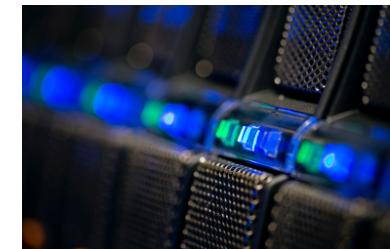




JASMIN (STFC/Stephen Kill)



Software Development at the Centre for Environmental Data Analysis

RAL Site Software Engineering Community Meeting

12 February 2018

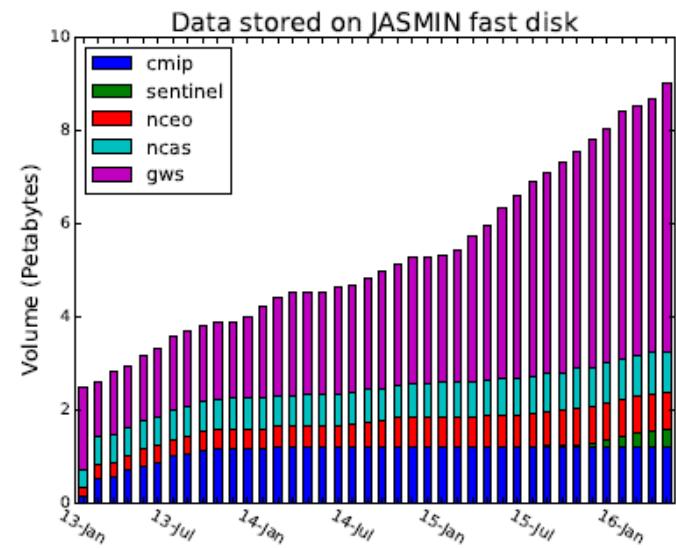
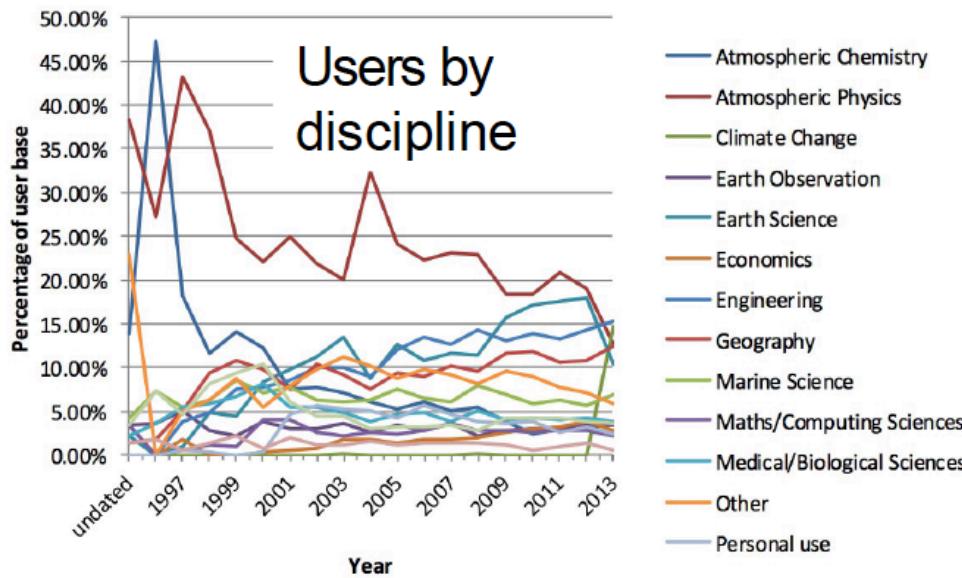
Philip Kershaw (on behalf of CEDA)

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

[Thanks and credit to STFC Scientific Computing Department who deploy and operate the JASMIN infrastructure on behalf of CEDA]



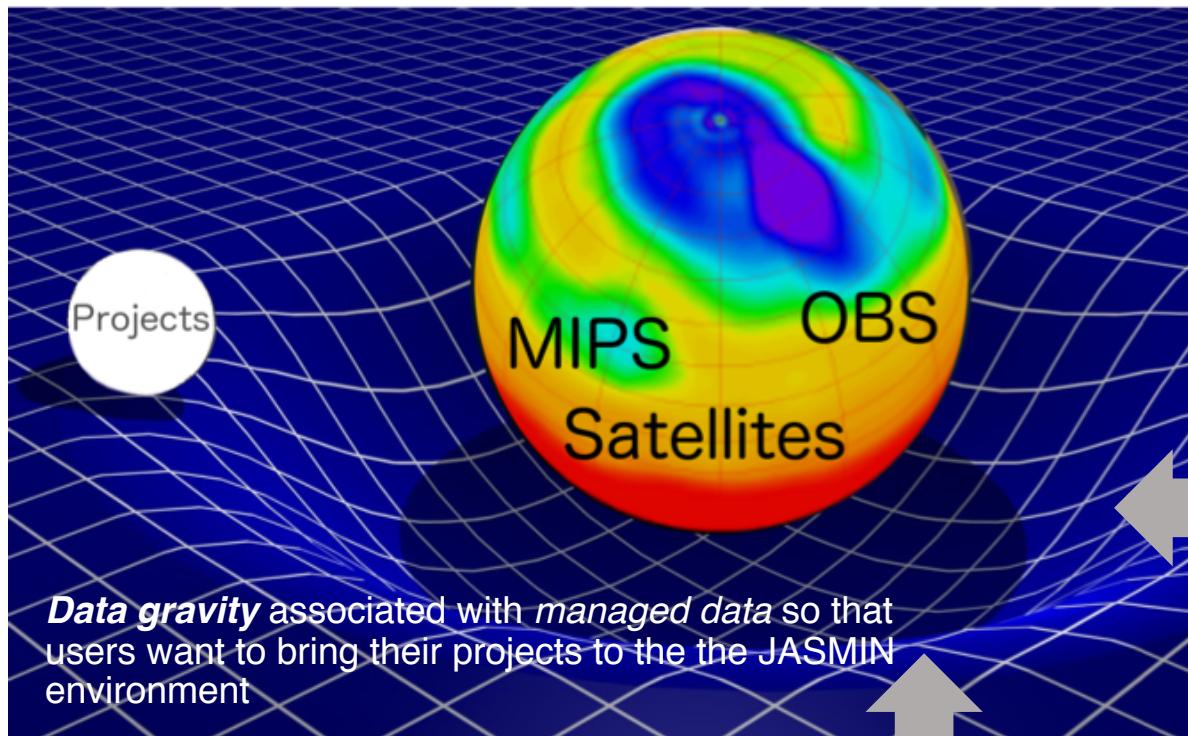
Data Growth and Diversification of user community



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



JASMIN as a Data Commons



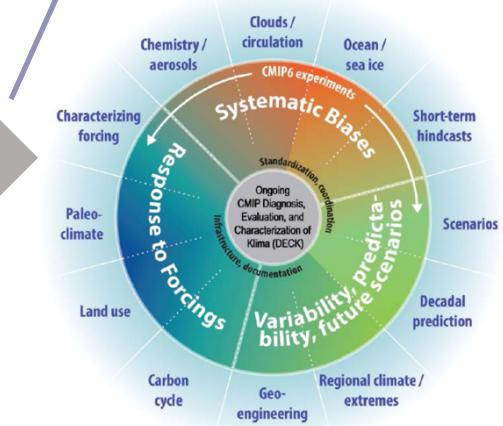
Sentinel Earth
Observation Data



Sentinel missions data rate: ~6PB/year

CEDA CREPP system to ingest
from MetOffice Hadley Centre

CMIP6



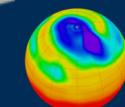
Climate Models
Couple-Model Intercomparison Projects



**National Centre for
Atmospheric Science**
NATIONAL ENVIRONMENT RESEARCH COUNCIL



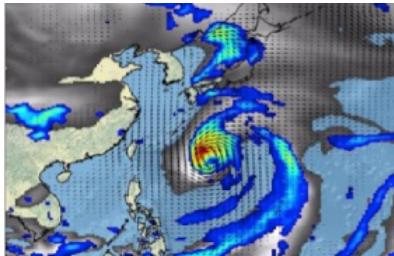
**National Centre for
Earth Observation**
NATIONAL ENVIRONMENT RESEARCH COUNCIL



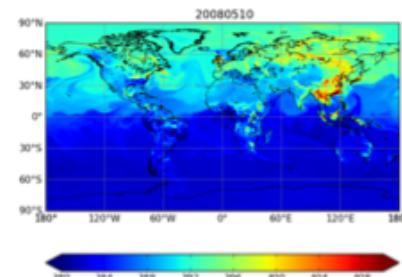
**Centre for Environmental
Data Analysis**
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATIONAL ENVIRONMENT RESEARCH COUNCIL



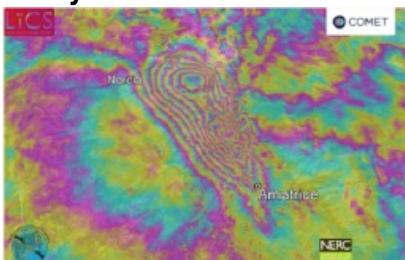
~150 Science projects on JASMIN to date



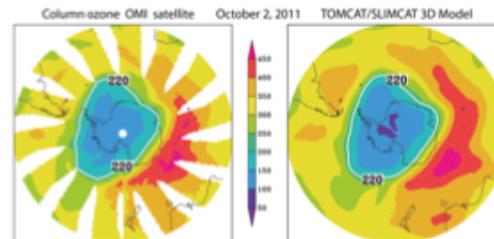
High Res Climate Model analysis



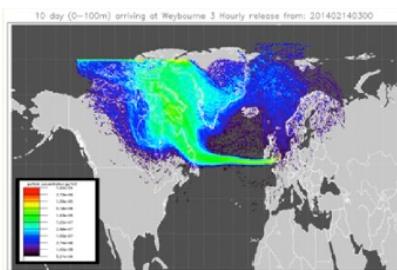
Regional carbon balance on a global scale



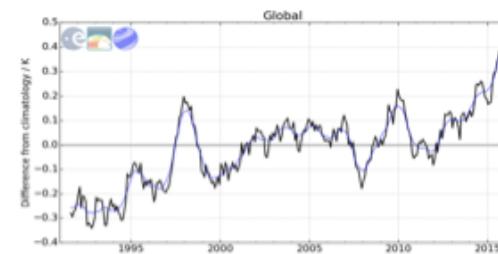
Fault analysis



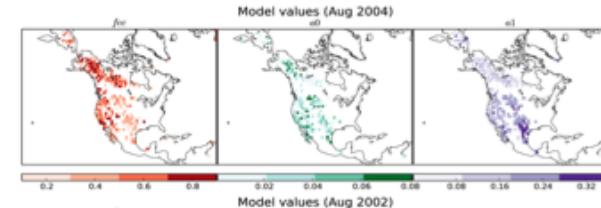
Antarctic Ozone hole: model vs. observations



Atmospheric dispersion



Sea Surface Temperature from satellite observations

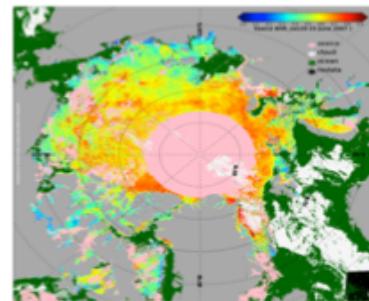
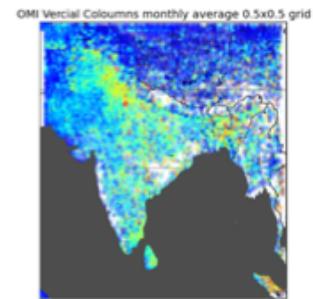


Model values (Aug 2004)



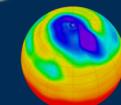
Model values (Aug 2002)

Deriving the impact of fire on vegetation from earth observation data



Climate variables from European and US instruments/satellites

Understanding oxidant chemistry over the Indian subcontinent



Centre for Environmental Data Analysis

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATIONAL ENVIRONMENT RESEARCH COUNCIL



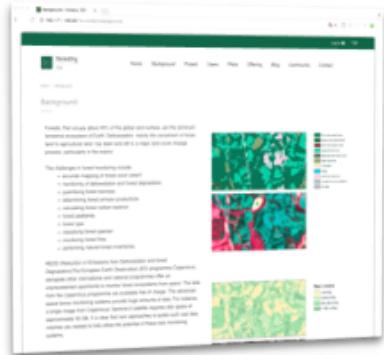
National Centre for
Atmospheric Science
NATIONAL ENVIRONMENT RESEARCH COUNCIL



National Centre for
Earth Observation
NATIONAL ENVIRONMENT RESEARCH COUNCIL



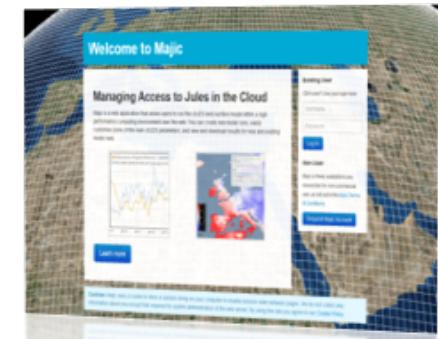
JASMIN usage: Cloud



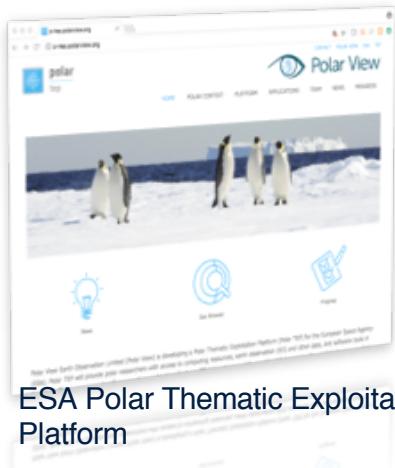
ESA Forestry Thematic Exploitation Platform



ESA Climate Change Initiative Open Data Portal



Majic interface to Jules Land-surface model on JASMIN

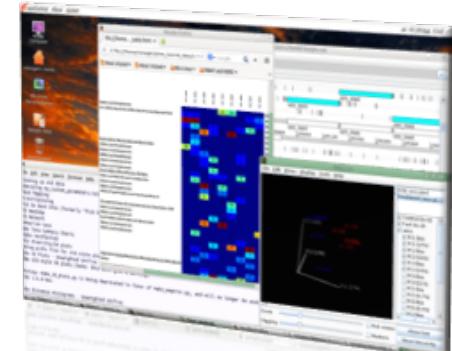


ESA Polar Thematic Exploitation Platform

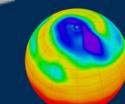


Attendees at ESA Summer school, ESRIN used OPTIRAD Jupyter Notebook environment

– Credit ESA

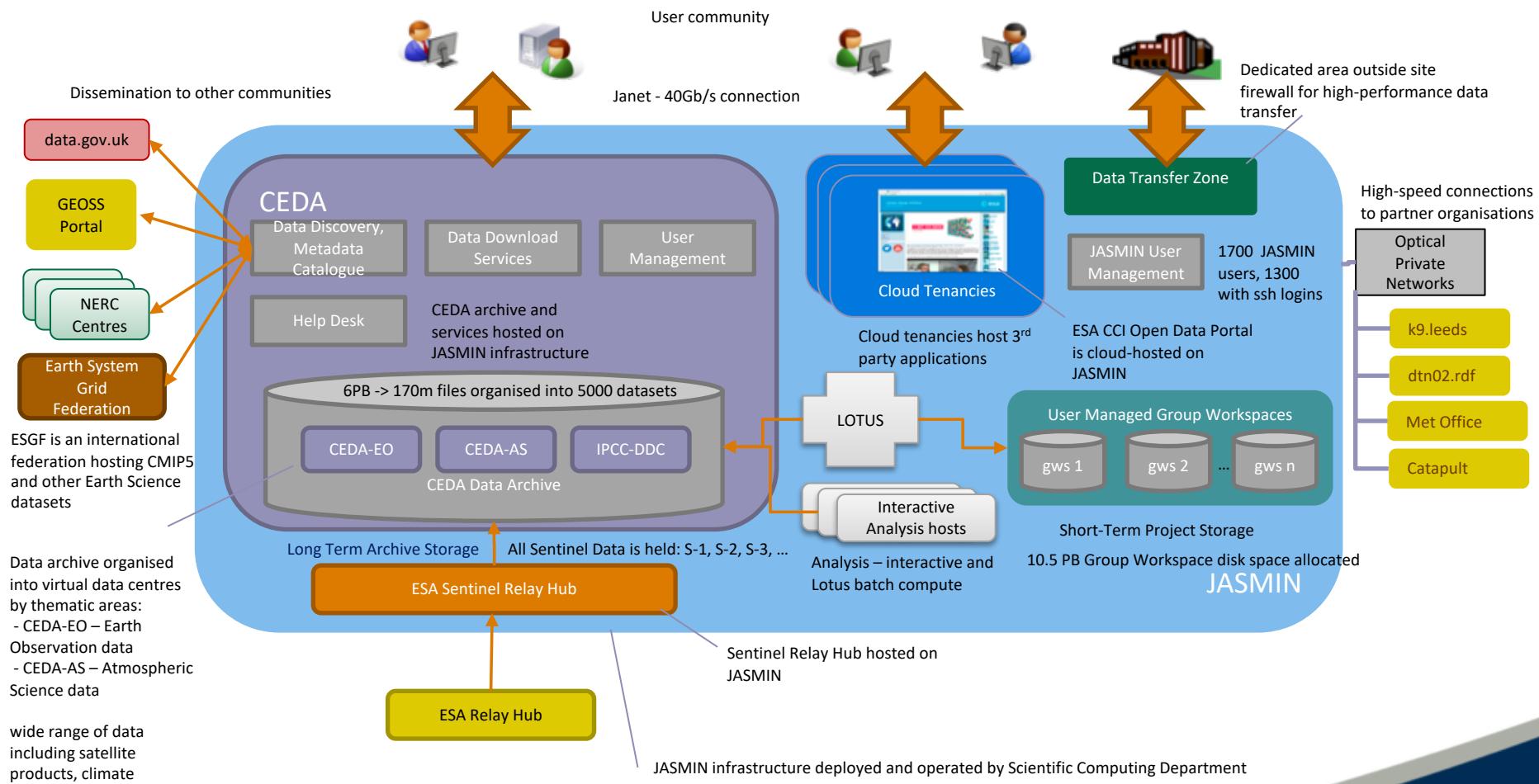


EOS Cloud – Desktop-as-a-Service for Environmental Genomics





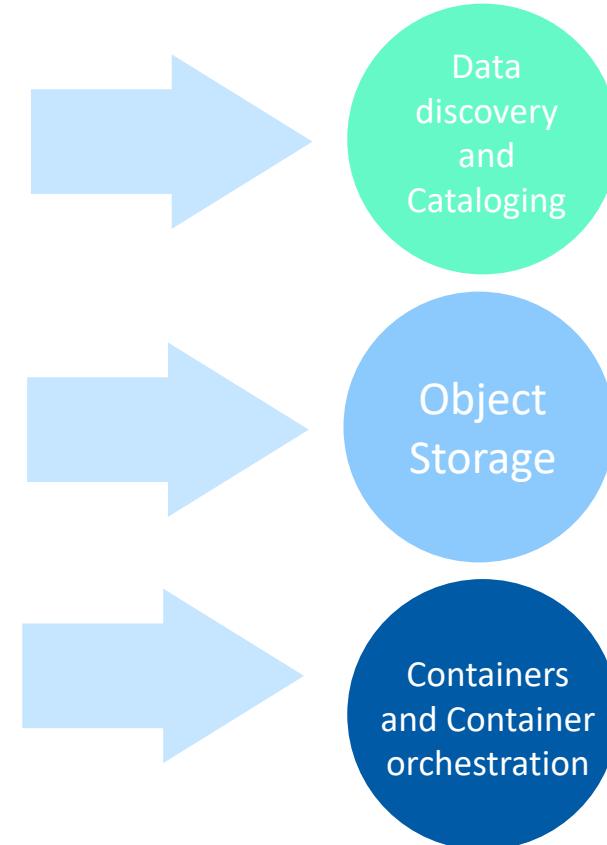
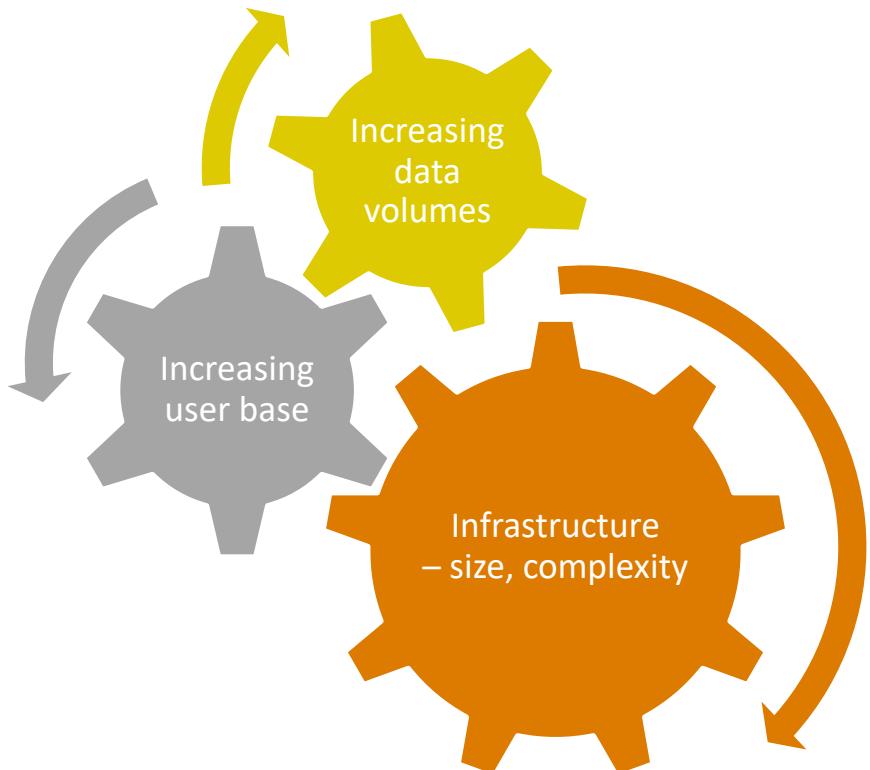
JASMIN and CEDA





Challenges and new developments to address them (1)

Key technologies to address specific challenges



On-prem ⇔ Public Cloud portability



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-ARIS, SAFIRE, AWE Polar5, K9-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** **ARIS**

Click an item to expand the panel.

Geographical Search

Temporal Filter

Refine Search via Time Range to

1k
100
10

Google

Only the first 1000 results will be plotted on 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

8141 hits with current selection.

123 milliseconds for response.

Mouse: Lat: 77.43, Lng: -18.63

Google

Key:

- Sentinel 1
- Sentinel 2
- Sentinel 3
- Landsat
- Other

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

Accessing the data collected on moving platforms such as research aircraft can be difficult – unless you were involved in the flight planning, or have access to the flight log files. The EUFAR Flight Finder was developed during the EUFAR14 project to maximise discovery, access and reuse of the EUFAR flight data. The EUFAR Flight Finder is now part of the CEDA Flight Data Catalogue, which includes flights and measurements, during aircraft research flights, which data are stored in the CEDA data archive in ESERO.

The output of the EUFAR Flight Finder was updated to include the full catalogue of flights from the Met Office CEDA archive and the CEDA Flight Data Catalogue. The CEDA Flight Data Catalogue is now stored in the CEDA archive, and now covers 1000s of flights, from 1990 to the present day.

Following the success of the EUFAR, the code was copied and modified to display satellite data in the CEDA Flight Data Catalogue. The EUFAR Flight Finder now includes a user friendly, image search, and image search, search capability for 2 million datasets in CEDA, and covers 10,000s of flights.

How the EUFAR Satellite Data Finder works

Location information, flight number, source organisation, target location, and date of the flight are required to search for flights. It is common for many flights to be carried out on the same day, so it is important to use a date range when searching for flights.

How to use the EUFAR Flight Data Catalogue

EUFAF flights, flights from the CEDA Flight Data Catalogue, and flights from the CEDA Satellite Data Catalogue are now available in the EUFAR Flight Data Catalogue. The EUFAR Flight Data Catalogue is a user friendly, image search, and image search, search capability for 2 million datasets in CEDA, and covers 10,000s of flights.

Challenges

There are many challenges in the EUFAR Flight Data Catalogue. One of the main challenges is to ensure that the data is accurate and consistent. Another challenge is to ensure that the data is accessible and easy to use. A third challenge is to ensure that the data is useful for research and development.

Conclusions and future plans

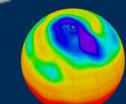
The EUFAR Flight Data Catalogue and CEDA Satellite Data Catalogue are now available for use in research and development. The EUFAR Flight Data Catalogue is a user friendly, image search, and image search, search capability for 2 million datasets in CEDA, and covers 10,000s of flights.

Future plans under consideration include:

- Improving the search functionality to make it faster and more accurate.
- Adding further satellite missions if they are supported by the flight and using orbital images.
- Adding further aircraft missions if they are supported by the flight and using orbital images.
- Adding a search for user-defined countries and flight days for validation studies.

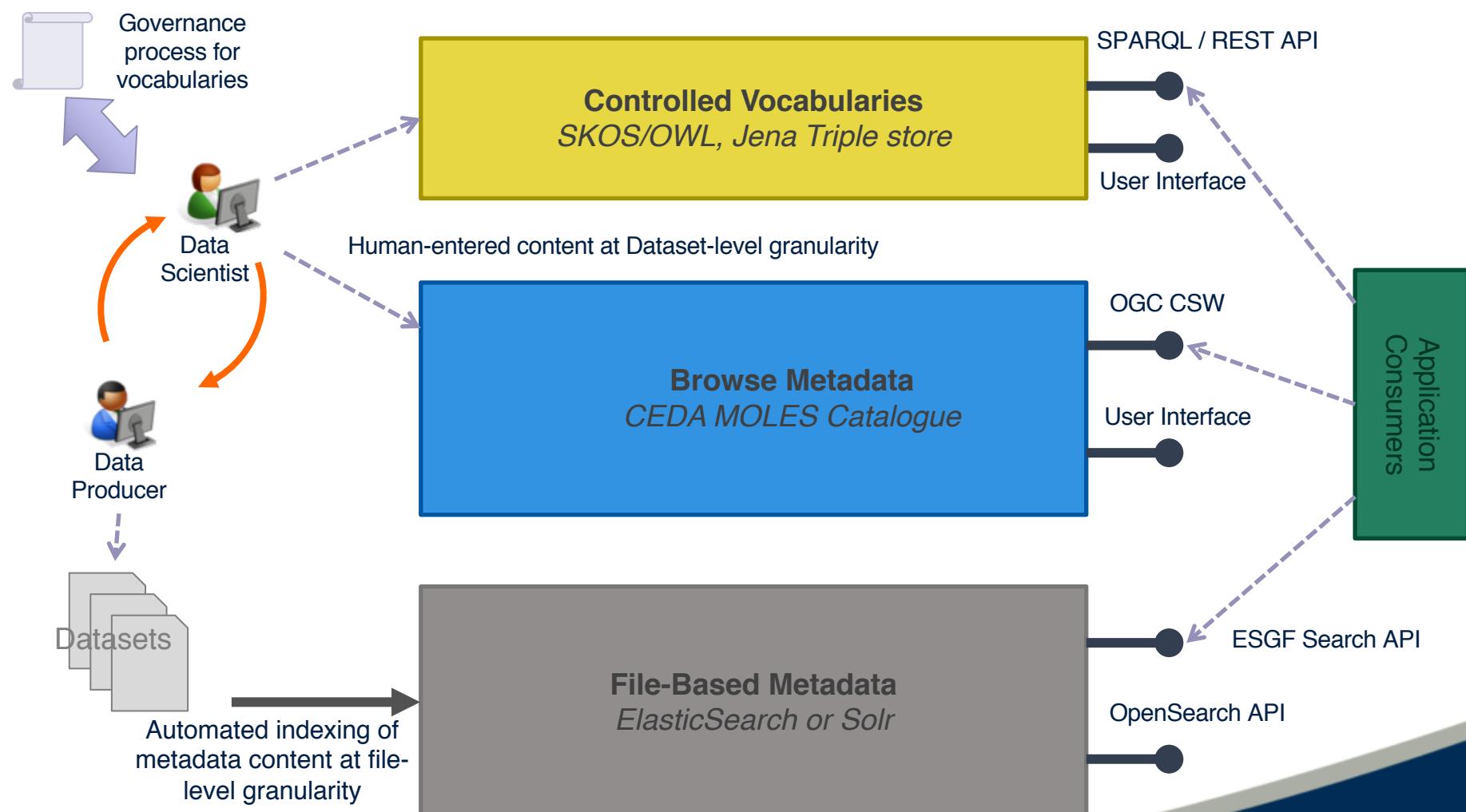
Contact: Wendy Garland
wendy.garland@ncas.ac.uk

<http://flight-finder.ceda.ac.uk>
<http://geo-search.ceda.ac.uk>



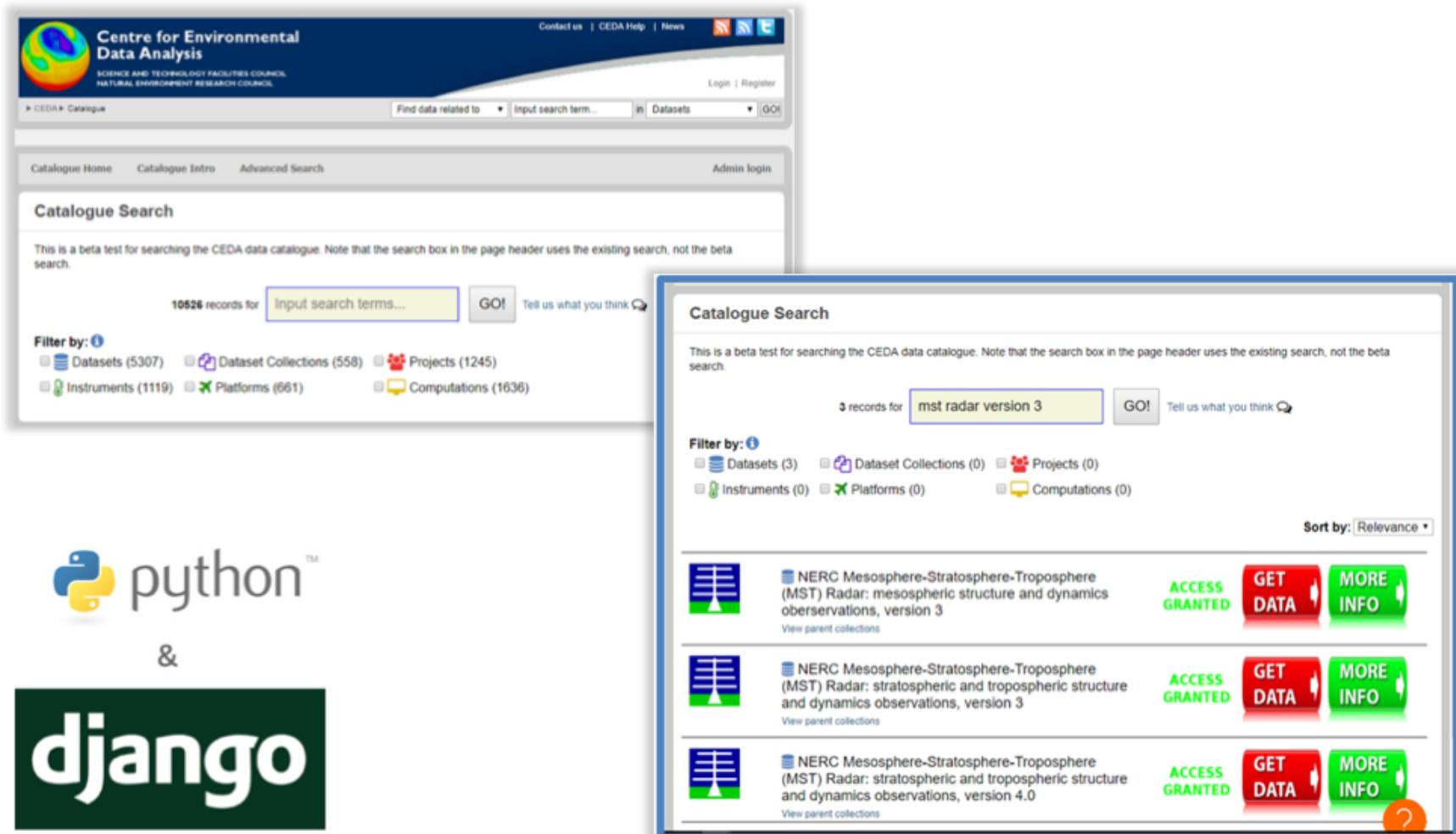


Sourcing information for Data discovery





Catalogue search



The image shows two screenshots of the CEDA Catalogue search interface. The left screenshot shows the search results for 'mst radar version 3' with 3 records found. The right screenshot shows the same search results with 10526 records found. Both screenshots include a search bar, filter options, and a list of results with 'GET DATA' and 'MORE INFO' buttons.

Catalogue Search

This is a beta test for searching the CEDA data catalogue. Note that the search box in the page header uses the existing search, not the beta search.

10526 records for Tell us what you think

Filter by:

- Datasets (5307)
- Dataset Collections (558)
- Projects (1245)
- Instruments (1119)
- Platforms (661)
- Computations (1636)

Catalogue Search

This is a beta test for searching the CEDA data catalogue. Note that the search box in the page header uses the existing search, not the beta search.

3 records for Tell us what you think

Filter by:

- Datasets (3)
- Dataset Collections (0)
- Projects (0)
- Instruments (0)
- Platforms (0)
- Computations (0)

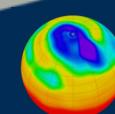
Sort by:

		ACCESS GRANTED	GET DATA	MORE INFO
	NERC Mesosphere-Stratosphere-Troposphere (MST) Radar: mesospheric structure and dynamics oberservations, version 3			
	NERC Mesosphere-Stratosphere-Troposphere (MST) Radar: stratospheric and tropospheric structure and dynamics observations, version 3			
	NERC Mesosphere-Stratosphere-Troposphere (MST) Radar: stratospheric and tropospheric structure and dynamics observations, version 4.0			

python™

&

django





Dataset and File-level metadata integration

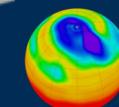
Details Related Datasets Process Variables Tools (3) Docs (2) Comments (0)

```
air_pressure (plev) [Pa]
  standard_name: air_pressure
  var_id: plev
  units: Pa
air_pressure_at_sea_level (psl) [Pa]
  standard_name: air_pressure_at_sea_level
  var_id: psl
  units: Pa
air_temperature (ta) [K]
  standard_name: air_temperature
  var_id: ta
  units: K
air_temperature (tas) [K]
  standard_name: air_temperature
  var_id: tas
  units: K
air_temperature (tasmax) [K]
  standard_name: air_temperature
  var_id: tasmax
  units: K
air_temperature (tasmin) [K]
  standard_name: air_temperature
  var_id: tasmin
  units: K
depth (depth) [m]
  standard_name: depth
  var_id: depth
  units: m
depth (lev) [m]
  standard_name: depth
  var_id: lev
  units: m
eastward_wind (ua) [m s-1]
  standard_name: eastward_wind
  var_id: ua
  units: m s-1
eastward_wind (uas) [m s-1]
  standard_name: eastward_wind
  var_id: uas
  units: m s-1
geopotential_height (zg) [m]
```

Dataset and File-level metadata integration

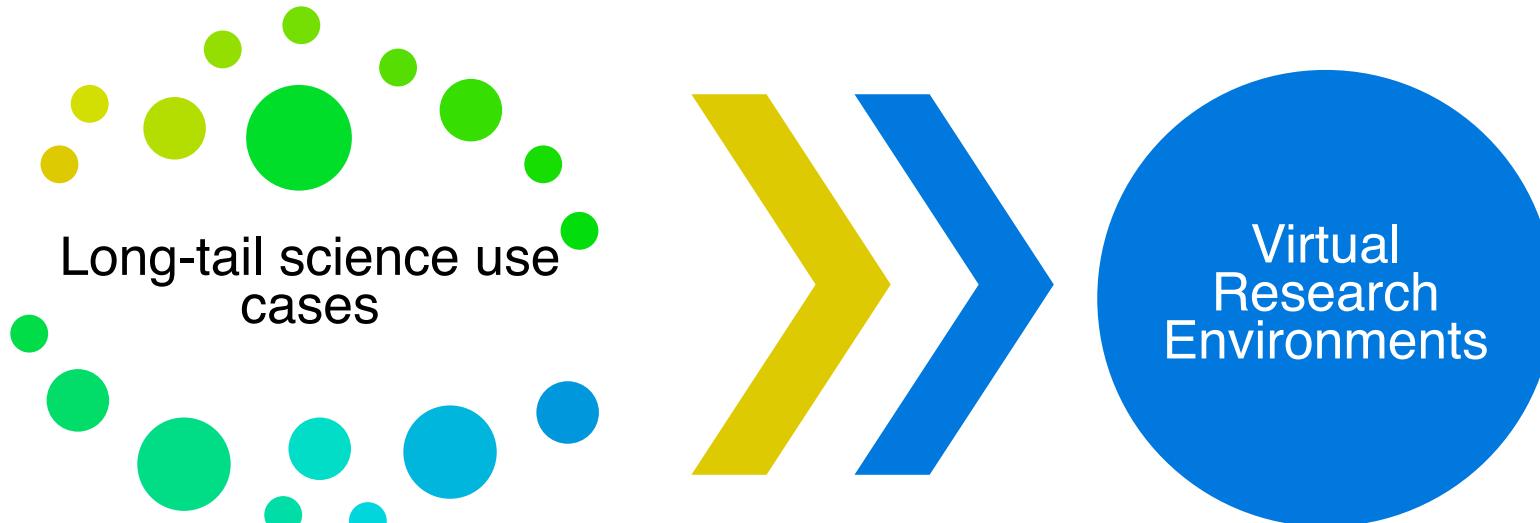


&



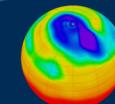


Challenges and new developments to address them (2)



- Need for effective exploitation of parallelism to deliver demonstrable benefit over user's desktop/laptop
- Need for ease of use
- Intuitive user interfaces

- Software-as-a-Service model
- Desktop app-style user experience
- dynamic provisioning of clusters
- Parallel programming libraries





Data Analytics

Boundaries and responsibilities for administration

Project admin



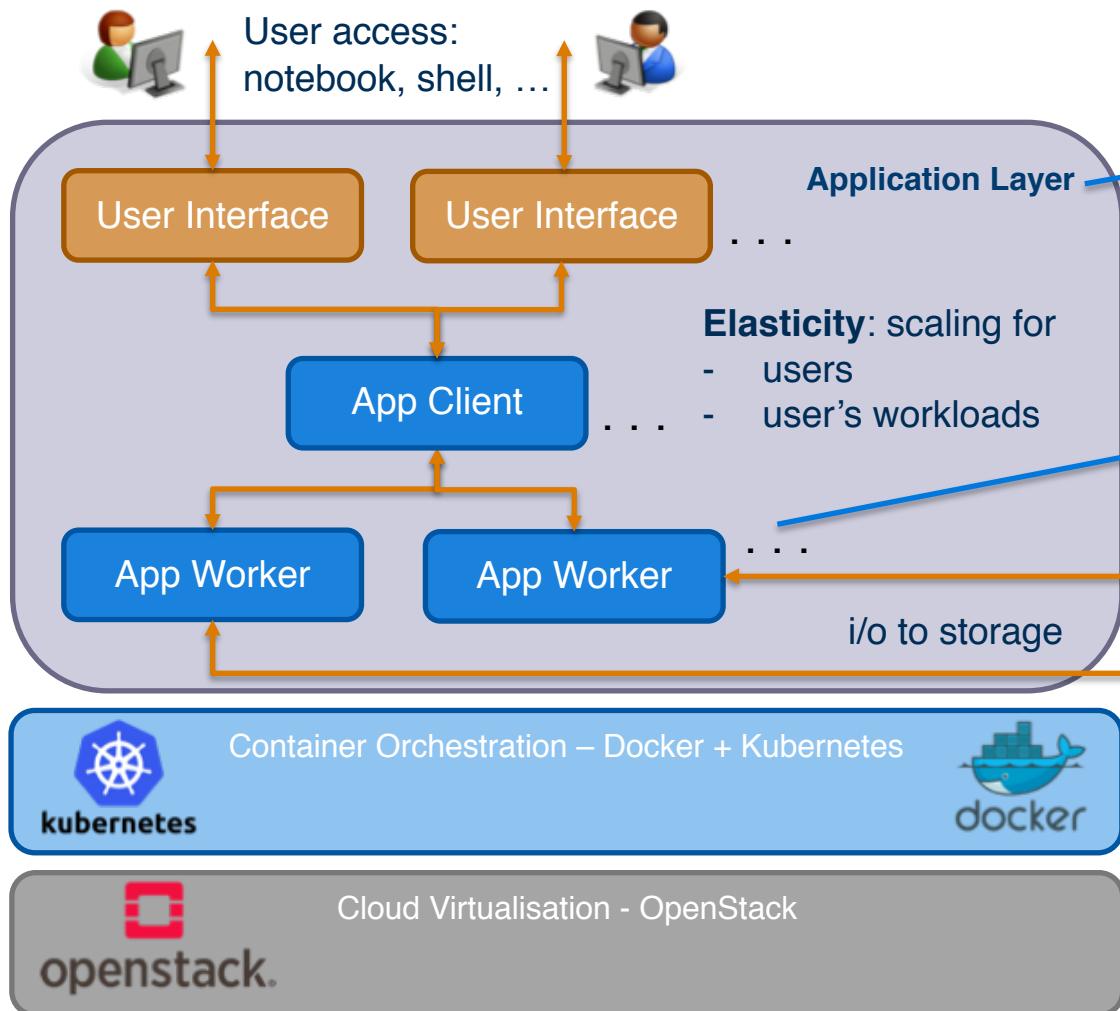
Application Management

JASMIN Admin



Container Management

Cloud Management



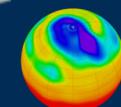
Parallelism:



DASK

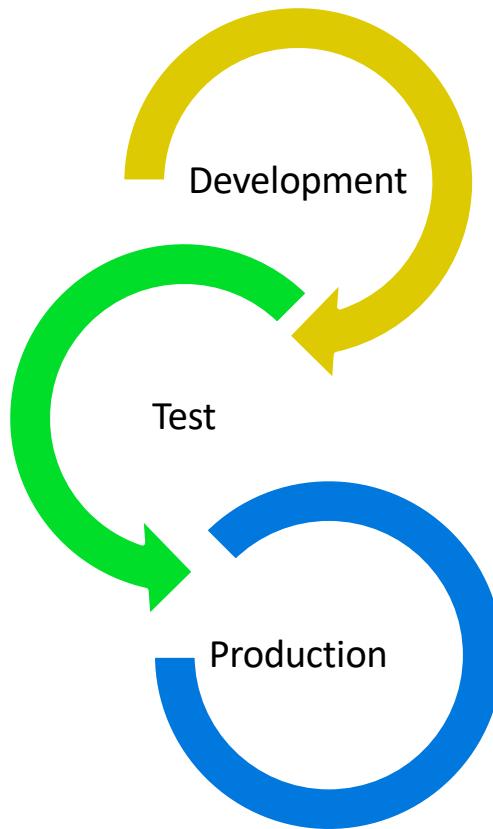


Storage





Development challenge: size and complexity of code and systems footprint



- CEDA Development team
 - 15 people includes data scientists, ops and managers – approx. 7 f/t s/w development
- Languages
 - Python predominates, all new projects Python 3
 - One Cython project
 - JS - React
 - Some Java
- Development environments: PyCharm, PyDev (Eclipse), CLI + editors
- All projects Open Source by default
 - <https://github.com/cedadev/>
 - Private git for deployment-sensitive content (e.g. Ansible playbooks)
- Training in the user community:
 - Introduction to Scientific Computing Course
- Build, test, integrate, operate
 - Vagrant + Ansible
 - Cloud dev tenancy
 - Standardised on RedHat 6/7
 - Planned: Docker + Kubernetes (OpenShift)
 - Production checkout process and documentation
 - Integrating code tests into Icinga/Nagios operational monitoring



Summary

- JASMIN: data gravity, a data commons for environmental sciences
- Challenges with respect to running at scale:
 - Data volumes
 - Numbers of users
 - Generation and indexing of content for effective discovery and understanding of data for users
 - Effective use of parallelism for long-tail of science users
 - Increasing footprint of code and systems to manage for development and operations
- New infrastructure services to address challenges:
 - Evolution of data discovery and cataloguing systems, AI exploitation?
 - Virtual Research Environments
 - Object store migration
 - Development and operations: Increasing Automation – virtualisation, containers and container orchestration





Further Information



CEDA team

- CEDA and JASMIN:
 - <http://www.jasmin.ac.uk/>
 - <http://www.ceda.ac.uk/>
- Github:
 - <https://github.com/cedadev/>
- 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
- philip.kershaw@stfc.ac.uk,
[@PhilipJKershaw](https://twitter.com/PhilipJKershaw)