CEMS: Building a Cloud-Based Infrastructure to Support Climate and Environmental Data Services

Philip Kershaw [philip.kershaw@stfc.ac.uk] (1), Mark Curtis (2), Ed Pechorro (3),
(1) STFC Rutherford Appleton Laboratory, NCEO / Centre for Environmental Data Archival, Didcot, United Kingdom, (2) Astrium GEO-Information Services Ltd., Farnborough, United Kingdom, (3) Logica Ltd., Leatherhead, United Kingdom

The facility for Climate and Environmental Monitoring from Space

CEMS, the facility for Climate and Environmental Monitoring from Space, is a new joint collaboration between academia and industry to bring together their collective expertise to support research into climate change and provide a catalyst for growth in related Earth Observation (EO) technologies and services in the commercial sector. A £3m investment by the UK Space Agency (UKSA) has made possible the development of a joint facility including **ISIC**, the **International Space Innovation Centre** and **STFC Rutherford Appleton Laboratory** at Harwell in the UK.

ISIC Building on the Harwell Campus

Key Goals

The vision has evolved and developed through collaboration work between the partners and set out in a Requirements Analysis report[1] and a business case set to the UKSA. Three main areas can be identified:

1) The provision of access to large-volume EO and climate model datasets co-located with high performance computing facilities;
2) A flexible infrastructure to support the needs of research projects in the academic community and new business opportunities for commercial companies.
3) Expertise and tools for **scientific data quality** and integrity giving users confidence and transparency in its data, services and products.

Phase 1 Project

5 month project started in December 2011 with an initial consortium including **Astrium-GEO**, **Logica**, **NCEO Reading** and **RAL Space**. This is tasked with:

- Purchase and deployment of the physical infrastructure;
- Design and preliminary development of supporting software system;
- A research study on Data integrity conducted by **VEGA Space**;
- The development of software demonstrators to explore science and business use cases

Use Cases

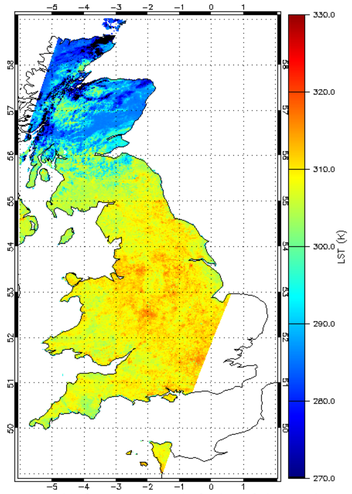
Science applications of CEMS have been identified in the following areas:

1) Operational or near-operational processing of satellite data;
2) Bringing multi-scale, multi-instrument data together for intercomparison and validation;
3) EO system design – observing system simulation experiments (OSSEs).

The ESA *Climate Change Initiative* is an important driver with many research institutes requiring processing capability for the generation of ECV (Essential Climate Variables) products. The benefits of CEMS in all of these use cases is the availability of significant processing capability close to the source data, removing the need to hold local archives at user sites (duplication of storage), and reducing overheads in data transfer. These advantages will become increasingly important into the future, considering the expected o20 increase in data volumes for the Sentinel instruments compared with the equivalent Envisat instruments.

Business applications: CEMS aims to provide an environment to stimulate innovation and growth in downstream applications. Examples included:

- Insurance sector: offshore assets and urban development in floodplains are susceptible to extreme weather events. Accurate scenario modelling is required on which to base premiums.
- Carbon Trading is another example market. EO data can be exploited for large scale analysis of carbon stock/flux.



ATSR Land Surface Temperature plot for the UK [John Remedios and Darren Ghent, University of Leicester]. Generation of LST products is one use case being explored for CEMS

Layered Architecture

The CEMS architecture is organised into layers. At the base, the underlying hardware is abstracted through to a virtualisation layer above. The latter enables consumption of resources by third parties via a cloud. The virtualisation layer also hosts the core system which manages the cloud services alongside the data management and curation functions of a data centre. At the top, an application and service provider layer provides interfaces to external user communities and partner organisations.

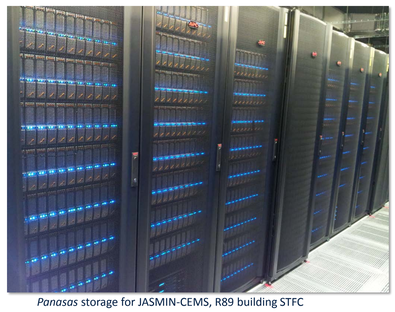
CEMS Cloud

Central to the development of this infrastructure is the utilisation of cloud-based technology: multi-tenancy and the dynamic provision of resources are key characteristics to exploit in order to support the range of organisations using the facilities and their varied use cases.

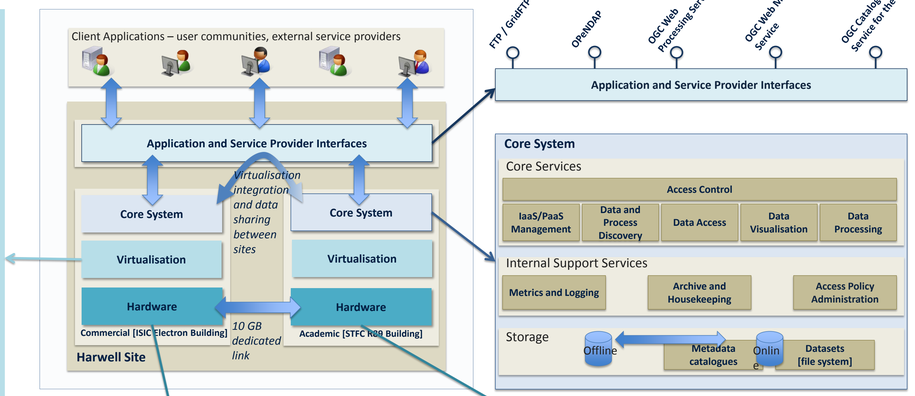
These characteristics will allow different research groups to set up and tear down portions of the storage, network and compute infrastructure tailored to their needs without the need for the upfront capital for hardware purchase.

It is expected that CEMS will support a spectrum of cloud models, from IaaS (Infrastructure as a Service) giving users the ability configure a complete custom environment to a SaaS (Software as a Service) model, exposing a library of useful processing algorithms through WPS. In addition, PaaS (Platform as a Service) based interfaces are envisaged to enable users to experiment, building applications and developing new processing algorithms and providing environments to better interact with the data.

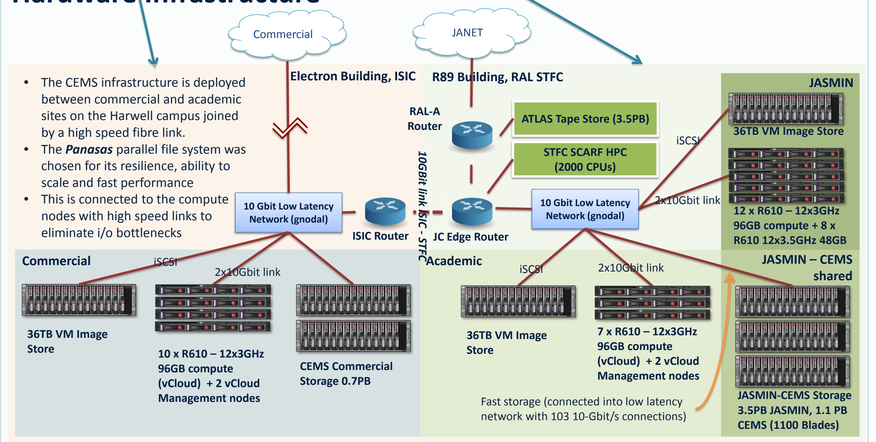
VMware vCloud has been selected to provide the software for virtualisation following a study conducted by Logica. Open source alternatives may be investigated in the future to augment this.



Panasas storage for JASMIN-CEMS, R89 building STFC



Hardware Infrastructure



JASMIN[2] is closely coupled to the CEMS infrastructure. Where CEMS focuses on the EO community, JASMIN is being deployed on behalf of the NCAS (National Centre for Atmospheric Science) to support the data analysis requirements of the UK and European climate and earth system modelling community.

References

[1] A User Requirements Analysis on a Facility for Climate and Environmental Monitoring from Space (CEMS), proposal to the Technology Strategy Board, Issue v1.0, 17th August 2011, Logica, NCEO, Astrium-GEO, RAL Space
[2] B.N. Lawrence⋆†§, V. Bennett†¶, J. Churchill‡, M. Juckes†§, P. Kershaw†¶, P. Oliver‡, M. Pritchard†§¶ and A. Stephens†§, **The JASMIN Super-data-cluster**, ⋆Department of Meteorology, University of Reading, Reading, U.K. †Centre for Environmental Data Archival, STFC Rutherford Appleton Laboratory, Didcot, U.K. ‡E-Science Department, STFC Rutherford Appleton Laboratory, Didcot, U.K. §National Centre for Atmospheric Science ¶National Centre for Earth Observation, Super Computing 2012 (submitted)