Science & Technology Facilities Council
Rutherford Appleton Laboratory

NCAS, CEDA and JASMIN: AN OVERVIEW

Thanks to all contributors:

Alison Pamment, Sam Pepler, Ag Stephens, Stephen Pascoe, Anabelle Guillory, Esther Conway, Alan Iwi, Matt Pritchard, Sarah Callaghan, David Hooper, Charlotte Pascoe

On behalf of the course team
(STFC/NERC:CEDA, NERC:NCAS CMS, NERC:NCAS Leeds)

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

National Centre for Atmospheric Science

NATURAL ENVIRONMENT RESEARCH COUNCIL

“National capability”

Computational Modelling Service
CMS

Facility for Airborne Atmospheric Measurements
FAAM

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Met Office

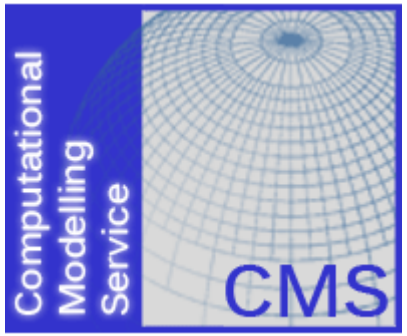
Facilities and Services

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

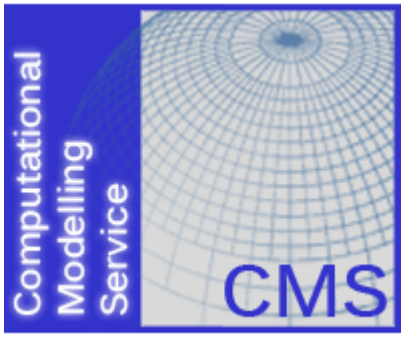
NCAS
atmospheric measurement facility
amf
excellence in measurement

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

NCAS Computational Modelling Services (CMS)



NCAS Computational Modelling Services (CMS)

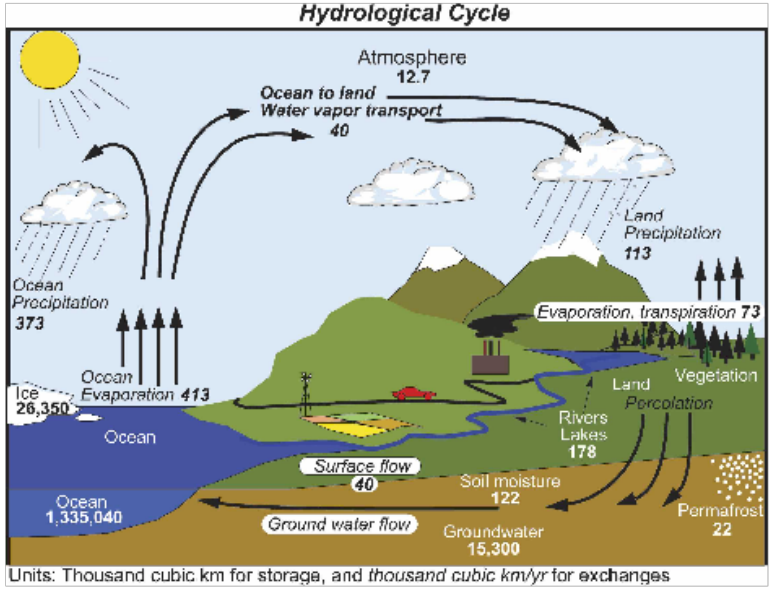


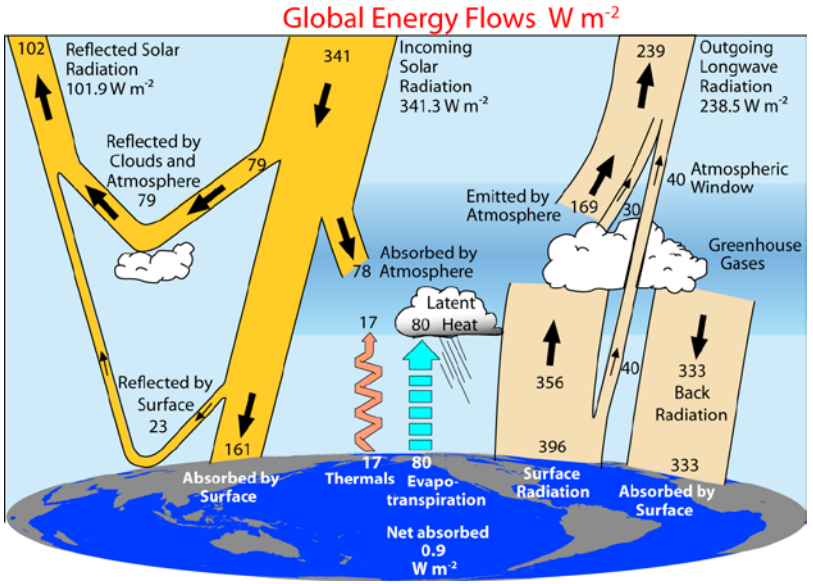
- NCAS activities in support of computational science (particularly High Performance Computing (HPC) and numerical modelling)

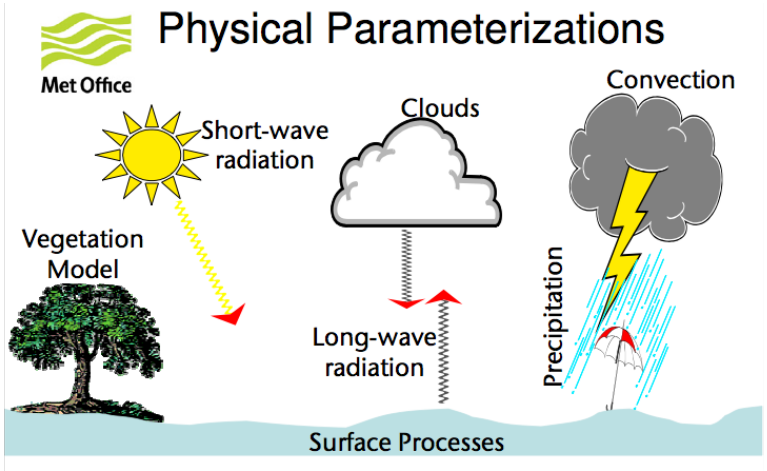
- Provides underpinning infrastructure for the UK academic atmospheric and polar science community to support climate, weather, and earth-system research.

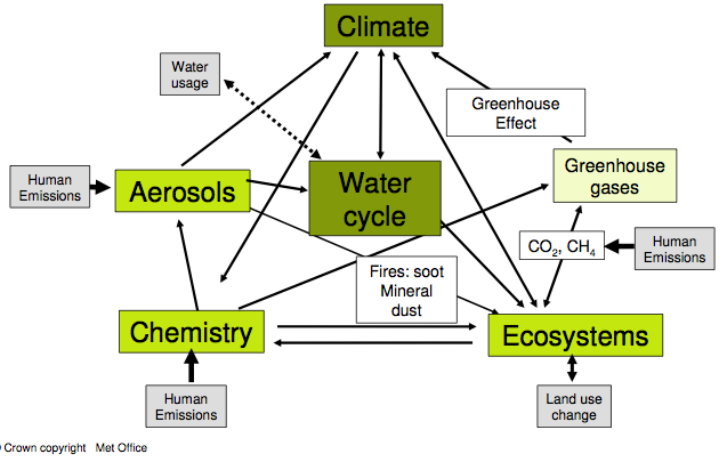
Climate Modelling

(Trenberth et al, 2007,2009)





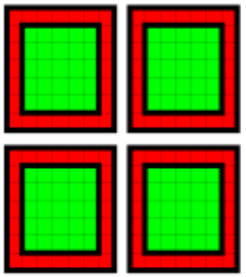


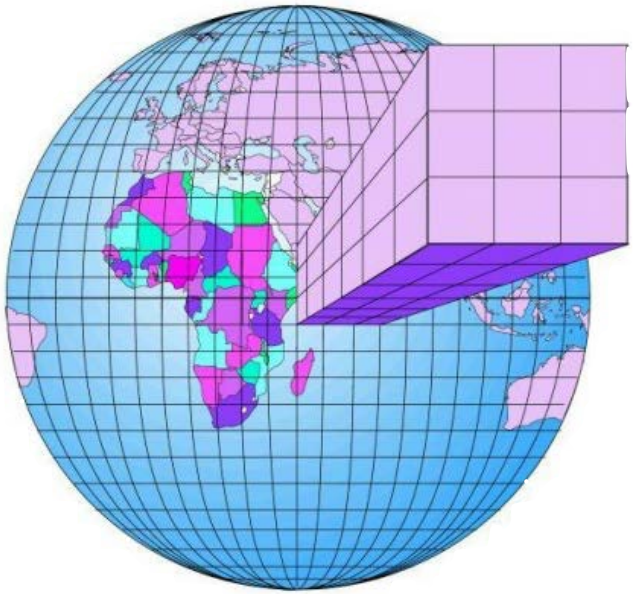


PARALLEL Implementation

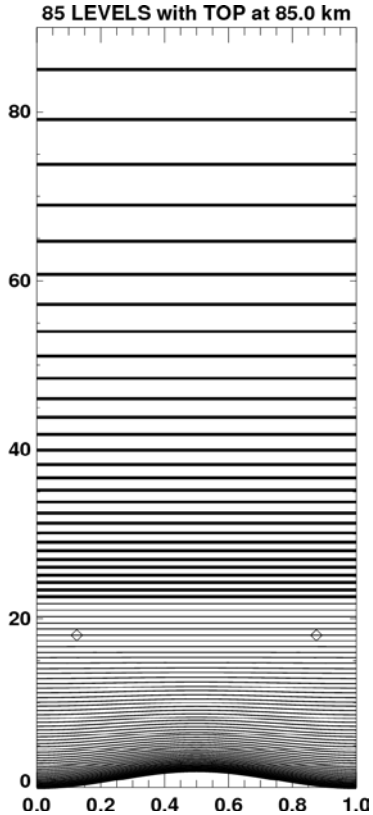
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- Regular, Static, Lat-Long Decomposition
- Mixed mode MPI/OpenMP
- Asynchronous I/O servers
- Communications on demand for advection
- Multiple halo sizes





Horizontal resolution



Land surface

Vertical resolution

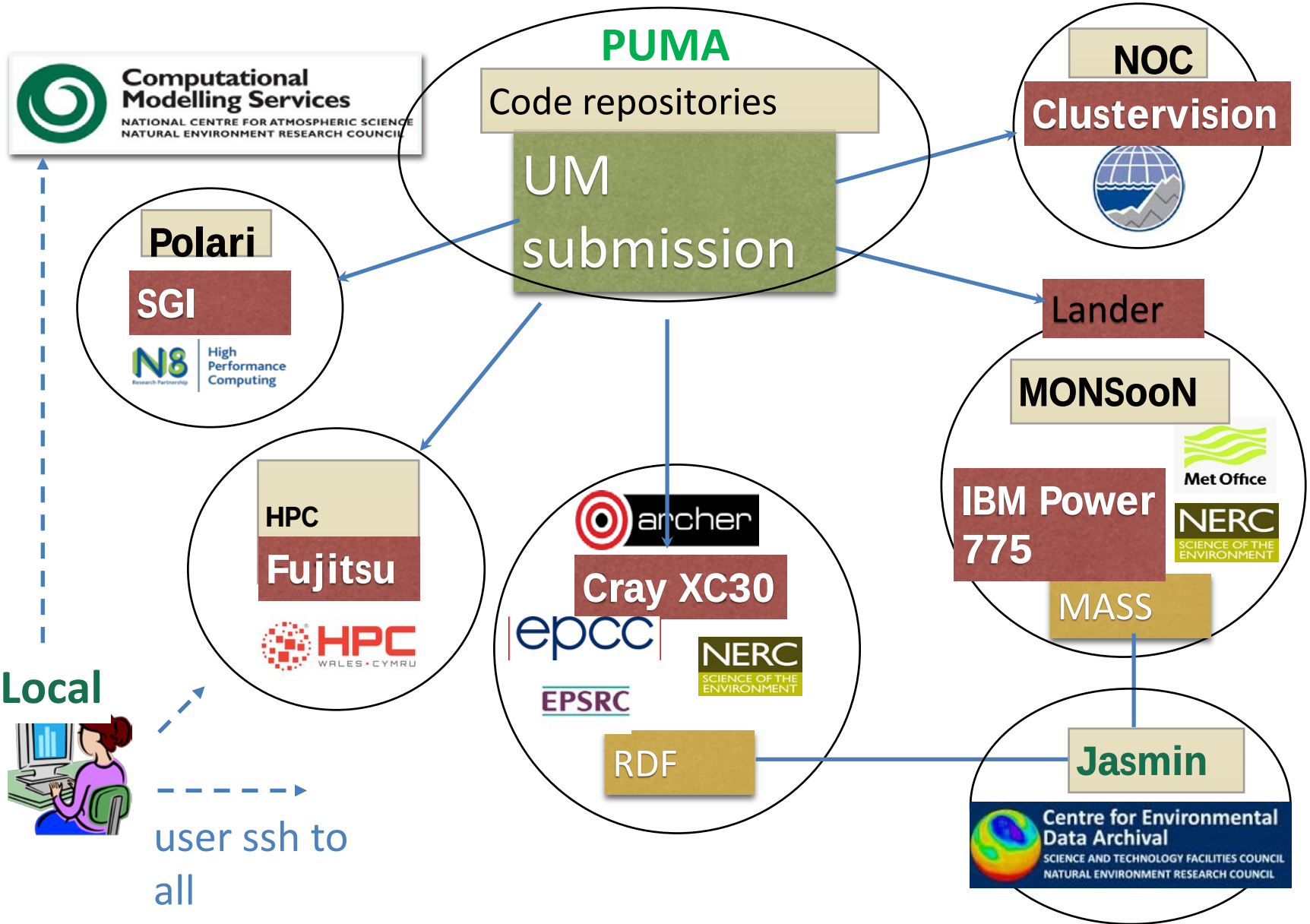
Global Models

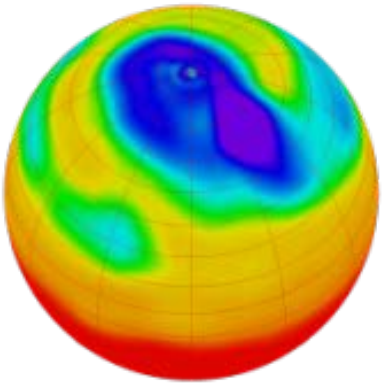
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

N96	N144	N216	N320	N512	N768	N1024	N2048
(192 x 145)	(288 x 217)	(432 x 325)	(640 x 481)	(1024 x 769)	(1536 x 1152)	(2048 x 1536)	(4096 x 3073)
~135 km	~90 km	~60 km	~40 km	~25 km	~17 km	~12 km	~6 km

	NWP	**Climate**
Run length	***10*** *day operational forecast,* *15 day ensemble forecast*	*Months (seasonal)* *Years, decades, centuries+*
Global resolution	*Testing:* *N320 (40 km) with 15 min ts* *Operational:* *N768 (17 km) with 7.5 min ts*	*Low resolution:* *N96 (135 km) with 20 min ts* *High resolution:* *N512 (25 km) with 15 min ts*
Dynamics	*Non-bit reproducible*	*Bit-reproducible*

NCAS supported MACHINES





Centre for Environmental Data Analysis

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

(CEDA)

NERC Data Centres

The UK's Natural Environment Research Council (NERC) funds seven data centres which between them have responsibility for the long-term management of NERC's environmental data holdings.

National Geological Data Centre

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

British Oceanographic Data Centre

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

The NERC Data Catalogue Service (DCS) allows data held by all NERC data centres to be located by users.

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

Science & Technology Facilities Council
Rutherford Appleton Laboratory

CEDA Data Centres

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

Contact us | CEDA Help | News

Login | Register

▸ CEDA ▸ Data Centres

Data search Search this site Go

About CEDA | Data Centres | Services | Projects | For Academics | For Business | Training | Contact Us | Help

Data Centres

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

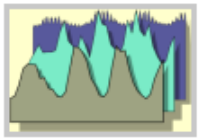
CEDA-Archive

The CEDA Archive operated the atmospheric and earth observation data centre functions on behalf of NERC for the UK atmospheric science and earth observation communities. It covers climate, composition, observations and NWP data as well as various earth observation datasets, including airborne
these functions we
Atmospheric Data
Centre (NEODC).

April 2016: Approx. 4 petabytes total

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
and



CEDA-Space

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

www.ceda.ac.uk

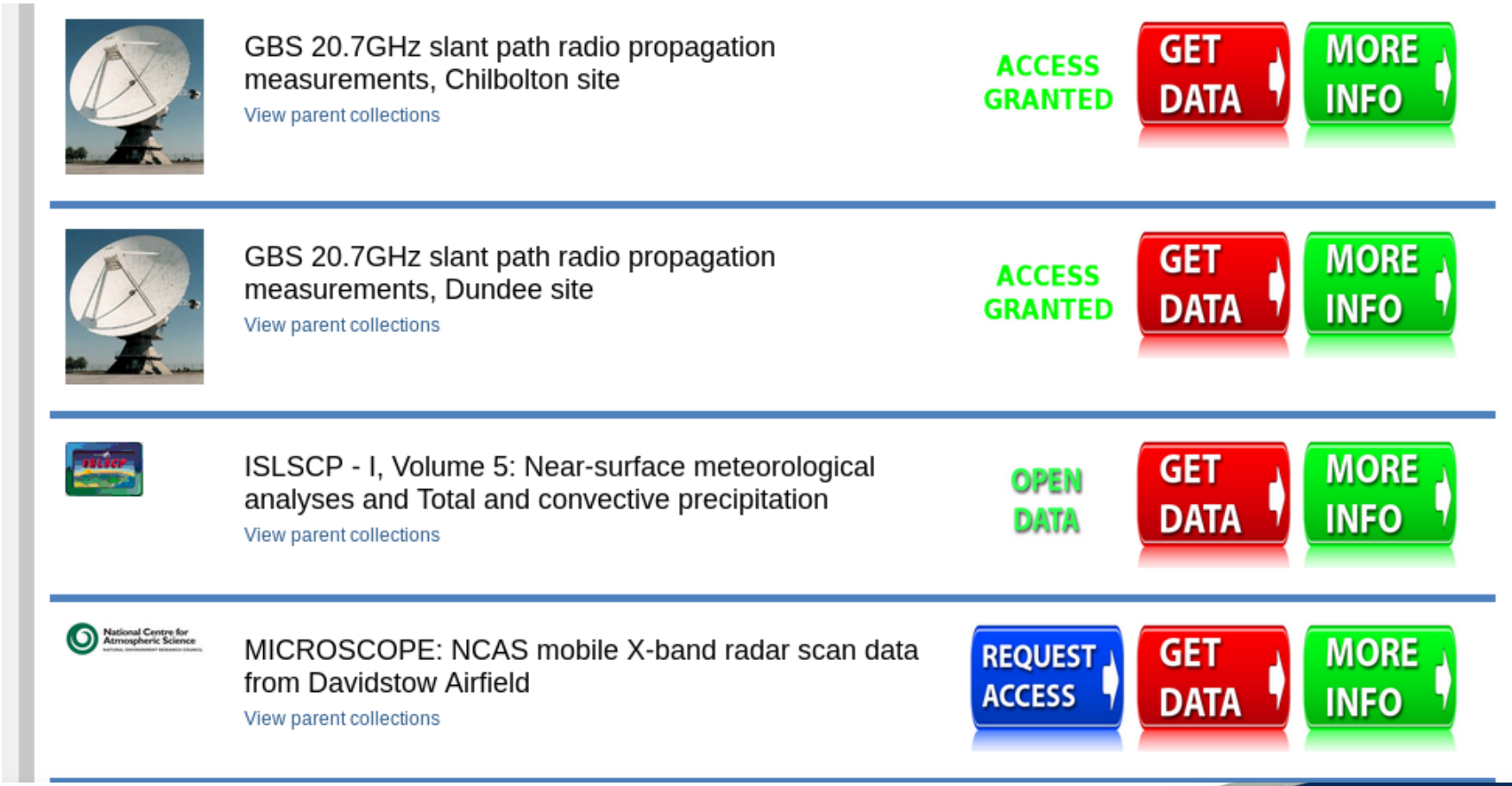
NATURAL ENVIRONMENT RESEARCH COUNCIL

NATURAL ENVIRONMENT RESEARCH COUNCIL

nmental
Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

CEDA "MOLES" catalogue

Search CEDA data holdings for atmospheric and EO data at catalogue.ceda.ac.uk



Separate catalogues for CEDA-Space and IPCC DDC

CEDA Projects

Projects

Characterisation of metadata to enable high-quality climate applications and services - CHARMe

CHARMe is a 2 year FP7 funded project aiming to link commentary metadata (e.g. annotations, supporting information about the data) and datasets. The project will deliver repositories of commentary metadata with interfaces for users to populate and interrogate the information. This will enable users to assess if the of climate data are fit for purpose.

CEDA is working with 8 other UK and European partners, and has key roles on the data model, software development, implementation in archives, and application to climate services.

InfraStructure for the European Network for Earth System Modelling - Phase 2 (IS-ENES II)

IS-ENES II is a FP7-Project, funded by the European Commission under

Climate Information Portal for Copernicus (CLIPC)

The CLIPC platform will complement exitsting GMES/Copernicus pre-operational components by providing access on decadal to centennial climate variability data to a wide variety of users. The data will include satellite and in-situ observations, climate models and re-analyses, transformed data products to enable impacts assessments and climate change impact indicators. Supporting data quality and related information will also be made available.

CEDA is leading the project, coordinating a consortium of 22 partners, and leads the access to climate data work package. This work package will provide the software infrastructure to a create a single point of access for climate model data from various sources: climate model data, in situ and satellite observations, and re-analyses.

ESPAS - Near-Earth Space Data Infrastructure for e-Science

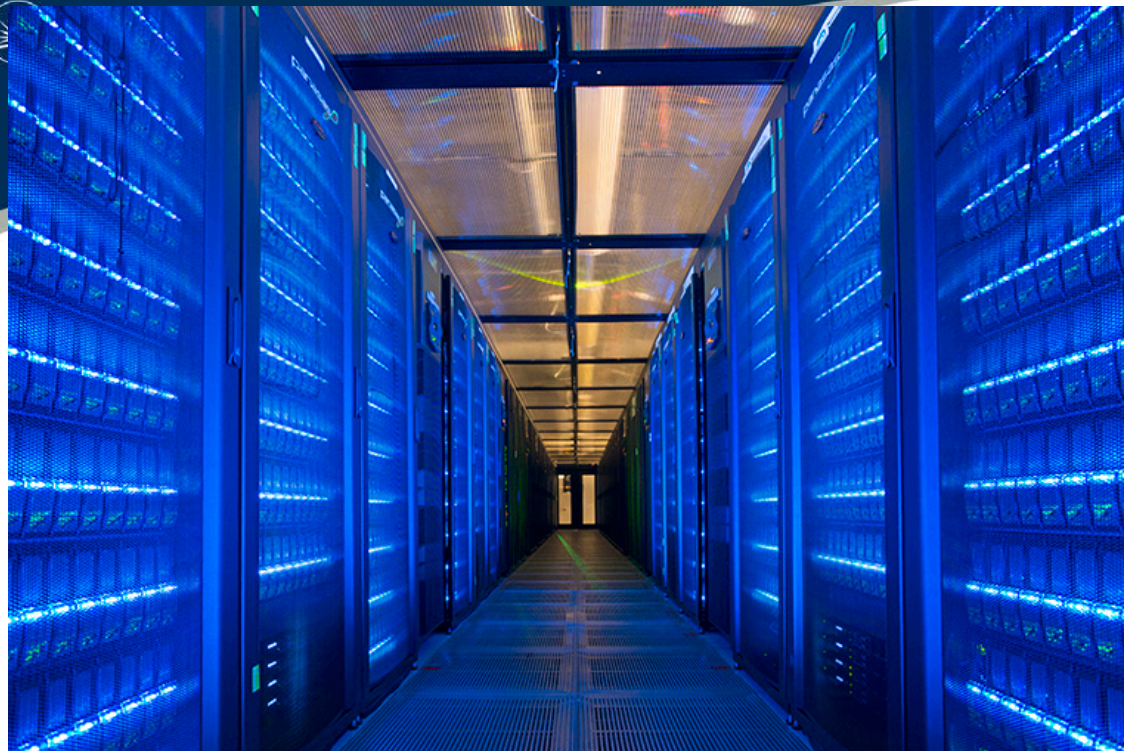
The ESPAS project aims to provide e-infrastructure necessary to support the access to observations, modelling and prediction of the Near-Earth

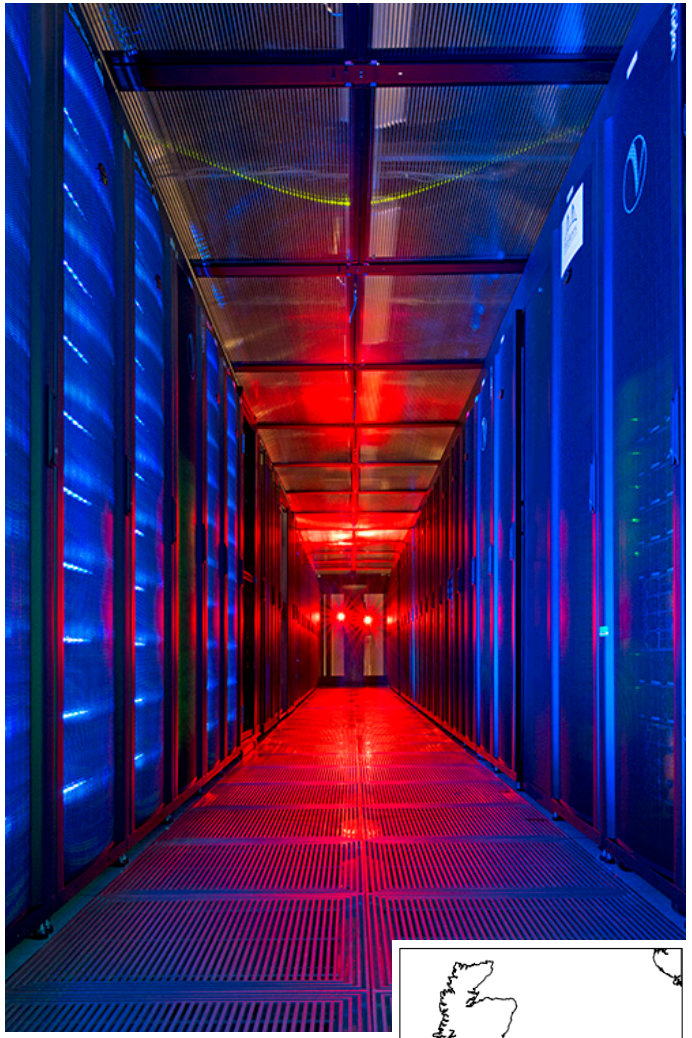
Lots more

JASMIN / CEMS Overview

The JASMIN facility is a "super-data-cluster" which delivers infrastructure for data analysis.

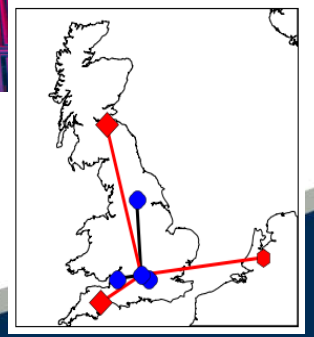
In technical terms it is half super-computer and half data-centre and it provides a globally unique computational environment.





JASMIN is a world leading, unique hybrid of:

- 16PB high performance storage (~250GByte/s)
- High-performance computing (~4,000 cores)
- Non-blocking Networking (> 3Tbit/sec), and Optical Private Network WAN's
- Coupled with cloud hosting capabilities



To address "one of NERC's most strategically important challenges: the improvement of predictive environmental science"

Prof. Duncan Wingham, NERC Chief Exec.

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

Science & Technology Facilities Council
Rutherford Appleton Laboratory

Hosted by STFC
Scientific Computing Department

"*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

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

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

Close working partnership with industry

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

Processing big data: the issues

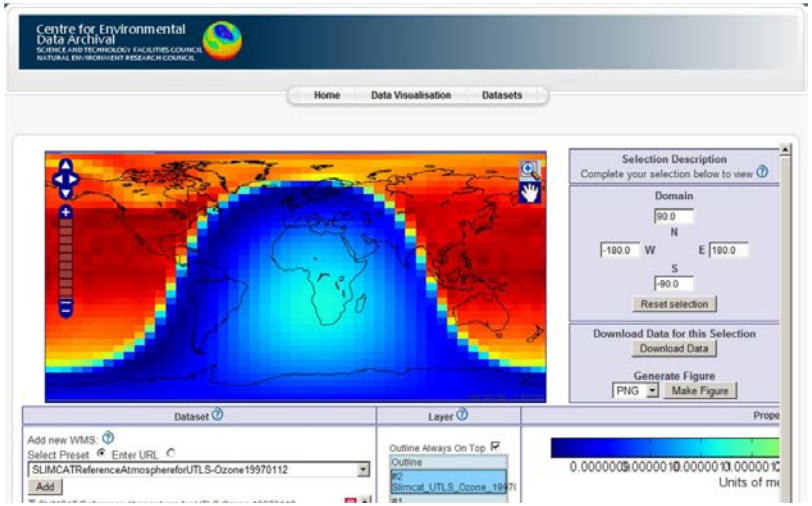
- Parallel processing in the Environmental Sciences has historically focussed on **highly-parallel models**

- Data analysis was typically run **sequentially** because:
  - It was a smaller problem
  - It didn't have parallel resources available
  - The software/scientists were not equipped to work in parallel

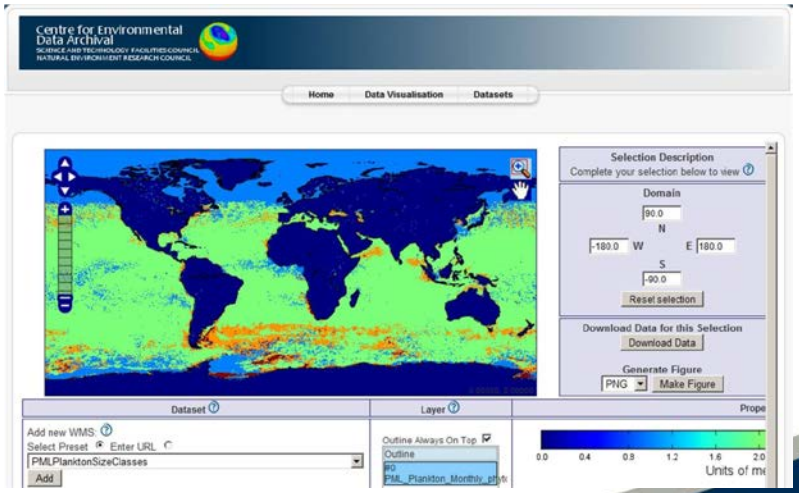
- Now we generate enormous datasets (e.g. UPSCALE - 300 Tb):
  - Processing big data **requires** a parallel approach
  - Platforms, tools, and programmers are becoming better equipped

JASMIN Use cases

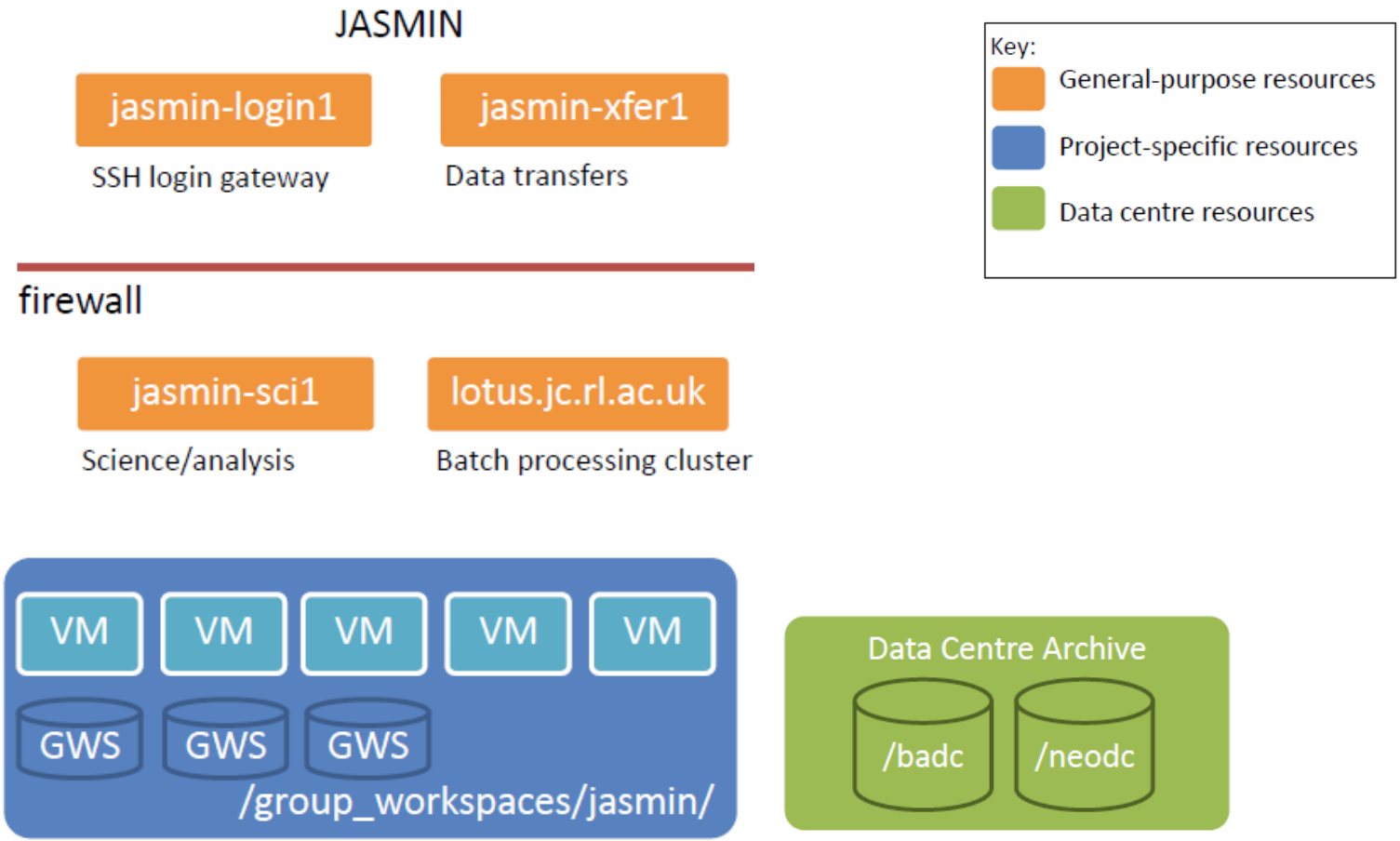
- Processing large volume EO datasets to produce:
  - Essential Climate Variables
  - Long term global climate-quality datasets



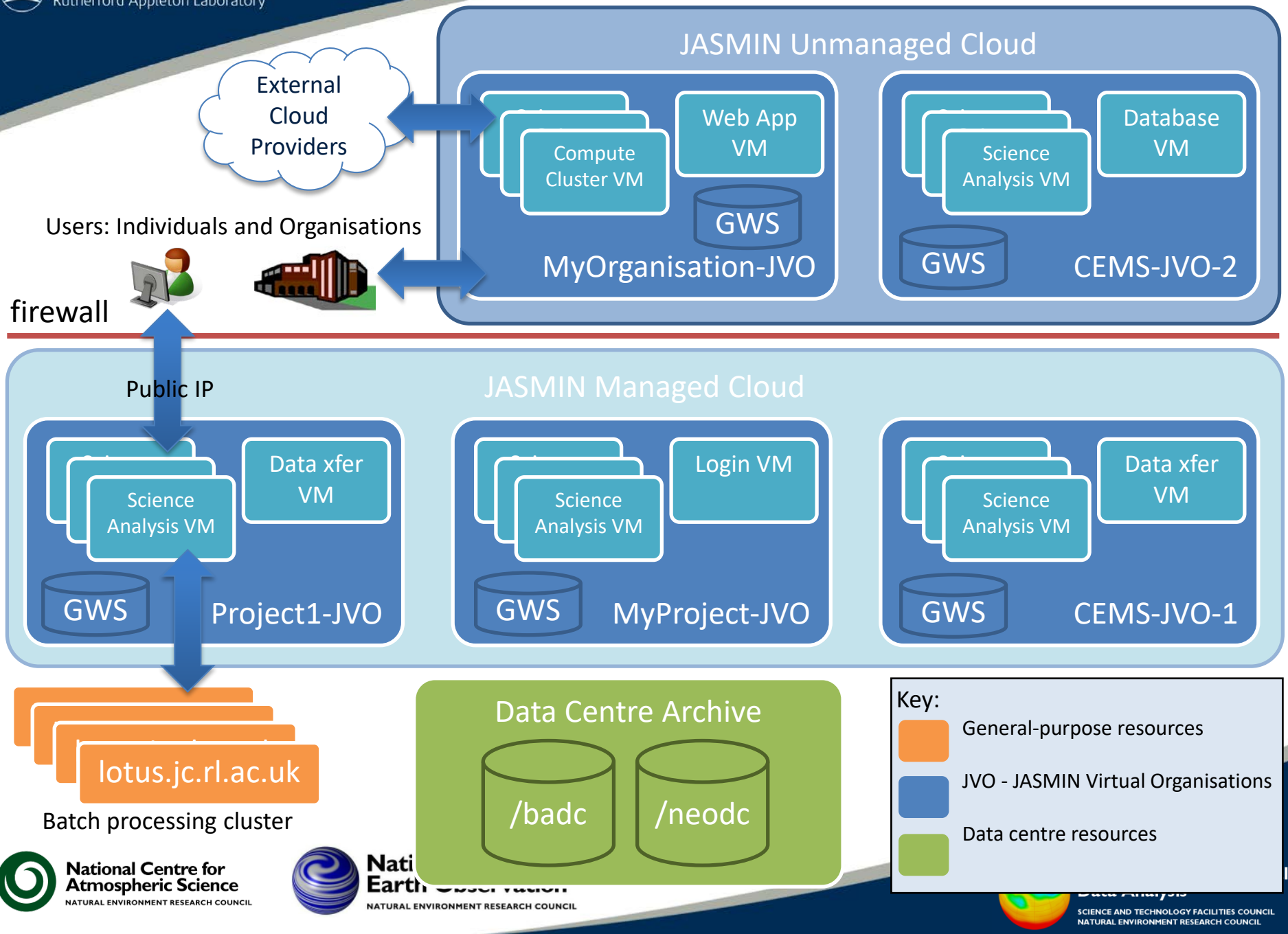
- Data validation & intercomparisons
  - Evaluation of models relying on the required datasets (EO datasets, in situ and simulations) being in the same place



JASMIN in pictures



JASMIN Cloud Infrastructure

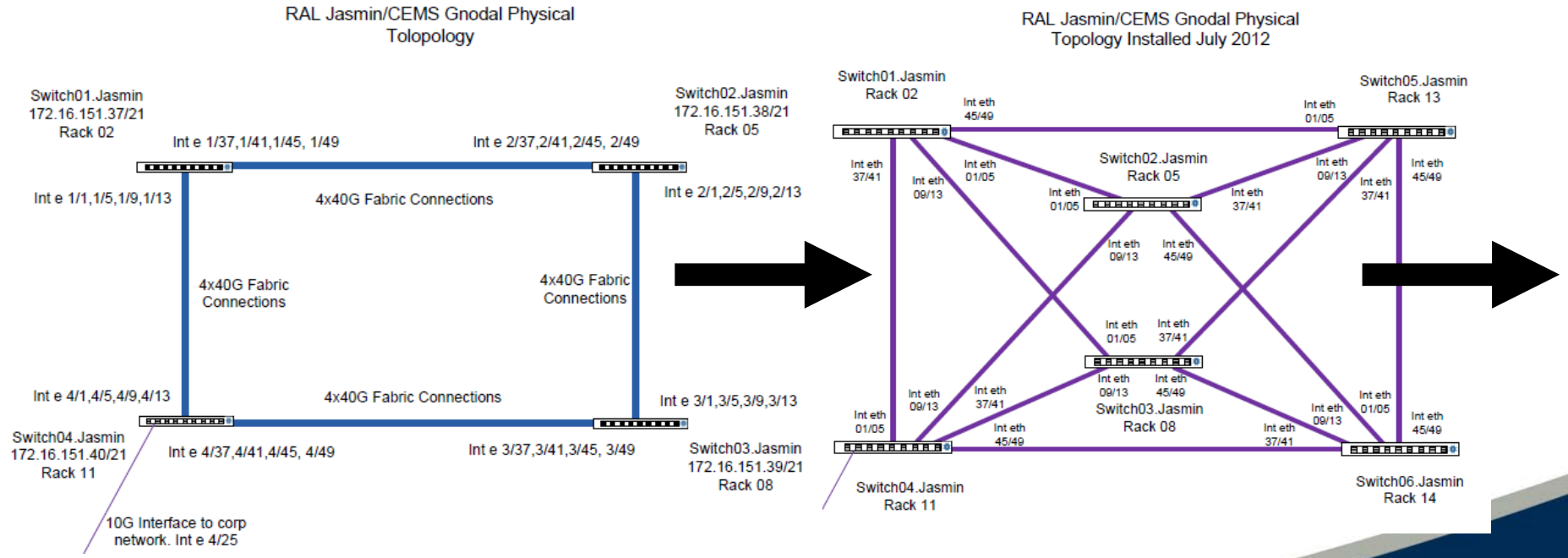


Internal network: vital to JASMIN / CEMS performance

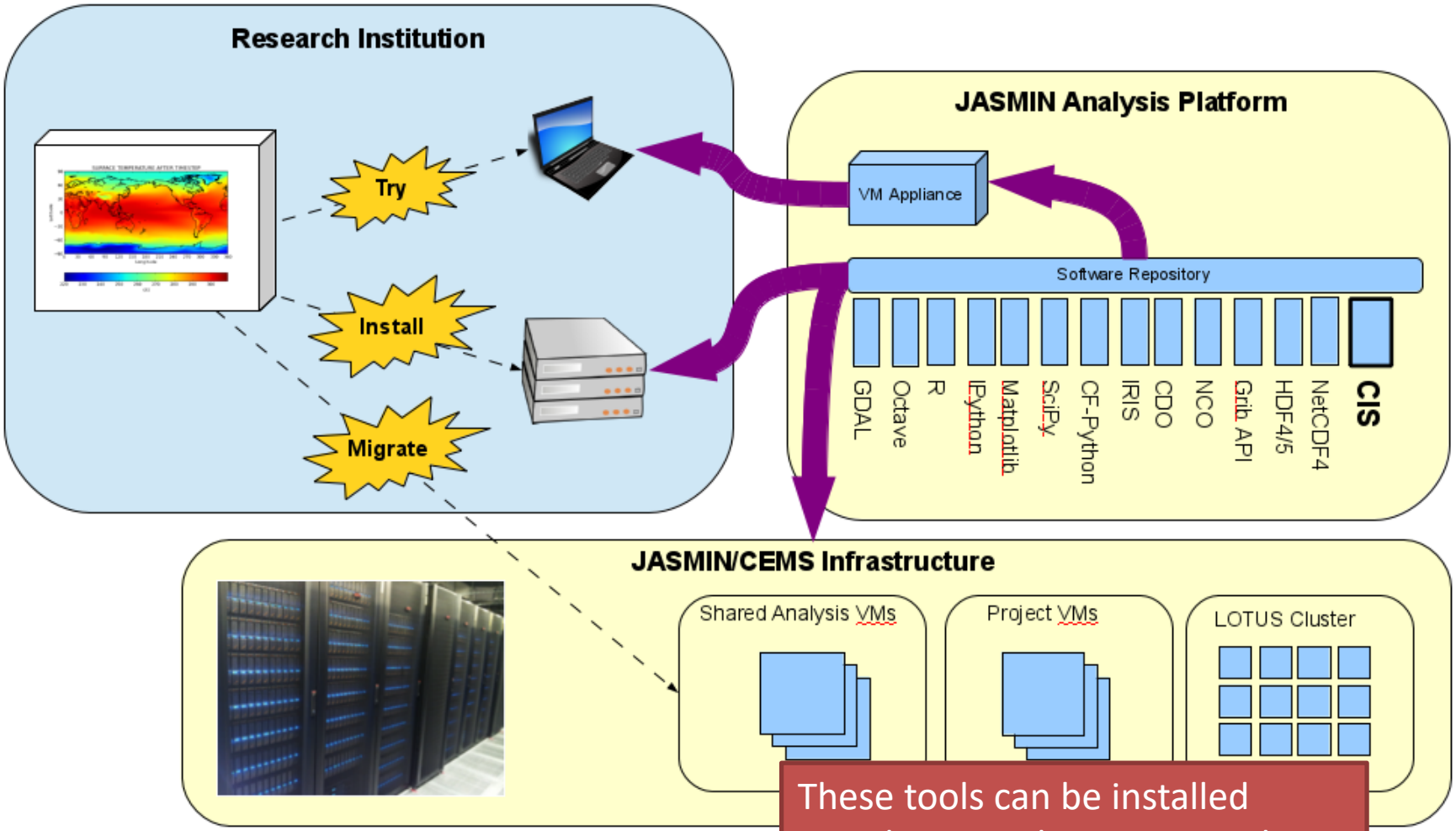
172.16.144.0/21 = 2,000 IPs

130.246.136.0/21

Flat Overlaid L2
160->240 Ports @ 10Gb



The "JASMIN Analysis Platform" – a re-usable, re-deployable bundle of common tools



These tools can be installed anywhere inside JASMIN with one or two commands!

Further Information

- NCAS website: ncas.ac.uk
- CMS website: cms.ncas.ac.uk
- CEDA website: ceda.ac.uk
- JASMIN website: jasmin.ac.uk