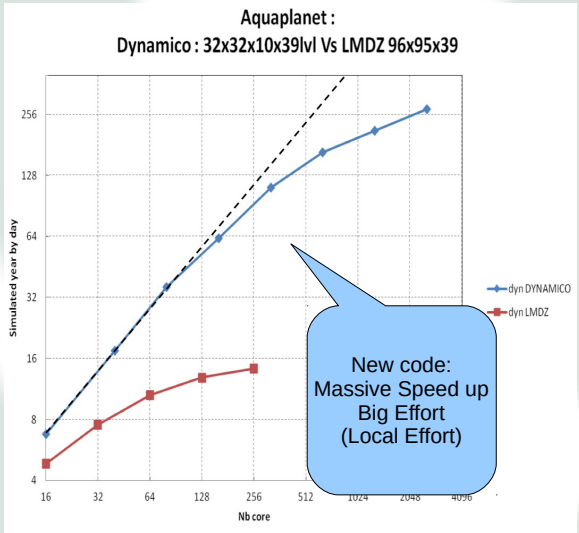
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

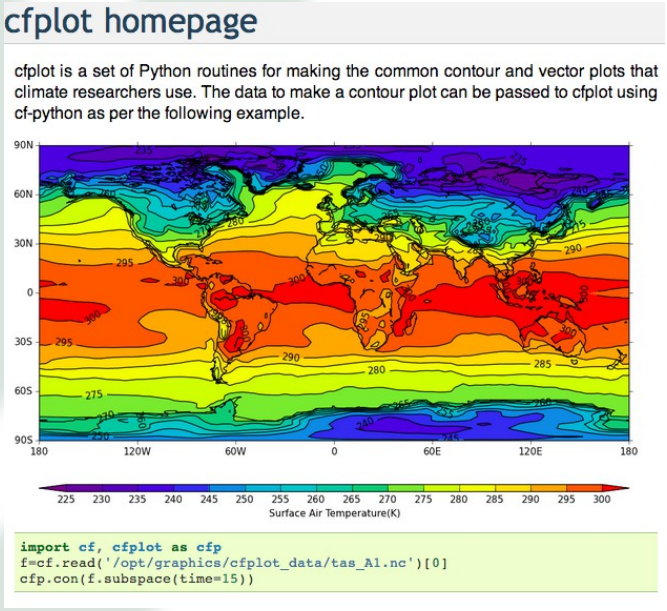
Four areas to consider:

- Workflow (e.g. CYLC)
- Simulation (The codes themselves)
- Analysis (CDO, NCO, IRIS, CF-Python etc)
- Data Management (I/O libraries, Tools to document data)

Software
Is
Infrastructure!







NIWA
Taihoro Nukurangi

T.Dubos, S.Dubesh, Yann Meurdesoif(LSCE-IPSL)
Results presented at IS-ENES2 workshop, March 2014

I have deliberately chosen Kiwi, French and British examples: Global activities!

Better Software (2): Semantics ...

CMIP5 (23/05/13):

101 experiments

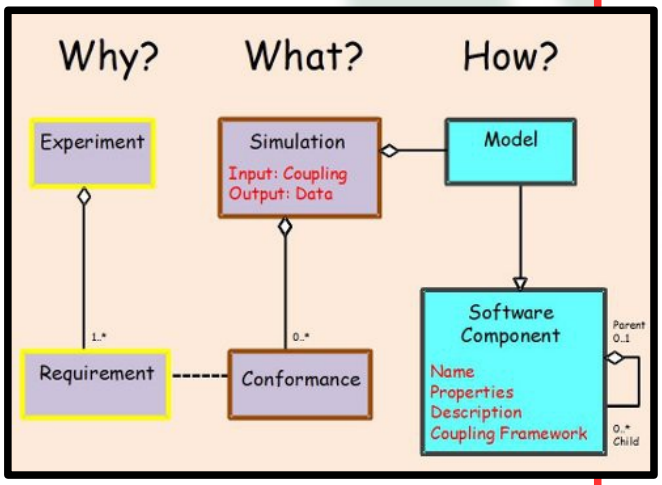
61 model variants

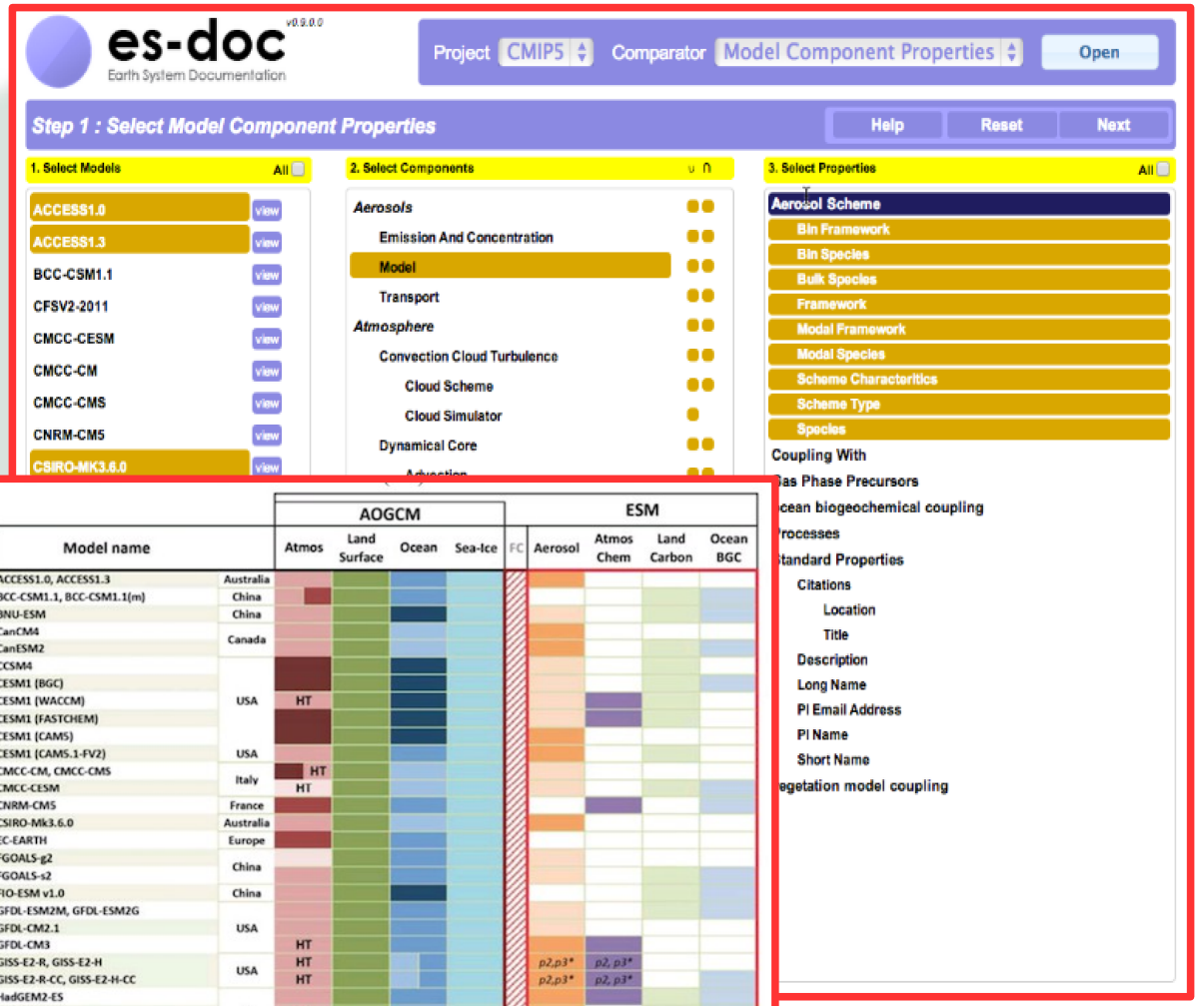
590,000 datasets!

4.5 million files

2 PB in global archive

Unknown PB locally!





Tools to "understand" datasets!

Funding Drivers!

Governments everywhere are seeing:
the **necessity** of “big” science
the **cost** of “big” science
(big is a local definition)

So they're desperate to

Consolidate Infrastructure
&
Introduce **efficiencies** of scale
but
We're already at scale! & We work globally!

Stocktake:

The science drivers are for more, bigger, faster.

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

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

Software is getting smarter, but we need to spend more on it (and the people who develop and maintain it).

Things are getting more complicated, and the money is (comparatively)
running out!

Solutions: Partnerships and e-infrastructures

Shared Infrastructure:
Contracts & Service Level Agreements

Formal Collaboration: MoU

Ad-HocCollaboration: Trust

e-Infrastructure south

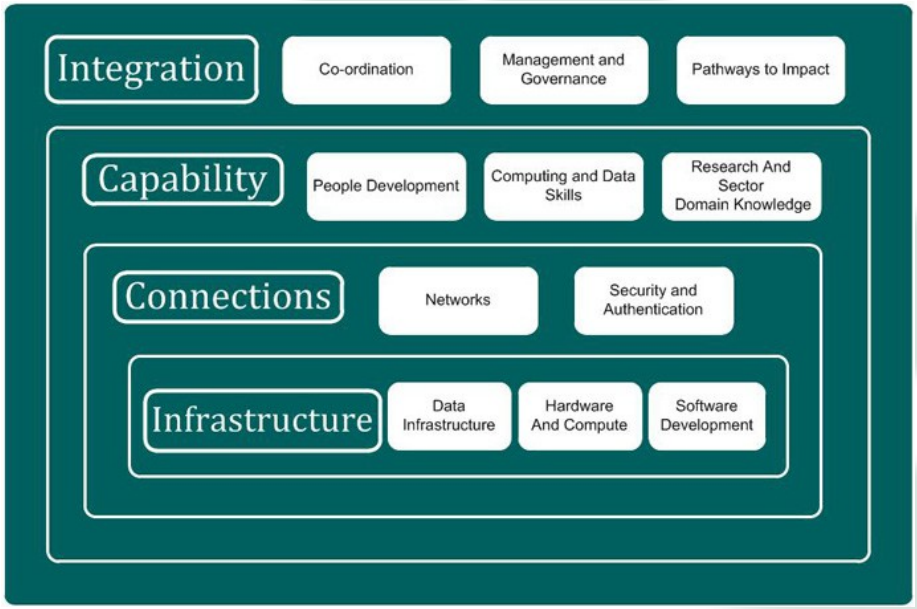
Met Office | NERC SCIENCE OF THE ENVIRONMENT

Joint Weather and Climate Research Programme
A partnership in climate research

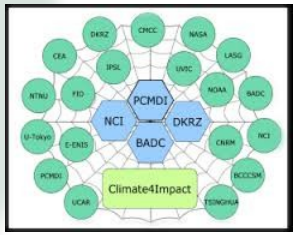
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

EUDAT



PRACE



ESGF
Earth System Grid Federation

Institutional Landscape

(Biotechnology and Biology)

Engineering and Physical Sciences Research Council

+ Universities, big and small ...

Building Partnerships: Some Case Studies

"NERC National Capability"
- Data

"Earth System Grid Federation"
- & European Network for Earth Simulation
- (all in the context of multiple collaborations)

& back to NERC National Capability
- And data again: JASMIN

Data Landscape – Consider where effort will yield results!

Big Data
Big Science
Ever larger!

(More)
Homogenous

?

Heterogeneous

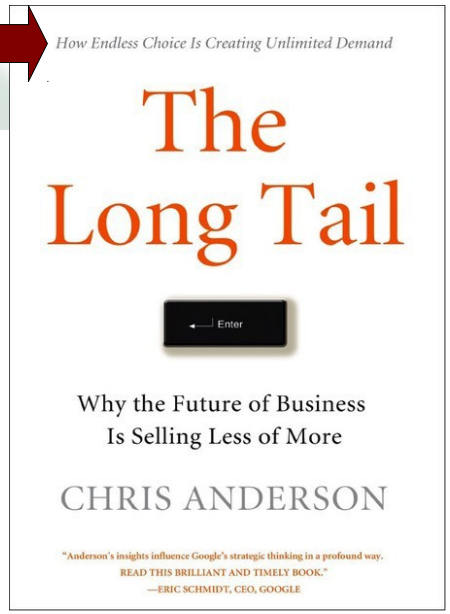
Petascale

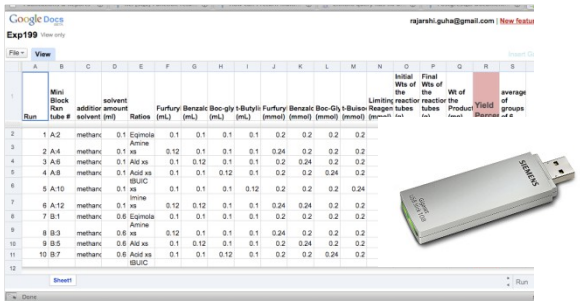
CMIP5
Archive

Earth Observation
(Climate Data)

Typical Departmental
Collection
One instrument archive

Fortran,
IDL,
Matlab,
Python
& (yuck)
Excel,
Access





Personal science:
Ever more complex, and
ever more of it!

Laptop Scale

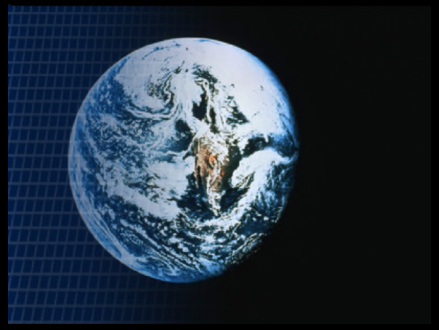
Slide concept: Carole Goble via Liz Lyons

NERC Data Centres

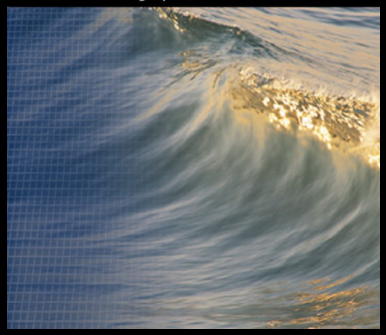
Hydrology:
National Water Archive

Atmosphere:
British Atmospheric Data Centre

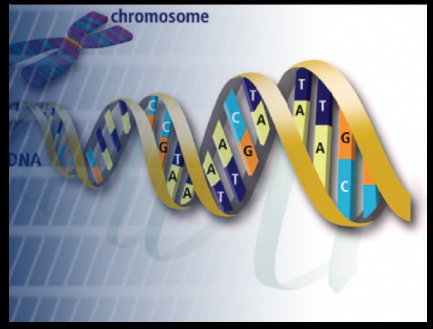
Earth observation:
NERC Earth Observation Data Centre



Ocean & marine:
British Oceanographic Data Centre



Bioinformatics:
NERC Environmental Bioinformatics Centre

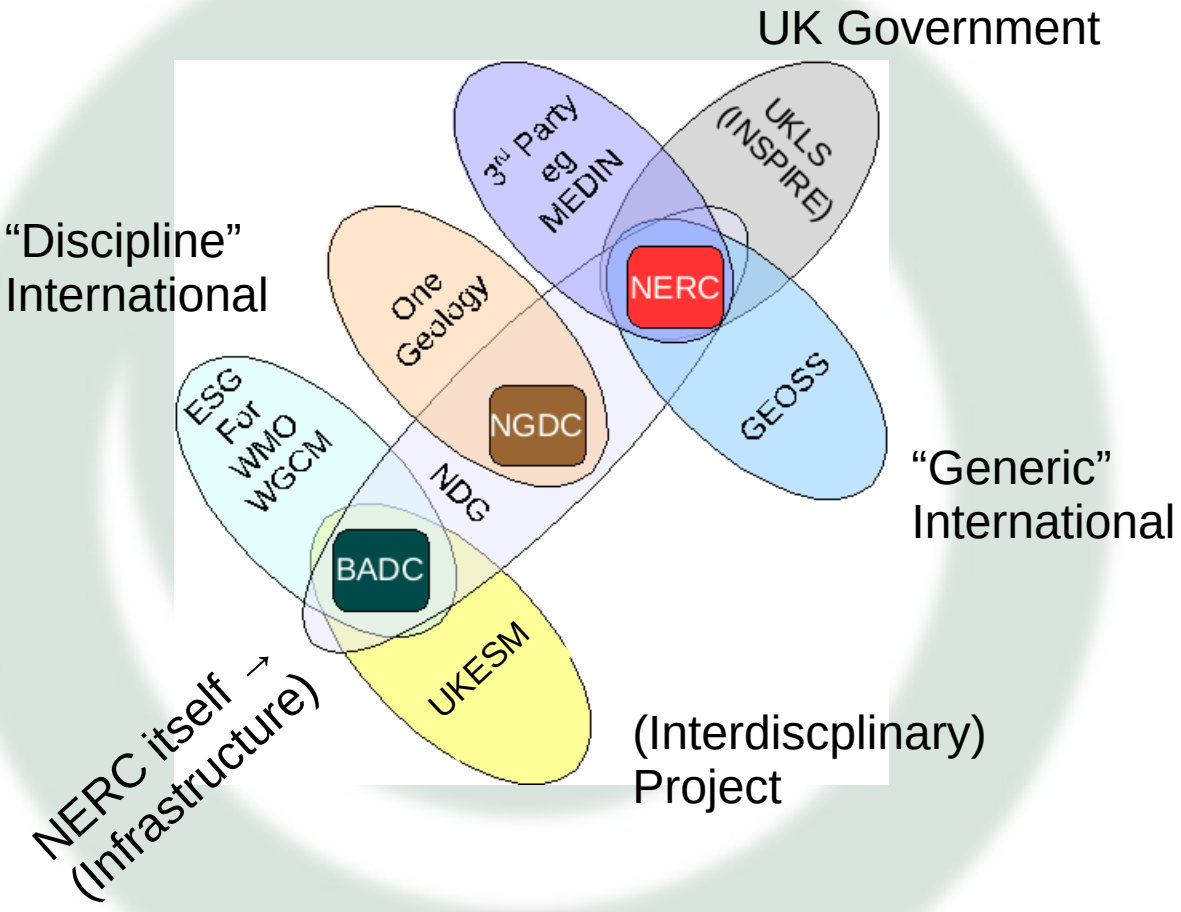


Earth:
National Geoscience Data Centre

Terrestrial & freshwater:
Environmental Information Centre

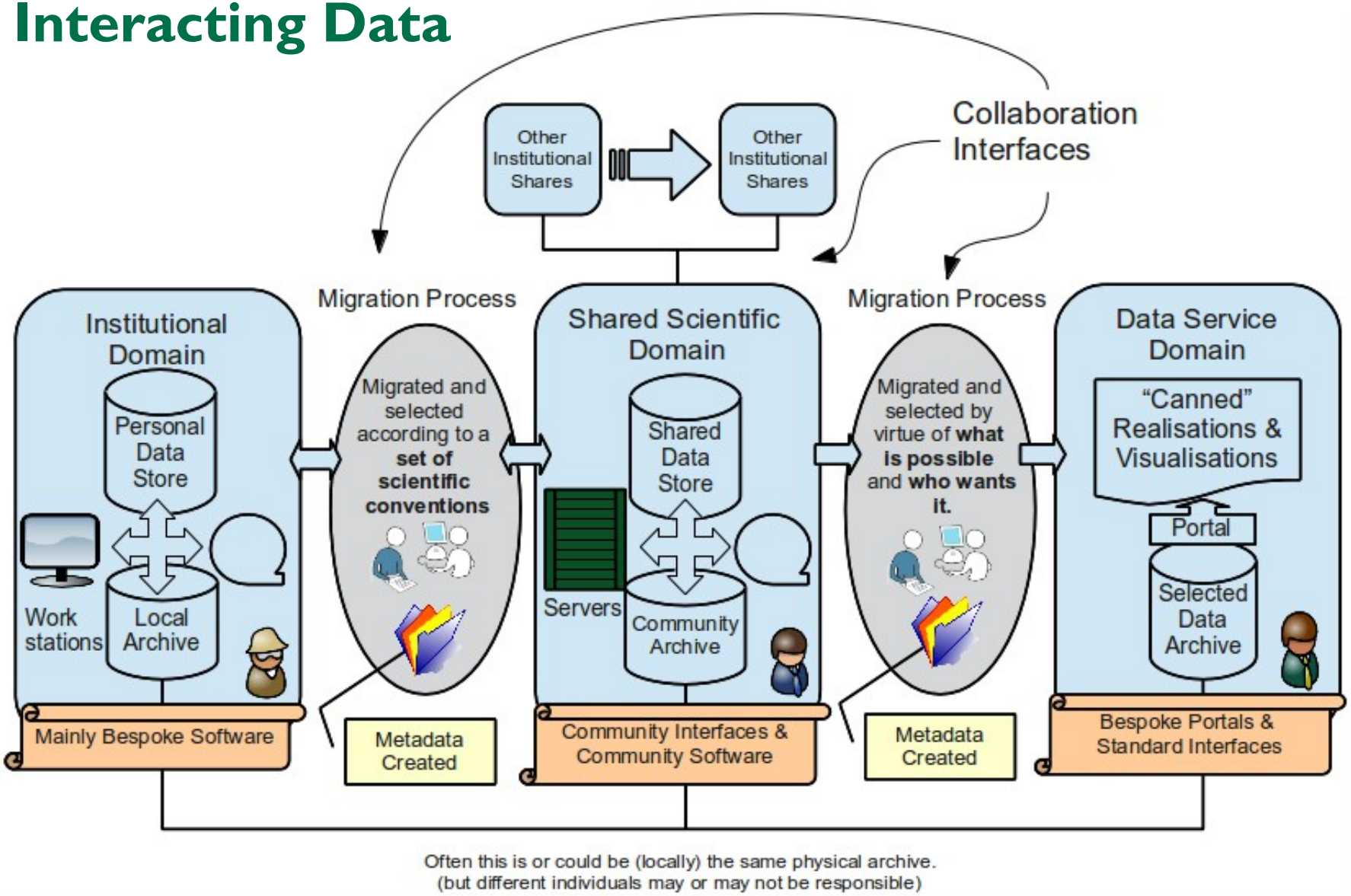
Polar:
Antarctic Environmental Data Centre

With whom do we build infrastructure?



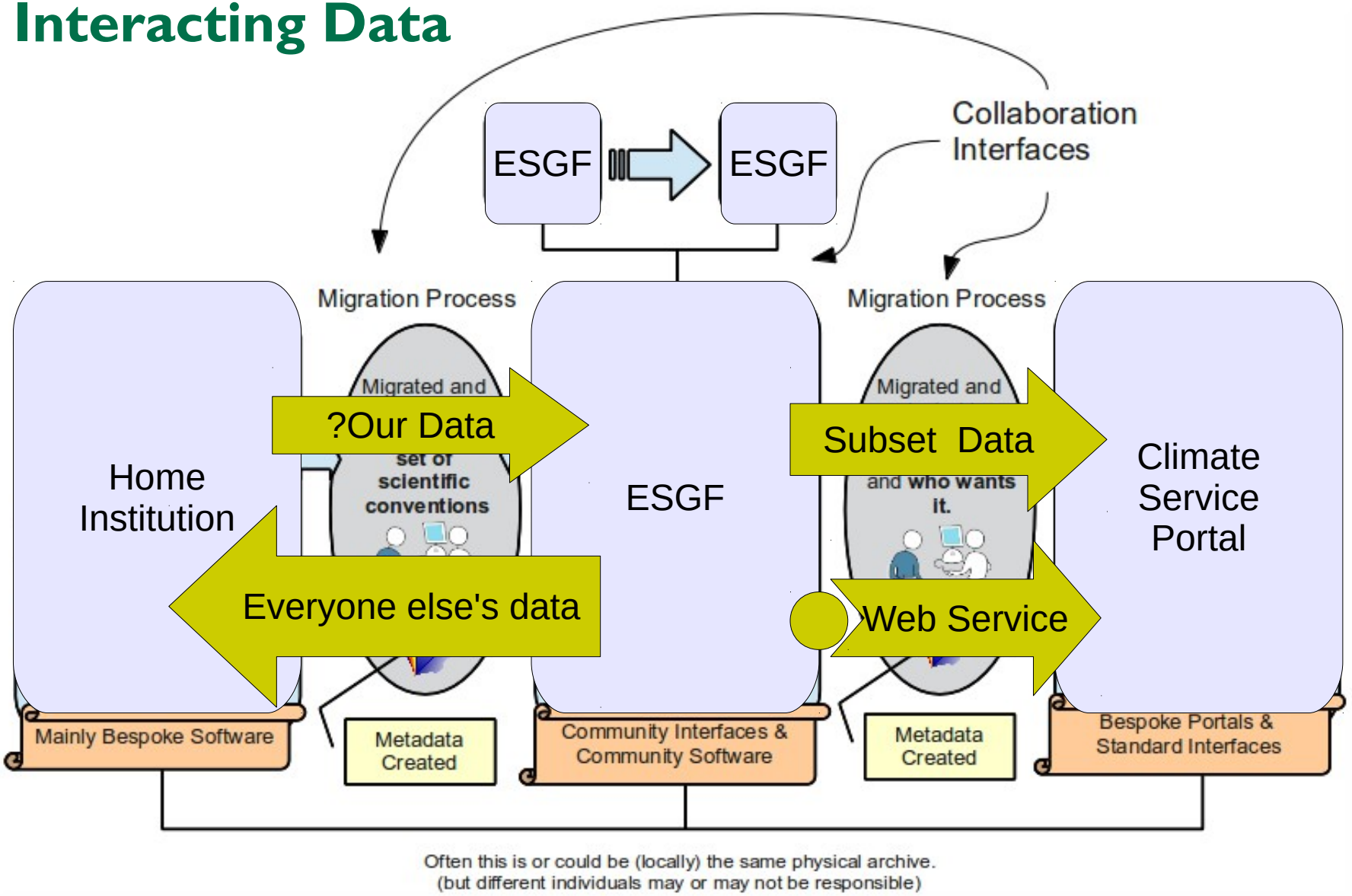
Different aims, technologies, business models, timescales ...

Interacting Data



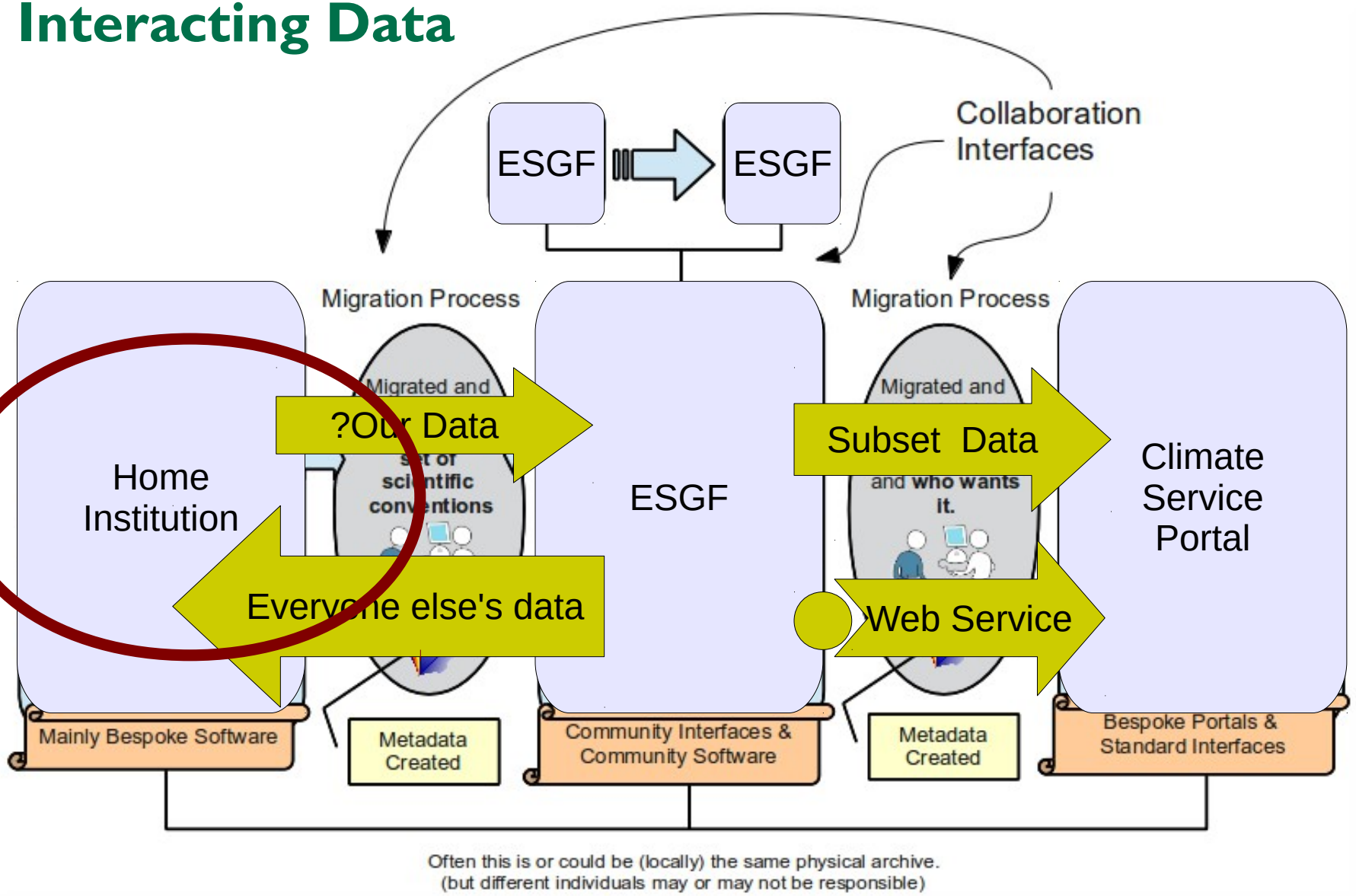
From concepts in Treloar et al 2007 (D-LI

Interacting Data



From concepts in Treloar et al 2007 (D-LI

Interacting Data



From concepts in Treloar et al 2007 (D-LI

My/Your Data Environment

At your home institution, you:

⟶ Have (some) control over your software environment

- Favourite packages, e.g. IDL
- Familiar Linux

→ Can buy/arrange more storage/compute on varying time-scales … can optimise ...

→ Are responsible for deleting/preserving your own data

→ Are likely to be duplicating data others have already downloaded *in your own institution* … let alone within a larger collaboration.

We all like control!

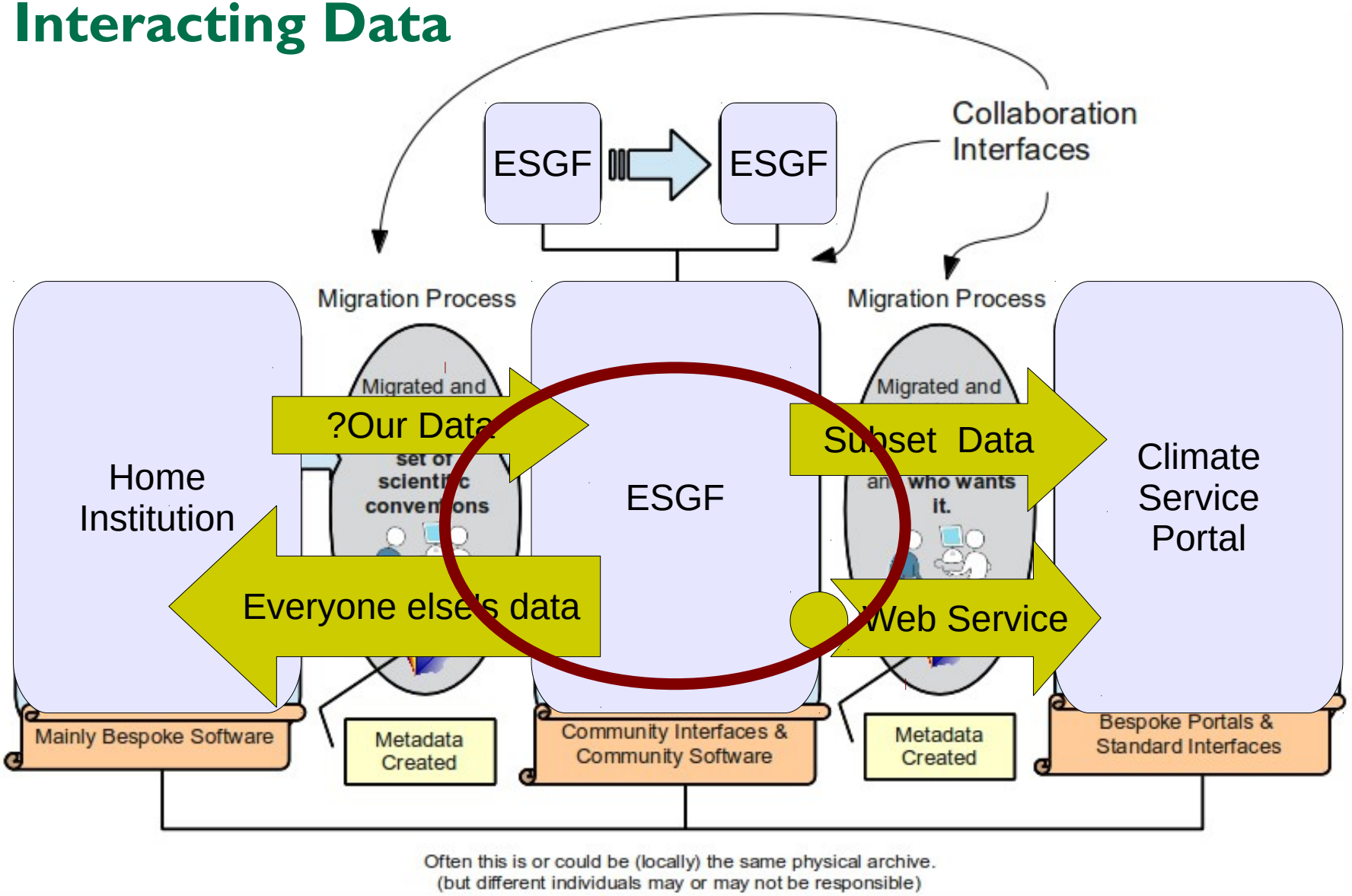
We all like the (illusion?) that we can scale our resources as necessary.

We all lose/destroy/duplicate data.

Most of us do our HPC remotely.

Some of us do our analysis remotely. Why not more of us?

Interacting Data



From concepts in Treloar et al 2007 (D-LI

A one slide guide to CMIP5 from a data perspective

Fifth Climate Model Intercomparison Project (CMIP5)

World Climate Research Programme WCRP- WGCM

Involves all the major climate modelling centres.

Original Timing:

o(2) PB of requested output from 20+ modelling centres finished early 2010!

Actual Timing?

Years late.

(23/05/13):

101 experiments

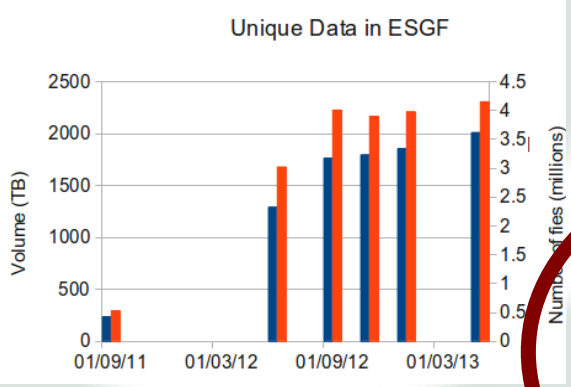
61 model variants

590,000 datasets!

4.5 million files

2 PB in global archive

Unknown PB locally!



Blue: Volume; Red: Files

(NB: replicas and versions!)

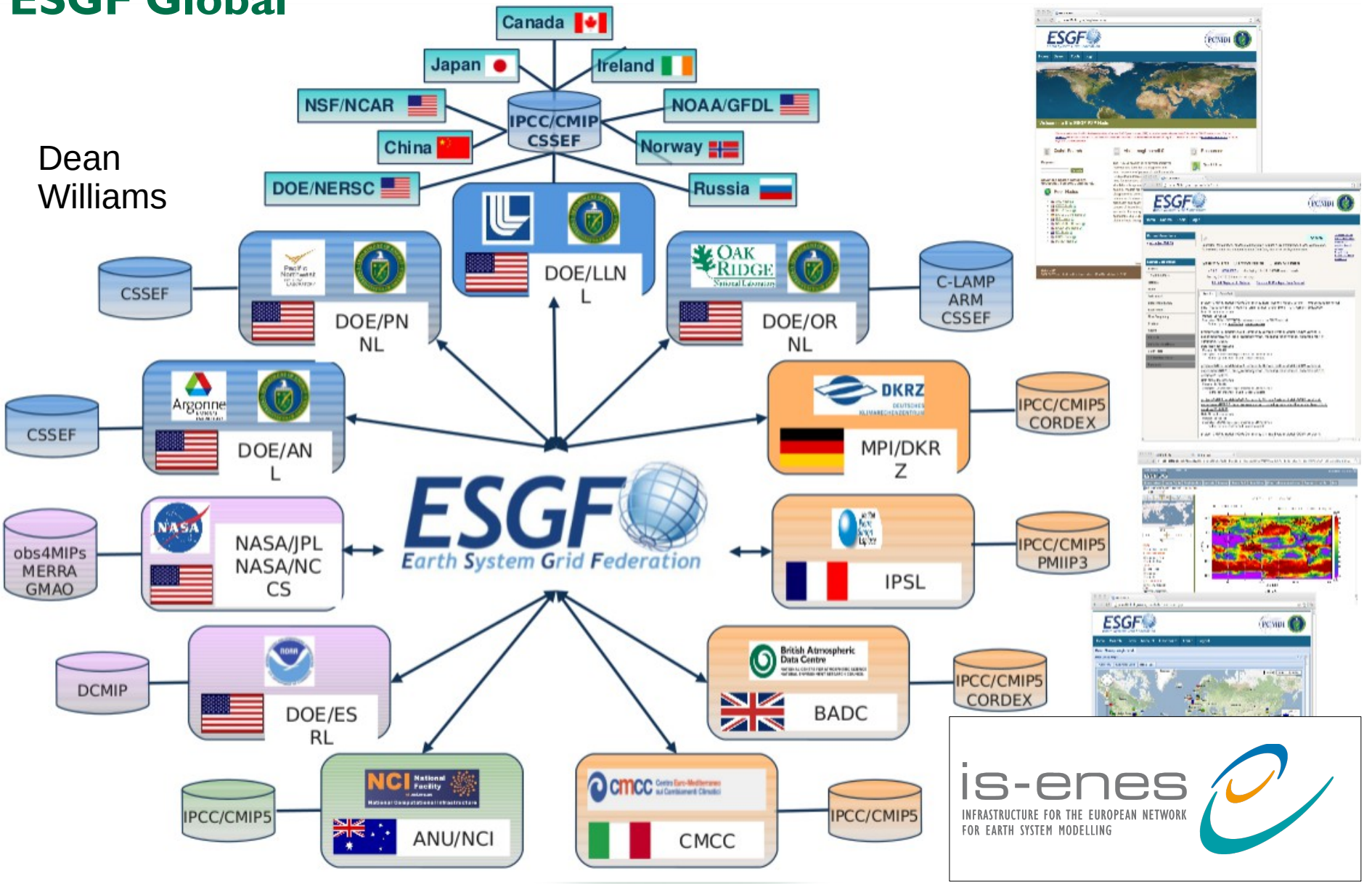
PCMDI-led, community developed (GO-ESSP)

s/w infrastructure for data delivery:

Earth System Grid Federation

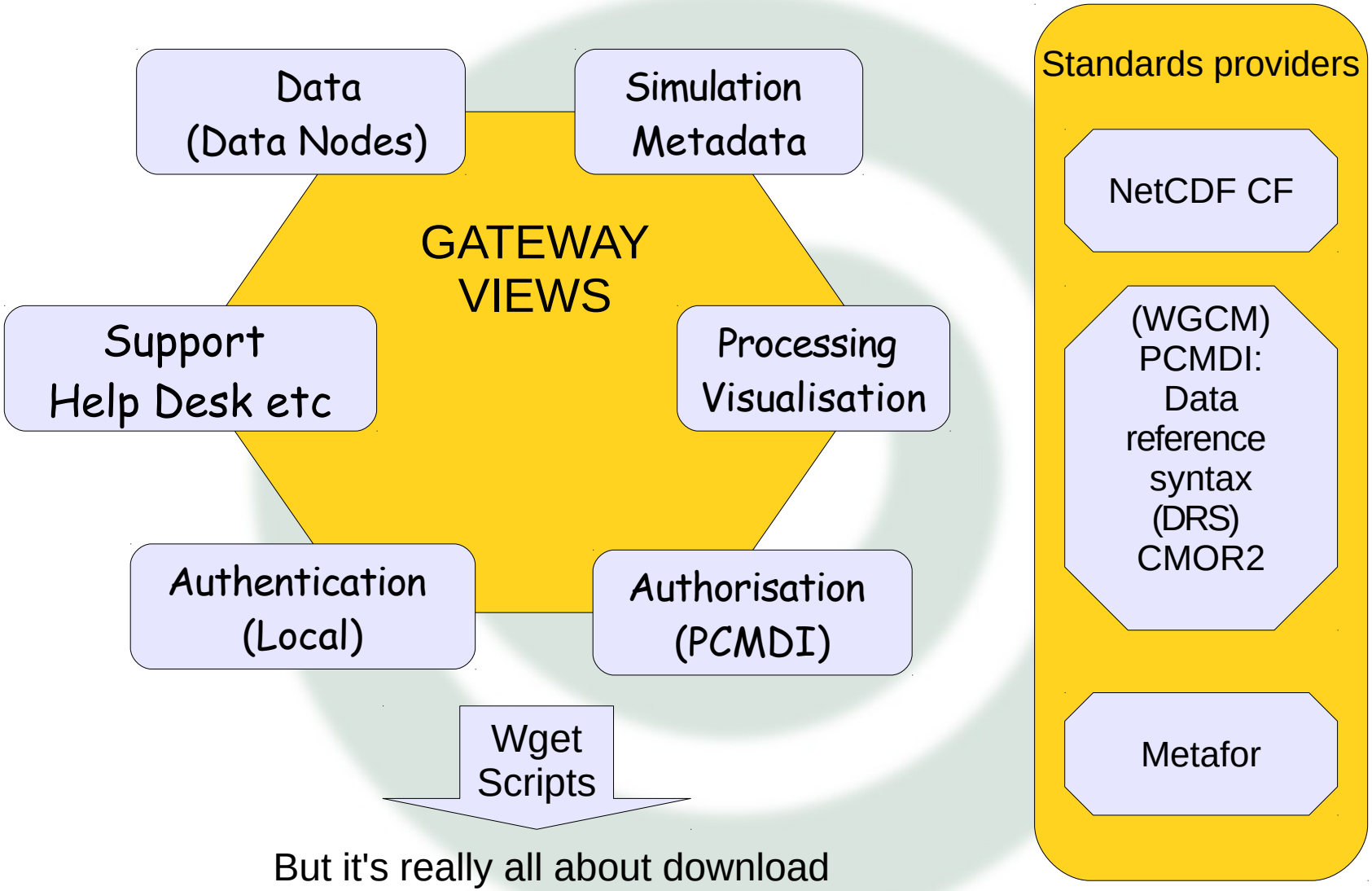
ESGF Global

Dean Williams



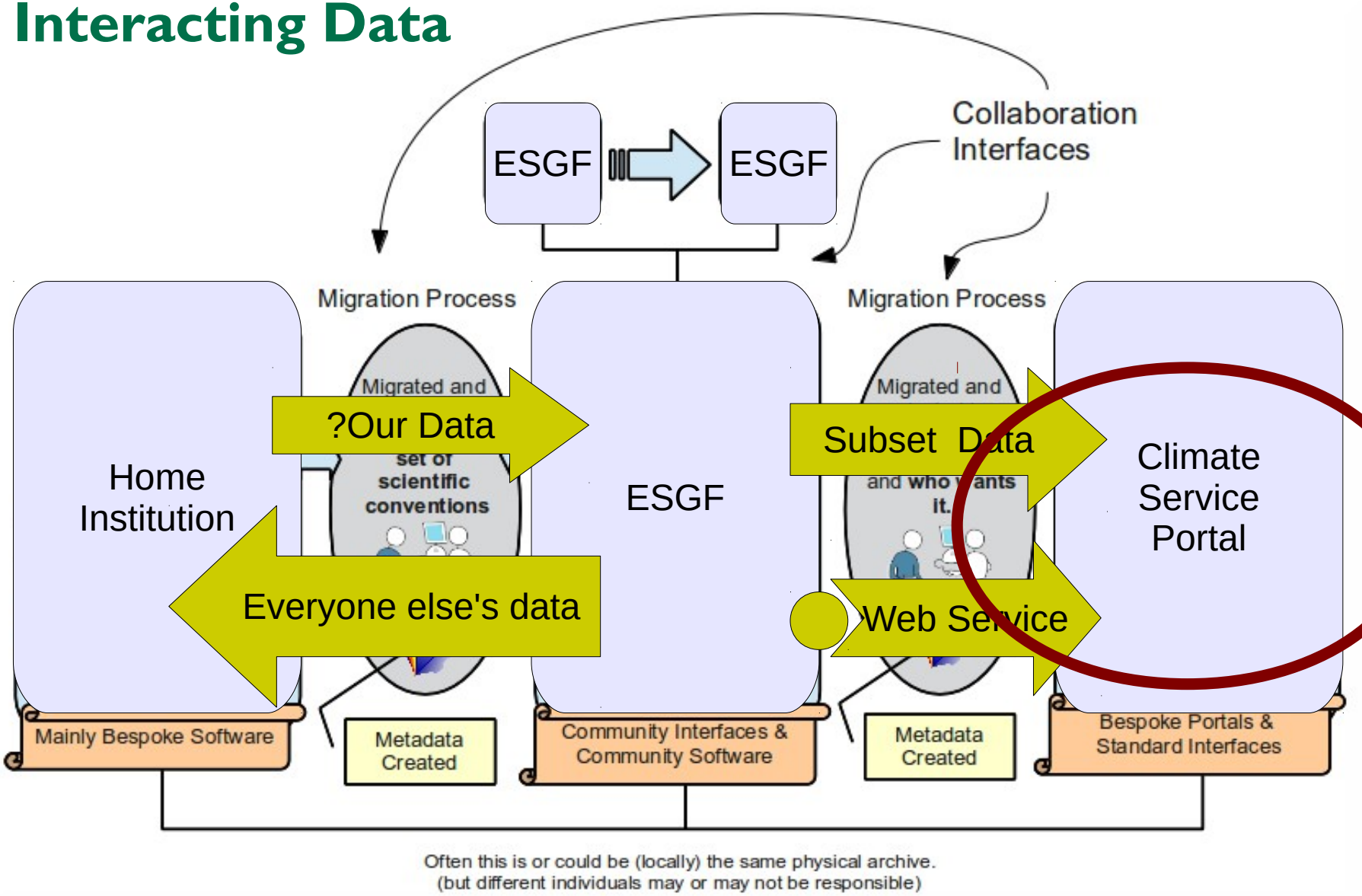
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

ESGF: abstract view



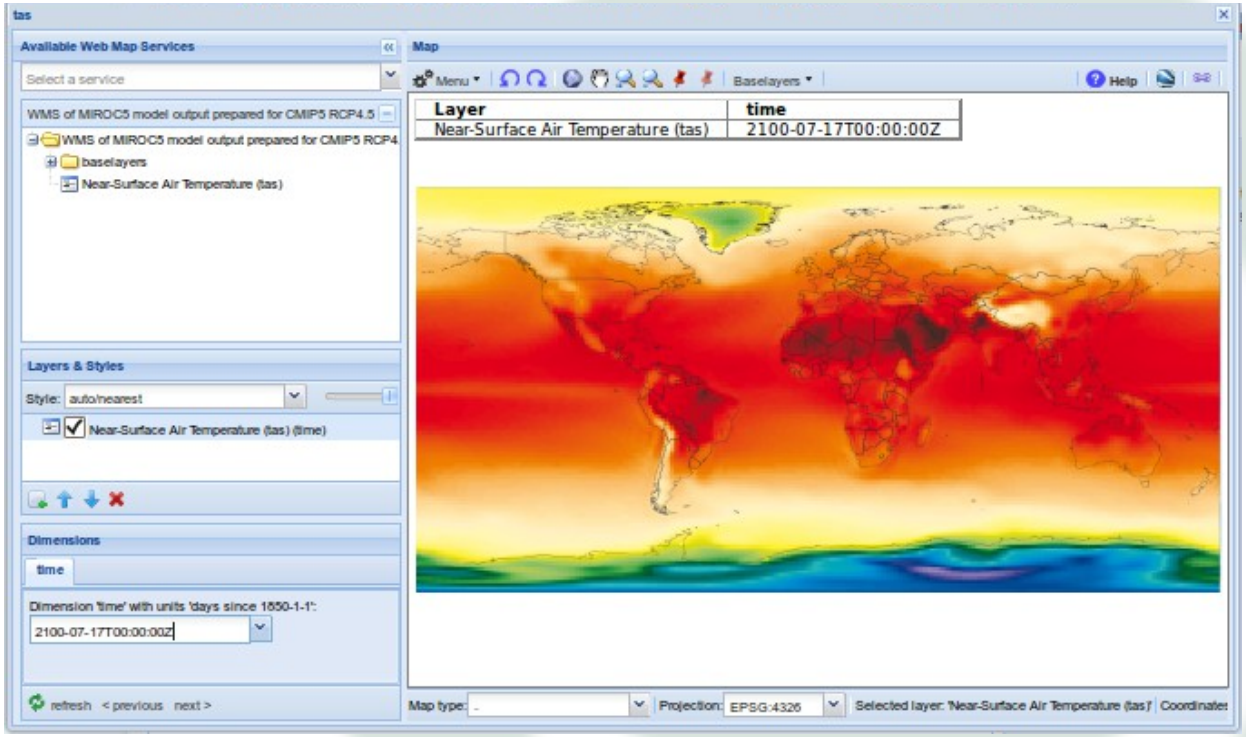
But it's really all about download (now)

Interacting Data



From concepts in Treloar et al 2007 (D-LI

ESGF: A federated data service infrastructure



Data provider: MIROC

Distribution: DIAS (JAPAN)

Identity provider: BADC

Authorisation: PCMDI

Quality control: DKRZ

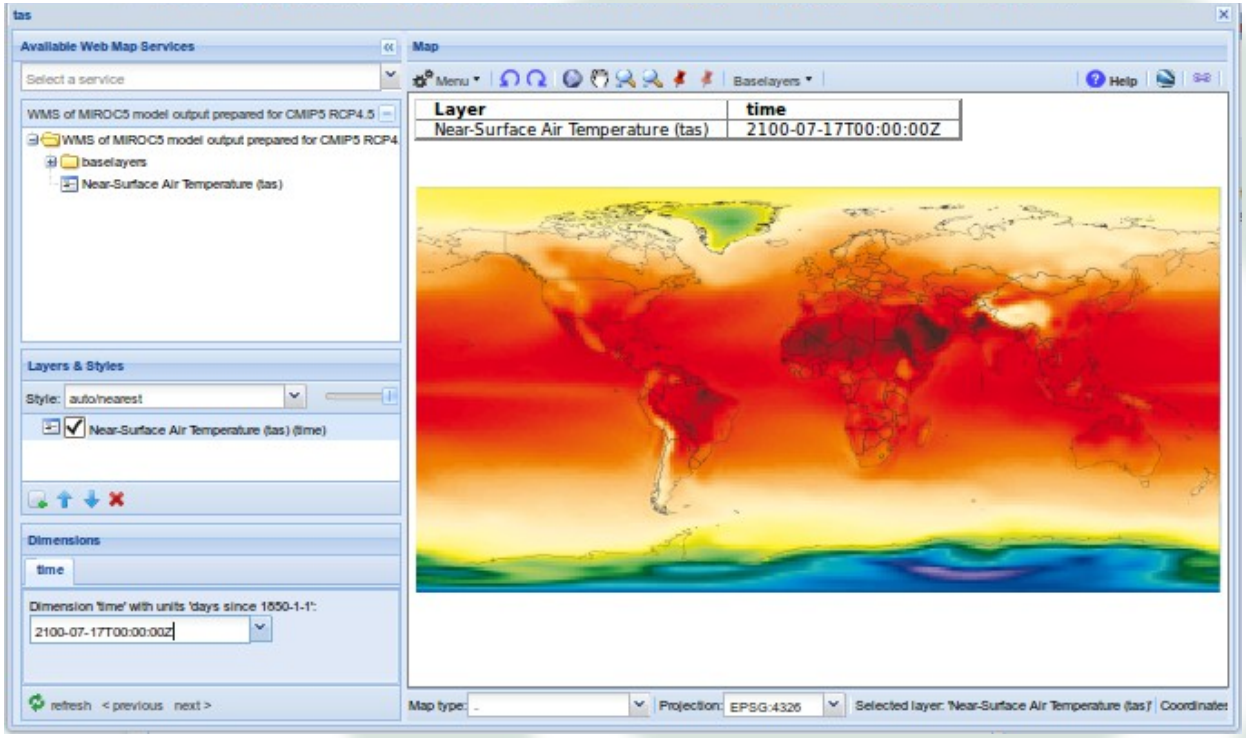
Visualisation: KNMI

A federated system optimises the use of the limiting resource: ***people.***

No institution can go it alone: data at scale is a global activity based around large national facilities....

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

ESGF: A federated data service infrastructure



Data provider: MIROC

Distribution: DIAS (JAPAN)

Identity provider: BADC

Authorisation: PCMDI

Quality control: DKRZ

Visualisation: KNMI

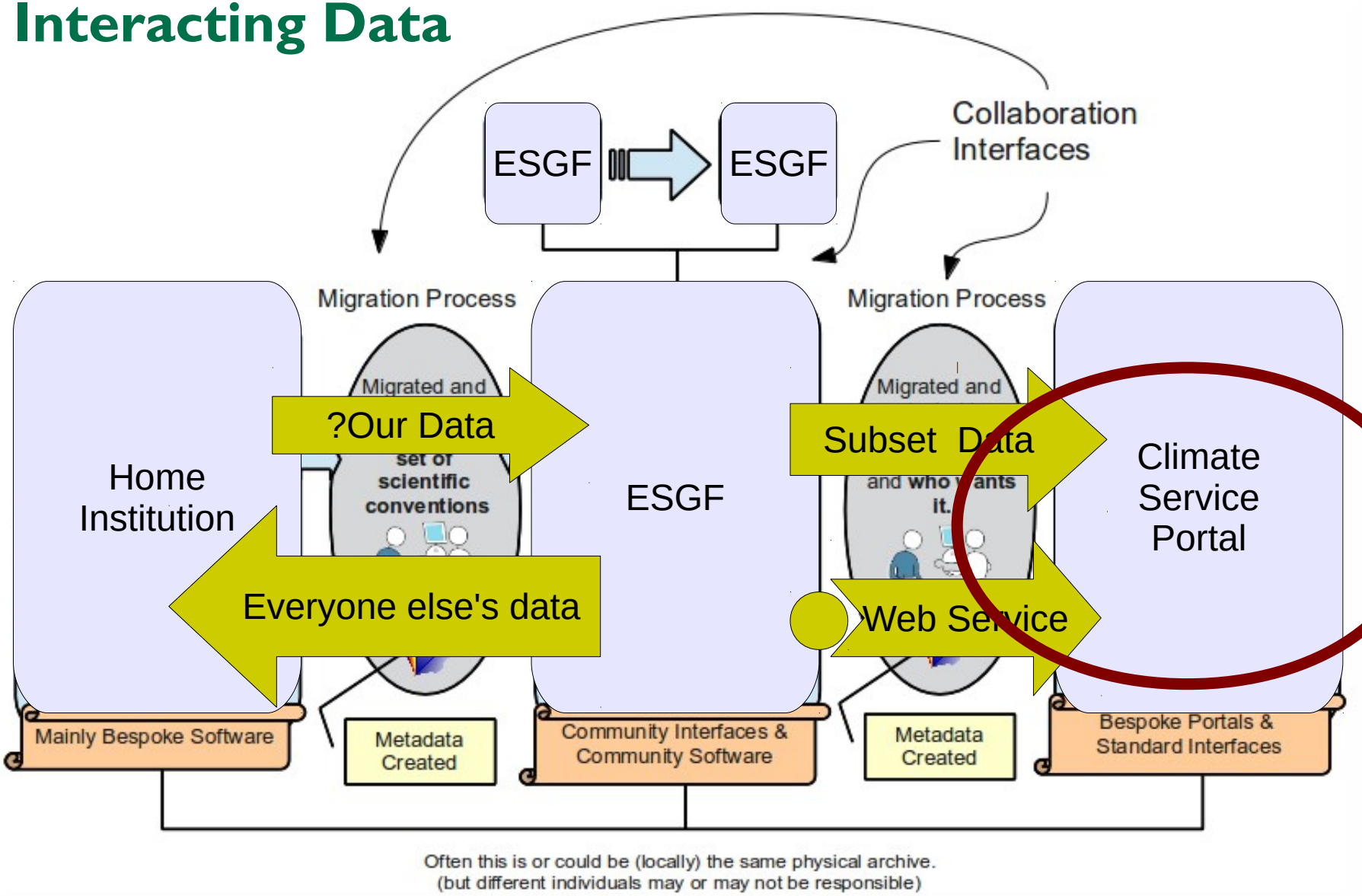
A federated system optimises the use of the limiting resource: ***people.***

No institution can go it alone: data at scale is a global activity based around large national facilities....

Oh really? Yes!
Very staff intensive!

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

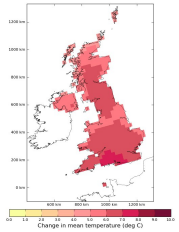
Interacting Data



From concepts in Treloar et al 2007 (D-LI

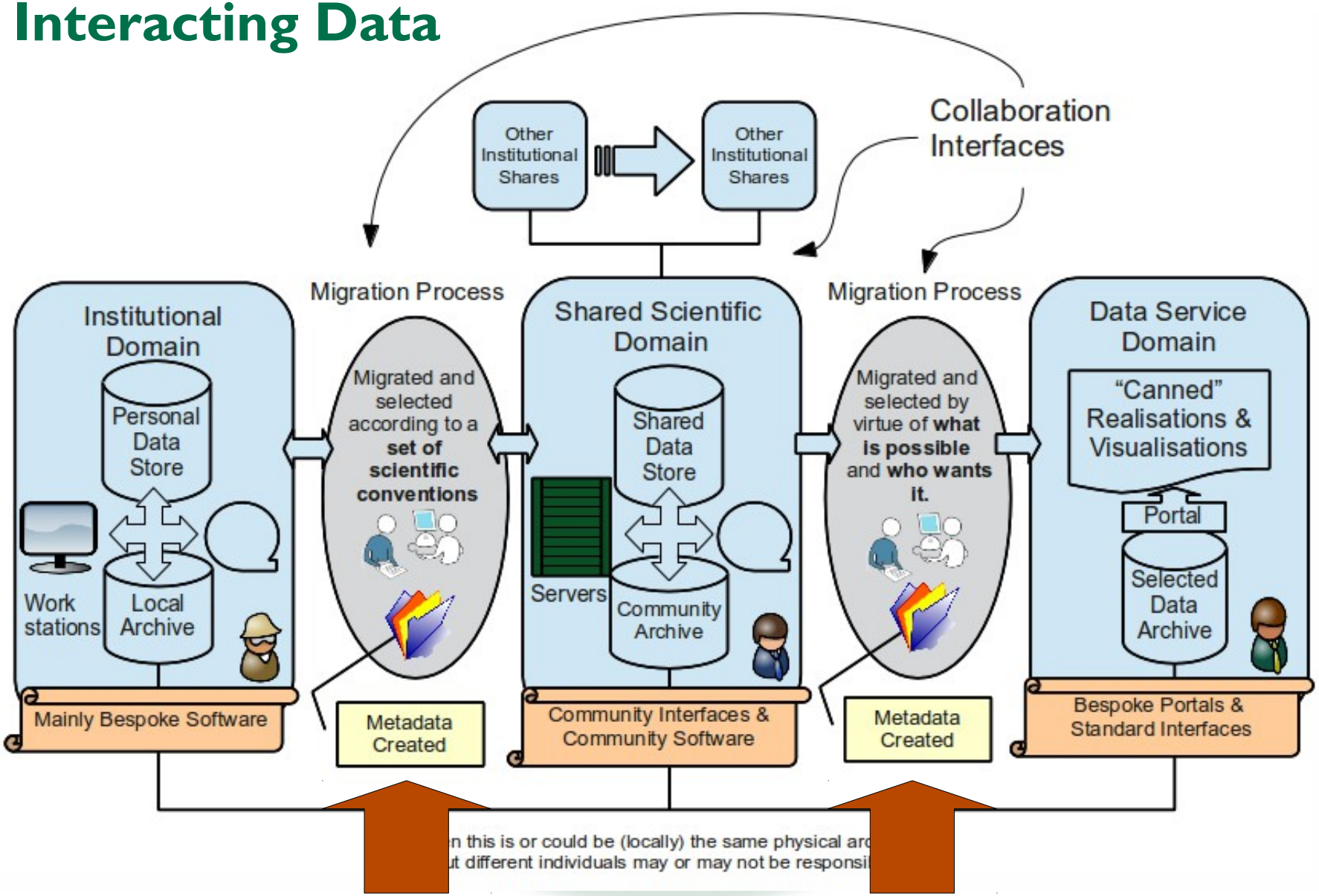
Bespoke Portals: E.G. UKCP09 User Interface

Special Data, Special Metadata …



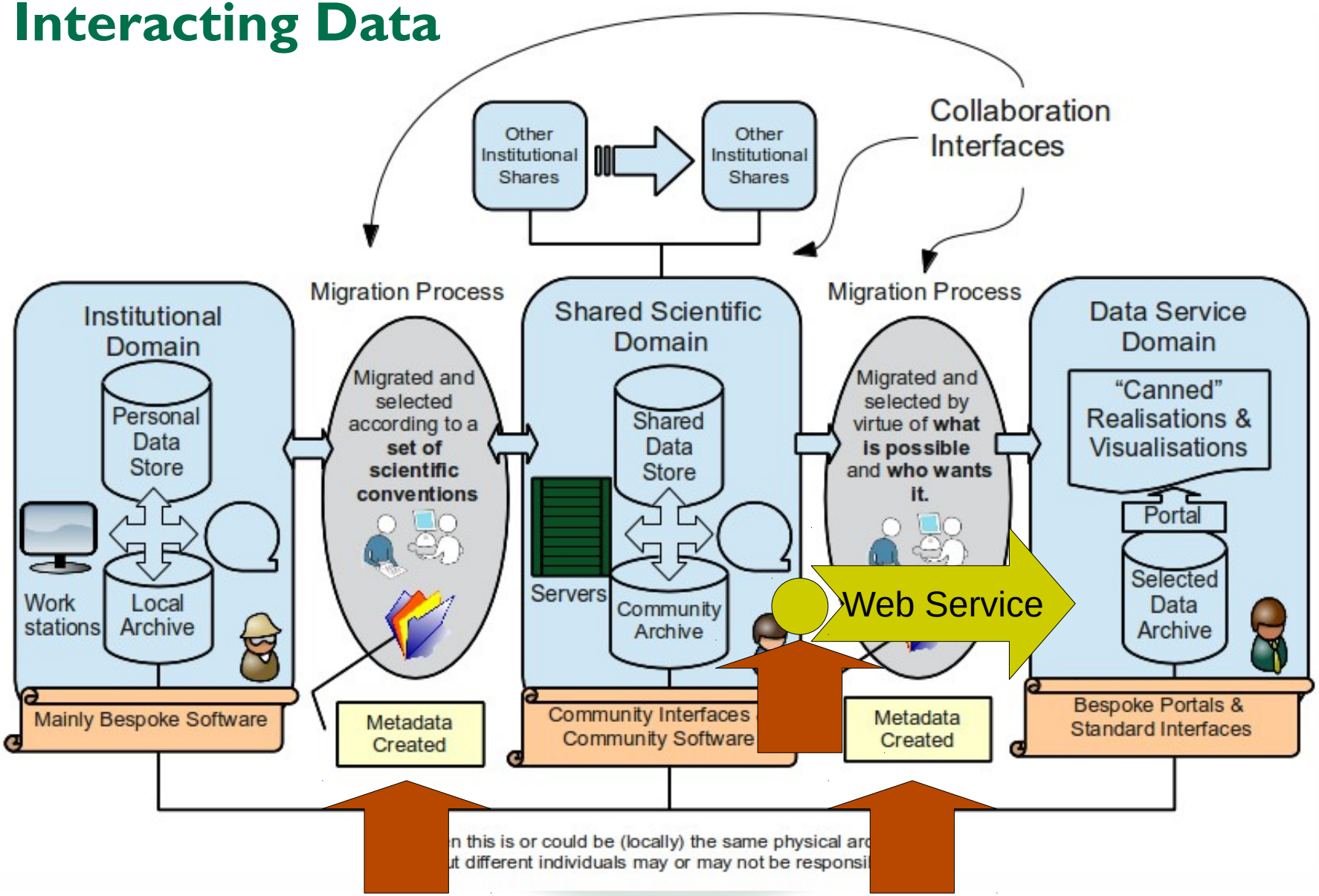
UK Climate Projections: Sophisticated User Interface, to support hundreds of simultaneous users dynamically interacting with data, organised for the specific purpose of supporting this user interface.

Interacting Data



From concepts in Treloar et al 2007 (D-LI

Interacting Data



From concepts in Treloar et al 2007 (D-LI

A one slide guide to CMIP5 from a data perspective

Fifth Climate Model Intercomparison Project (CMIP5)

World Climate Research Programme WCRP- WGCM

Involves all the major climate modelling centres.

Original Timing:

o(2) PB of requested output from 20+ modelling centres finished early 2010!

Actual Timing?

Years late.

(23/05/13):

101 experiments

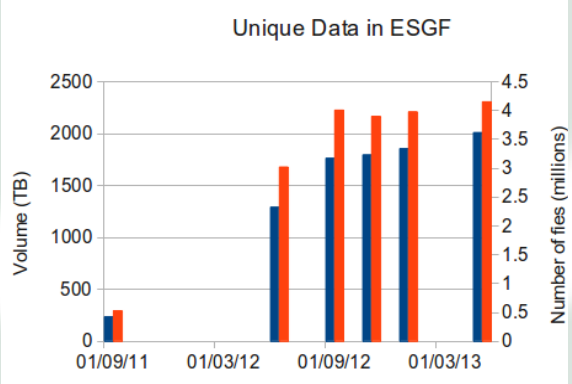
61 model variants

590,000 datasets!

4.5 million files

2 PB in global archive

Unknown PB locally!



Blue: Volume; Red: Files

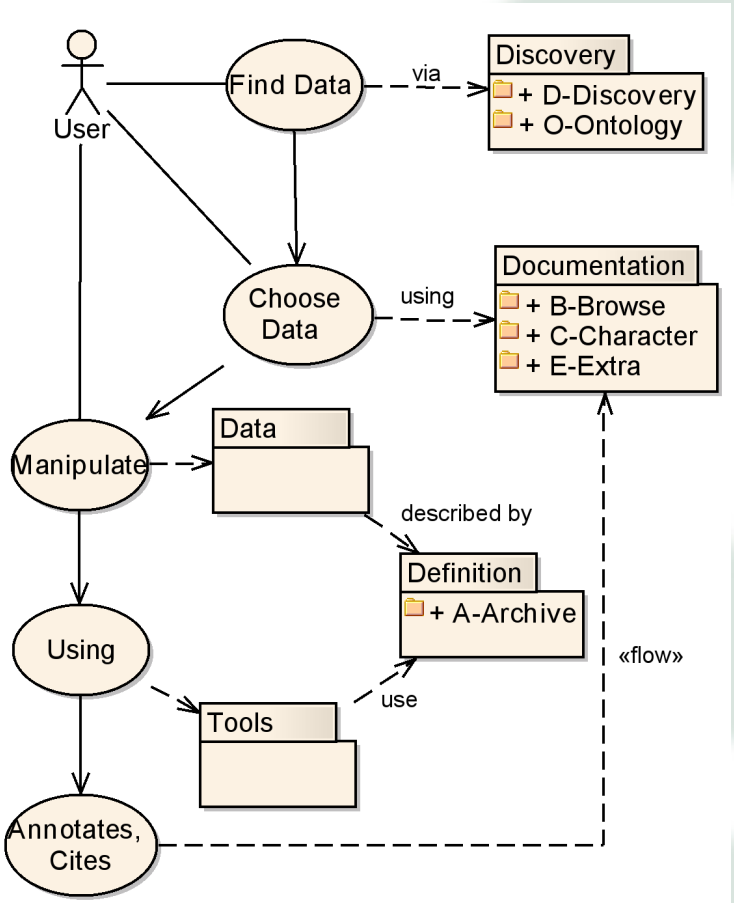
(NB: replicas and versions!)

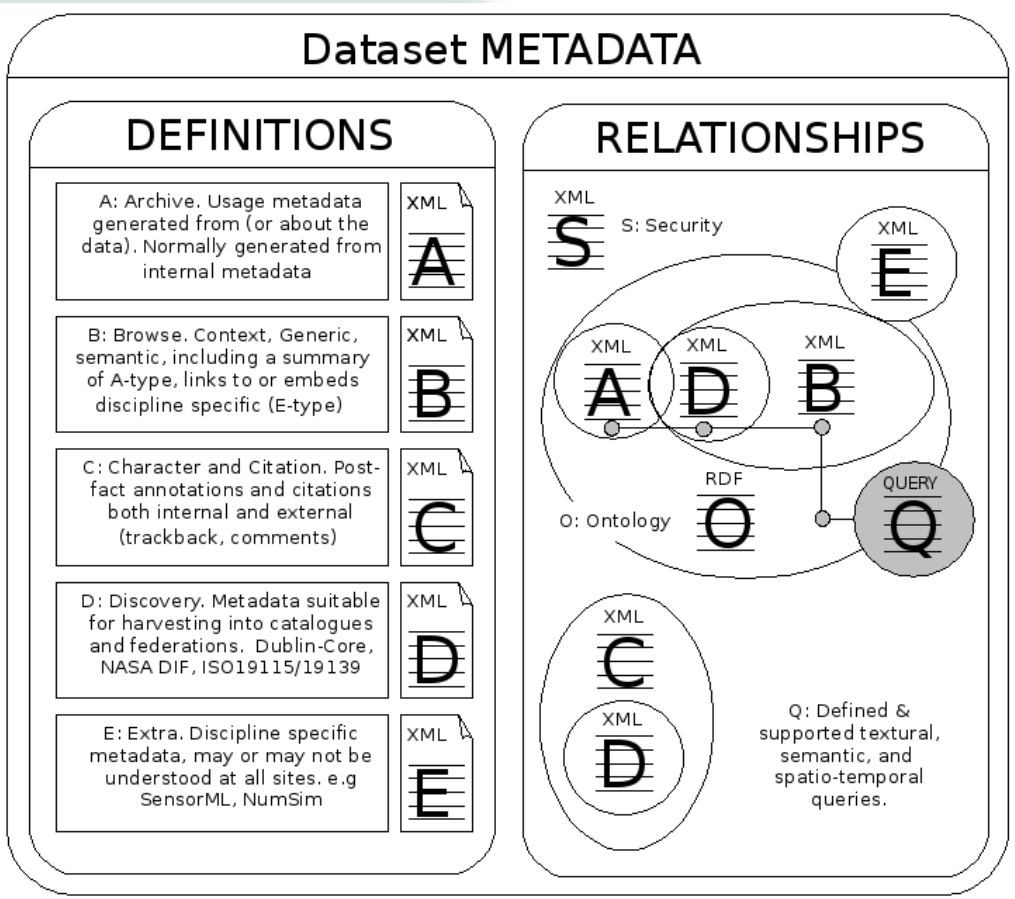
PCMDI-led, community developed (GO-ESSP)

s/w infrastructure for data delivery:

Earth System Grid Federation

Metadata Taxonomy





Lawrence et al 2009, doi:10.1098/rsta.2008.0237

My A, B, and C metadata will be very different from yours.

I do not see mileage in (much) semantic standardisation across disciplines in this space
(as opposed to navigable interoperability a la “linked data”)

(Standard Names, Cross-Discipline, Yes, Slowly.
International Intra-Discipline, Yes, Quickly.)

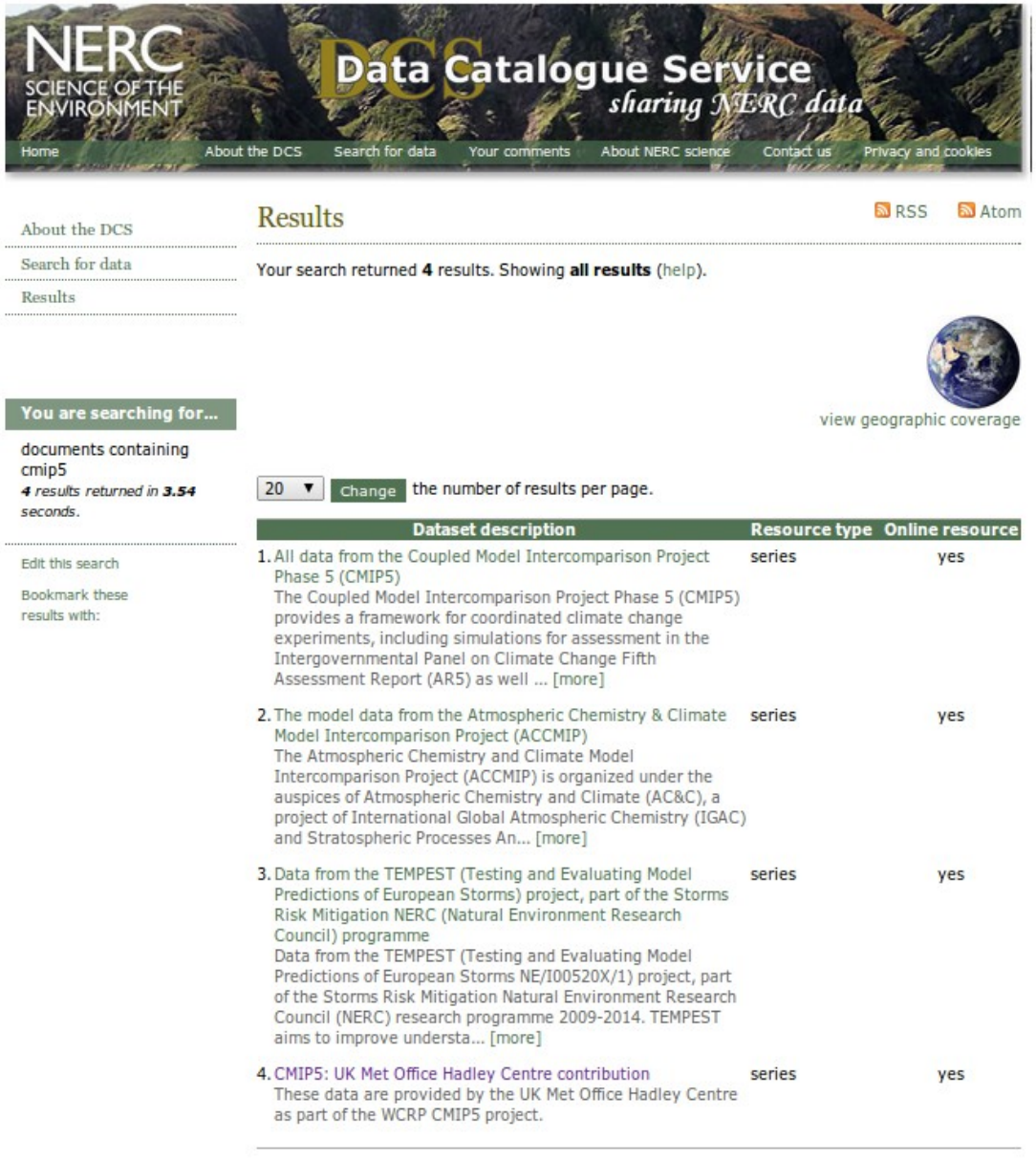
Discovery Metadata

Well, at least you can find that we hold the CMIP5 data via catalogue servers …

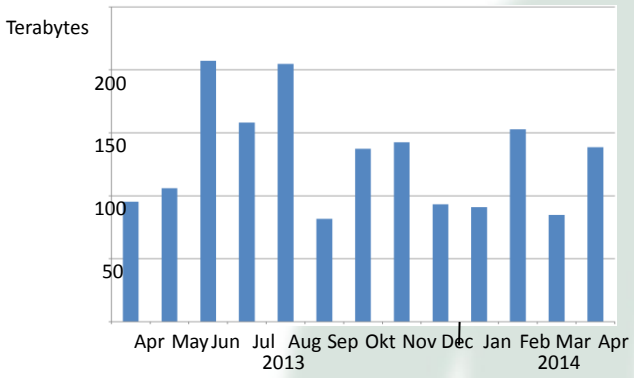
A single entry to a multi-petabyte dataset, and because we're the MOHC node, we have a specific entry for them.

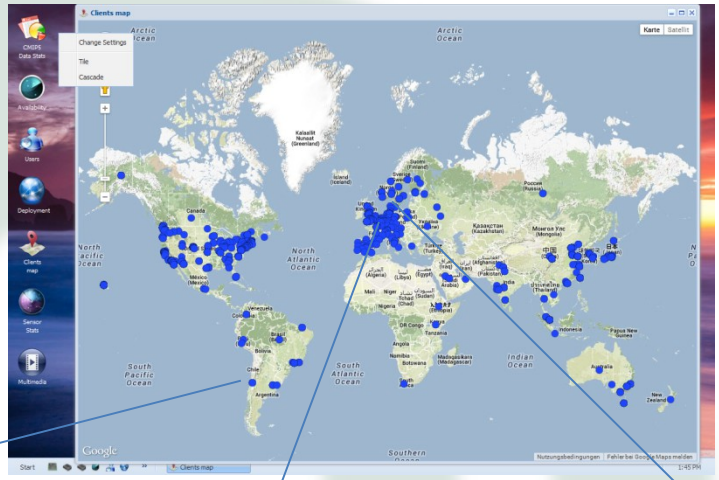
Should we have more?

("What is a dataset?")



ESGF Usage





is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Slide courtesy of Stefan Kindermann, DKRZ and IS-ENES2

Individual End Users

- Limited resources (bandwidth, storage,..)

Organized User Groups

- Organize a local cache of required files
- Most of group don`t access ESGF but cache

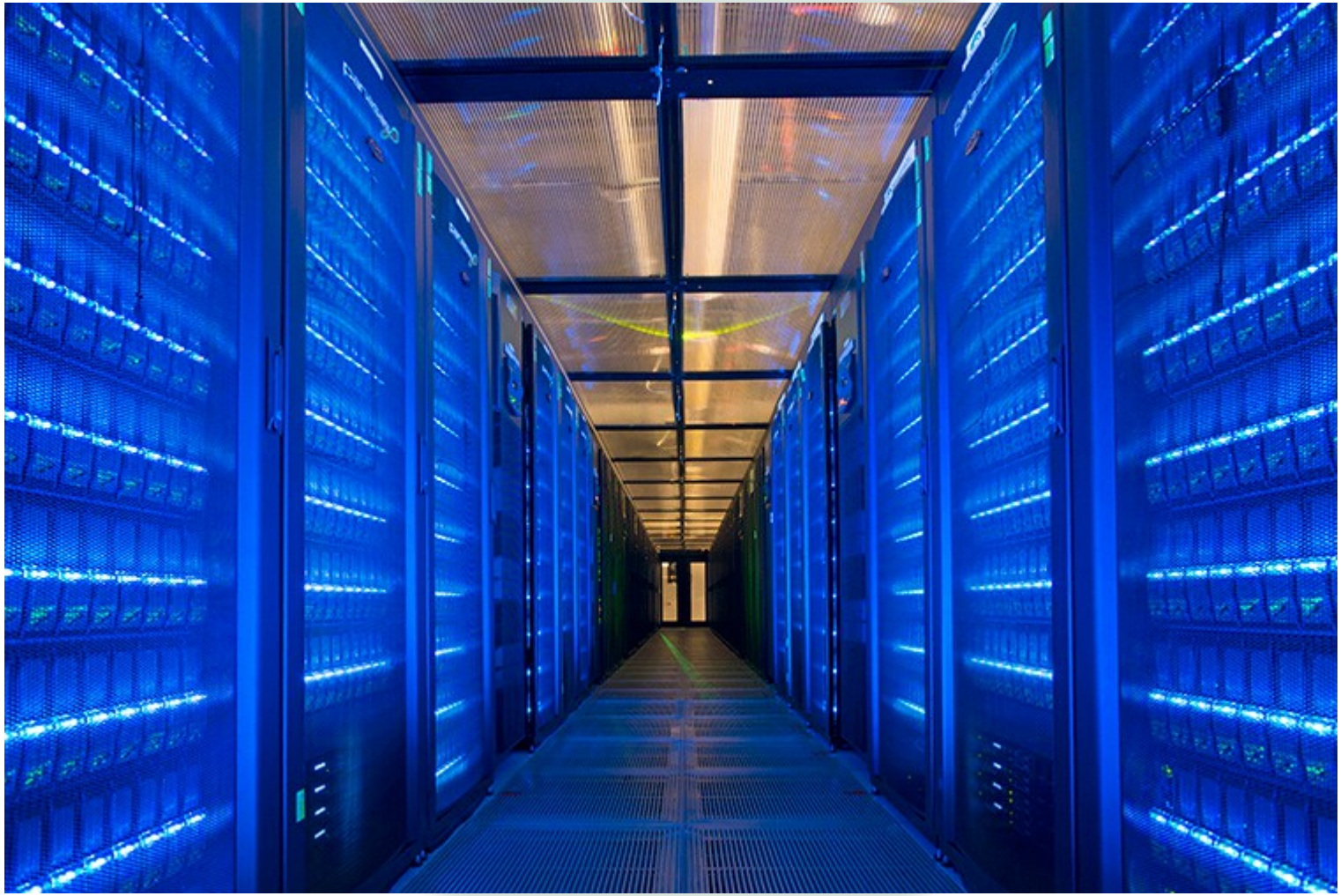
Data Centre Service Group

- Provides access to ESGF replica cache
- May also provide access to data near compute resources
- (BADC, DKRZ, IPSL, KNMI, UC)

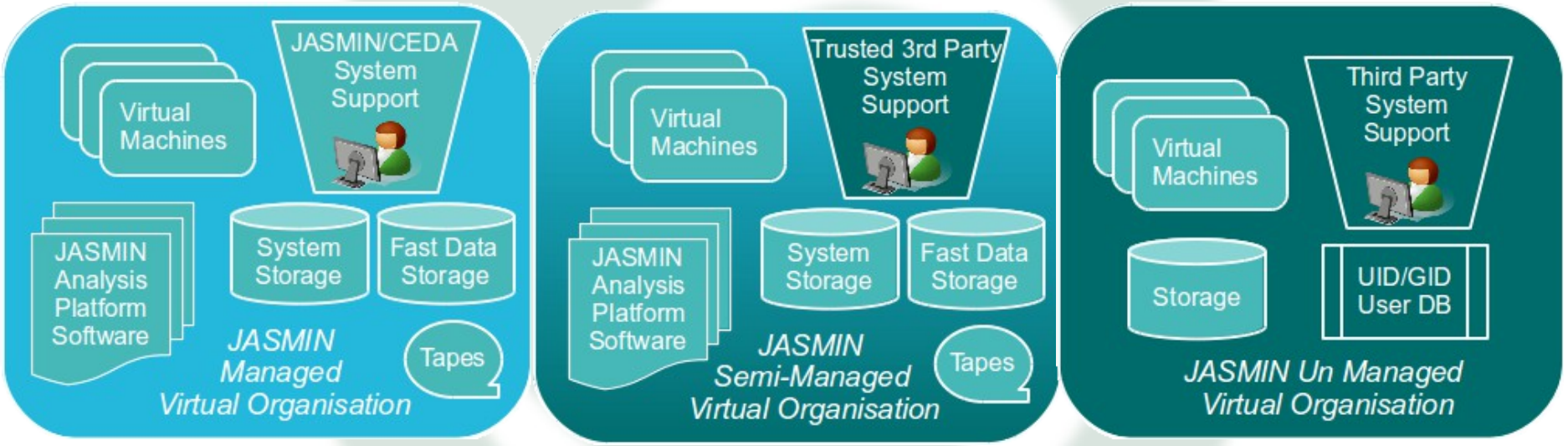
Trend

Needed: Replacement for „*Download and Process at Home*“ Approach

Dedicated Analysis Facilities: JASMIN



Managed, Semi- and Un-managed Organisations



Platform as a Service (Paas) ----------> Infrastructure as a Service (IaaS)

Some Special VOs

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.

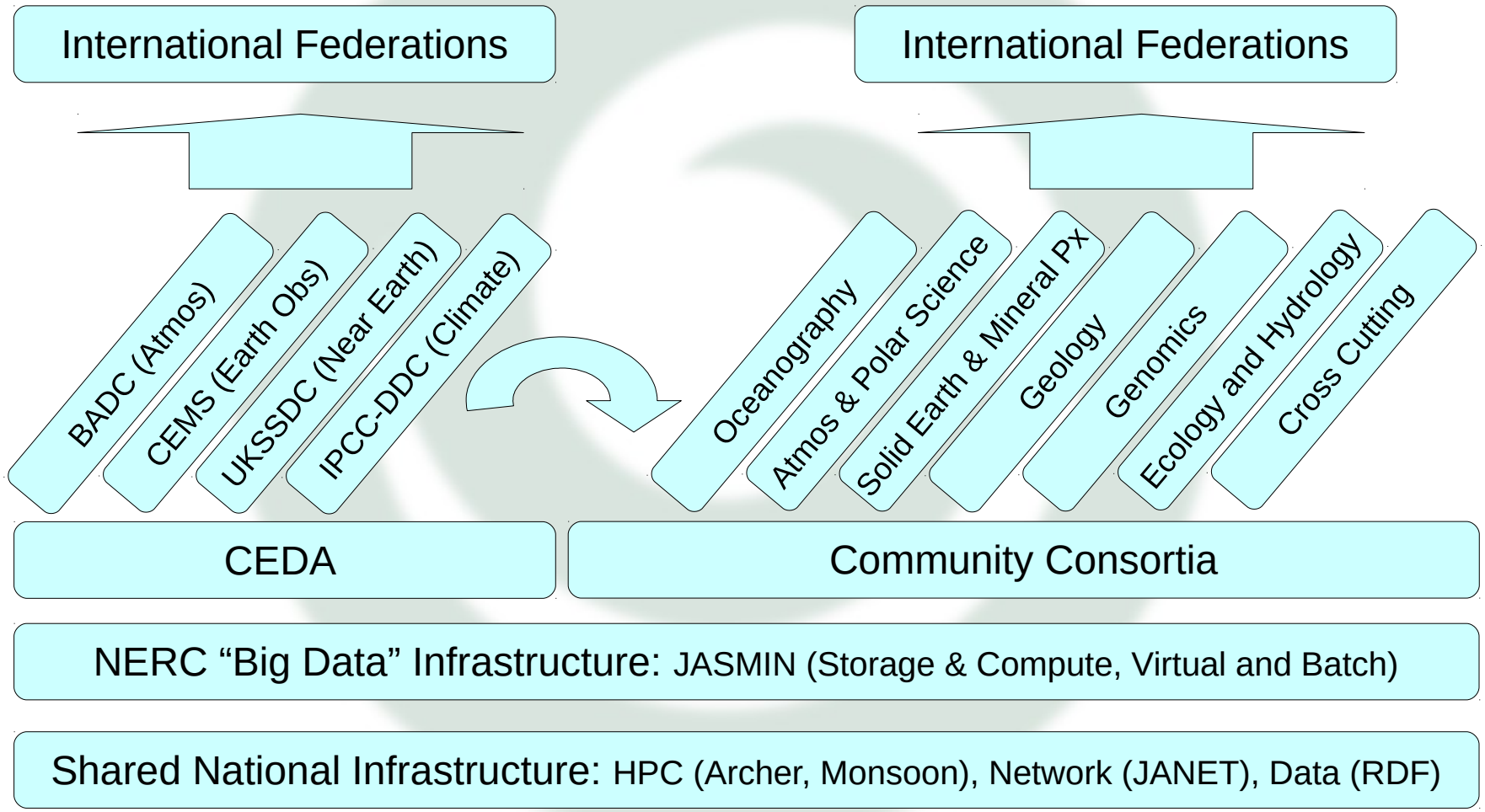
CEMS: The facility for Climate, Environment and Monitoring from Space

- 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.

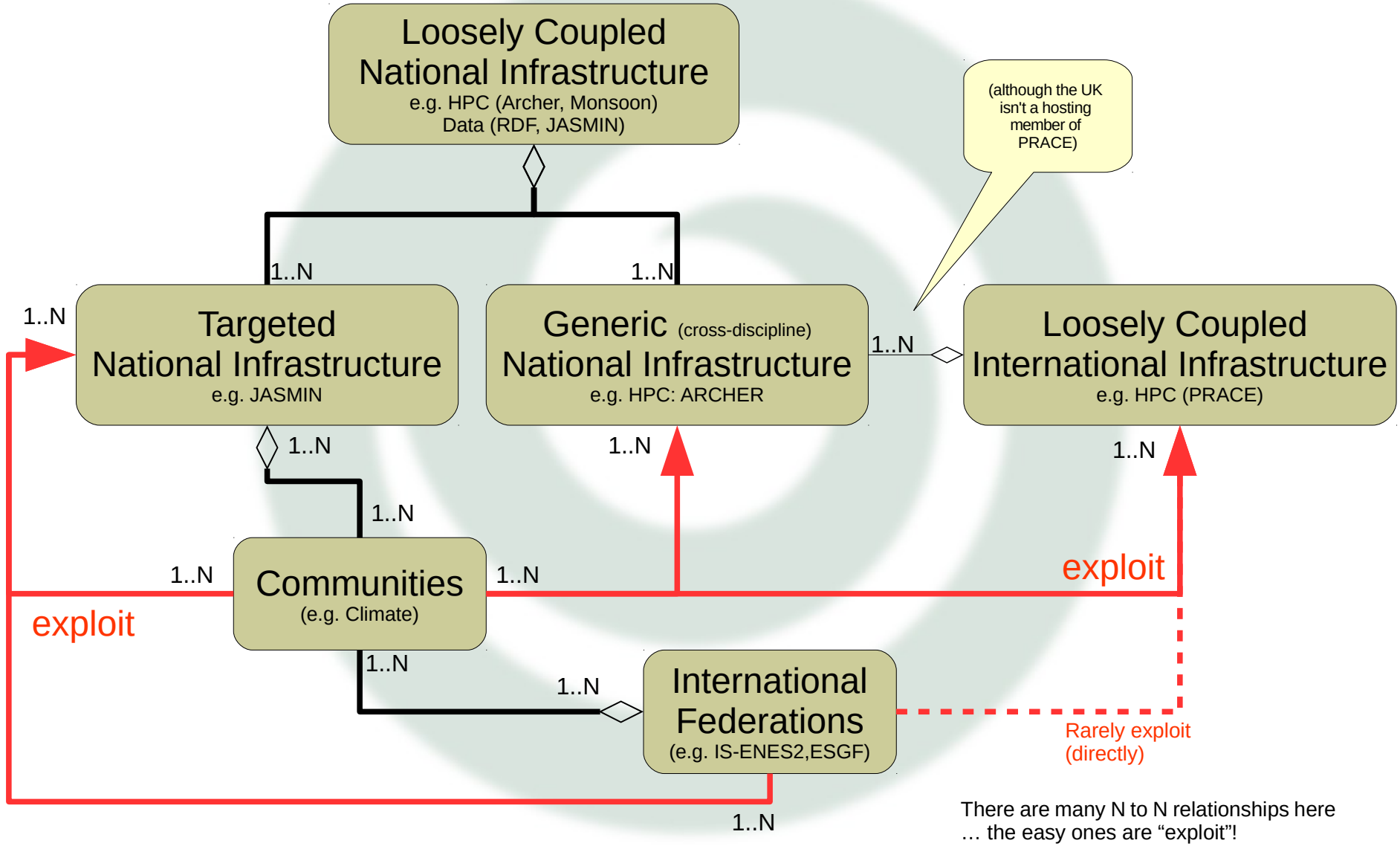
EOS Cloud

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

UK e-Infrastructure (from a NERC perspective)



Infrastructural Relationships



There are many N to N relationships here … the easy ones are "exploit"!

ExArch – Pushing towards exascale data handling

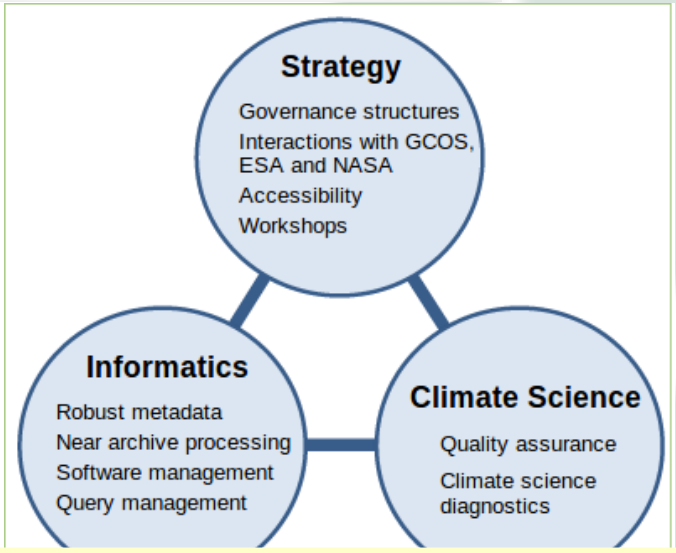
ExArch: Climate analytics on distributed exascale data archives (Juckes PI, G8 funded)

NERC Centres for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

UCLA · Institut Pierre Simon Laplace · DKRZ

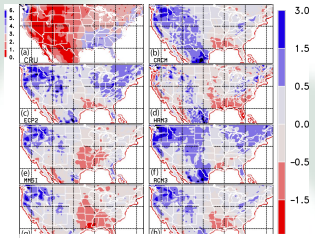
UNIVERSITY OF TORONTO · Princeton University

Centro Euro-Mediterraneo per i Cambiamenti Climatici · Science & Technology Facilities Council

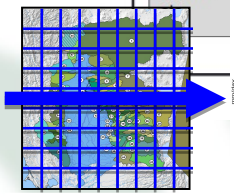


Martin Juckes, V. Balaji, B.N. Lawrence, M. Lautenschlager, S. Denvil, G. Aloisio, P. Kushner, D. Waliser, S. Pascoe, A. Stephens, P. Kershaw, F. Laliberte, J. Kim, S. Fiore

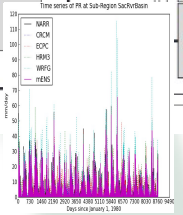
Regional Climate Model Evaluation System (RCMES)



Observation/Model rainfall

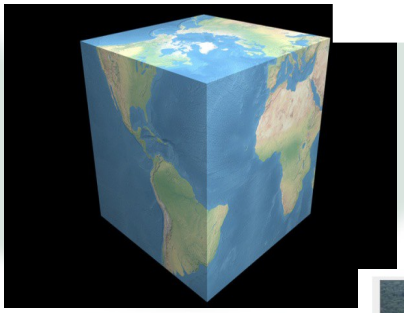


Map over a basin using an area-matching method

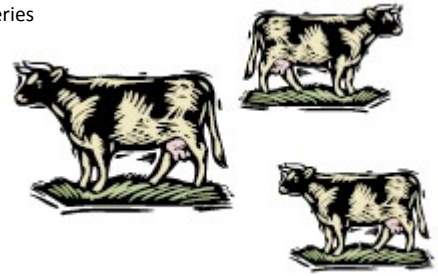


Get basin-mean time series

NOAA – PMEL "Live Access Server"



CMCC parallel data analytics framework



CEDA OGC Web Services

http://climate4impact.eu/

Crossing borders or crossing disciplines, can one do both at the same time?

Yes, but only if you limit yourself to specific activities, and not everything!

This talk has been about “hardware” infrastructure, we could have had a very similar conclusion if we discussed software Infrastructure!

One last thought: Credit and Kudos in shared activity?

Or?

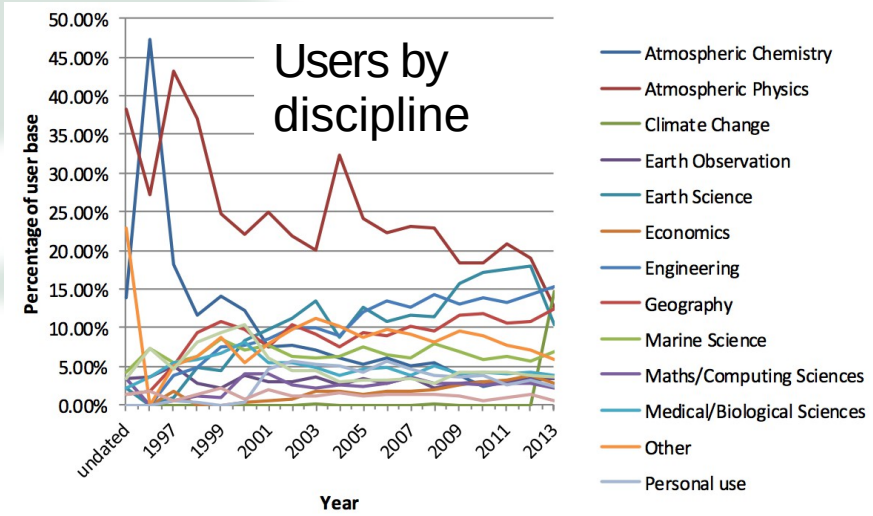
SEAT

Everyone contribute to the “single entity” or everyone contribute to a “common platform”?

Centre for Environmental Data Archival

Exist: *"to support environmental science, further environmental data archival practices, and develop and deploy new technologies to enhance access to data."*

-> Curation and Facilitation



Curation: Four Data Centres

- British Atmospheric Data Centre
- NERC Earth Observation Data Centre
- IPCC Data Distribution Centre
- UK Solar System Data Centre

(BADC, NEODC, IPCC-DDC, UKSSDC)

Over 23,000 registered users!

+ active research in curation practices!

Facilitation:

- Data Management for scientists (planning, formats, ingestion, vocabularies, MIP support, ground segment advice etc)
- Data Acquisition

  (archiving 3rd party data for community use)
- JASMIN Support

  (Group Workspaces, JASMIN Analysis Platform, Cloud Services, Parallelisation)

STFC Scientific Computing Department (SCD)

"Computing Expertise across length scales from processes within atoms to environmental modelling"

- Applications development and support,
- Compute and data facilities and services
- Research and Training
- Numerical Analysis

Data Services

- STFC: Facility Archives (ISIS, Diamond)
- LHC: UK Hub (Tier 1 archive)
- BBSRC: Institutes data archive
- MRC: Data Support Service
- NERC: CEDA backup and JASMIN elastic tape

High Performance Computing

- Emerald GPU cluster for Oxford, UCL, Southampton, Bristol.
- SCARF HPC for RAL
- Hartree: Blue Joule bluegene HPC
- Hartree: Blue Wonder idataplex HPC
- JASMIN: NERC super data cluster

Close working partnership with industry