

# JASMIN: Managing Variety in a Climate Data Community Platform

Richard Smith, Victoria Bennett, Philip Kershaw, Brian Lawrence

Center for Environmental Data Analysis

RAL Space, STFC

# Overview

---

What is JASMIN?

Sources of Variety

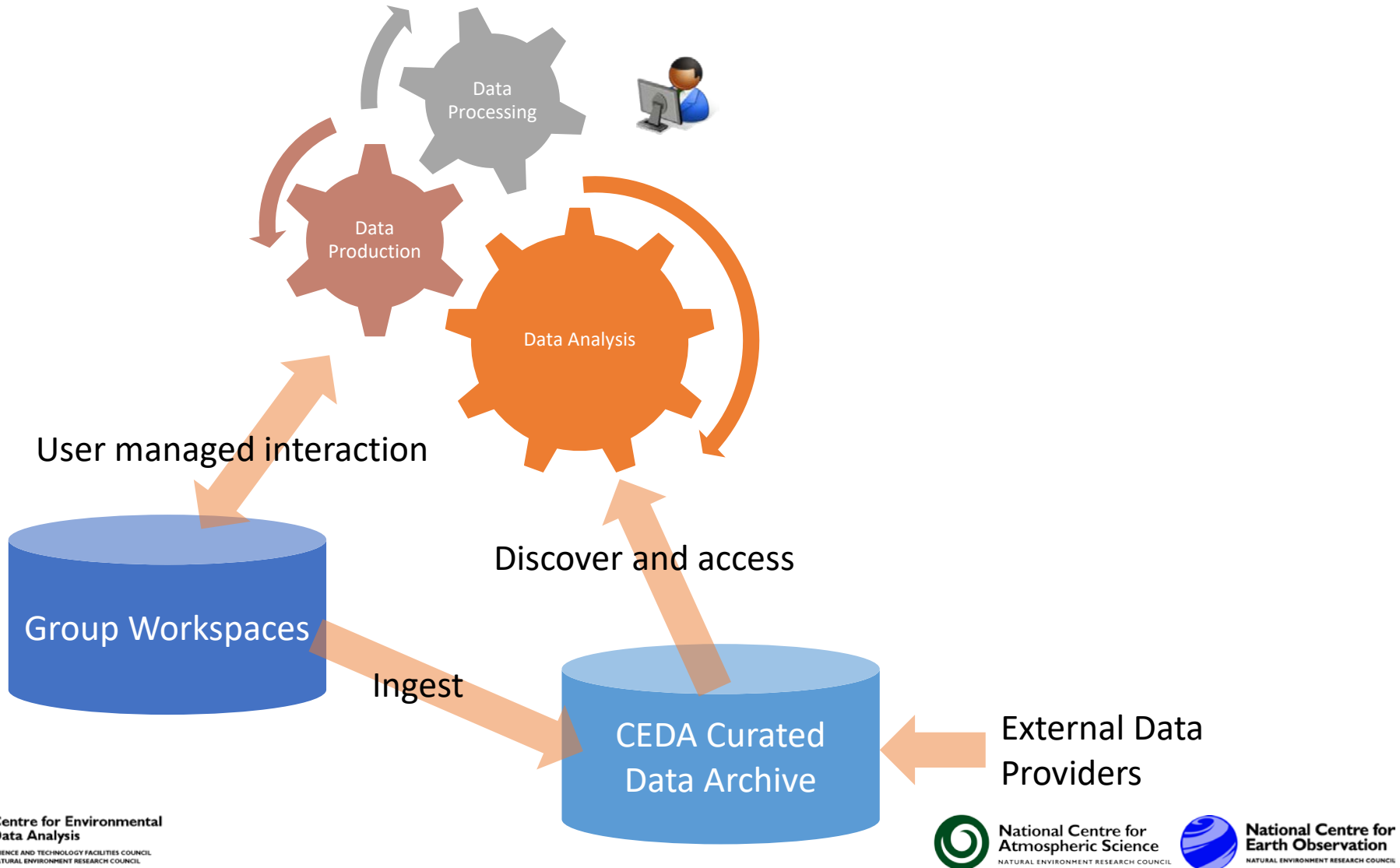
Challenges

Solutions: Creating and index

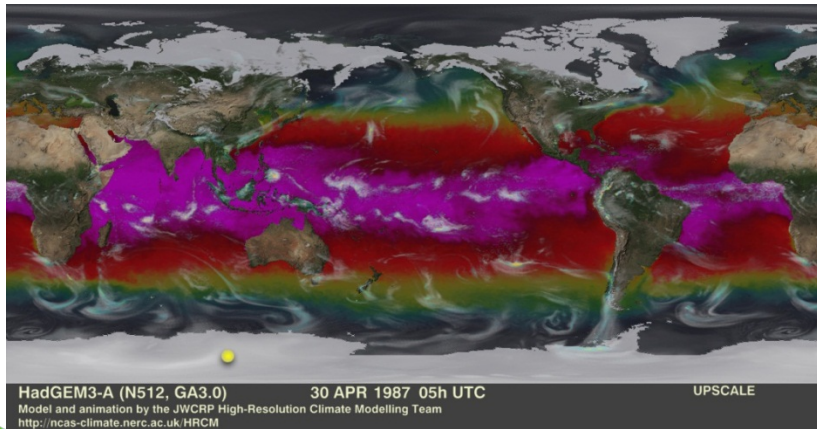
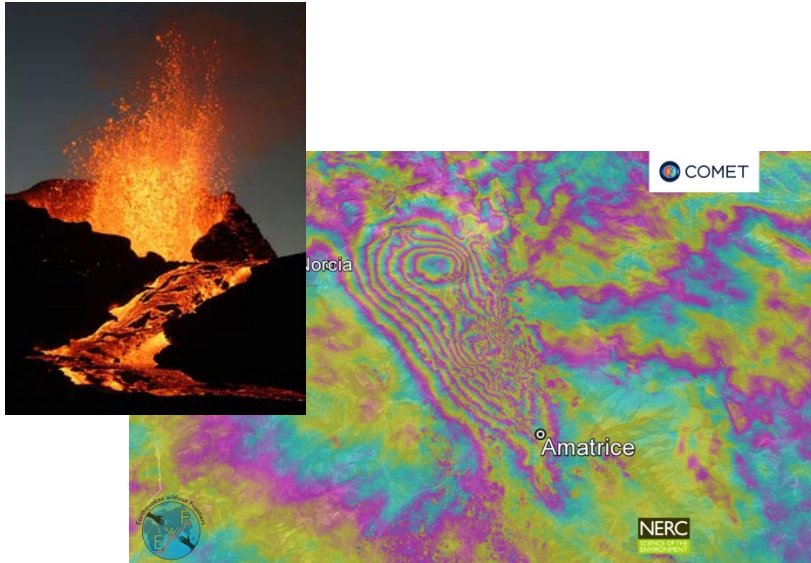
Solutions: Evolution of the service

# What is JASMIN?

# JASMIN: Bring the compute to the data



# JASMIN



## Biases in *ad-hoc* data



Virtual Environments deployed on JASMIN

Thematic Exploitation Platform for ESA

OCI Open Data Portal for ESA

MAHC interface to JULES model

EOS Cloud — Desktop-as-a-Service for Environmental Geomatics

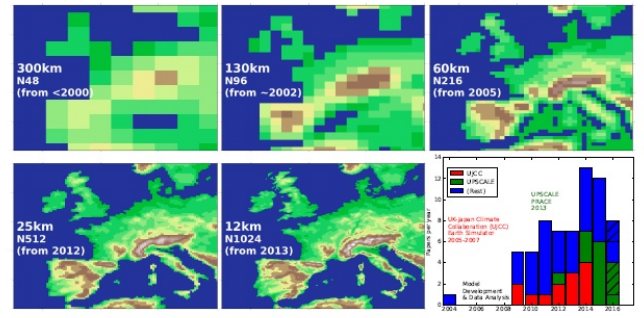
Hosted Jupyter Notebooks

NERC Environmental Workbench

Box 9: Exemplar projects deployed in the JASMIN infrastructure zone — “the unmanaged” cloud.

# JASMIN: The missing piece

## Growing Need - High Resolution Climate Programme!



Just one example, of the *many* axes of growing scientific demand in simulations and observation:

- ▶ From 7K to 3.1M points (0.05 MB to 25MB) for a single timestep of a single level of a global field.
- ▶ Multi-year data management campaigns support the data analysis (which needs to include similarly high-resolution observations).



MetOffice supercomputer



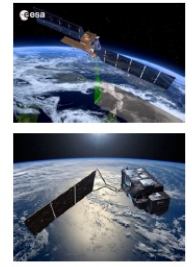
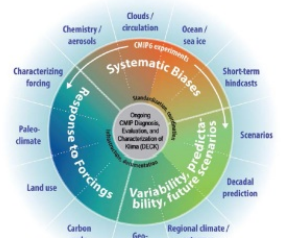
ARCHER supercomputer (EPSRC/NERC)



The UK JASMIN Environmental Commons: Now and into the Future  
Bryan Lawrence - RAL, 27th June 2017

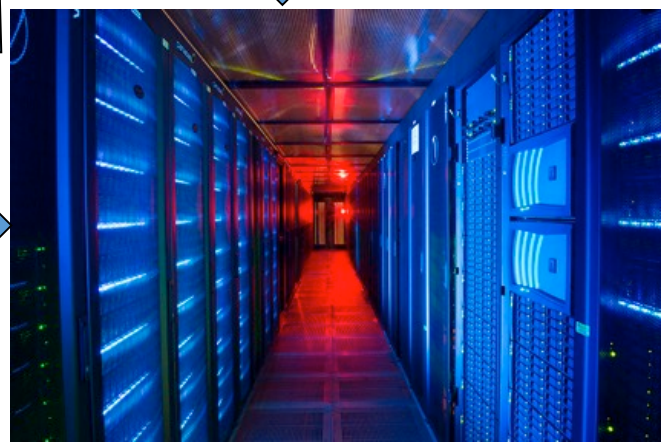


## The Organised Data Deluge



Sentinel 1A (2014), 1B (2016)  
Sentinel 2A (2015) 2B (2017?)  
Sentinel 3A (2016) 3B (2018?)  
Data rate: o(6) PB/year

- aerosol (CCI)
- cloud (CCI)
- fire (CCI)
- ghg (CCI)
- glaciers (CCI)
- antarctic ice sheet (CCI)
- ice sheets greenland (CCI)
- land cover (CCI)
- ocean colour (CCI)
- ozone (CCI)
- sea ice (CCI)
- sea level (CCI)
- sst (CCI)
- soil moisture (CCI)
- cmug (CCI)



JASMIN (STFC/Stephen King)

CMIP6 data volumes and data rates not yet known, but the European contribution to HiresMIP alone is expected to exceed 2 PB.



The UK JASMIN Environmental Commons: Now and into the Future  
Bryan Lawrence - RAL, 27th June 2017



National Centre for Atmospheric Science  
NATURAL ENVIRONMENT RESEARCH COUNCIL



National Centre for Earth Observation  
NATURAL ENVIRONMENT RESEARCH COUNCIL

# Sources of Variety

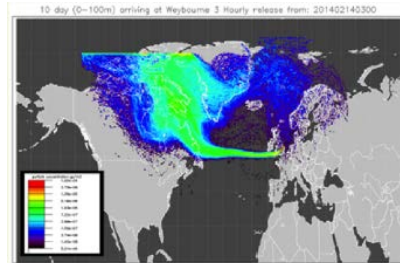
# Challenges: Project Variety (150+ projects)



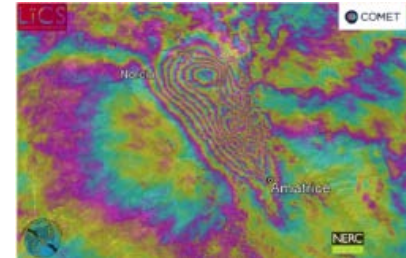
Trend analysis for 1000's of UK species. Unprecedented scope and complexity



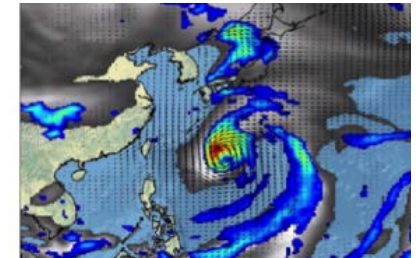
COMET-CPOM: Near real time monitoring of all active volcanos



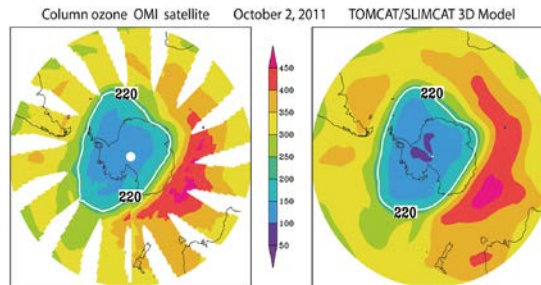
Atmospheric Dispersion



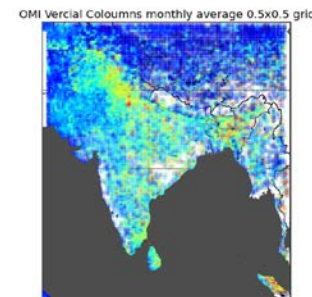
Fault Analysis



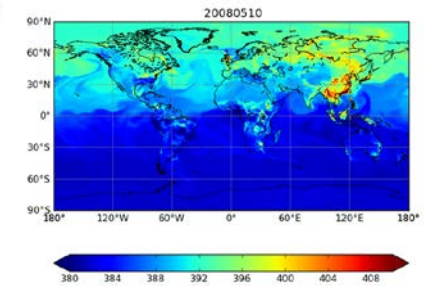
High Res Climate Model analysis



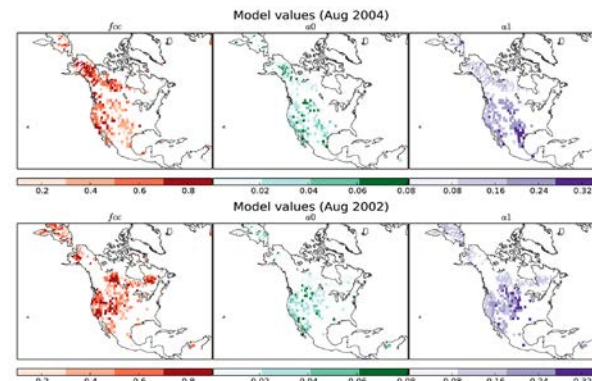
Antarctic Ozone Hole: model vs obs



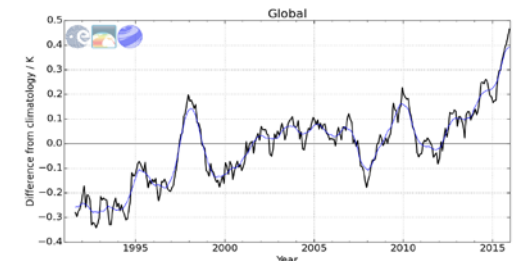
Oxidant chemistry, India



Regional carbon balance on global scale



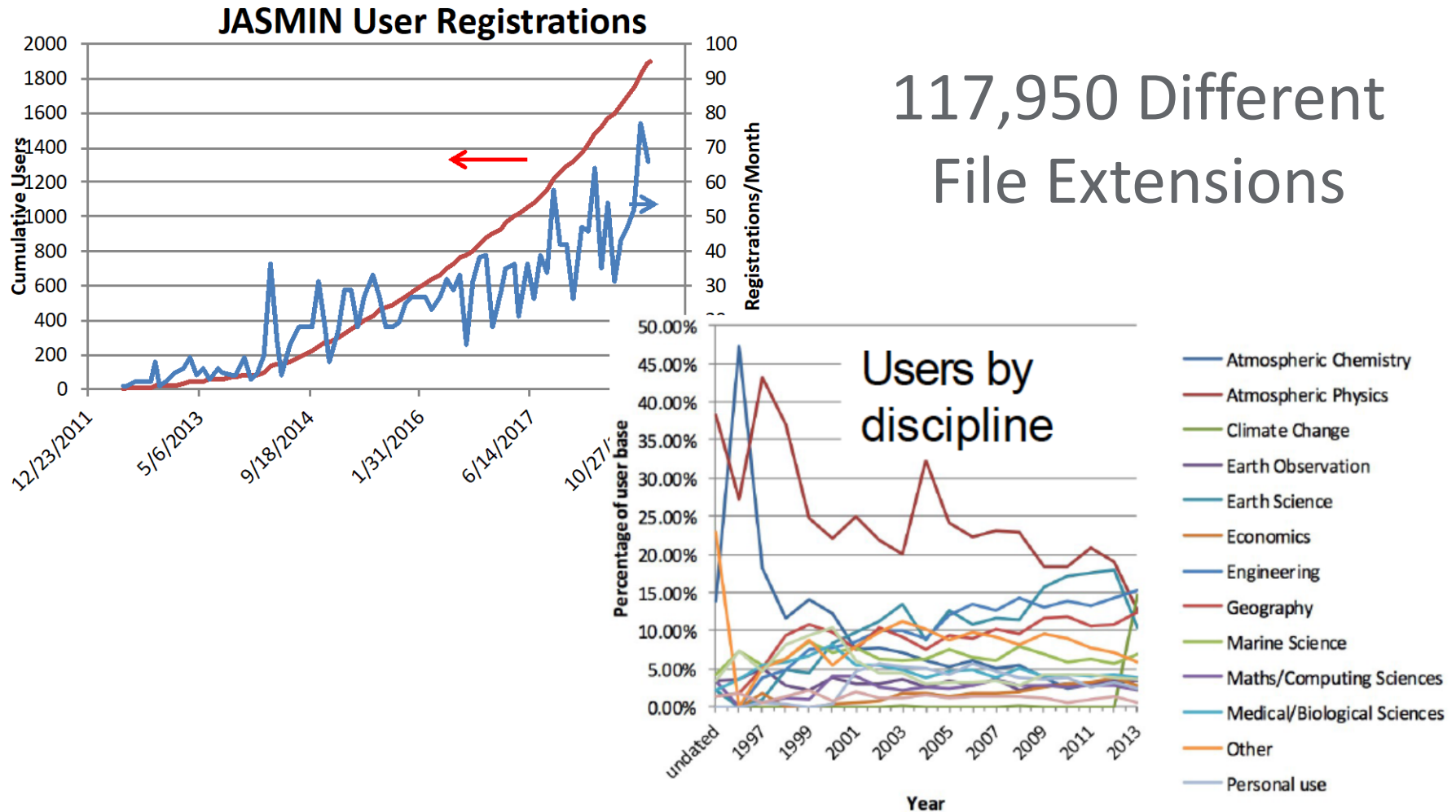
Fire Impact Maps



Sea Surface Temp from satellite obs



# Challenges: User Domain Variety



# Challenges: Hardware Variety

JASMIN has been built in stages

Each phase adds storage, compute and network improvements



Storage Media for JASMIN Phase 4

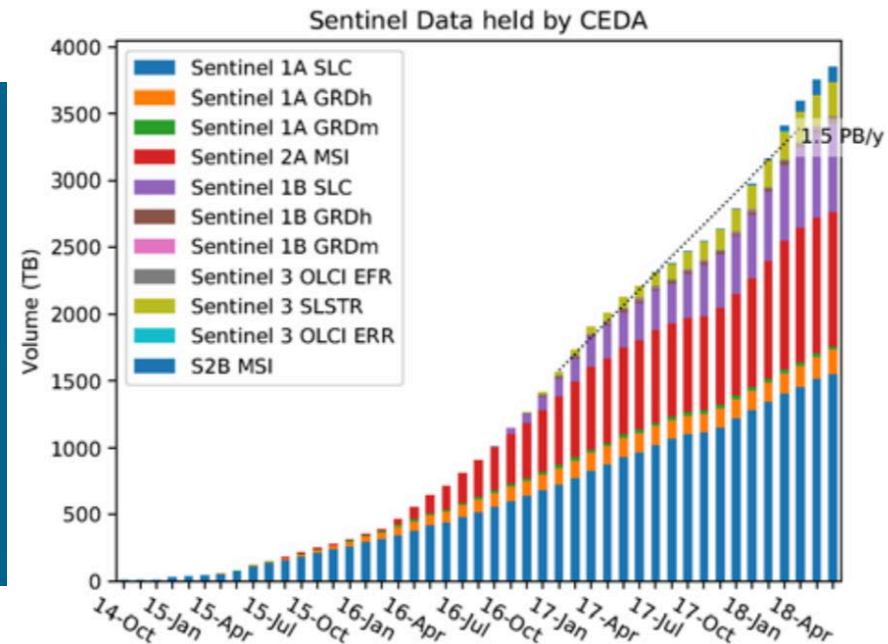


Figure 1: Sentinel data growth 2014-2018 in CEDA archive on JASMIN

# Challenges

---

How do we make the best use of the data held on JASMIN?

How do we display what is available to potential users?

How do we provide for the many different user skill levels and domains which want to use our platform?



# Solutions: Creating an Index

```
31 def __init__(self, path):
32     self.file = None
33     self.fingerprints = set()
34     self.logdupes = True
35     self.debug = debug
36     self.logger = logging.getLogger(__name__)
37     if path:
38         self.file = open(os.path.join(path, "requests.txt"), "w")
39         self.file.seek(0)
40         self.fingerprints.update(requests)
41
42 @classmethod
43 def from_settings(cls, settings):
44     debug = settings.getbool("debug", False)
45     return cls(job_dir(settings), debug)
46
47 def request_seen(self, request):
48     fp = self.request_fingerprint(request)
49     if fp in self.fingerprints:
50         return True
51     self.fingerprints.add(fp)
52     if self.file:
53         self.file.write(fp + os.linesep)
54
55 def request_fingerprint(self, request):
56     return request_fingerprint(request)
```

# Solutions: File Variety

---

211,000,000+  
Files

10  
File Parsers

118,000  
“File Extensions”

125+  
CPU Days

200,000  
Files/day

# Solutions: Elasticsearch Indexing

```
{
  "_index" : "ceda-fbi-2018-08-26",
  "_type" : "file",
  "_id" : "b3e2c5c0fce2bea0b15107857ae017391879dd22",
  "_score" : 1.0,
  "_source" : {
    "info" : {
      "name_auto" : "Image00249.jpg",
      "size" : 3962794,
      "name" : "Image00249.jpg",
      "location" : "on_disk",
      "type" : ".jpg",
      "spot_name" : "corral",
      "directory" : "/badc/corral/images/metobs/indian_ocean/India
        /IndianDailyWeatherReports/OCT01-DEC31_1941",
      "last_modified" : "2015-02-13T14:41:56",
      "user" : "badc",
      "group" : "open",
      "md5" : "7dc80418c9c988a543408887a6f424d3"
    }
  }
},
```

# Solutions: Elasticsearch Indexing

```
{
  "best_name" : "GLOBAL MEAN LW radiation TOA (OLR)",
  "var_id" : "gm_rlut",
  "long_name" : "GLOBAL MEAN LW radiation TOA (OLR)",
  "names" : [
    """"GLOBAL MEAN LW radiation TOA (OLR)""""
  ],
  "units" : "W/m^2",
  "agg_string" : """"long_name":"GLOBAL MEAN LW radiation TOA (OLR)"
    , "names":"GLOBAL MEAN LW radiation TOA (OLR)", "units":"W/m^2", "var_id"
    : "gm_rlut""""
},
{
  "best_name" : "GLOBAL MEAN SURFACE SH FLUX FROM SEA (GBM) W/M2",
  "var_id" : "gm_field3228",
  "long_name" : "GLOBAL MEAN SURFACE SH FLUX FROM SEA (GBM) W/M2",
  "names" : [
    """"GLOBAL MEAN SURFACE SH FLUX FROM SEA (GBM) W/M2""""
  ],
  "units" : "",
  "agg_string" : """"long_name":"GLOBAL MEAN SURFACE SH FLUX FROM SEA (GBM
    ) W/M2", "names":"GLOBAL MEAN SURFACE SH FLUX FROM SEA (GBM) W/M2"
    , "units":"","var_id":"gm_field3228""""
}
```

# Solutions: Elasticsearch

**CEDA Archive** Search Catalogue Get Data Help Tools Deposit News Sign in

Abbreviation: Not defined  
Keywords: CMIP5, WCRP, climate change, NCAR, CCSM4, historicalExt, atmos, fx

Related Records Details / Docs (0) Process **Variables (5)** Tools (3)

**Surface Altitude** (orog) [m] ↑

**lat\_bnds** (lat\_bnds) ↑

**lon\_bnds** (lon\_bnds) ↑

**Co-ordinate Variables**

**latitude** (lat) [degrees...] ↑

**longitude** (lon) [degrees...] ↑

Map data © 2015

Related parties

**Authors (1)**

- National Center for Atmospheric Research (NCAR)

**Publishers (1)**

? Help

**CEDA Archive** Search Catalogue Get Data Help Tools Deposit My Account News

## Browse Archive

This is a cached view of the archive, some more recent changes may not be displayed. [View most recent listing.](#)

/ badc/

Download multiple files: \* Depth: 1 Go

240 dirs 0 files	Description	Size	Actions
<b>ARCHIVE_INFO</b>			
<b>CDs</b>	Copies of datasets distributed via CDROM		
<b>abacus</b>	ABACUS Work Package 1: Plants Data using plant harvesting equipment in Abisko (Sweden) and Kevo (Finland)		
<b>accacia</b>	Aerosol-Cloud Coupling And Climate Interactions in the Arctic (ACCACIA) Measurement Campaign		

**Centre for Environmental Data Analysis**

### CEDA Satellite Data Finder

Collapsable Header  
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

1373052 hits with current selection.  
1051 milliseconds for response.

Map Satellite

Rectangle search: [ ]

Key:

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

Map data © 2015 Google, INEGI Terms of Use

**EUFAR** **Centre for Environmental Data Analysis**

### 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 Tutorial

Choose an index:

EUFAR FAAM ARSF

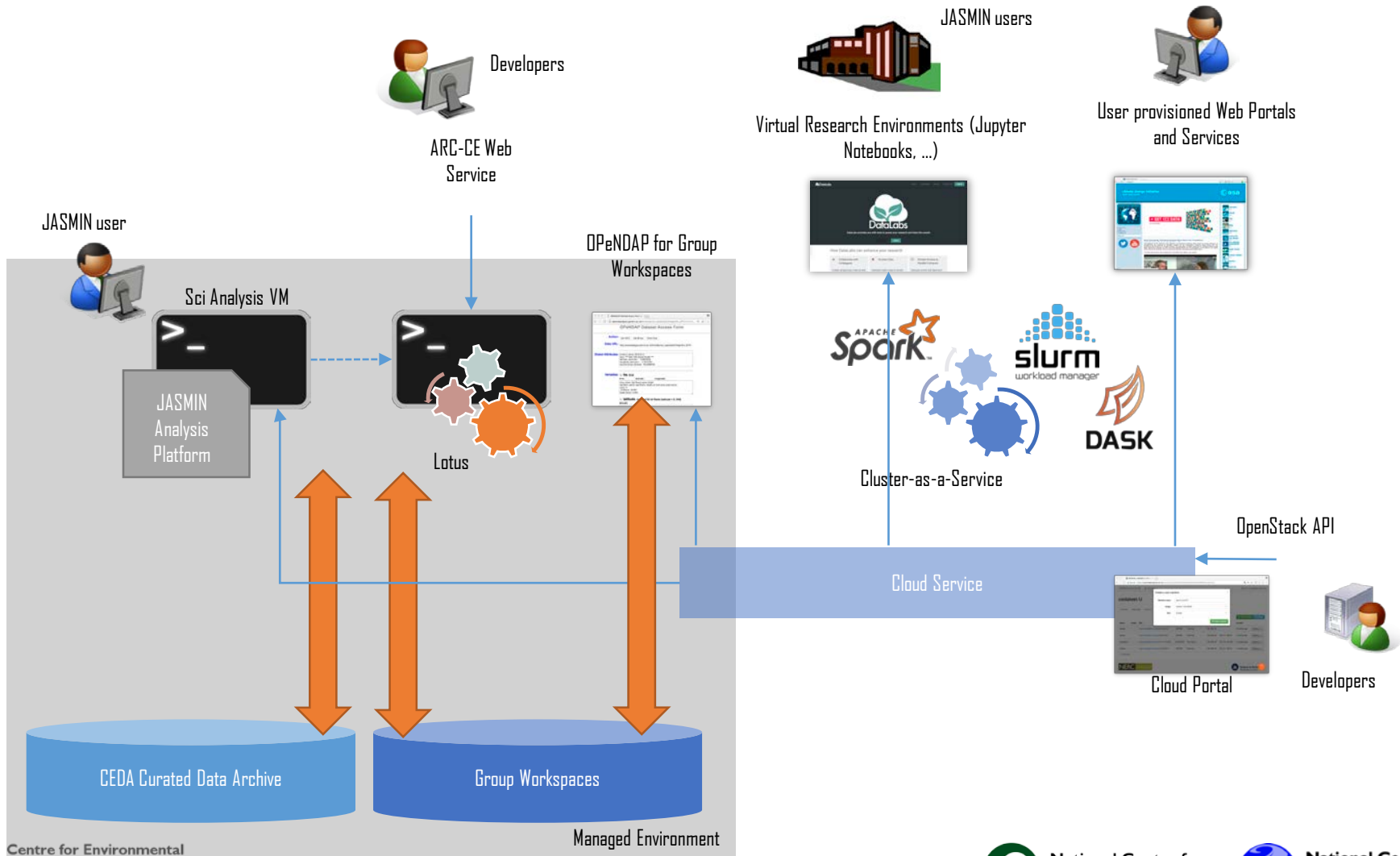
Map Satellite

Map data © 2015 Google, INEGI Terms of Use



# Solutions: Service Evolution

# Solutions: Access Patterns



# Summary

---

- JASMIN as a platform provides a variety of options to use and process data
- Elasticsearch underpins our move to object storage and building a register of the archive
- JASMIN evolves as the needs of our user community evolves and more technologies become available.