JASMIN (STFC/Stephen Kill)

CEDA, JASMIN and Managing User Communities

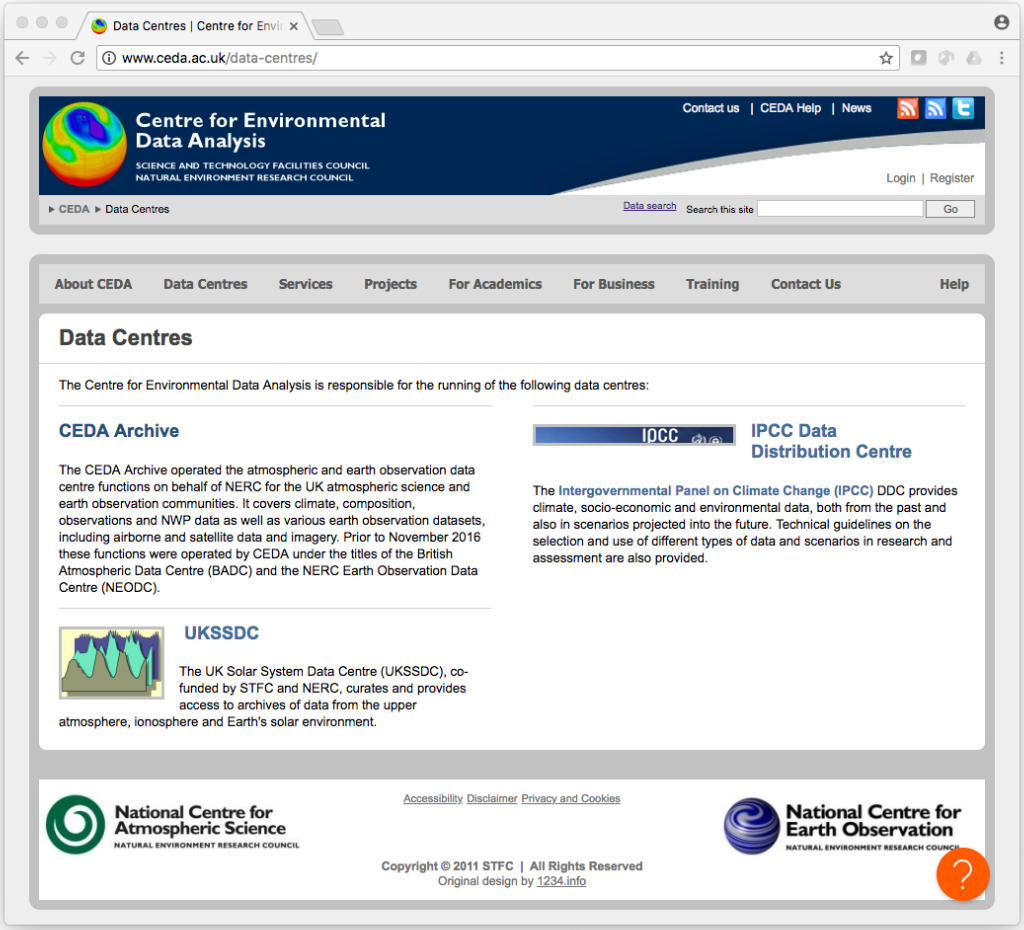
UK Tier 0 Collaboration Meeting

14-16 March 2018

Victoria Bennett

NCAS/NCEO, Centre for Environmental Data Analysis, RAL Space

CEDA Context



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

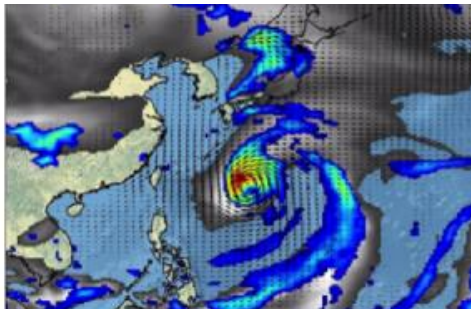
➔ Curation & Facilitation

JASMIN Context

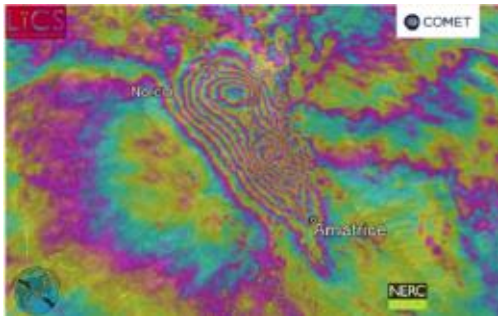
- JASMIN provides a Data Commons for environmental sciences community
- Hosted & operated by STFC for NERC
- *An STFC cross-department collaboration*
  - STFC Scientific Computing Department
    - Hardware architecture, cloud software architecture & operation
  - CEDA (RAL Space)
    - User services & applications development, cloud software architecture
- Both partners involved from the outset in architectural design and implementation

Science Goals

- Support the science & impact objectives of major stakeholders
  - NERC : NCAS & NCEO
- Help community deal with demands of 21st century environmental science
  - Growth in numerical simulation …& simulated data
  - Increasing complexity of algorithms needed to exploit data
  - Managing & preserving increasing volumes of NERC community data
  - Address digital skills gap
  - Curating digital information products of national significance

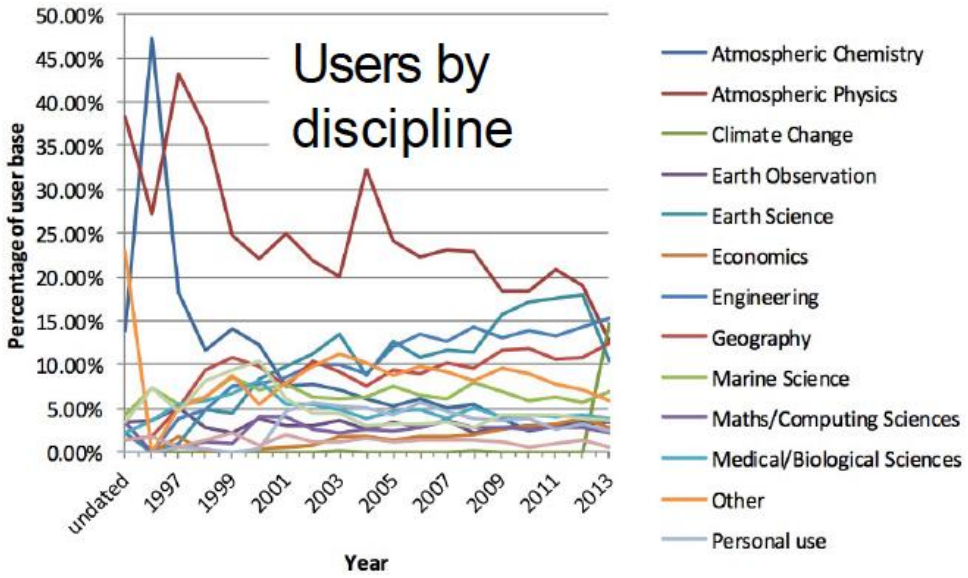


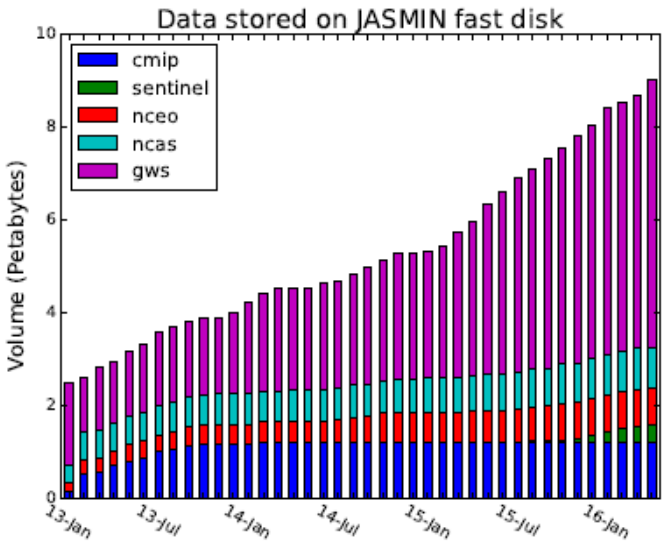
High Res Climate Model



Seismic Fault analysis

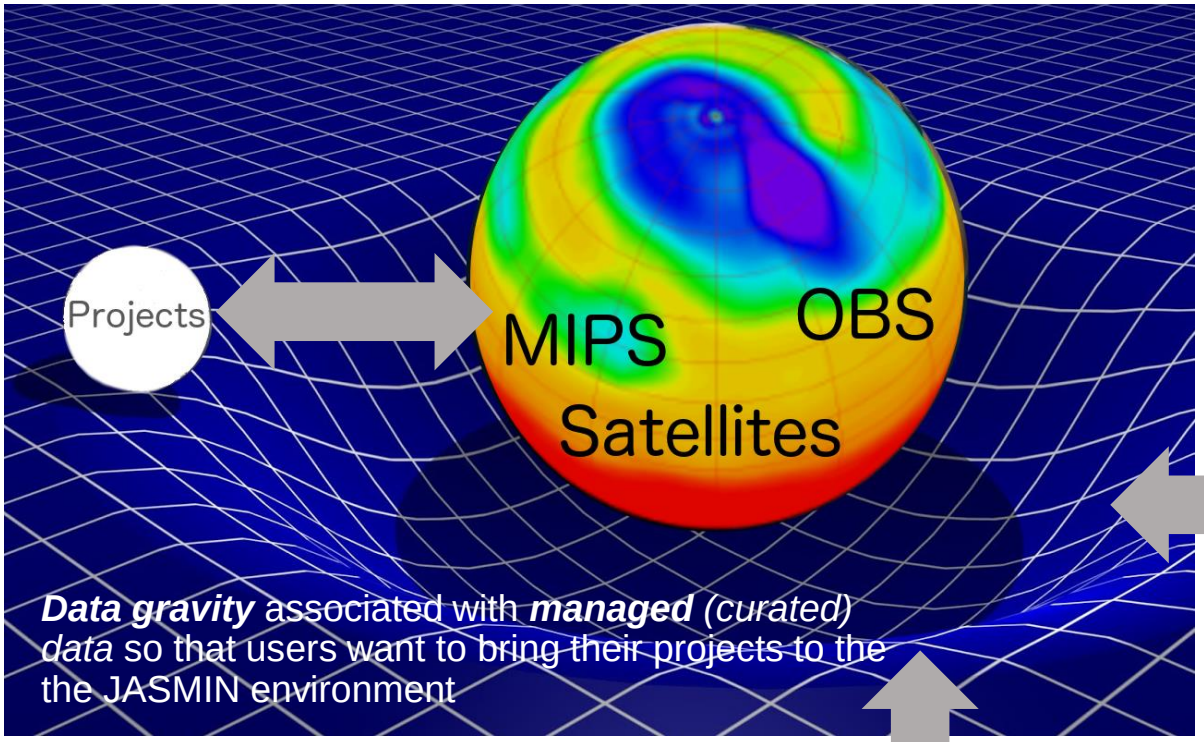
Data Growth and Diversification of user community





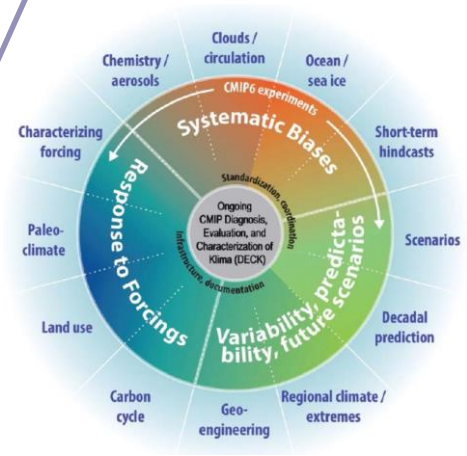
2013-2016 increasing data storage on JASMIN, in Group Workspaces (GWS) and archive

JASMIN as a Data Commons



CEDA CREPP system to ingest from MetOffice Hadley Centre

CMIP6



Sentinel Earth Observation Data

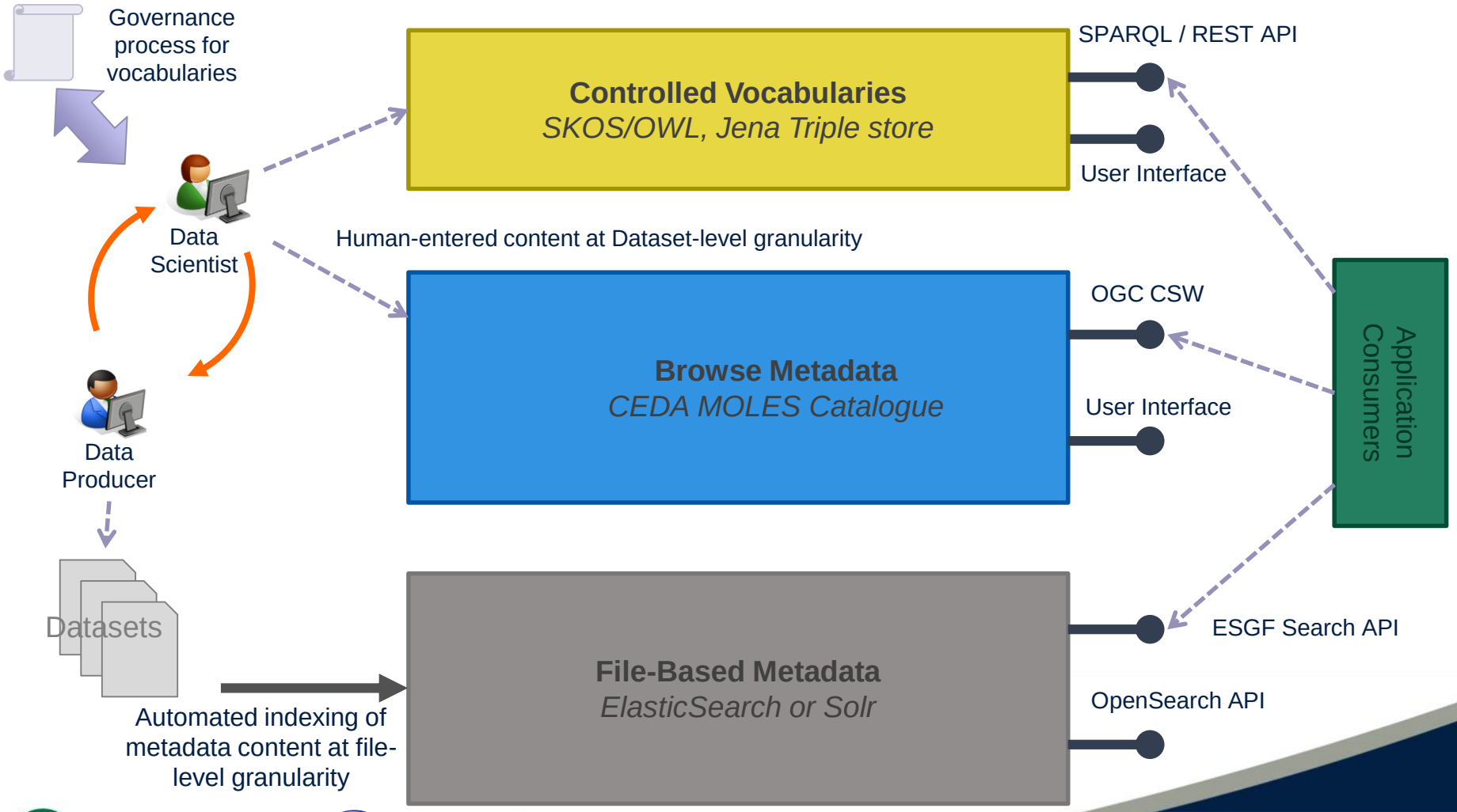
Sentinel missions data rate: ~6PB/year

Climate Models

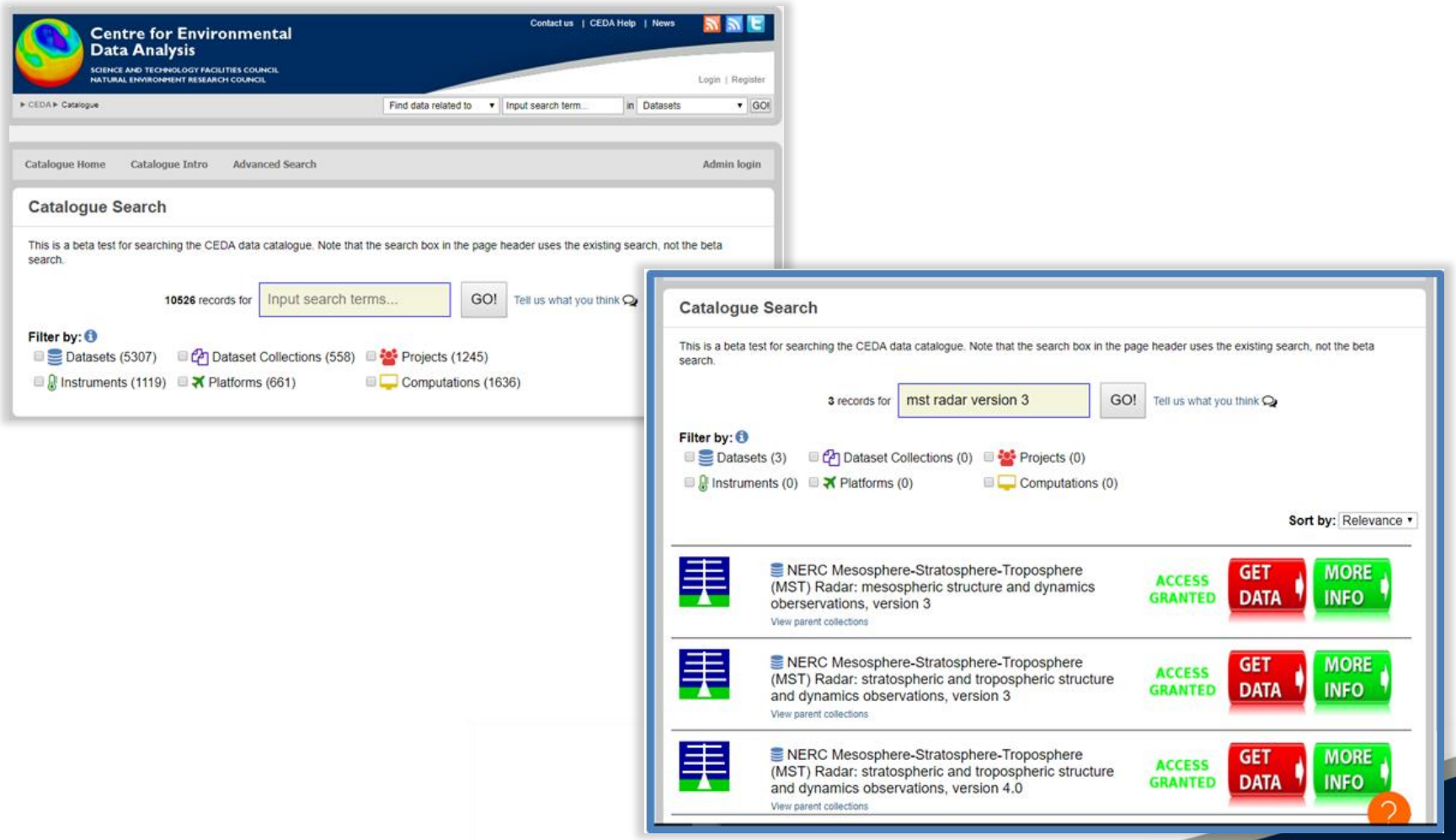
Couple-Model Intercomparison Projects

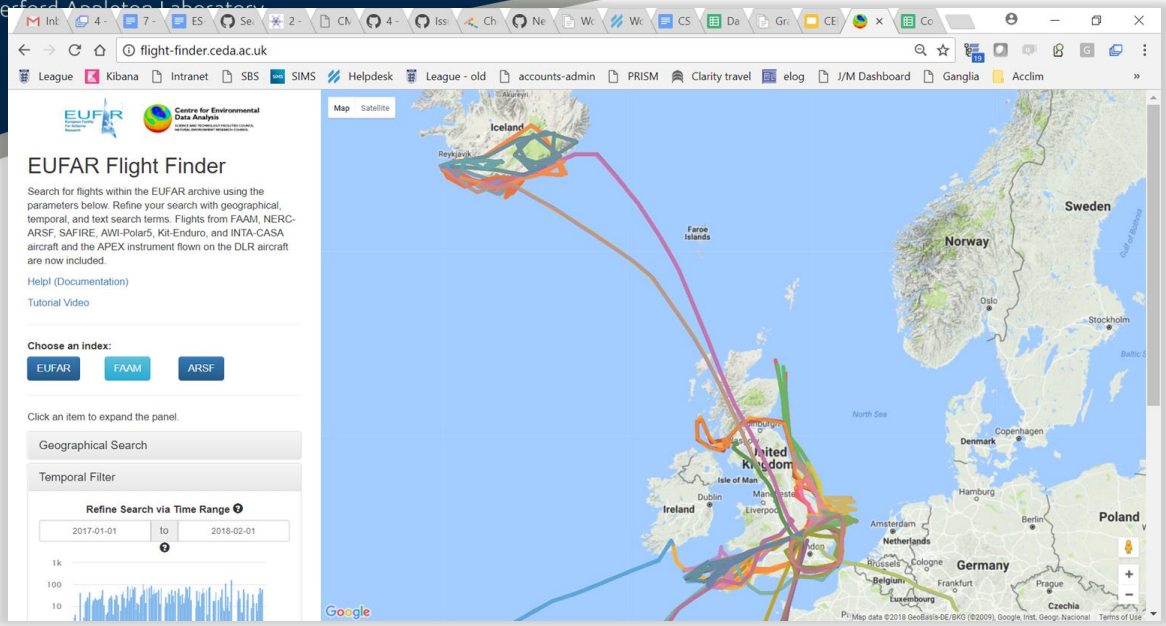
Copernicus
Europe's eyes on Earth

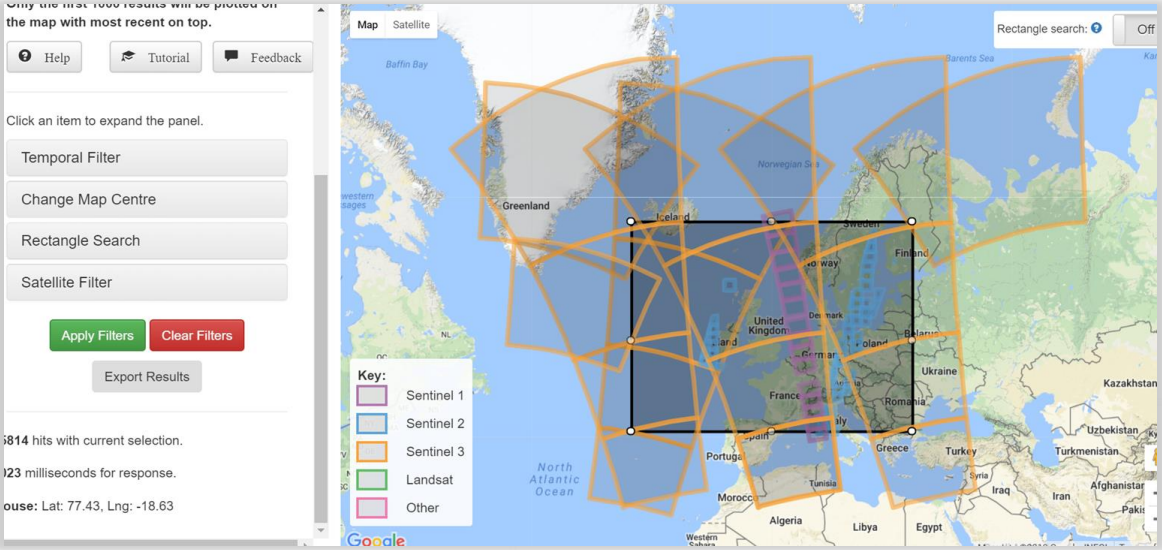
Managed Data: Sourcing information for Data discovery



MOLES – Catalogue Search




flight-finder.ceda.ac.uk
EUFAR Flight Finder
Search for flights within the EUFAR archive using the parameters below. Refine your search with geographical, temporal, and text search terms. Flights from FAAM, NERC-ARSF, SAFIRE, AWI-Polar5, Kit-Enduro, and INTA-CASA aircraft and the APEX instrument flown on the DLR aircraft are now included.
Help! (Documentation)
Tutorial Video
Choose an index:
EUFAR
FAAM
ARSF
Click an item to expand the panel.
Geographical Search
Temporal Filter
Refine Search via Time Range
2017-01-01
to
2018-02-01
Map
Satellite
Iceland
Reykjavik
Faroe Islands
Norway
Sweden
Oslo
Stockholm
Denmark
Copenhagen
United Kingdom
Isle of Man
Dublin
Ireland
Netherlands
Amsterdam
Belgium
Brussels
Cologne
Germany
Frankfurt
Hamburg
Berlin
Poland
Prague
Czechia
Luxembourg
Google


the map with most recent on top.
Help
Tutorial
Feedback
Click an item to expand the panel.
Temporal Filter
Change Map Centre
Rectangle Search
Satellite Filter
Apply Filters
Clear Filters
Export Results
5814 hits with current selection.
123 milliseconds for response.
ouse: Lat: 77.43, Lng: -18.63
Map
Satellite
Rectangle search:
Off
Baffin Bay
Greenland
Iceland
Norwegian Sea
Barents Sea
Finland
United Kingdom
Ukraine
France
Romania
Kazakhstan
Uzbekistan
Turkmenistan
Greece
Turkey
Syria
Iraq
Iran
Afghanistan
Portugal
Morocco
Tunisia
Algeria
Libya
Egypt
North Atlantic Ocean
Key:
Sentinel 1
Sentinel 2
Sentinel 3
Landsat
Other
Google

EUFAR Flight Finder &
CEDA Satellite Data Finder
Wendy Garland, Richard Smith and Ag Stephens
NCAS/NCEO, Centre for Environmental Data Analysis, RAL Space, STFC, UK
Introduction
How the EFF and Satellite Finder works
http://flight-finder.ceda.ac.uk
http://geo-search.ceda.ac.uk
Challenges
Conclusions and future plans
www.eufar.net
http://flight-finder.ceda.ac.uk
http://geo-search.ceda.ac.uk
Contact: Wendy Garland
wendy.garland@ncas.ac.uk

Logical View

CEDA Data Centres

IPCC Data Distribution Centre

ESGF

Other …

NERC Managed Analysis Compute

NERC Cloud Analysis Compute

Other compute …

CEDA Archive Services

Data Centres, Curation, DB systems

User management, External Helpdesk

Analysis Environment

Compute:

Science VMs + User Management, PaaS and IaaS Cloud,
Group Workspaces: Fast Disk & Elastic Tape

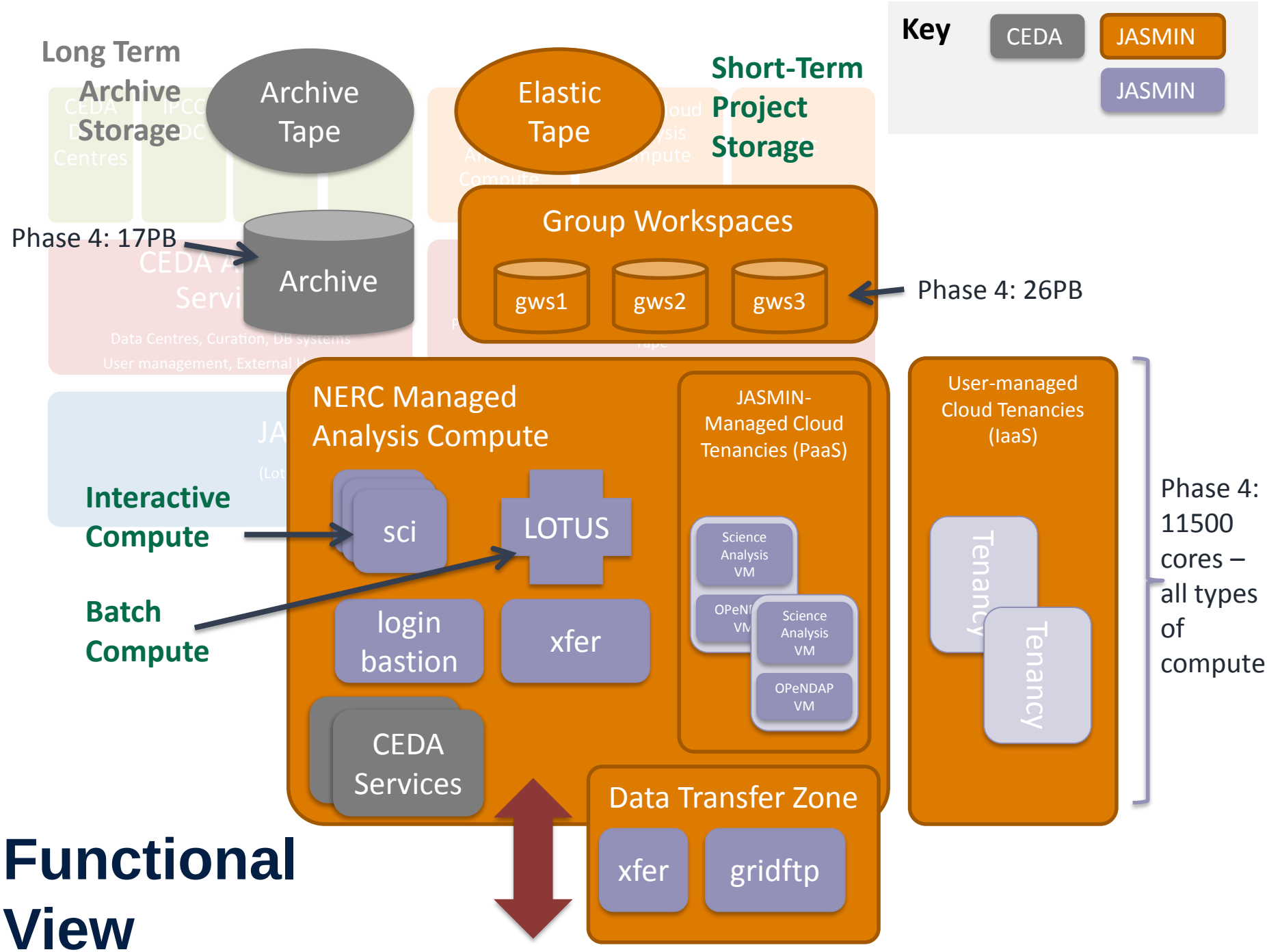
External Helpdesk

JASMIN Compute and Storage

Managed Compute (Lotus and GWSs, Tape Store + Data Transfer Zone),

Community Cloud + Internal Helpdesk

NERC SCIENCE OF THE ENVIRONMENT



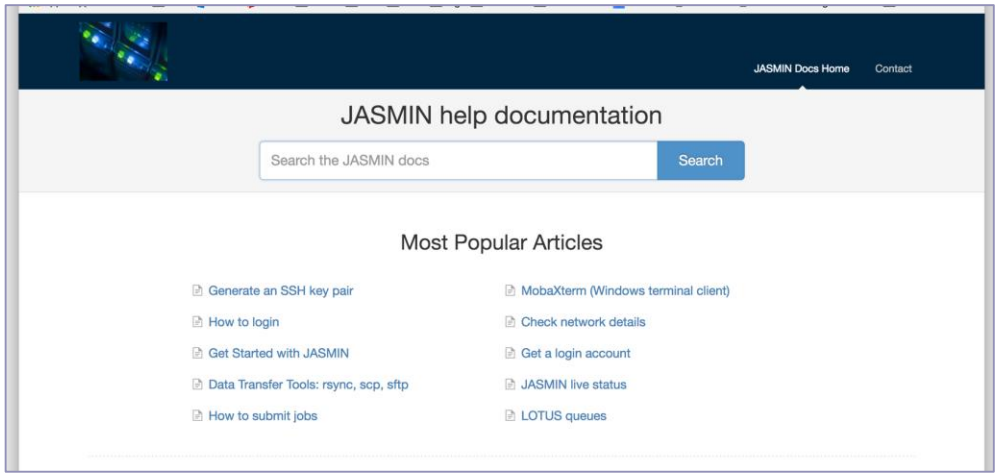
Functional View

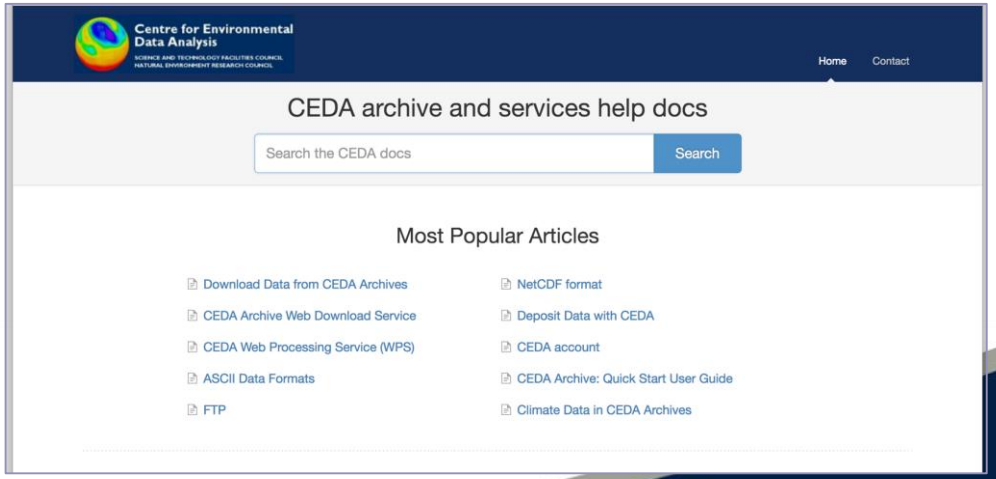
Main service options for JASMIN users

- Direct POSIX-based access (majority of JASMIN use)
  - Group workspaces
  - Lotus batch compute
  - Direct access to the archive
- Cloud
  - Platform-as-a-Service:
    - tenants deploy pre-defined VM templates
    - VMs have direct POSIX access to group workspace and data archive
    - VMs are patched by the JASMIN team
  - Infrastructure-as-a-Service:
    - tenants deploy pre-defined VM templates
    - But have full root access to VMs
    - Can allocate their own storage
    - VMs managed by the tenancy owner
    - HTTP and FTP access only to GWS and data archive
- *Object storage and dual S3/POSIX interface storage being deployed this year for JASMIN Phase 4*

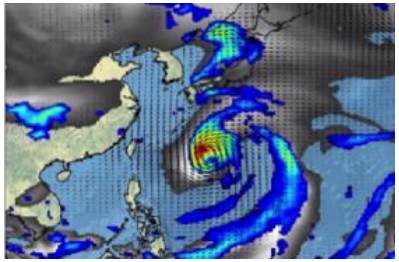
User support

- Helpdesk service for CEDA and JASMIN users
  - Approx. 50 queries/week
  - Initial response within 1 working day
- Documentation – continuously evolving
- Data provider support via team of data scientists
  - Formats
  - Metadata
  - Data upload
  - Data policy

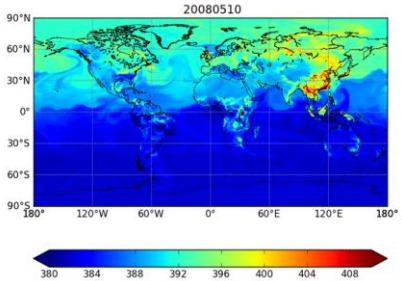




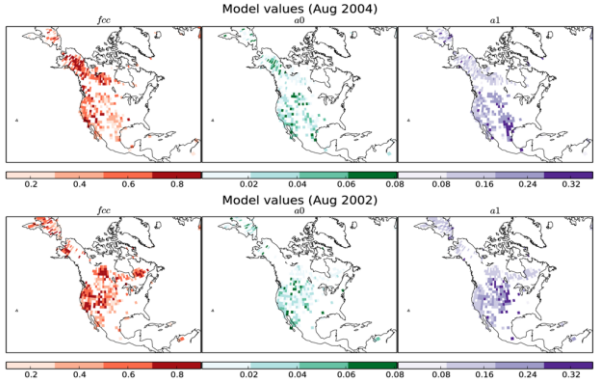
~150 Science projects on JASMIN to date



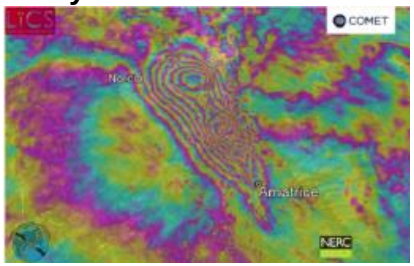
High Res Climate Model analysis



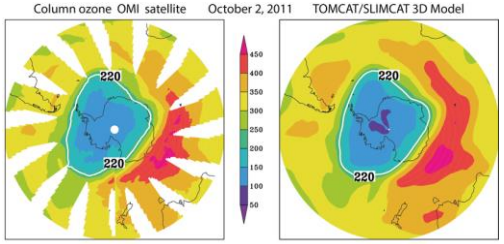
Regional carbon balance on a global scale



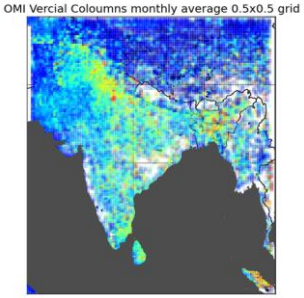
Deriving the impact of fire on vegetation from earth observation data



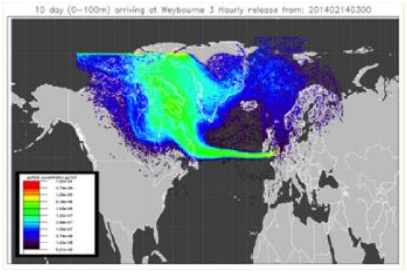
Fault analysis



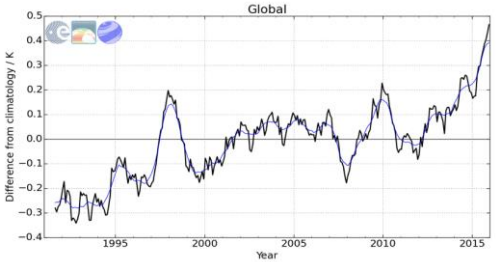
Antarctic Ozone hole: model vs. observations



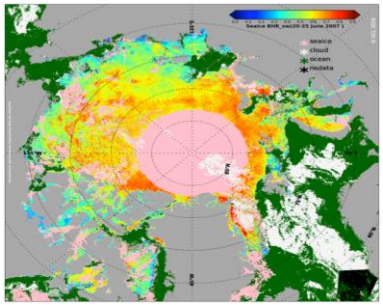
Understanding oxidant chemistry over the Indian subcontinent



Atmospheric dispersion

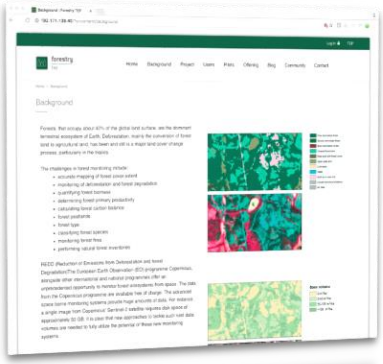


Sea Surface Temperature from satellite observations

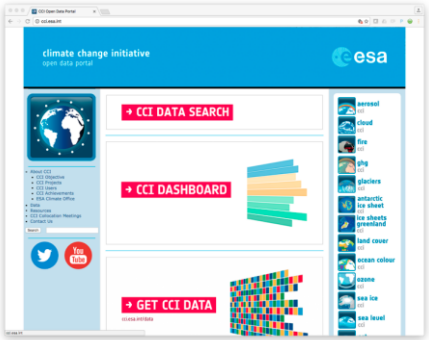


Climate variables from European and US instruments/satellites

JASMIN usage: Cloud

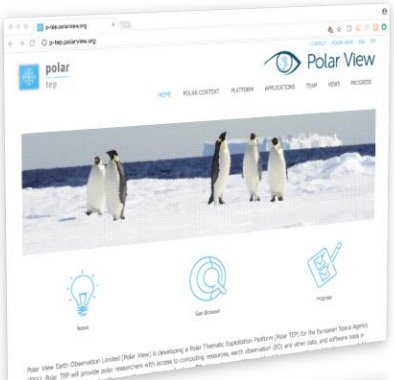


ESA Forestry Thematic Exploitation Platform



ESA Climate Change Initiative Open Data Portal

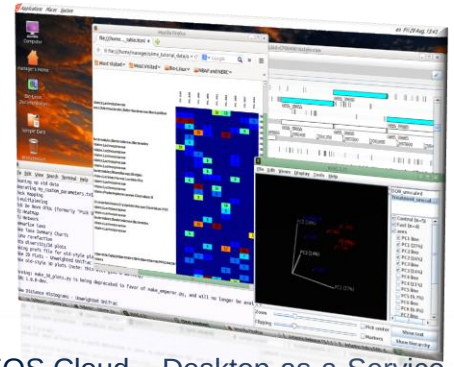
Majic interface to Jules Land-surface model on JASMIN



ESA Polar Thematic Exploitation Platform

Attendees at ESA Summer school, ESRIN used OPTIRAD Jupyter Notebook environment

– Credit ESA



EOS Cloud – Desktop-as-a-Service for Environmental Genomics

Funding & Governance

- Funding
  - Capital
    - "Periodic" capital investment
      - Phase 1, 2, 3 : 2011 - 2015 (~£12M total)
      - Phase 4 currently underway (£8M)
  - Recurrent
    - £650K/year via NERC SLA
  - New governance arrangements April 2018 as part of NERC's National Capability commissioning / funding route
    - NERC, NCAS & NCEO have primary role in approving JASMIN strategy

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Science & Technology Facilities Council

NERC SCIENCE OF THE ENVIRONMENT

Issues & Gaps

- Data intensive computing facility
  - Storage is a major cost
  - Phase 4 introduces more cost effective storage technologies for future scaling: object storage and scale-out file system
- Balancing "bleeding edge" technology with demands of operational service
  - Research infrastructure, not 99.999% available
  - Small team (capital-heavy)
- Diverse user community
  - Wide range of workflow requirements
  - Wide range of user skill level
  - Training, user education, engagement

Further Information

- CEDA and JASMIN:
  - http://www.jasmin.ac.uk/
  - http://www.ceda.ac.uk/

- JASMIN paper

  Lawrence, B.N. , V.L. Bennett, J. Churchill, M. Juckes, P. Kershaw, S. Pascoe, S. Pepler, M. Pritchard, and A. Stephens. **Storing and manipulating environmental big data with JASMIN.** *Proceedings of IEEE Big Data 2013, p68-75,* doi:10.1109/BigData.2013.6691556