JASMIN, Common Data Issues and more…

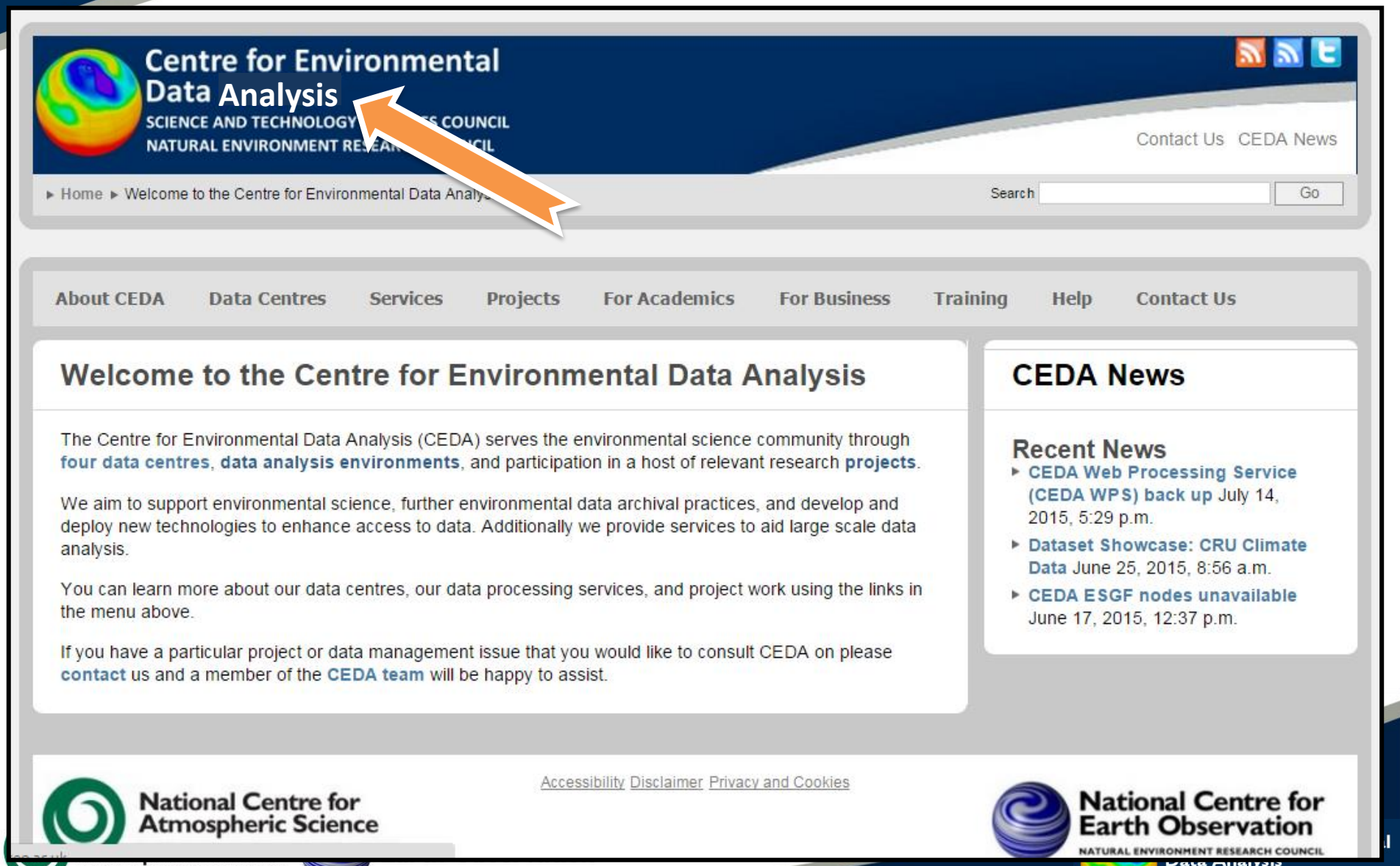
Numerical Simulations at scale are all about logistics!

November 25, 2015

Ag Stephens and Bryan Lawrence

(STFC Centre for Environmental Data Analysis)

www.ceda.ac.uk



Our data centres

Data Centres

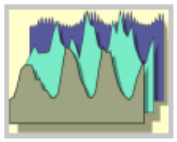
The Centre for Environmental Data Archival is responsible for the running of the following data centres:

The British Atmospheric Data Centre

The British Atmospheric Data Centre (BADC), NERC's designated data centre for the UK atmospheric science community, covering climate, composition, observations and NWP data.

NERC Earth Observation Data Centre

The NEODC is NERC's designated data centre for Earth Observation data and is part of NERC's National Centre for Earth Observation.



The UK Solar System Data Centre

The UK Solar System Data Centre, co-funded by STFC and NERC, curates and provides access to archives of data from the upper atmosphere, ionosphere and Earth's solar environment.

IPCC Data Distribution Centre

The Intergovernmental Panel on Climate Change (IPCC) DDC provides climate, socio-economic and environmental data, both from the past and also in scenarios projected into the future. Technical guidelines on the selection and use of different types of data and scenarios in research and assessment are also provided.UK Climate Projections.

Outline

- JASMIN
- Common data issues
- Big data and scale
- Where to put the data?
- ESGF
- Constraints and how they influence what we can do

JASMIN - what is it?

JASMIN: A platform for data processing

Storage

- **15 PB high-performance disk** storage:
  - CEDA archive (5 PB)
  - Over 100 Group Workspaces for projects/groups (10PB).
- Equivalent capacity in near-line **tape** system

Compute

- ~4000 compute nodes, interchangeably deployable in either:
  - batch compute cluster (**LOTUS**), or
  - cloud hosting environment
- Individual scientific analysis servers

Science & Technology Facilities Council
Rutherford Appleton Laboratory

http://www.jasmin.ac.uk

JASMIN | What is JASMIN | Services | Users | Get Access | How to Use | Help | FAQ | Search | Go

Data-intensive computing

JASMIN provides the UK and European climate and earth-system science communities with an efficient data analysis environment. Many datasets, particularly model data, are too big to be easily shipped around: JASMIN enables scientists to bring their processing to the data.

Flexible data access

JASMIN provides new ways for scientists to collaborate in self-managing group workspaces, enabling models and algorithms to be evaluated alongside curated archive data, and for data to be shared and evaluated before being deposited in the permanent archive.

Scalable future

JASMIN enables CEDA to carry out its mission of data curation and facilitation more efficiently. Fast, parallel, scalable storage provides a home for in-demand archive data, while a virtualised server infrastructure provides a more capable base for delivery of CEDA's data centre services to the science community.

NERC SCIENCE OF THE ENVIRONMENT

Science & Technology Facilities Council

Powered by Mezzanine and Django | Themed with Bootstrap

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

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

A quick tour of JASMIN

- Services: Overview diagram, GWS, JAP, LOTUS, HPDT
- Users: brief
- Get Access: signing up, managing SSH keys, IP reg, Login
- How to Use: navigate - and others.
- Help: contact us

JASMIN analysis servers (open to all)

Server	CPUs	Memory
jasmin-sci1.ceda.ac.uk	8	32Gb
jasmin-sci2.ceda.ac.uk	8	32Gb
jasmin-sci3.ceda.ac.uk	**48**	**2Tb**
cems-sci1.cems.rl.ac.uk	8	16Gb
cems-sci2.cems.rl.ac.uk	**48**	**2Tb**

JASMIN Analysis Platform (JAP)

- Software stack for scientific analysis on JASMIN (and elsewhere)
- Common packages for climate science and geospatial analysis
- O/S: RHEL6, CentOS6.
- Deployed on JASMIN
  - Common "Science" servers
  - LOTUS
  - Sci Virtual Machine "template"

Common data issues

Standards - how they help/hinder us

- CMOR, CF-netCDF, ISO19115, DRS etc…
- They clearly have great value:
  - CMIP5 demonstrated how such efforts could have significant benefits
  - E.g. "everyone called tas 'tas' and rcp45 'rcp45"
- And they have great cost:
  - CMIP5 demonstrated that we could expend a lot of effort agreeing standards
  - And a lot more effort trying to implement them
- Do we have a choice?
  - Without standards we'll get nowhere
  - But we need to **communicate** the standards are and **provide tools to help people comply.**

Example Tasks and Deliverables in PRIMAVERA and CRESCENDO

Data Management Plans

Within PRIMAVERA we will produce and implement the Data Management Plan (DMP) as part of participation in the Open Research Data Pilot.

DMPs are useful for:

- Making us plan for data issues
- Communicating the lifecycle of different data within a project:
  - Third-party data used
  - Data generated
  - Formats, metadata, longevity, storage etc.
  - Final dissemination

They work best as living documents!

Example tasks

PRIMAVERA Data Specification:

- The specification of the Required Inputs and Outputs is a pre-requisite for all Stream 1 simulations in WP6, based on CMIP5.

PRIMAVERA Data Conversion:

- Establish the methodology and workflow for bringing data to JASMIN and **converting to the standard data format specified** in the DMP.

Both projects: Publishing to ESGF

- Making key data sets available via ESGF interfaces.

CRESCENDO: Improvements to ESMValTool

- Common code to validate model outputs (including comparison with other data) - *code near the data*

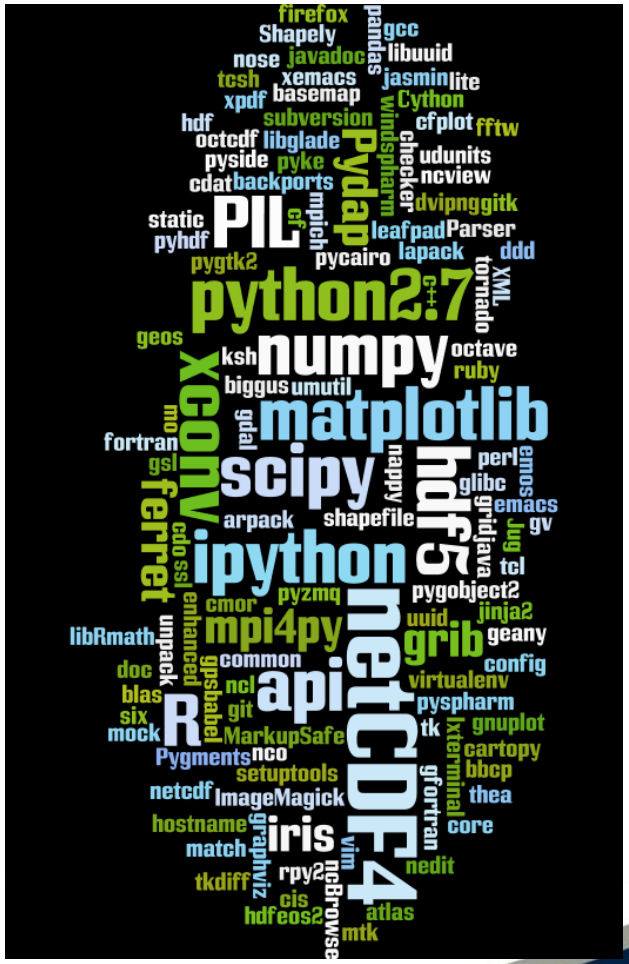
Colocation with other data

- "Bring the code to the data" - is a useful approach when working with climate simulations.

- An advantage of platforms like JASMIN is that your project data is available alongside existing archived, and current project, data sets such as:
  - CMIP5
  - CORDEX
  - ERA-Interim

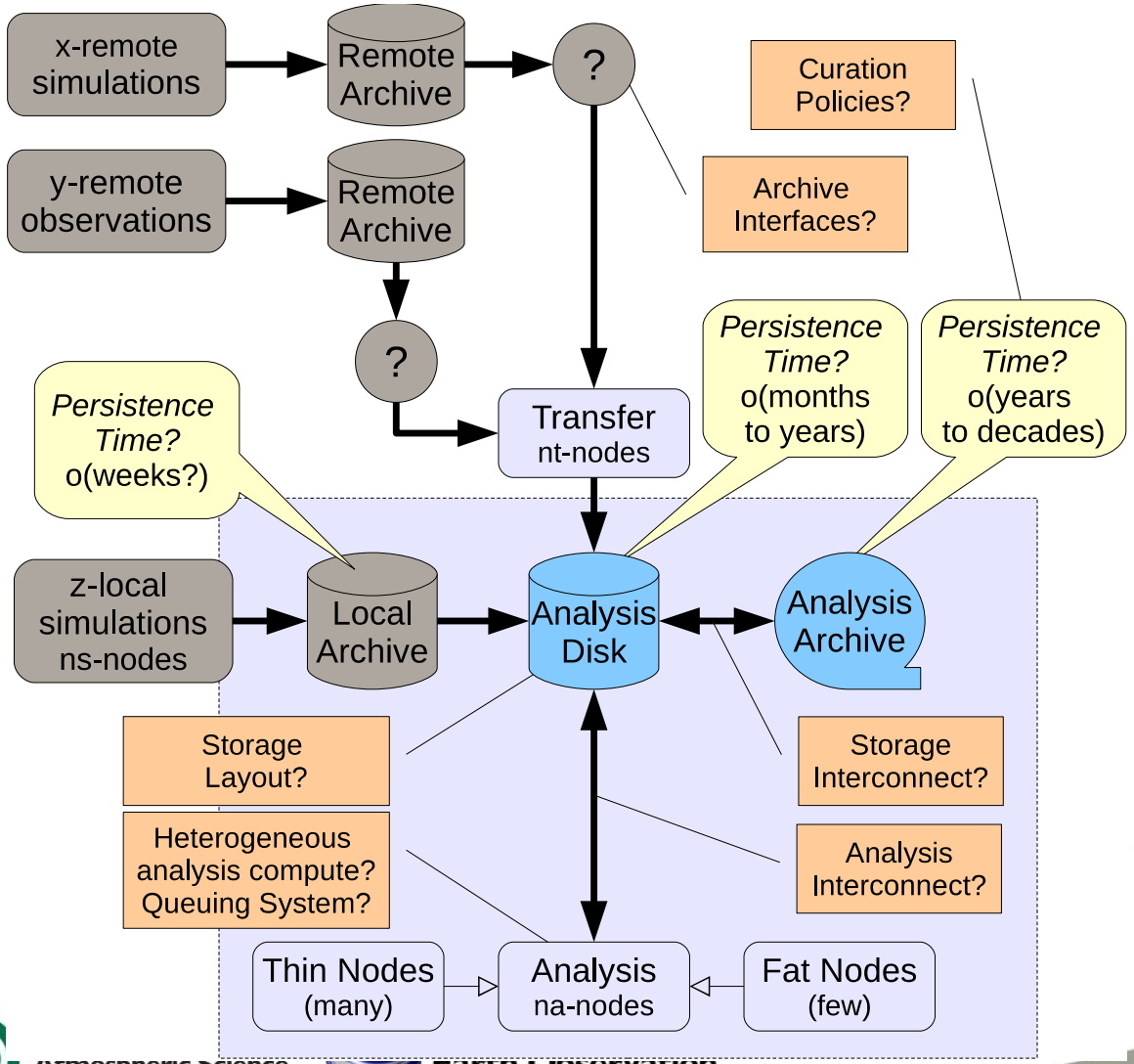
Colocation with common toolkit

Hence our JASMIN Analysis Platform



But what if you need other codes?

Common project elements (PRIMAVERA, CRESCENDO and beyond)



In both projects we have **remote simulations** in many places.

In both projects we want a **common analysis environment** and **downstream services**.

In both projects we have to deal with **differing persistence of differing data**.

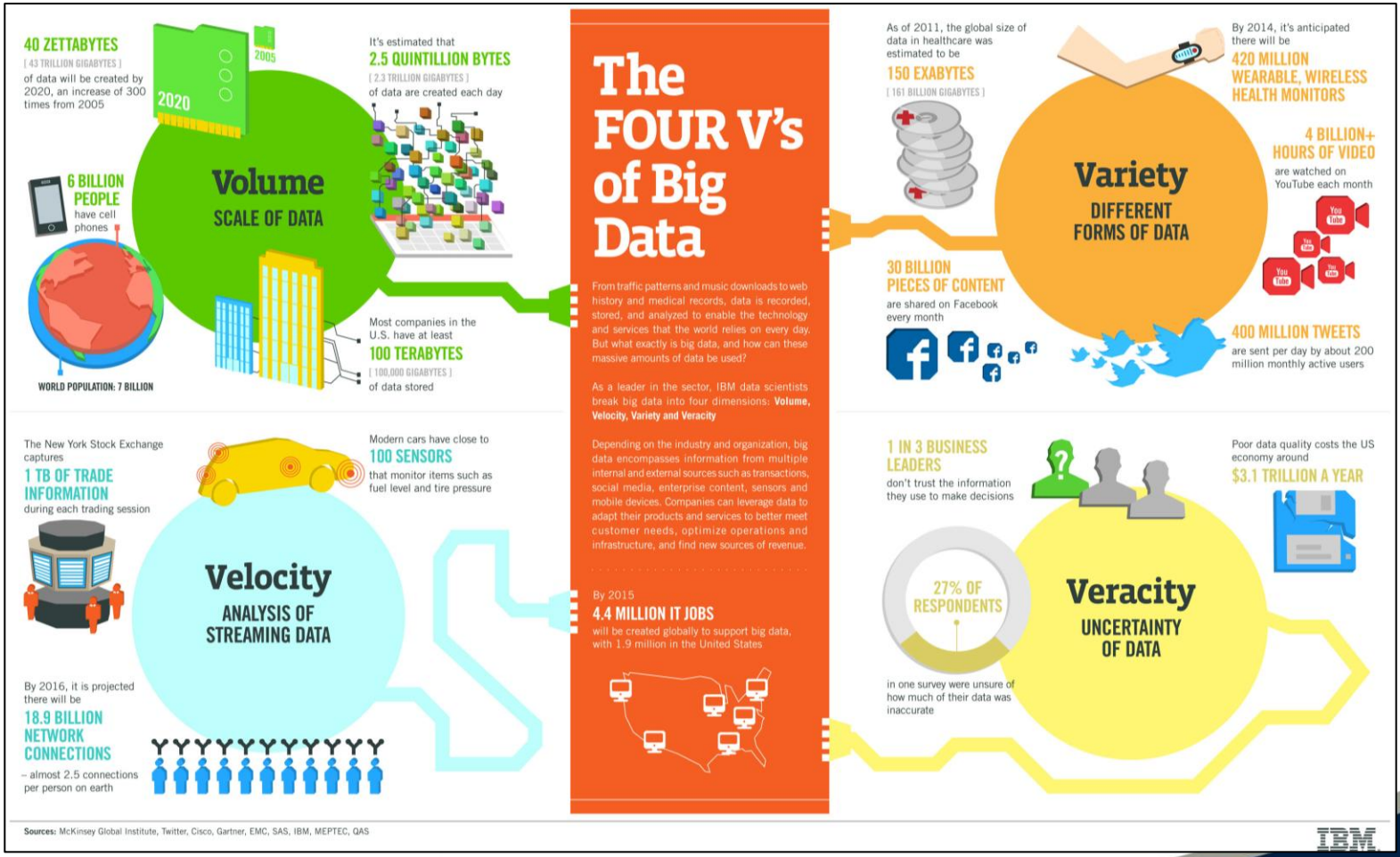
Quality checking

There are different levels of quality checking that we need to be clear about:

1. The Scientist runs his/her own quality assurance on the simulations to check the **scientific integrity** of the results.
2. The Data Manager runs his/her **compliance checks** on the data files to ensure that the metadata, file-naming and data structures meet **an agreed common specification**.
3. What is *intrinsic* data quality (that the data producer worries about) and what is *extrinsic* data quality (that the end-user is concerned with)?

The problem with extrinsic quality

Is climate data "Big Data"?



Thinking about scale (CMIP5)

Fifth Climate Model Intercomparison Project (CMIP5)

World Climate Research Programme

WCRP- WGCM

Production involved all the major climate modelling *centres*

Original Timing:

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

Actual Timing?

- Years late.

(23/05/13):

101 experiments

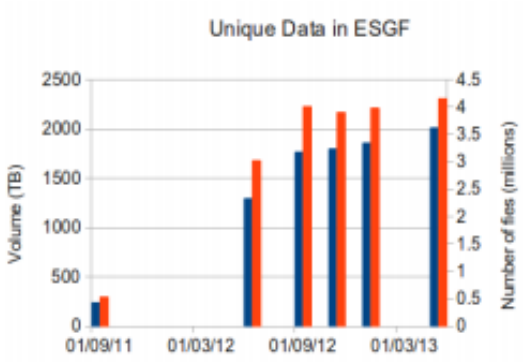
61 model variants

590,000 datasets!

4.5 million files

2 PB in global archive

Unknown PB locally!



Blue: Volume; Red: Files

(NB: replicas and versions!)

Data Delivered by:

PCMDI-led, community developed (GO-ESSP) s/w infrastructure for data delivery:

Earth System Grid Federation (ESGF)

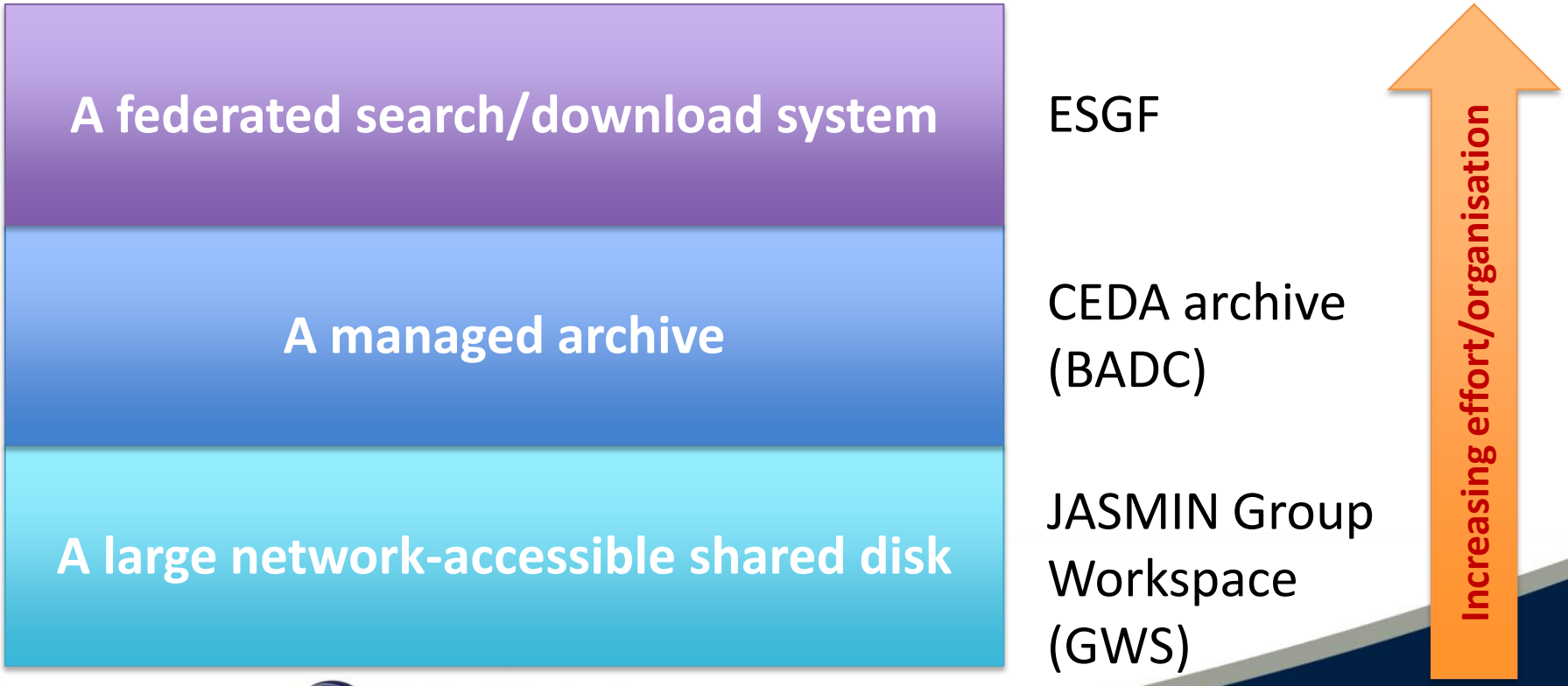
And now we have to deal with this!

	Short name of MIP	Long name of MIP
1	AerChemMIP	Aerosols and Chemistry Model Intercomparison Project
2	C4MIP	Coupled Climate Carbon Cycle Model Intercomparison Project
3	CFMIP	Cloud Feedback Model Intercomparison Project
4	DAMIP	Detection and Attribution Model Intercomparison Project
5	DCPP	Decadal Climate Prediction Project
6	FAFMIP	Flux-Anomaly-Forced Model Intercomparison Project
7	GeoMIP	Geoengineering Model Intercomparison Project
8	GMMIP	Global Monsoons Model Intercomparison Project
9	HighResMIP	High Resolution Model Intercomparison Project
10	ISMIP6	Ice Sheet Model Intercomparison Project for CMIP6
11	LS3MIP	Land Surface, Snow and Soil Moisture
12	LUMIP	Land-Use Model Intercomparison Project
13	OMIP	Ocean Model Intercomparison Project
14	PMIP	Palaeoclimate Modelling Intercomparison Project
15	RFMIP	Radiative Forcing Model Intercomparison Project
16	ScenarioMIP	Scenario Model Intercomparison Project
17	VolMIP	Volcanic Forcings Model Intercomparison Project
18	CORDEX*	Coordinated Regional Climate Downscaling Experiment
19	DynVar*	Dynamics and Variability of the Stratosphere-Troposphere System
20	SIMIP*	Sea-Ice Model Intercomparison Project
21	VIACS AB*	VIACS Advisory Board for CMIP6

Where to put data?

ESGF is not the right answer to every question

We can answer the "where can I put my simulations?" question in three layers:



It all depends what you need to do

Who do you want to share with?

- Small group of collaborators → GWS
- The world → Archive and/or ESGF

How visible do you want to data to be?

- Partially - through SSH or simple web file list → GWS
- Visible through some catalogues/discovery systems → Archive
- Visible alongside high-profile climate projects/datasets → ESGF

How long should your data persist?

- Months to a few years → GWS
- Longer → Archive and/or ESGF - which will live longer?

What can a GWS offer?

- For "projects" that want to:
  - Share a LARGE network-accessible disk.
  - Allow access from a number of institutions.
  - Pull data from external sites to a common cache.
  - Push data to external sites.
  - Process/analyse the data (on batch compute).
  - Process/analyse the data in conjunction with other *archived* or *group workspace* datasets.

A large network-accessible shared disk

JASMIN Group Workspace (GWS)

Danger of "useful" shared disk solution

The problem with providing an effective shared disk approach (GWS) is the data providers can think:

"My data is on CEDA = Job done!"

"This is much easier than spending days/weeks producing metadata and documenting my method."

But data will NOT persist here!

A managed archive (e.g. CEDA)

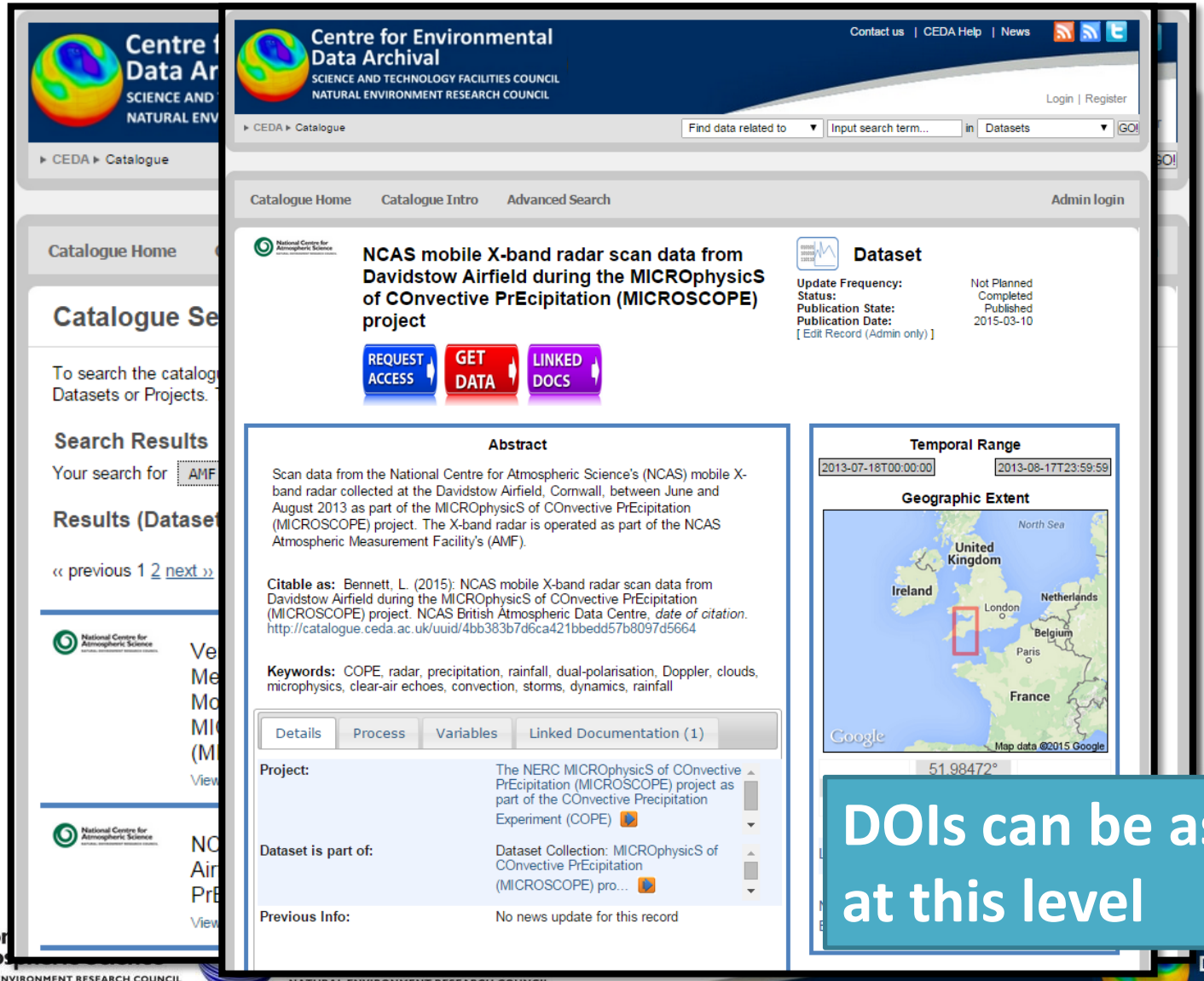
- Agree formats and metadata
- High level metadata records: Project, people, collection, model, computation, instrument, platform etc.
- Data managers involved in ingestion process - *at least now!*

A managed archive

CEDA archive (BADC)

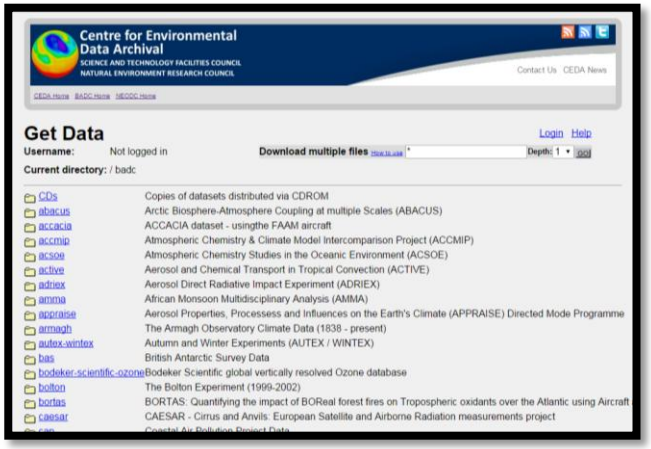
Data Catalogue (metadata)

http://catalogue.ceda.ac.uk



DOIs can be assigned at this level

Data Access Tools



Generic services:

- HTTP-based "**Browse**" service →
- **FTP** access
- **OpenDap** service now launched: **http://dap.ceda.ac.uk**

Publishing to ESGF

A federated search/download system

ESGF

The MIP "Data Workflow"

(courtesy of Guillaume Levavasseur - IPSL)

We are working closely with ESGF partners.

Automation is essential

ESGF (through a browser)

ESGF is NOT funded!

- No single project is paying the real costs of their storage, for the physical disk/tape media, or for the people involved. Not even the true cost of the people to do the massaging …
- The entire ediface depends on the careful construction of infrastructure that can be built from **contributions of many different projects** for many different audiences:
  - Hardware, Software, Metadata
  - Tools and utilities
  - No one project has control over any of those things:
    - **Not even WGCM/CMIP.**
    - **Not Primavera and Crescendo.**

Versioning

CMIP5 made great advances in the management of versioning:

- Multiple data set versions can be made documented and accessible side-by-side
- Preserves provenance
- Follows good curation principles

Within CMIP5, versioning was managed **inconsistently across the federation**, resulting in:

- Inability to determine whether data has changed (on server)
- Different nomenclature at different sites ("v1" vs "v20110424")

The impact of **Petascale**

Both PRIMAVERA and CRESCENDO will produce

large volumes of data.

Too much data, not enough disk

- Hitting Petascale has implications…
- The cost of (high-performance) disk continues to be significantly higher than tape.
- At CEDA, we have survived a long time with a disk-based solution.
- At this scale we need a tape-based solution with caching of "popular" data - *currently under development*.
- **This becomes relevant to all projects delivering big data!**
- We may ask:
  - What is the most important data (that should stay in the cache)?

Too many users, not enough compute - A word or two about compute platforms

- Our 4,000 core LOTUS cluster (on JASMIN) has been a game-changer for some scientists analysis/inter-comparing data.
- However, with capability comes new requirements …
- How can we continue to deliver compute power as demand grows? Do we need a focussed solution?

The wrong compute?

- "Bring the compute to the data" has been our mantra for the last couple of years.
- But sometimes it's “your compute” not mine. How can we enable seamless shipping of code between different physical environments.
- The next phase is multiple deployment environments
  - **Batch clusters** - already in place
  - **Scaling containers** (e.g. using Docker) in the future.
- IaaS (Infrastructure as a Service) versus Paas (Platform as a Service) in the (local/private) Cloud.

Coming back to MIPs

Model Intercomparison Projects are a response to two factors:

1. The need to evaluate our assumptions and our models. Evaluation requires comparison, and differing ways of doing simulation.
2. The need to share information between groups. Organising and delivering this at scale is hard.

All of this is bigger than any one group can do alone. In the UK, the response has been the UKESM programme, bringing many more communities into the ESM frame.

As move to more societally relevant science (the "grand challenge"), we will have more and **more communities in play, and more evaluation to do.**

Scale, Information, Collaboration and MIPs

- Both the computing and the human trends are towards scale!
- At scale, we need to formally codify information.
- Dealing with the volume, variety, and velocity of data, being shared by multiple communities, requires more than just unstructured documents to capture information and requirements. **Need metadata (targeted at specific parts of workflow), controlled vocabularies, tools and automation.**
- There is a necessary inertia in developing, populating, and exploiting information systems at scale. Agility, at scale, is hard!
- We all need to get used to "**conforming**" if we want to collaborate, and future MIPs are going to require even more information constraints, from definitions, to data outputs.

Effective data management at scale

What can/do we do to manage data at scale?

- Employ standards and conventions:
  - Our catalogue is based on ISO standards for describing environmental science data and geospatial information…
  - …but we need to adapt it to our communities' needs.
- Employ consistent methods of naming (e.g. using common vocabularies) and describing data.
- Promote decent data formats (for re-usability and tool development)…
  - I.e. NetCDF and the CF Metadata Conventions

We need more Data Managers (inside projects)

Given this situation:

- The only way forward is to employ 100s more Data Managers at Data Centres.

unlikely

- The only way forward is for data providers to become more active in the data management:
  - Organising/communicating data standards and specifications
  - Actively available disk/tape allocation
  - Establishing methods of data reduction (e.g. reducing precision or sub-setting)
  - Migrating to tape to free up space.

Conclusions

- PRIMAVERA and CRESCENDO are projects that contribute to and exploit the “CMIP era”:
  - They will generate data that will be “available in ESGF”
  - The project teams will be analysing their (big) data
  - Downstream users will include
    - Expert users of modelling data
    - Novice users of modelling data
- There is a long way between the HPC platform where the data is generated and the data appearing in ESGF.
  - Not all data should travel that route.
  - Most data will need massaging to make it all the way.

Conclusions

Everything is scaling up: number of experiments, models, projects, files, size of files, requirements for post-processing etc.

- Data Management facilities are also scaling (which buys us time)…
- …but the Data Managers are not multiplying.

So we need to evolve:

- Automate everything!
- Engage the scientists to:
  - Prepare data/metadata according to agreed standards
  - Become part of the data management process