Science Drivers: Why JASMIN?

Bryan Lawrence

NERC SCIENCE OF THE ENVIRONMENT

Science & Technology Facilities Council

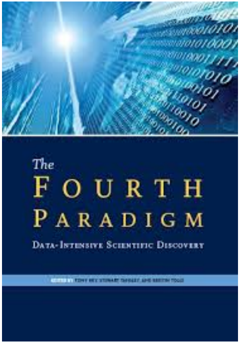
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

New ways of thinking: Data Intensive Science

The four paradigms of science:

1. **E**xperimental Science - Repeatable experiments observing and/or interfering with and observing nature.
2. **T**heoretical Science - Mathematics and Abstraction and analytical solutions compared with experiment.
3. **C**omputational Science - Numerical solutions of equations and simulations of structure and evolution.
4. **D**ata Intensive Science - All the above but mediated by so much data that the science starts ab-inito from (often) someone else's data.

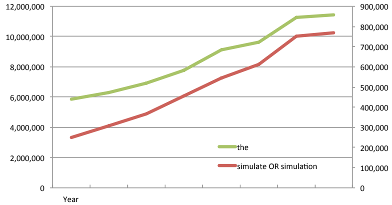


Jim Gray
(Tony Hey et al ...)

We are still working through the consequences of the advent of data intensive science, but JASMIN is one!

The growth of simulation

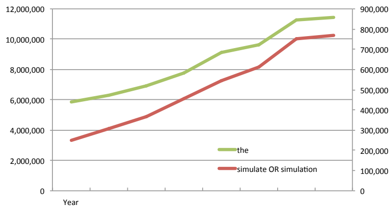
A proxy for the use of simulation in science: the growth in the numbers of (scientific) papers which mention *simulate* or *simulation* in their abstracts (via ISI).



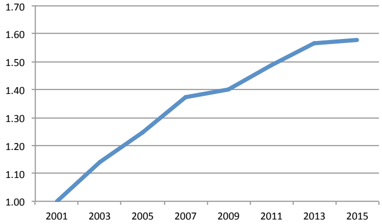
In the last fifteen years (2001-2015), the number of papers published have doubled, but the number of papers with simulation in the abstract have tripled.

The growth of simulation

A proxy for the use of simulation in science: the growth in the numbers of (scientific) papers which mention *simulate* or *simulation* in their abstracts (via ISI).

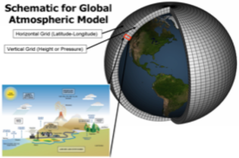


In the last fifteen years (2001-2015), the number of papers published have doubled, but the number of papers with simulation in the abstract have tripled.



Even if we conclude that a doubling of papers doesn't reflect a doubling in science of the community, we can still conclude that the proportion of the community doing, or exploiting, simulation has grown by 50%!

Core Science Requirements



Today:	Observations	Models
Volume	20 million = **2 x 10^7**	5 million grid points 100 levels 10 prognostic variables = **5 x 10^9**
Type	98% from 60 different satellite instruments	physical parameters of atmosphere, waves, ocean

Soon:	Observations	Models
Volume	200 million = **2 x 10^8**	500 million grid points 200 levels 100 prognostic variables = **1 x 10^{13}**
Type	98% from 80 different satellite instruments	physical and chemical parameters of atmosphere, waves, ocean, ice, vegetation

→ Factor 10 per day

→ Factor 2000 per time step

→ but many more time steps needed

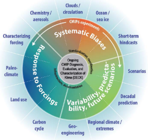
Big International Drivers:

WWRP World Weather Research Programme

GAW

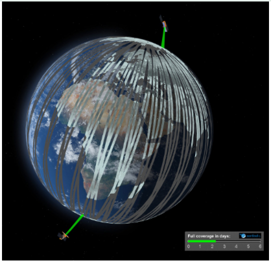
Copernicus Europe's eyes on Earth

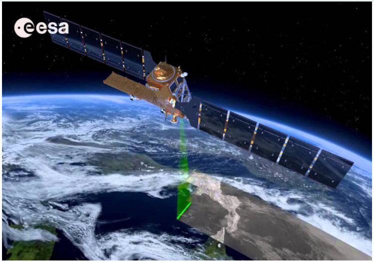
WCRP World Climate Research Programme



- aerosol cci
- cloud cci
- fire cci
- ghg cci
- glaciers cci
- antarctic ice sheet cci
- ice sheets greenland cci
- land cover cci
- ocean colour cci
- ozone cci
- sea ice cci
- sea level cci
- sst cci
- soil moisture cci
- cmug cci

The Sentinels: Big EO data crucial to NERC science!





Sentinels

Sentinel 1A (2014), 1B (2016)

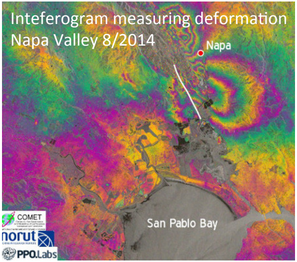
Sentinel 2A (2015) *2B (2017?)*

Sentinel 3A (2016) *3B (2018?)*

Data rate: o(6) PB/year

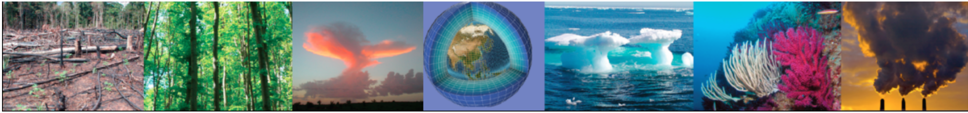
NERC SCIENCE OF THE ENVIRONMENT

COMET: Centre for Observation and Modelling of Earthquakes, Volcanoes, and Tectonics



(Picture credits: ESA, Arianespace.com, PPO.labs-Norut–COMET-SEOM Insarap study, ewf.nerc.ac.uk/2014/09/02/new-satellite-maps-out-napa-valley-earthquake/)

Infrastructure Requirements



Infrastructure Strategy for the European Earth System Modelling Community 2012-2022

Key science questions

1. How predictable is climate on a range of timescales ?
2. What is the sensitivity of climate and how can we reduce uncertainties ?
3. What is needed to provide reliable predictions of regional climate changes ?
4. Can we model and understand glacial-interglacial cycles ?
5. Can we attribute observed signals to understand processes ?

Recommendations:

1. Provide a blend of high-performance computing facilities . . .
2. Accelerate the preparation for exascale computing . . .
3. Ensure data from climate simulations are easily available and well documented, especially for the climate impacts community.
4. Build a physical network connecting national archives with transfer capacities exceeding Tbits/sec.
5. Strengthen the European expertise in climate science and computing . . .

enes
EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

The Propagation of Direct Numerical Simulation

Primarily mathematical representation of a complex system of processes

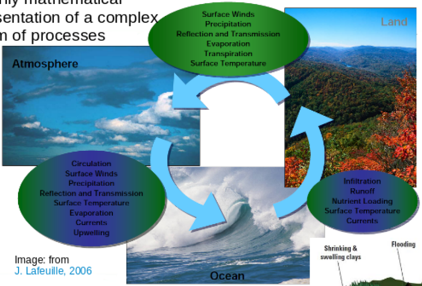
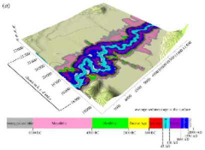
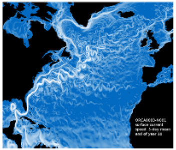
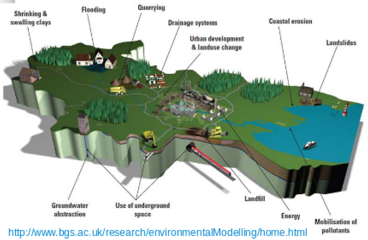


Image: from J. Lafeuille, 2006



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



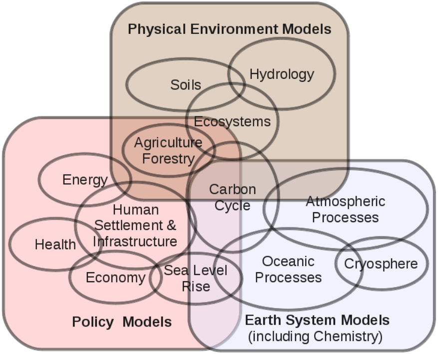


http://www.bgs.ac.uk/research/environmentalModelling/home.html

More communities want to observe and simulate the world at ever higher resolution!

More complexity!

Communities



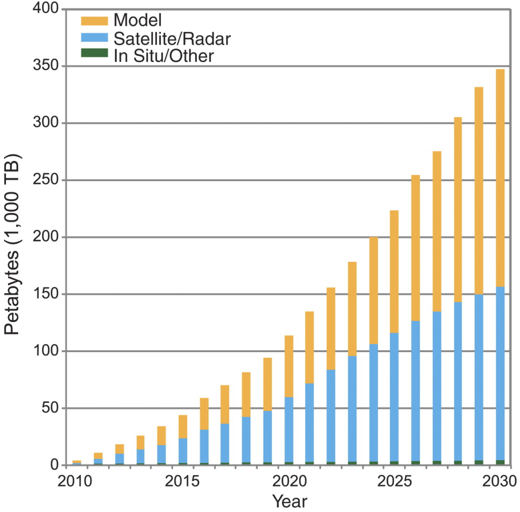
Many interacting communities, each with their own software, compute environments, observations etc.

Figure adapted from Moss et al, 2010

More Data

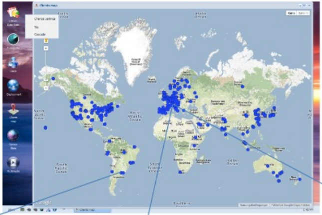
Fig. 2 The volume of worldwide climate data is expanding rapidly, creating challenges for both physical archiving and sharing, as well as for ease of access and finding what's needed, particularly if you're not a climate scientist.

(BNL: Even if you are?)



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

The trend



is-enes

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, use cache instead!

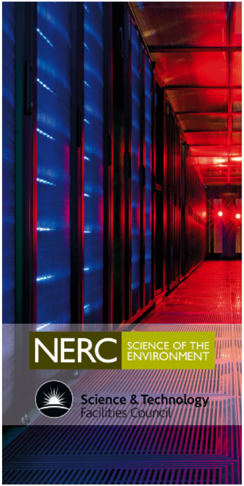
Data Centre Service Group

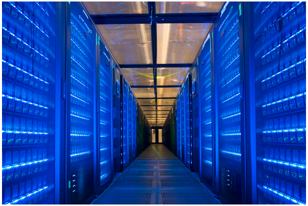
- Provides access to ESGF replica cache
- May also provide access to data near compute resources
- (BADC, DKRZ, IPSL, KNMI, UC)

Trend →

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

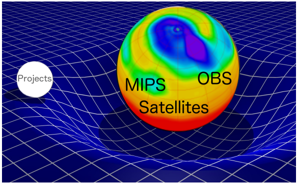
So we have built a Data Intensive Computing System: JASMIN





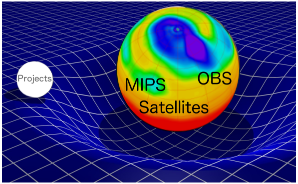
- 16 PB Fast Storage
  (Panasas, many Tbit/s bandwidth)
- 1 PB Bulk Storage
- Elastic Tape
- 4000 cores: half deployed as hypervisors, half as the "Lotus" batch cluster.
- Some high memory nodes, a range, bottom heavy.

JASMIN — The Data Commons

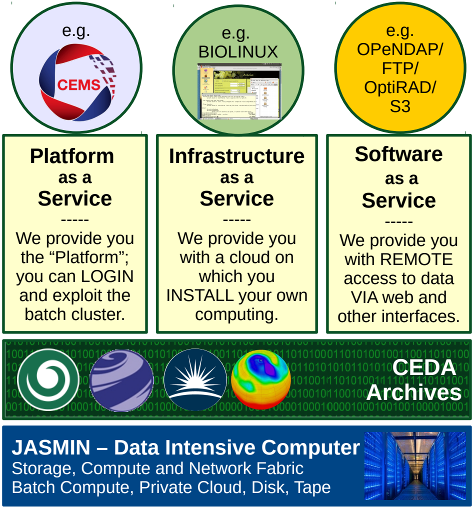


- Provide a state-of-the art storage and computational environment
- Provide and populate a managed data environment with key datasets (the "archive").
- Encourage and facilitate the bringing of data and/or computation alongside/to the archive!
- Provide flexible methods of exploiting the computational environment.

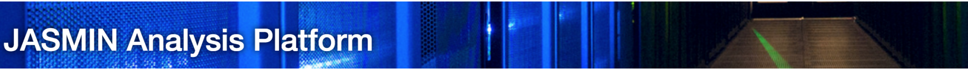
JASMIN — The Data Commons



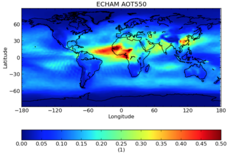
- Provide a state-of-the art storage and computational environment
- Provide and populate a managed data environment with key datasets (the “archive”).
- Encourage and facilitate the bringing of data and/or computation alongside/to the archive!
- Provide flexible methods of exploiting the computational environment.

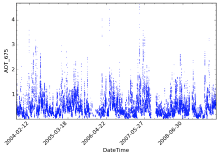


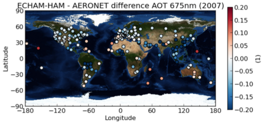
Tools



Community Intercomparison Suite:

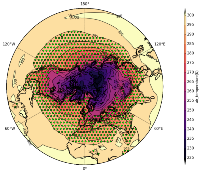




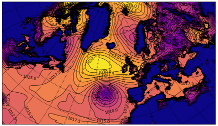


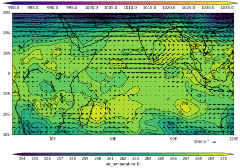
See Duncan Watson-Parris et al, 2016 (doi:10.5194/gmd-2016-27)

CF-python, CF-plot, cfview:



```
import cf, cfplot as cfp
f=cf.read('/opt/graphics/cfplot_data/tas_A1.nc')
g=f.subspace(time=15)
cfp.gopen()
cfp.cscale('magma')
cfp.mapset(proj='npstere')
cfp.con(g)
cfp.stipple(f=g, min=265, max=295, size=100, color='#00ff00')
cfp.gclose()
```





http://www.cistools.net/
http://cfpython.bitbucket.org
http://ajheaps.github.io/cf-plot
...and many more ... all shared and (hopefully) kept up to date on the JAP.

Common Software/Algorithm Patterns

Supporting a wide variety of algorithms and workflows: (but much to do to exploit parallelism)

"Big Data Ogres" by analogy with the Berkely Dwarves for computational patterns.

Different Problem Architectures, e.g:

1. Pleasingly Parallel (e.g. retrievals over images)
2. Filtered pleasingly parallel (e.g. cyclone tracking)
3. Fusion (e.g. data assimilation)
4. (Space-)Time Series Analysis (FFT/MEM etc)
5. Machine Learning (clustering, EOFs etc)

Important Data Sources, e.g:

1. Table driven (eg. RDBMS + SQL)
2. Document driven (e.g XMLDB + XQUERY)
3. Image driven (e.g. GeoTIFF + your code)
4. (Binary) File driven (e.g. NetCDF + your code)

Sub-Ogres: Kernels & Applications, e.g:

1. Simple Stencils (Averaging, Finite Differencing etc)
2. 4D-Variational Assimilation/ Kalman Filters
3. Data Mining Algorithms (classification/clustering) etc
4. Neural Networks

Modified from Jha et al 2014 arXiv:1403.1528[cs]

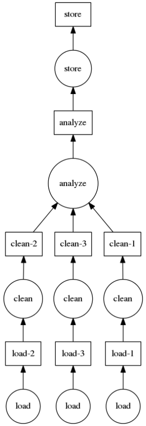
Uncommon software solutions

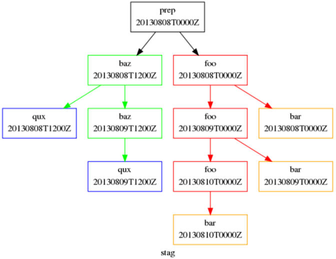
Plethora of parallel architectures and tools out there

Contrast between two very different parts of our workflow:

- Many of our analysis tasks are build once, use once, throwaway. No room for optimisation (or MPI), yet
- Much of our workflow is repeatable: “build”, “run”, “move”, “reduce/reformat”, “analyse”. Much room for automation.

Whatever tools we use, we’ll need to get use to generating, understanding, and exploiting concurrency in more complicated ways:



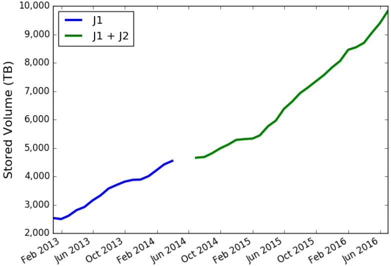


Much to do, as infrastructure providers, and users, to harness these tools to accelerate our workflows!

(These two examples: dsk, and cylc, representing analysis and scheduling, reduction and proliferation.)

Local growth thus far?

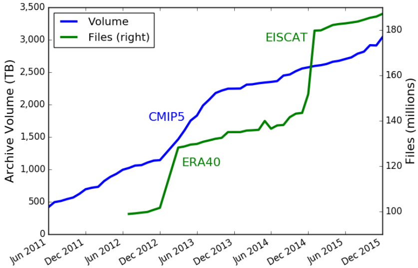
JASMIN phase 1 and phase 2: total disk storage:



Note growth rates:

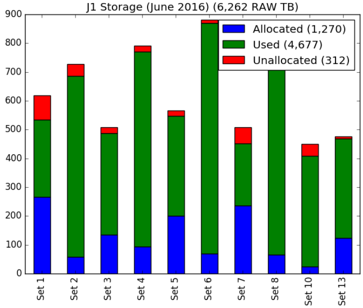
- from early 2013 to early 2015: 1.25 PB/year
- since early 2015: 2.5 PB/year (ARCHER upgrade?)

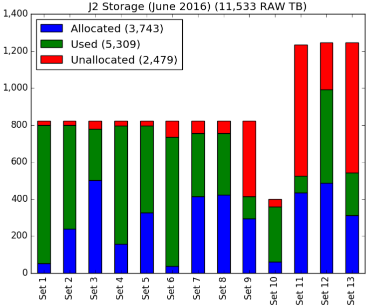
Within that, the archive growth rate:



- Note the steep rise in 2012/2013 associated with CMIP5 (we expect CMIP6 to be ten times as big and arriving in the same sort of time duration).

Capacity for future growth?



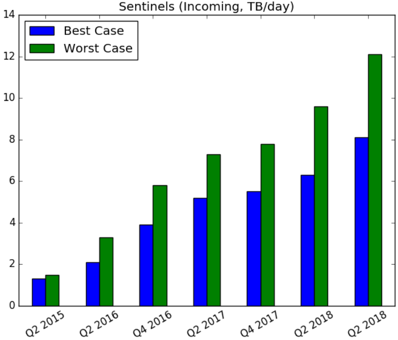


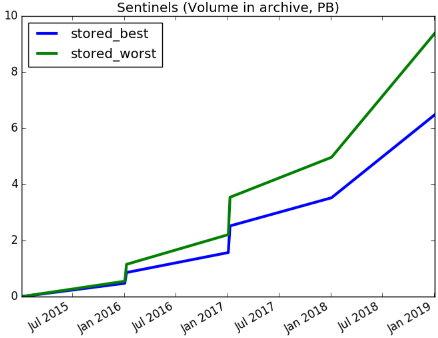
- J1 bladesets are more or less full, with a little headroom for user data analysis.
- 2 PB (Raw) of room for new allocations (≈ one year until JASMIN is full!)

It's worth noting that while we (have) need(ed) parallel file systems for performance and ease of management at the petascale, they don't come without problems in terms of managing and using them efficiently.

Sentinel Data Rates

If we just consider Sentinels 1 to 3

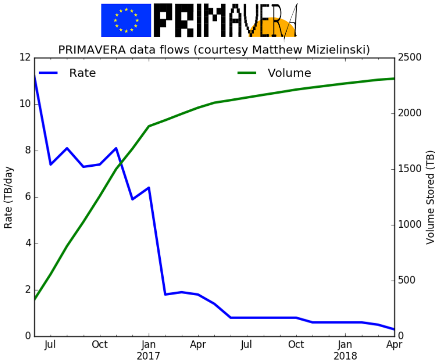




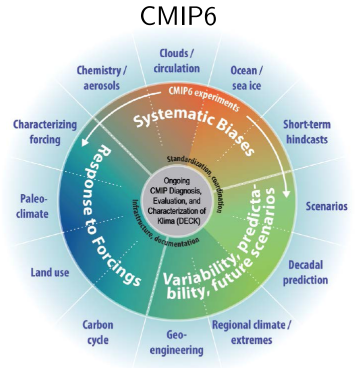
- Uncertainty in data rates grows with time. Not surprisingly we consider lower data growth rates "best" case scenarios.
- Uncertainty plays out in the archive with differences of petabytes in the volume stored in the not too distant future.

It's clear that we won't be storing all Sentinel data on disk, and while we have the tape (and library) capacity to store the Sentinel data, we don't yet have the software systems in place to make it easy for the user community to exploit a rolling cache for whole mission processing.

Model Data Rates



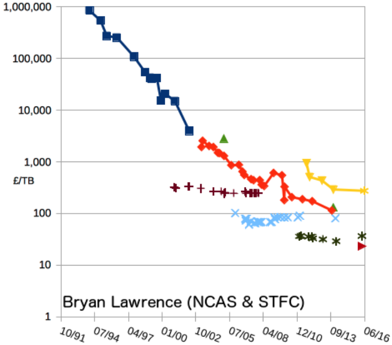
- Typical modelling project: o(1 year) running primary simulations, but more data arriving up to two years.
- Data will be used for many years thereafter by both the original modelling community, and many others.



- Data rates and volumes still unknown, but at least 10 PB over the 2017/2019 period, and possibly much much more.
- . . . as much of which needs to be on disk as we can manage!

Storage Costs

The cost of storage at STFC over 25 years



(With thanks to Peter Chiu, Jonathan Churchill, and Tim Folkes)

Kryder's Law is definitely slowing!

- Disk (Blue and red lines)
- Parallell Disk (Yellow lines)
- Tape Generations show as unconnected points (often same tapes, different drives!)
- Tape is likely to be cheaper for the foreseeable future (disk technology advances slowing rapidly)
- (The worry is that market forces may drive tape and even disk into extinction!)

Note that as the volume increases, the cost of software to manage the large volume storage needs to be added to the raw cost of disk.

We might be able to move to object stores, which will bring our raw cost per TB down, but we're all going to have to learn to live without file systems!

Drivers and Choices

Massive increase in data production driven by Moore's Law from

- Growth in the use of simulation,
- Growth in resolution and complexity of simulation,
- Growth in the number and frequency of observations,

Compounded by

- Complex interactions between communities

Resulting in massive increase in

- Number and volume of datasets in the CEDA archive,
- Number of communities, and volumes of data they need to handle, and
- Interactions and the need for quality documentation of data and computational provenance.

Drivers and Choices

Massive increase in data production driven by Moore's Law from

- Growth in the use of simulation,
- Growth in resolution and complexity of simulation,
- Growth in the number and frequency of observations,

Compounded by

- Complex interactions between communities

Resulting in massive increase in

- Number and volume of datasets in the CEDA archive,
- Number of communities, and volumes of data they need to handle, and
- Interactions and the need for quality documentation of data and computational provenance.

All this accompanied by a complex and varied computational requirements:

- People need access to data, and to share data, and to provide their own services on their data.
- Data volumes are getting so large that parallelised workflows are necessary, but it's hard to build "throw-away parallel codes".
- "Bringing compute to the data" requires support for complex computational environments.

Leading to

- Complex infrastructure requirements, the death of old software friends (e.g filesystems), the rise of new paradigms (containerisation), and
- Much new learning for us, the community of scientists who want exploit data to address society's pressing problems!