Introduction to JASMIN

What is JASMIN?

- Half super-computer and half data-centre

- Unique computing environment

- Central node at RAL is referred to as the "super-data-cluster"

- Initially installed early 2012

- Used primarily by the atmospheric science and Earth observation communities (so far)

- JASMIN phases 2 and 3 - 2014

- Will provide compute resources for other environmental sciences

JASMIN Funding

JASMIN locations

JASMIN-North
University of Leeds
150 Tb

JASMIN-Core (phase 1)
STFC RAL
5 Pb + compute

JASMIN-West
University of Bristol
150 Tb

JASMIN-South
University of Reading
500 Tb + compute

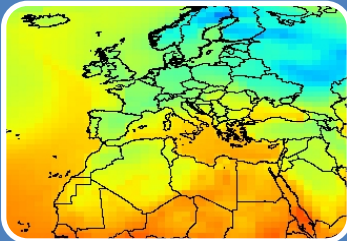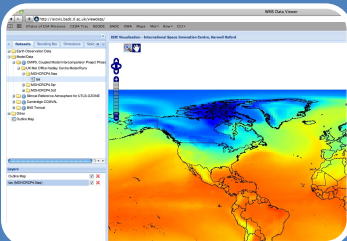# JASMIN links

JASMIN functions
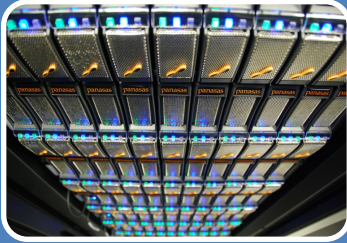
CEDA data storage & services

- Curated data archive
- Archive management services
- Archive access services (HTTP, FTP, Helpdesk, …)

Data intensive scientific computing

- Global / regional datasets & models
- High spatial, temporal resolution
- Private cloud

Flexible access to high-volume & complex data for climate & earth observation communities

- Online workspaces
- Services for sharing & collaboration

What can you do on JASMIN?

WHAT CAN YOU DO ON JASMIN?

1. SHARE "LARGE" DATA WITH COLLABORATORS

Many projects are making JASMIN a focal point for collaboration by setting up a "Group Workspace". This is a shared disk made accessible to a chosen group of scientists working on the same project.

Group Workspaces currently **range between 5Tb and 380Tb** in size. Typically they are used for:
- Sharing large data – pushing from remote sites
- Accessing data – retrieving to remote sites
- Analysing data – space for storing big outputs
- Intercomparison – comparing data with established archived datasets, such as CMIP5 and ERA-Interim
- Overflow from HPC – staging outputs from costlier storage environments

2. ANALYSE DATA

Once logged into JASMIN, users can make the most of the **co-location with the NCAS BADC and CEMS Academic (NEODC) archives** as well as large Group Workspaces supporting specific projects. All data is stored on the high-performance storage platform.

Additionally, **generic scientific analysis virtual machines** are available for users to login and run code next to the data. The JASMIN Analysis Platform (Pascoe, 2013) is configured with a suite of **common software tools for data manipulation** (such as NetCDF libraries, NCO, CDO and the Python-Numpy stack).

3. EFFICIENT DATA TRANSFERS

JASMIN is built upon **4.6Pb (petabytes)** of PANASAS disk connected into the low latency network with 115 x 10Gbit/s connections. Workflows and data transfers can make use of the parallel networking to increase performance. Dedicated transfer servers are available for connection via SCP, RSYNC and GridFTP protocols. A 1Gbit/s "Lightpath" link is operational to the UK Met Office/MONSooN and a 2Gbit/s link to the HECTOR HPC site is currently being installed.

4. RUN YOUR OWN VMs (Root included!)

Specific projects may require dedicated computing resources and/or specific software environments. JASMIN **provides root access to project-specific virtual machines** so that external administrators can build and maintain specialised environments with minimum overhead and support. Projects, such as those discussed below (UPSCALE, NAME and PRECIS), typically marry this provision with a Group Workspace to store their inputs and outputs.

5. PARALLEL PROCESSING

The **"LOTUS" computing cluster** is available to users with processing needs that are computationally expensive. LOTUS is managed through batch queues which allow splitting of large jobs to run on many nodes or submission of model code that utilises parallel I/O functionality directly (such as OpenMPI). Centrally installed software is visible on all nodes and both archives and Group Workspaces can be accessed where permitted.

JASMIN Use cases

- Processing large volume EO datasets to produce:

  - Essential Climate Variables

  - Long term global climate-quality datasets

- EO data validation & intercomparisons

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