



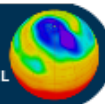
# NERC Big Data

## And what's in it for NCEO?

June 2014

Victoria Bennett

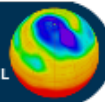
CEDA  
(Centre for Environmental  
Data Archival)



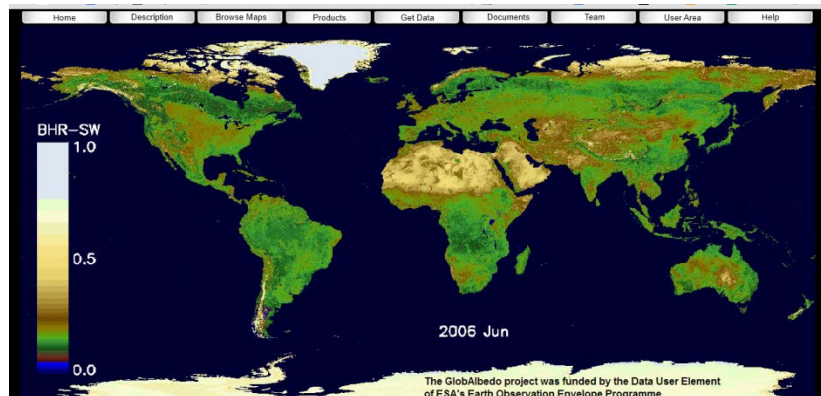


# Outline

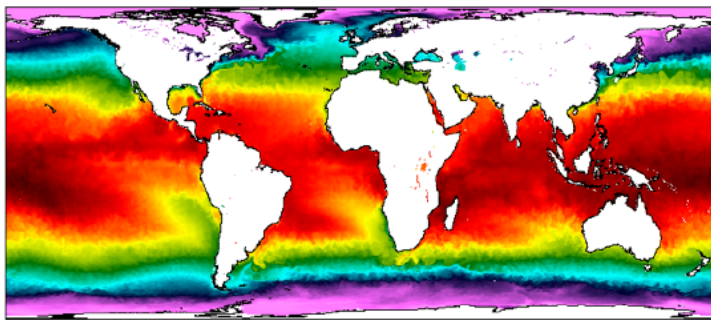
- CEDA and EO Data evolution
- NERC Big Data
- NERC's Big Data Facilities
  - JASMIN and CEMS



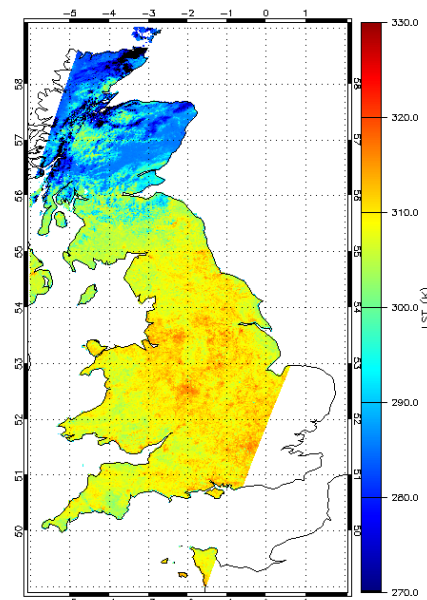
# UK Earth Observation scientists use super-data-cluster for Big Data processing and analysis



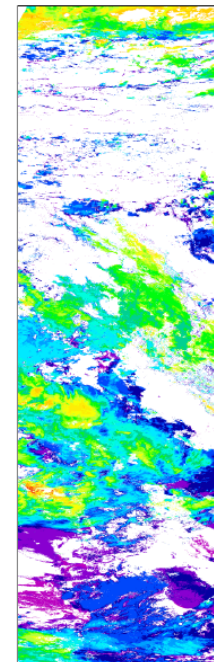
GlobAlbedo (UCL)



CCI SST (Reading)



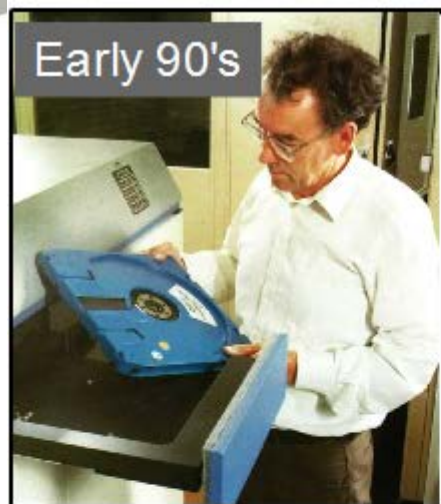
LST (Leicester)



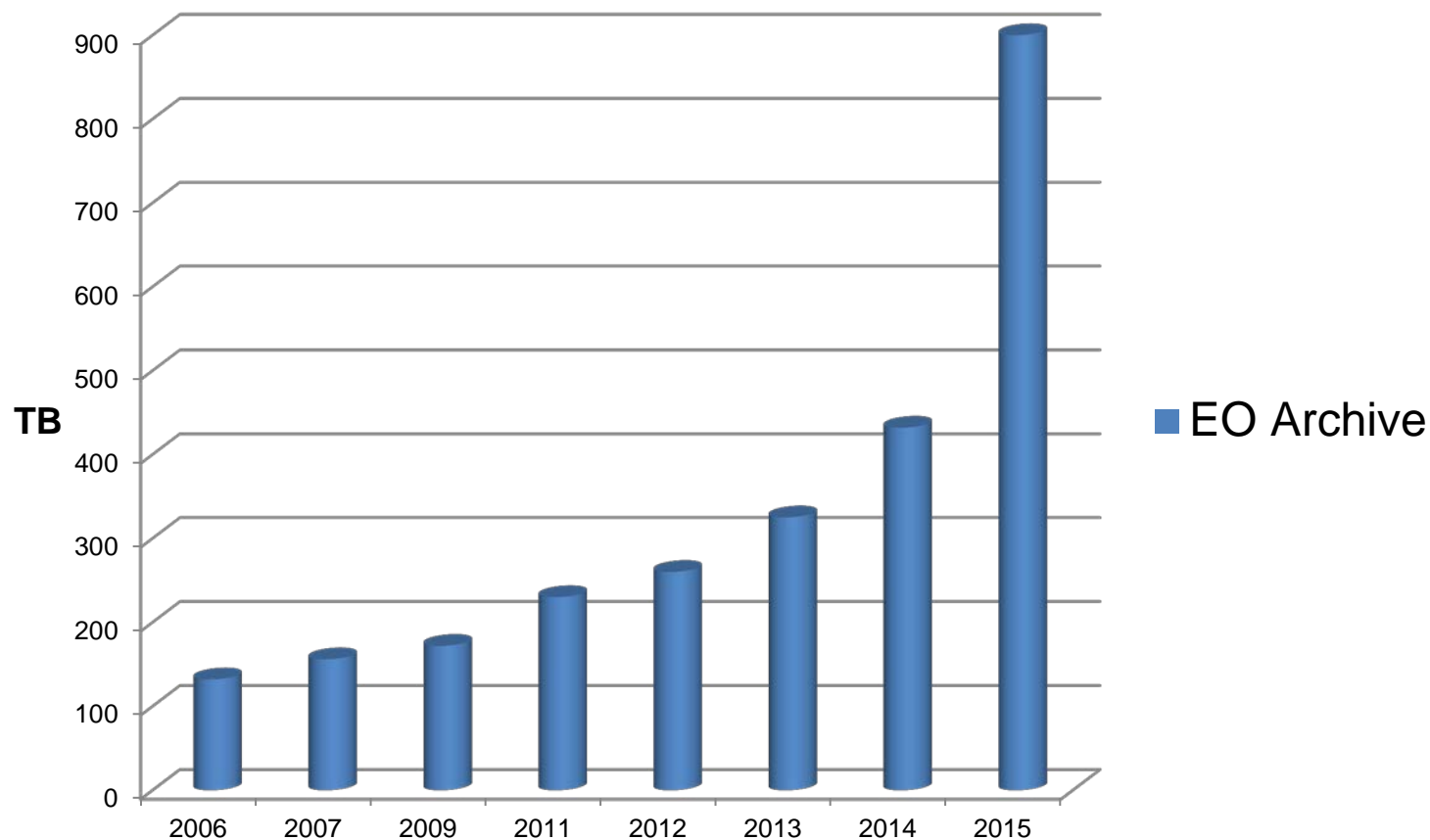
CCI Cloud (RAL)



# CEDA Evolution



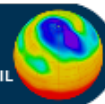
# EO Data Volumes: CEDA





# Big EO Data

- Datasets are getting bigger
- AATSR Level 1 + Level 2
  - Day: 15 GB
  - Year : ~5.5 TB
- Sentinel-3A core products (land + marine) L1+L2
  - Day: 2170 GB
  - Sentinel 3: ~790 TB
    - .... 172,000 DVDs...
    - And there's Sentinel 1-A/B, 2-A/B, .. etc



# NERC Big Data

---

NERC Environmental Big Data;

BIS allocated £13m capital funding to support 'Big Data'

Between 2013-2015 NERC is investing in:

- Compute and storage capacities of JASMIN
- Development of the academic component of CEMS
- Cloud-based software infrastructure to support environmental science (NERC Environmental Workbench)
- Environmental Big Data capital assets across the research community:
  - New digital assets, equipment for new data, processing and storage hardware, software to share, explore and visualise data

<http://www.nerc.ac.uk/funding/available/nationalcapability/envinfo/>



# Access: NERC Data Centres

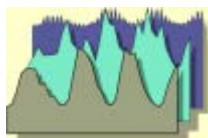


**British Atmospheric  
Data Centre**

NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE  
NATURAL ENVIRONMENT RESEARCH COUNCIL



**NERC Earth Observation  
Data Centre**



**Solar System Data  
Centre**



**British Oceanographic  
Data Centre**



**National Geoscience Data  
Centre**



**Polar Data Centre**



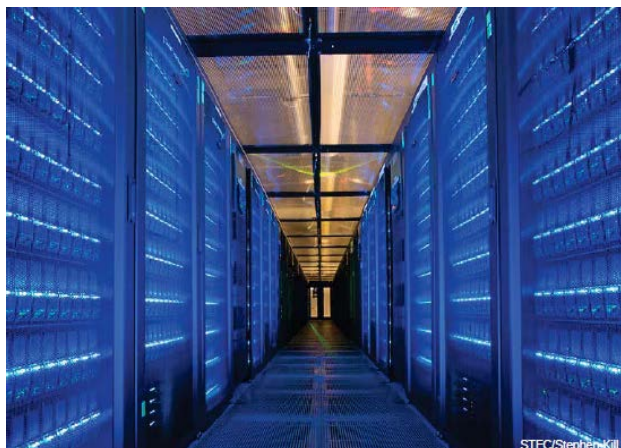
**Environmental  
Information Data Centre**

**Further Information & data discovery service:**

<http://www.nerc.ac.uk/research/sites/data/>



# JASMIN & CEMS: Big Data Facilities



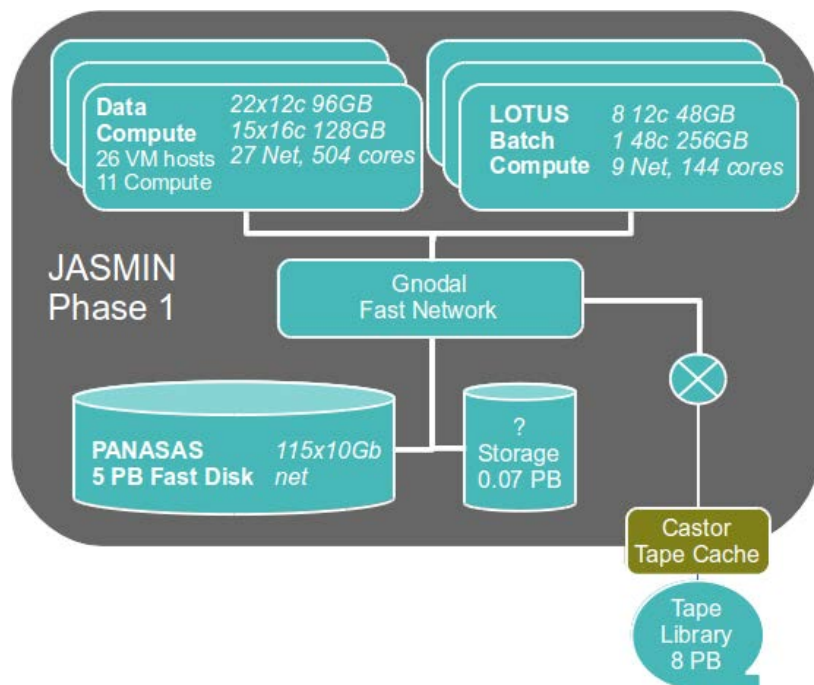
- JASMIN (super data cluster)
  - storage & services (CEDA)
  - scientific computation
  - access to high volume & complex data



- CEMS facility – Climate and Environmental Monitoring from Space



# JASMIN-1

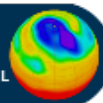


- JASMIN is configured as a storage and analysis environment
- Two types of compute:
  - a virtual/cloud environment, configured for flexibility
  - a batch compute environment, configured for performance
- Both sets of compute connected to 5 PB of parallel fast disk

GWS

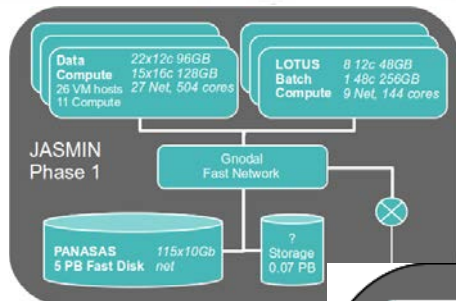
Lotus

Bespoke VMs

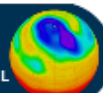
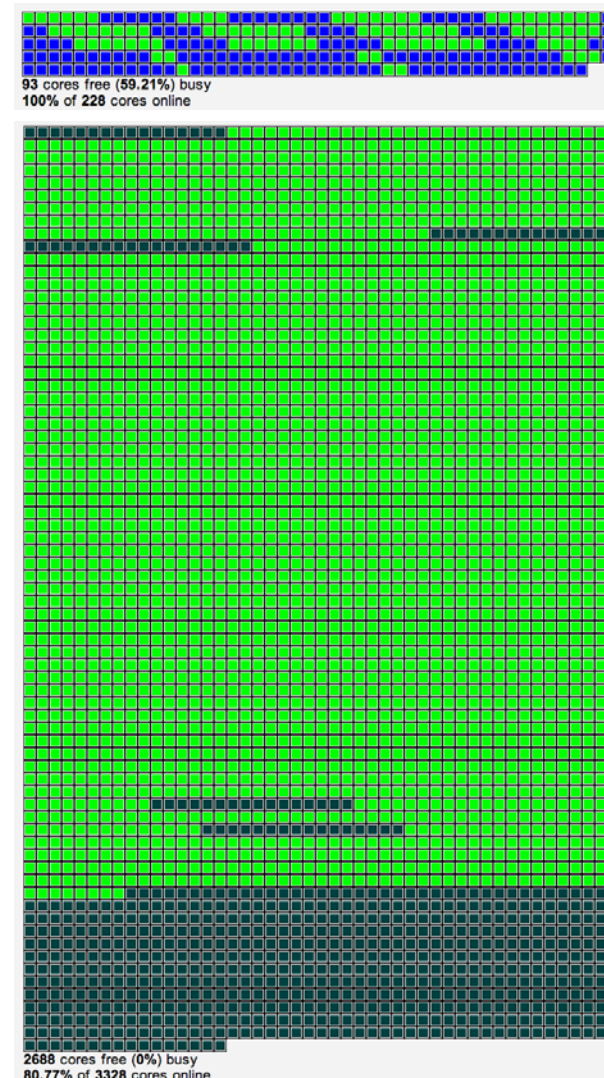
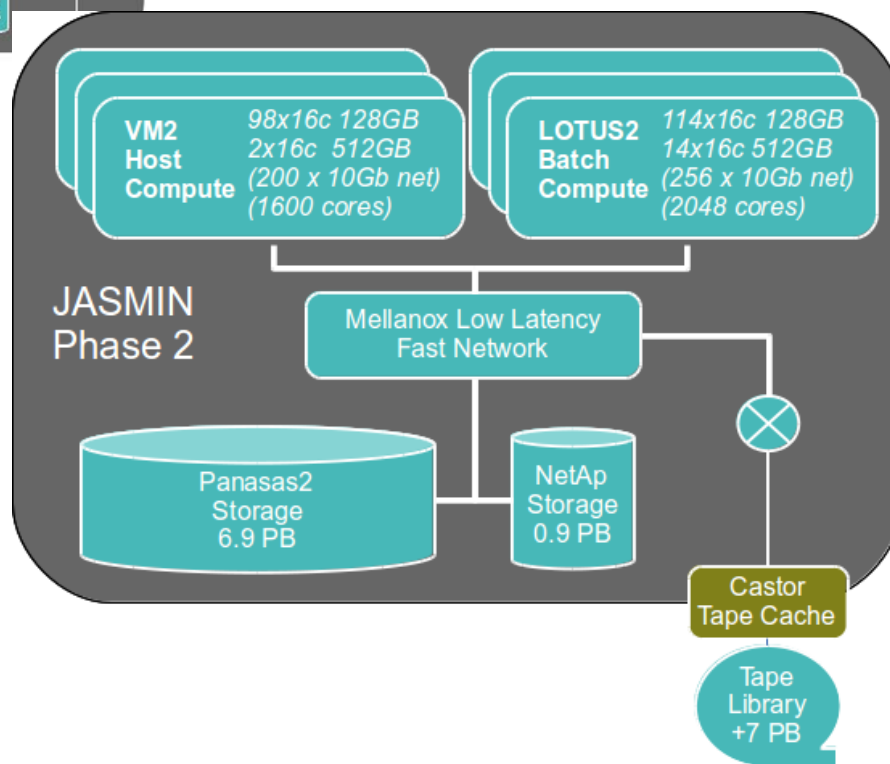




# JASMIN-2



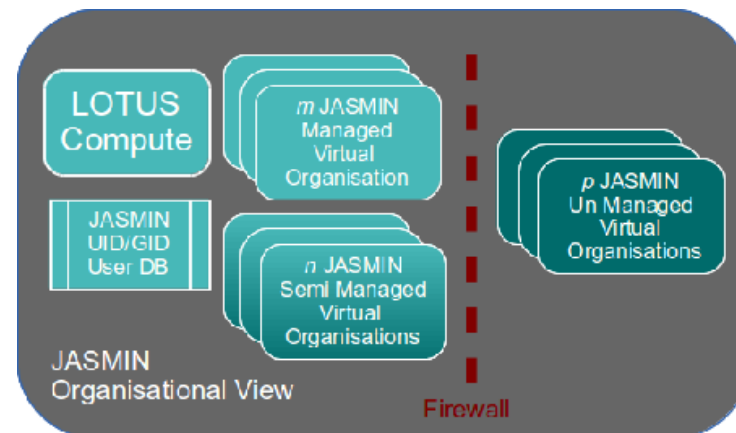
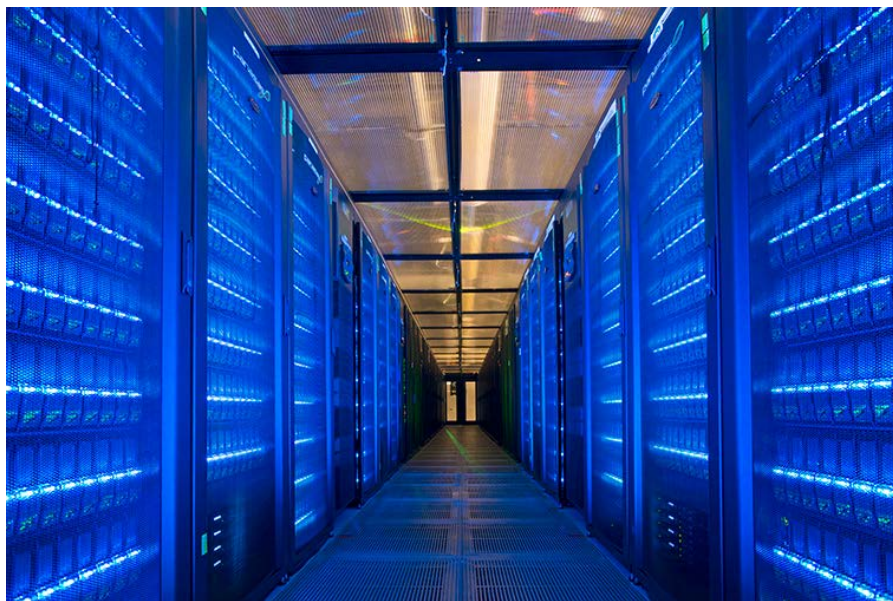
5+7+1=13 PB Disk  
Batch Compute  
Host Compute





# JASMIN-2

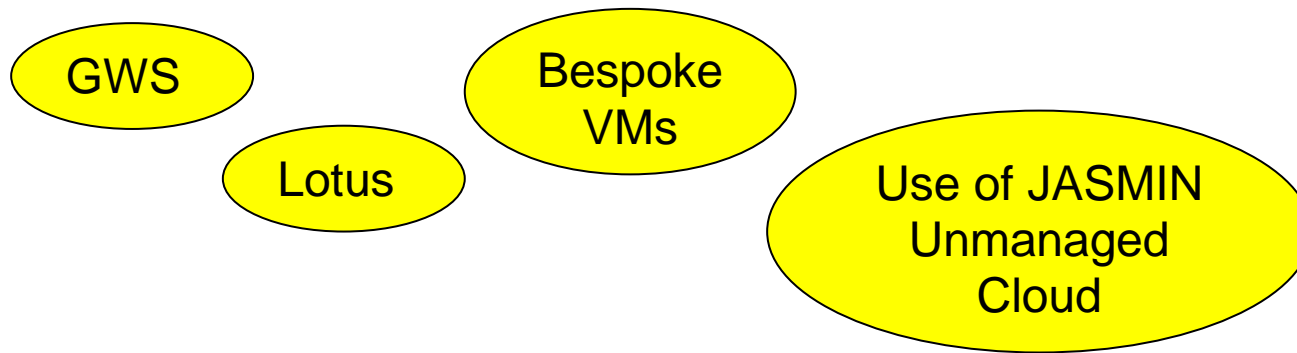
- Supporting NERC-wide science
  - NERC community and Met Office
- Virtual Organisations
  - “Managed” and “Un-managed” Cloud



- **NCEO's Academic CEMS** is a Virtual Organisation on JASMIN
  - Data
  - Services
  - Link to Sat Apps Catapult



# Using CEMS on JASMIN-2



- JASMIN is carved up into consortia for different areas of NERC science
- Consortium managers are responsible for approving resource requests
  - Similar process for NERC HPC allocation
- “EO and Climate Services” is one of 8 consortia

# Who is using CEMS on JASMIN?

- CEDA/NCEO Data Centre
  - Long term curation and dissemination of NCEO datasets
  - Third party datasets needed by science community
    - Please complete our survey!
- NCEO projects
- ESA and EC projects in NCEO community

Processing, storage, analysis and dissemination of EO Big Data: typically global long term environmental data from satellites

Academic CEMS Usage (June 2014)	
GWS	22 ; 1500 TB
VMs	48
Login users	71
Data download users	360 ; 130 TB (1 yr)

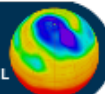
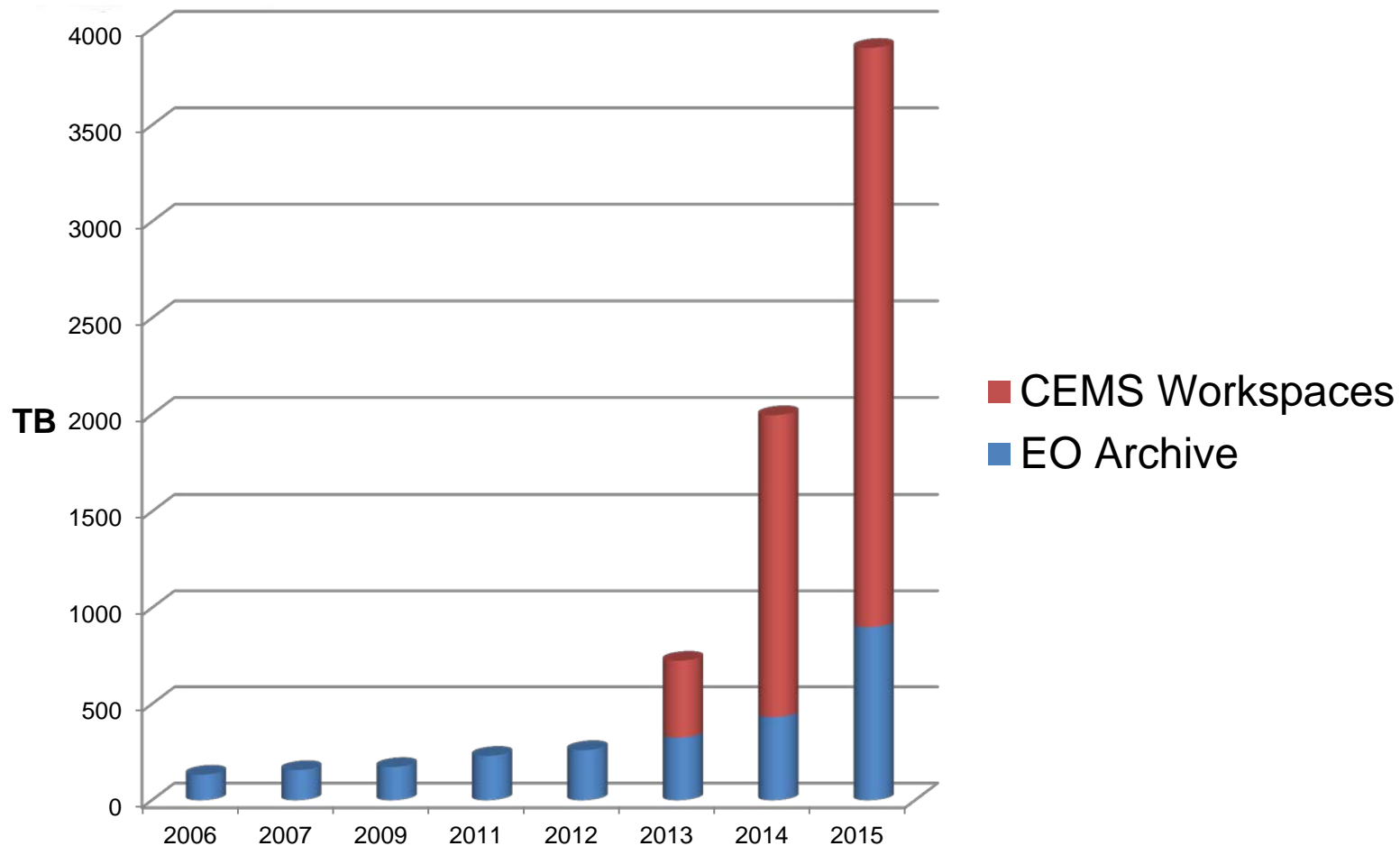
Talks/posters at this conference: 7



February 2014: 1,000,000<sup>th</sup> job run on Lotus  
Said Kharbouche, UCL, GlobAlbedo project



# EO Data Volumes: CEDA and CEMS





# Why use CEMS on JASMIN?

- Storage
- Processing
- Fast I/O\*
- Data: CMIP5 archive (>1 PB), CEDA archives (> 1PB) – BADC, NEODC all on the same hardware : *SCIENCE*
- Satellite Applications Catapult Link: innovative applications, commercial services, exploitation of research data products, collaboration opportunities : *IMPACT*



**CATAPULT**  
Satellite Applications

\* #1 in the world for I/O performance?

# Sat Apps Data Discovery Hub

**Data Discovery Hub**

Log in Register

Datasets Publishers About

/ Datasets

Clear

nceo

6 datasets found for "nceo" Order by: Relevance ▼

**NCEO atmospheric profiles retrieved from the Michelson Interferometer for Pas...**

This dataset contains atmospheric profiles from the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) on Envisat, processed with the Oxford L2 retrieval...

**Phytoplankton Size Class**

This dataset contains a 10 year monthly climatology and monthly composites of the fractional contributions of three phytoplankton size classes (micro-, nano- and picoplankton)...

**SEVIRI (Spinning Enhanced Visible and InfraRed Imager) Fire Radiative Power (...)**

This dataset contains Fire Radiative Power (FRP) data over Africa from Meteosat Second Generation (MSG) Spinning Enhanced Visible and InfraRed Imager (SEVIRI) data. Fires are...

**Vertical Profiles of Ozone and other Trace Gases Data from the Global Ozone M...**

Ozone profiles derived by the remote sensing group at National Centre for Earth Observation (NCEO, at the Rutherford Appleton Laboratory, Oxfordshire, UK) from radiances...

Map data © OpenStreetMap contributors  
Tiles by MapQuest

Date of Collection Clear

Topic Category

- Environment (6)
- Imagery / Base Maps... (4)
- Climatology / Meteo... (2)

Show More Topic Category

Data Format

There are no Data Format that match this search

Data Cost and Access

Public - Free with ... (4)



# Thanks for your attention

[victoria.bennett@stfc.ac.uk](mailto:victoria.bennett@stfc.ac.uk)

