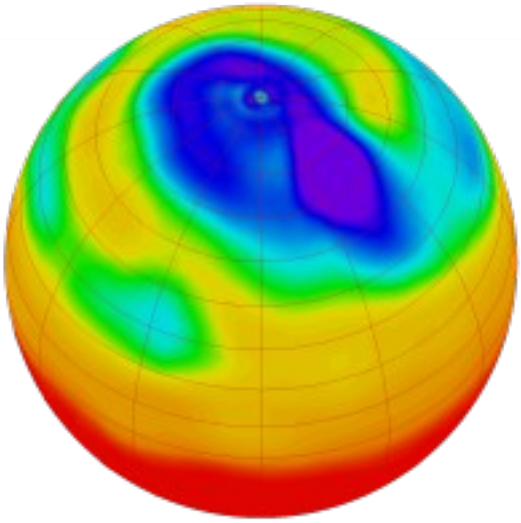
Science & Technology Facilities Council



STFC
Centre for Environmental Data Archival (CEDA)
Annual Report
2010
(April 2009-March 2010)

CEDA delivers the
British Atmospheric Data Centre
for the National Centre for Atmospheric Science
and the
NERC Earth Observation Data Centre
for the National Centre for Earth Observation
and the
IPCC Data Distribution Centre
for the IPCC

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Introduction from the Director


Bryan Lawrence

The mission of the Centre for Environmental Archival is to deliver long term curation of scientifically important environmental data at the same time as facilitating the use of data by the environmental science community.

CEDA was established by the amalgamation of the activities of two of the Natural Environment Research Council (NERC) designated data centres: the British Atmospheric Data Centre, and the NERC Earth Observation Data Centre. The process began with administrative functions (in 2005) and has proceeded steadily since, as new activities have and continue to be accreted into CEDA. Until 2008, the constituent parts of CEDA reported independently to NERC, but in 2009 we produced the first public report for CEDA. We are pleased to present here in our second annual report, evidence of the range of activities undertaken in CEDA, from methodical management of data, through innovative software development, to a close engagement with the scientific community – gathering and delivering the requirements of 21st century digital environmental science.

Digital curation involves ensuring that data remains both safe and fit for purpose. For most of us it is intangible, we're not really aware of it except for when things go wrong. We have considerable experience of personal curation failings: we all know someone (or have had it happen to us) who has suffered a hard-drive failure, but just as important is recognising the inability to understand the contents of a spreadsheet we created a few months ago as a curation failure resulting in unneeded angst and repetition of effort. Hopefully, we have less experience of service curation failure, even as we rely on remotely managed data (some of which we create, some of which we consume). From banks to insurance companies, from facebook to flickr, as our world of digital data expands, so too does our dependency. All service providers carry out curation; a delicate balancing act between preserving the bits and bytes, and evolving the way those bits and bytes are stored and understood to support new products and services. So too in science, the journal article is no longer the only important product of science – more and more of the underlying data is being preserved both as evidence for what was done, and as raw material for new and interesting syntheses of information. Digital curation is now part and parcel of doing science, ensuring we understand what was done and can repeat it, even as we have forgotten the doing of it.

CEDA plays its part in curation and evolution by endeavouring to preserve for posterity what has been done and facilitating new and interesting ways of exploiting data for the future. Such facilitation now includes both building new delivery services (which is what the users see) and new information services to ensure the data is always fit for purpose and available for reuse (the hard part of which most users never see). It also includes acquiring third party data as necessary and engaging with producers and users of data alike. In this, the second annual report of CEDA, we present some of our curation and facilitation accomplishments from the last year, beginning with a summary of important events and collaborations. We follow that with some selected highlights, and some statistical reports, including a financial summary. We then present our targets for next year. CEDA has done much, well, and of necessity; only some of what we have done is presented here, and we trust that you find something of interest.

Bryan Lawrence
Director

Summary

CEDA continues to support the atmospheric science community in the UK and abroad through the provision of data management and discovery services, and has continued to develop tools and services to aid data preservation, curation, discovery and visualisation.

In this year CEDA delivered in excess of 100 TB of data in 17 million files to nearly 3000 distinct users. CEDA also became "petascale", with in excess of 1 PB of spinning disk backed up by the Atlas Tape Store.

Major international collaborations built around two European projects,Metafor and IS-ENES, have been strengthened, even as an significantly larger global collaboration to deliver an "Earth System Grid Federation" to support the upcoming fifth Coupled Model Intercomparison Project (CMIP5) has been put together under the auspices of the Global Organisation for Earth System Science Portals (GO-ESSP).

In additional to the core remit of serving the Natural Environment Research Council's National Centres of Atmospheric Science and Earth Observations (NCAS and NCEO), CEDA has delivered major projects in support of both Defra and DECC[1] (providing the data systems for the UK Climate Projections 09 and IPCC[2] Data Distribution Centre respectively). A number of other projects with funding from a range of other bodies were also carried out, including work for the European Space Agency, the Joint Information Services Committee and others.

Notable Events

1. CEDA delivered the UK Climate Projections (UKCP09) user interface and underlying data systems, with a "go-live" in June 2009, which followed independent quality testing by IBM, who tested the systems to ensure that they could support up to 1000 simultaneous users (a number never even closely tested during the actual launch).

2. A science support information tool has been developed to allow BADC staff to more effectively scope and track the data management activities for the various NERC programmes. This tool partially automates the writing of data management plans as well as providing statistics to show how complete the data management activities are for a project. The tool is also being made available to other NERC data centres, providing the potential for better data centre interaction for joint projects, and hence a better user experience.

3. Mr Maurizio Nagni joined the CEDA staff as a software developer. His initial work has been on developing metadata systems to support the NERC data discovery service, the Marine Environment and Data Information Network (MEDIN), and CEDA itself. He has also been working on the installation and deployment of the BADC portal into the CMIP5 Earth System Grid Federation.

4. An improved collation of publications and reports available from the BADC has been produced, providing improved data documentation (both in terms of data usage and data production).

[1] UK Department of Energy and Climate Change
[2] Intergovernmental Panel on Climate Change

Major Collaborations

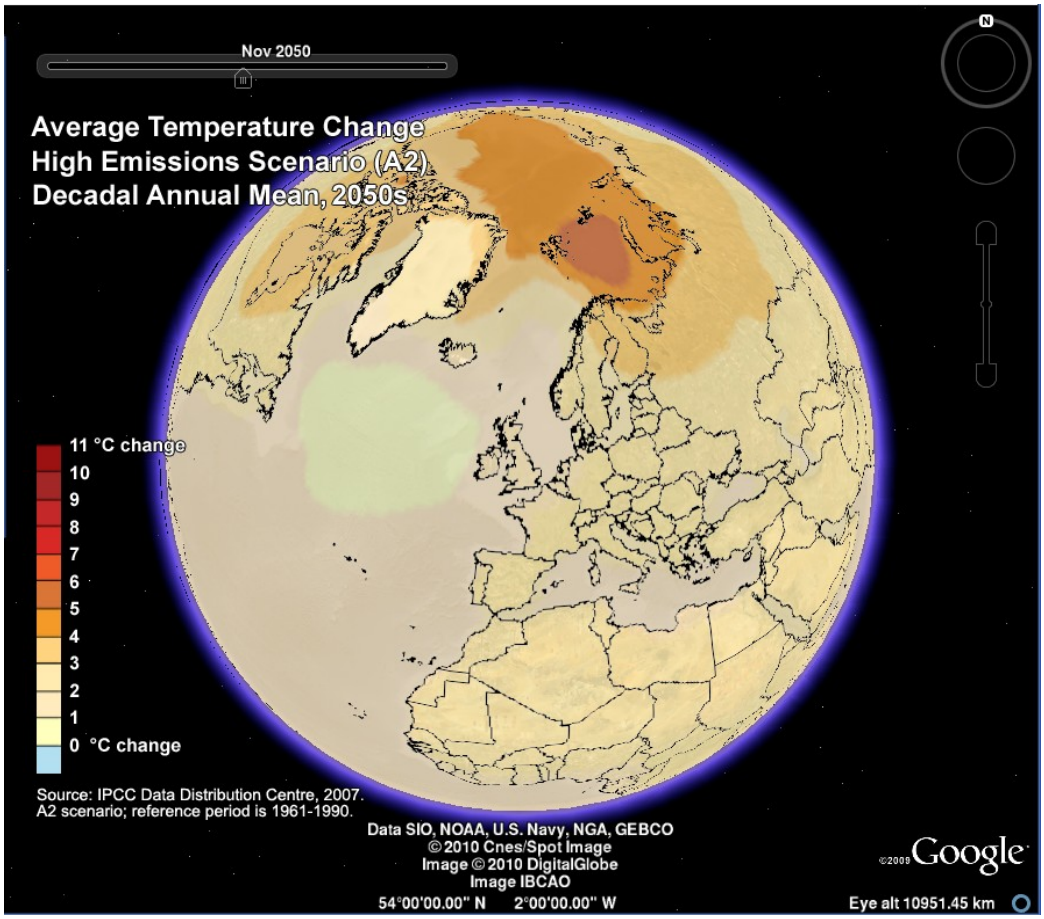
In 2009/2010, significant national and international collaborations have been continued and/or begun. On the national scale, CEDA itself reflects a collaboration between the earth observation community and the atmospheric sciences community (via NCEO and NCAS). Additionally, CEDA is:

- Working closely with the other NERC centres, under the auspices of the implementation plan for the NERC Science Information Strategy.
- Partnering with the UK Met Office on a number of projects, in particular, delivering the UK Climate Projections (UKCP09) User Interface and the Climate Impacts Link project (concentrating in 2009/2010 both on data delivery and putting systems in place to support Met Office involvement in CMIP5).
- Building the Earth System Grid Federation in partnership with the US Programme for Climate Model Diagnosis and Intercomparison and their US Earth System Grid partners (particularly those at NCAR[3] and GFDL[4]) on software to support the forthcoming fifth Coupled Model Intercomparison Project (CMIP5).
- Hosting the archive for the International Coupled Chemistry Climate Modelling Validation project (CCMVAL).
- A leading partner in two major European projects: Metafor (documenting climate codes and their resulting simulations to unprecedented levels of clarity) and IS-ENES (developing an InfraStructure for a European Network for Earth system Simulation).
- Using UK Department of Energy and Climate Change (DECC) funding to lead the delivery of the IPCC data distribution centre (http://www.ipcc-data.org in partnership with the DKRZ[5] hosted World Data Centre for Climate and Center for International Earth Science Information Network (CIESIN) at Columbia University).
- Working with the European Space Agency to extend earth observation metadata standards.
- Delivering a key role in evolution of the Climate Forecast NetCDF metadata conventions via standard name management.
- Providing data discovery services for the Marine Environment Data Information Network (MEDIN).
- Working with the wider UK atmospheric science and earth observation communities, via a range of projects, with NCAS and other NERC funding.

[3] National Center for Atmospheric Research
[4] Geophysical Fluid Dynamics Laboratory
[5] German Climate Computation Center

Highlight:Support for the climate model data user community (including the IPCC)

Martin Juckes, Charlotte Pascoe & Kevin Marsh



IPCC climate projections supplied by the IPCC Data Distribution Centre viewed through Google Earth

The IPCC Data Distribution Centre continued to provide data to a wide range of users. In addition to routine activity, data was provided to the Google Outreach team who had been engaged to provide support material for the Copenhagen Climate Conference (December, 2009). We provided Google Outreach with multi-model ensemble means of projected temperature and precipitation changes which they distributed as Google Earth files (Figure 1). During the conference the page describing the product received 3 million hits. In addition to the data, we provided an accompanying data description prepared in consultation with IPCC Secretariat and Bureau members.

The Defra-funded Climate Impacts LINK Project currently has 179 registered users of which 46 have been added in the past 12 months. During this period, over 10Tb of LINK data have been downloaded by users. The new systems developed under LINK to transfer data from the Hadley Centre to the BADC have also been used for UK Climate Projections model data (vital for the launch of the UKCP09 user interface) and associated detailed metadata records. A number of additional datasets have been requested by users and thanks to the systems developed under LINK, it has been possible to recover these data to the BADC very quickly - in some cases within a few days of the request being made. The LINK data archived at the BADC are also being used by the Hadley Centre themselves to make data easily available to their own researchers.

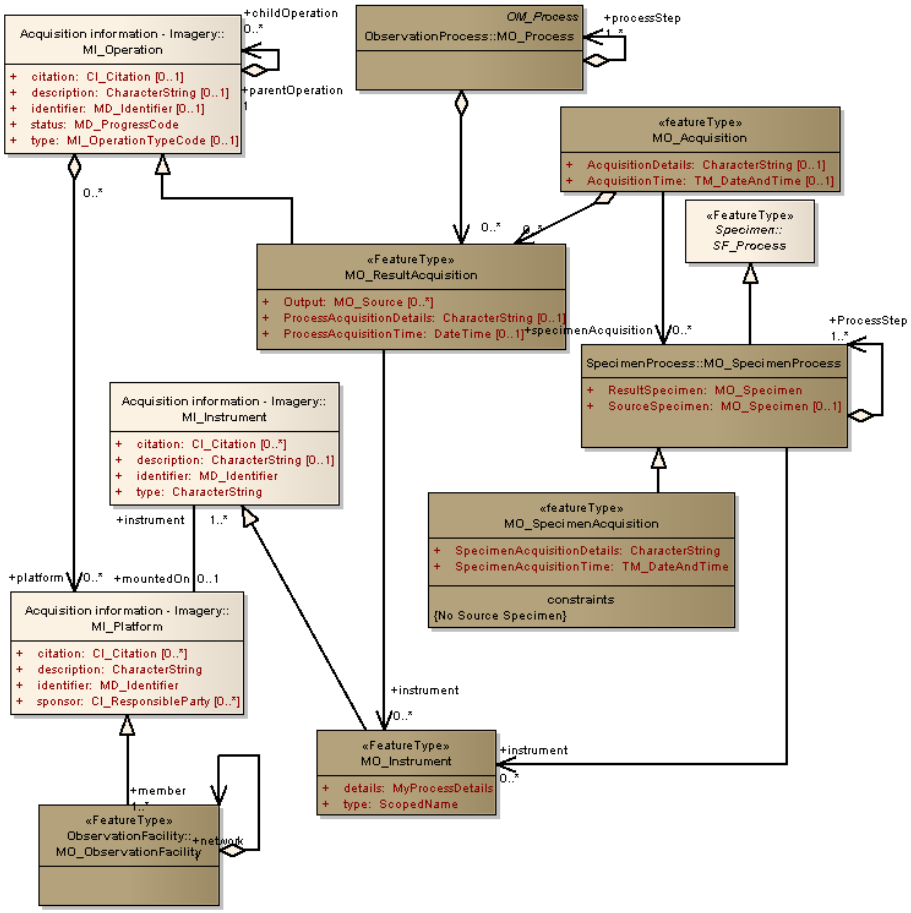
Highlight: Progress with MOLES

Spiros Ventouras, Bryan Lawrence, Andrew Woolf (STFC e-Science), Simon Cox (CSIRO, Australia)

The **M**etadata **O**bjects for **L**inking **E**nvironmental **S**ciences (**MOLES**) model was developed within the NERC DataGrid project (Lawrence et.al., 2009) to fill a missing part of the 'metadata spectrum'. It is a framework to encode the relationships between the tools used to obtain data, the activities which organised their use, and the datasets produced. MOLES is primarily of use to consumers of data, especially in an interdisciplinary context, to allow the establishment of provenance, and the intercomparison of such information without recourse to discipline-specific metadata or private communications with the original investigators. MOLES is also of use to the custodians of data, providing an organising paradigm for storage.

A typical sequence of data capturing involves one or more projects under which a number of activities are undertaken, using appropriate tools and methods to produce the datasets. Following this typical sequence, the relevant metadata can be partitioned into the following main sections:

- Project section
- Activity section (including both observation acquisition and numerical computation)
- Observation section (metadata describing the methods used to obtained the data, the spatial and temporal sampling regime, quality etc.)
- Observation collection section (the data themselves). The structure of individual observation collections may be domain-specific. The MOLES model investigates the use of CSML (Climate Science Modelling Language) for atmospheric data.



MOLES classes associated with the initial acquisition of results and/or specimens (from a draft of V3.4 of MOLES). Darker boxes are MOLES classes, lighter boxes are those from the standards (in this case ISO19115-part2 and the draft ISO19156 proposal.

The model has been tested as a desk exercise by constructing object models for scenarios from various disciplines. In 2009/2010 V3.2 was released, and substantial further progress towards later revisions made.

Ventouras, S., B.N. Lawrence, S. Cox (2009) , The MOLES-v3 Information Model, Proc. of the Data Management Workshop , October, University of Cologne, Germany

Lawrence, B.N., R. Lowry, P. Miller, H. Snaith, and A. Woolf (2009): Information in environmental data grids. Phil. Trans. R. Soc. A, 367, 1003 - 1014. doi:10.1098/rsta.2008.0237.

Highlight: First steps to implementing the NERC Science Information Strategy

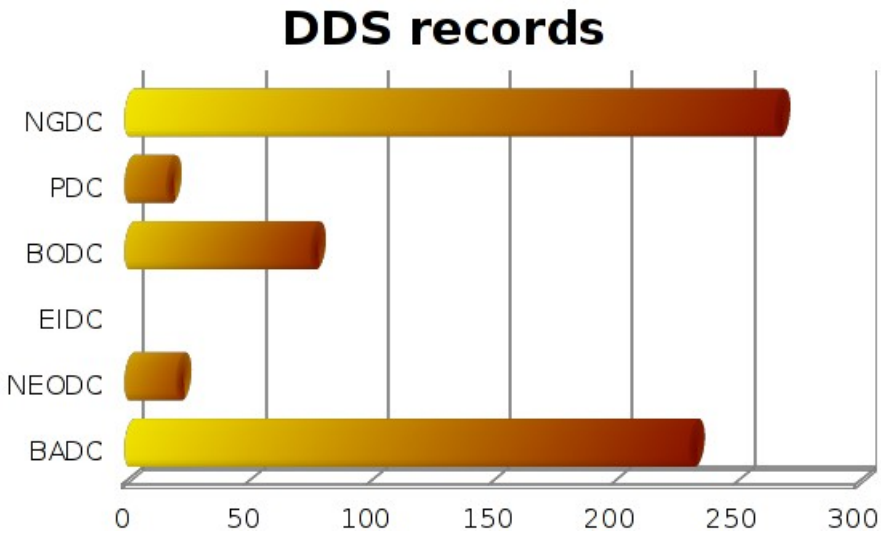
Sam Pepler

The NERC Science Information Strategy (SIS) has been created to provide the framework for NERC to work more closely and effectively with its scientific communities, both internal and external, in delivering data and information management services to support its 5 year science strategy, the Next Generation Science for Planet Earth (NGSPE). The NERC Executive Board approved the Science Information Strategy in July 2009 and the Implementation Plan in February 2010. The implementation programme is expected to run for 3 years.
The strategy focuses on the continuing development of NERC's information management processes and sets the context under which all of its science data and information activities, especially the operations of the data centres, will be carried out in the future. Emphasis is placed on managing and curating data with potential for reuse and on data that are of significant evidential benefit to the research record. The strategy applies to both digital and analogue information but with an initial focus on managing digital data.

The objectives of the SIS are to:

- Actively facilitate closer co-operation between NERC's 6 Environmental Data Centres (EDCs)
- Ensure greater consistency of service provided to the customers and suppliers of the EDCs
- Enable more efficient use of resources within the EDCs, rationalising activities and infrastructure where appropriate

CEDA staff have contributed both to the strategy and the implementation plan. The projects that make up the programme start from April 2010. While staff in CEDA will be involved in all the projects, we have a particular interest in some of the phase 1 projects



Number of NERC records in the NERC data discovery service by data centre – not a very discriminating metric!

- Data Citation – This aims to tackle the problem of data citation and introduce mechanisms for data identification
- Data Centre Metrics – This project aims to define useful data centre metrics to monitor the effectiveness of data centres. The most obvious and simple metric is the number of datasets per data centre, (see the figure above) but this does not provide much useful information about effectiveness and/or efficiency.
- Grant Information - Information contained in the grants system is the starting place for data management planning, but this is not accessed in an efficient or consistent manner. This project aims to resolve this.
- Data Value Checklist – Selection of data is not consistent across the data centres. A statement of data value will give credit to proposals that will produce valuable data.

Highlight: Results from CEDA Research Breaks.

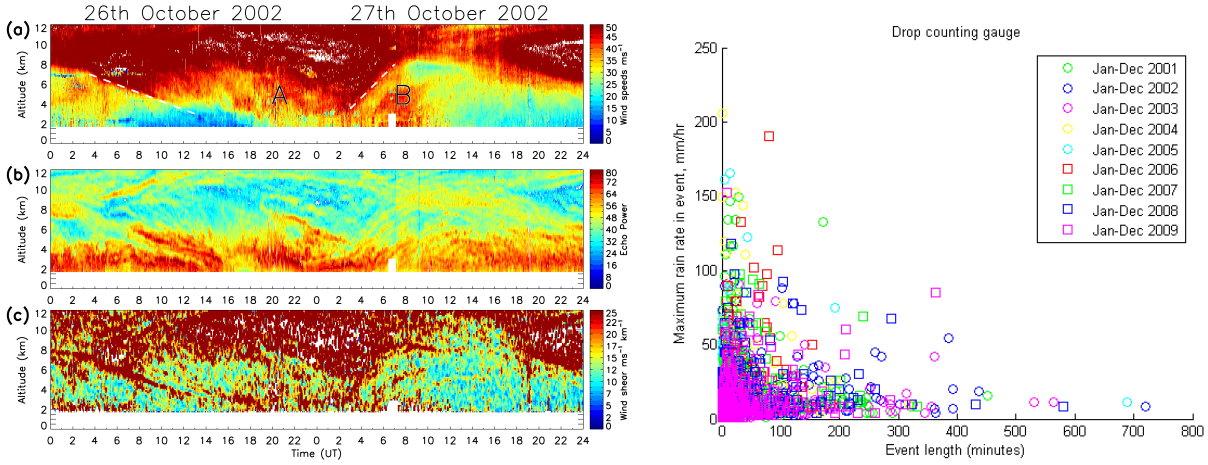
Graham Parton &Sarah Callaghan

In addition to other research activities within CEDA, a "research breaks" scheme was developed to enable CEDA team members to undertake short term, targeted research projects. These are important to ensure deliverable results from a dedicated period of research either conducted individually or in collaboration with external parties, and help to ensure that team members remain abreast with present research across the environmental sciences and enhance the research contribution to the wider science community by the group. During 2009-10 two research breaks were taken: one examining the origin of the most damaging winds to affect the UK (Graham Parton), while the second carried out an analysis of rain gauge data from the Chilbolton observatory (Sarah Callaghan).

Previous work by Parton et al (2009) provided observational evidence of mesoscale strong wind features such as Sting Jets resulting in damaging winds at the surface, demonstrating the potential importance of such mid-tropospheric wind events in wind storms at the surface. In 2009-10 Parton developed a mid-tropospheric wind climatology using the MST radar located near Aberystwyth. The resulting paper (Parton et al., 2010) presented a climatology from 7 years worth of high temporal resolution data, leading to an analysis of the most damaging events affecting the UK, including sting jets. Comparison with radiosonde data demonstrated the British Isle wide applicability of the work and this lead to collaboration on a second paper which used MST radar derived winds as an improvement for aircraft descent paths into East Midlands airport (Liling et al., 2010).

Callaghan analysed 9 consecutive years of drop counting rain gauge data collected at Chilbolton Observatory in Hampshire to investigate the fast fluctuations of rainfall rate over a long time period. Knowledge of the fine scale variability of rain (both in the spatial and temporal domains) – possible with the Chilbolton data - is important for the development of accurate models used in small-scale forecasting, implementation and operation of rain affected systems (e.g. microwave radio communications) and flood mitigation. Continuing rain gauge measurements at Chilbolton ensure that these datasets will become increasingly valuable, providing a "ground-truth" for comparisons with climate and other models.

Parton G.A., Dore A.J., Vaughan G. *A climatology of mid-tropospheric mesoscale strong wind events as observed by the MST radar, Aberystwyth*. Meteorological Applications. 2010, in press.
Ren L, Reynolds T.G., Clarke J-P. B., Hooper D.A., Parton G.A., Dore A.J. *Meteorological influences on the design of advanced aircraft approach procedures for reduced environmental impacts.* Meteoroogical Applications, 2010, in press.



Left Panel: MST radar time-height plots showing frontal surfaces, warm sector winds and sting jet (labelled B). Right Panel: Comparison between the maximum rain rate recorded in a given event and the duration of that event for the nine years of drop counting rain gauge data.

Highlight: Results from CEDA Research – Data Fusion.

Martin Juckes, Alan Iwi & Jamie Banks (AOPP, Dept. Physics, University of Oxford)

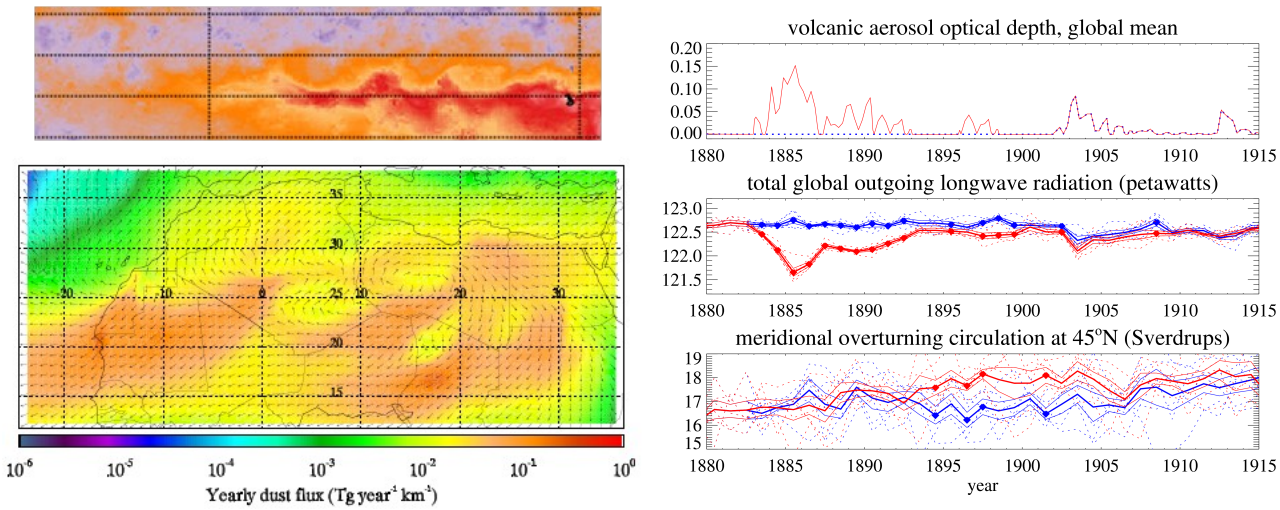
In addition to targeted, short research projects carried out under the CEDA research break scheme members of the CEDA have been involved with a number of other research projects, covering numerical weather prediction (NWP), climatology and data assimilation.

CEDA's data fusion software, which creates a continuous field by running a Kalman smoother over sparse data, has been applied to a re-analysis of sea surface temperature data generate by the (A)ATSR Reanalysis for Climate (ARC) project. The ability of the reanalysed data to capture small scale features such as tropical instability waves to the west of the Galapagos Islands is shown below. The optimal interpolation generated by the Kalman Smoother allows such features to be captured in the consolidated product without the loss of details which results from less sophisticated filling algorithms.

In March 2010 Jamie Banks, a PhD student from the University of Oxford co-supervised by Martin Juckes, submitted a PhD thesis on modelling of Saharan dust emissions and transport (Banks, 2010). The thesis provides new insight into the seasonal cycle of dust emissions and demonstrates how assimilation of limited satellite observations can greatly improve agreement between the transport model and independent observations. The work also shows there is still considerable uncertainty arising both from poor knowledge of the soil properties across the Sahara and from uncertainties in the near surface winds.

Work has continued during 2009-10 under the NERC-funded GCEP project in collaboration with Reading University. The effect of volcanic aerosol on the Atlantic Meridional Overturning Circulation (MOC) was modelled using the HadCM3 climate model. The figures (below, right) shows a comparison of results from ensembles forced with or without the volcanic eruptions of the late 19th century. Whilst the radiation responds rapidly to the volcanic aerosol, a lagged response in MOC is also seen. This research found that the MOC increase was due to a reduction in high-latitude precipitation, leading to saltier surface waters in the Greenland Sea, which in turn increased convection and southward flow at depth through the Denmark Straits.

Banks, J. PhD Thesis. *Modelling the Emission and Transport of Saharan Dust.* University of Oxford. Submitted. 2010.



Top Left: tropical instability waves in the Pacific to the west of the Galapagos islands (the small island group in the bottom right corner). Bottom Left: yearly dust fluxes across North Africa, along with prevailing directions of transport (March 2006 to February 2007). Right Panel: Imposed volcanic aerosol forcing and response in two ensembles of HadCM3: one with full reconstructed historical forcings, and one with the volcanic eruptions of the late 19th century removed (blue lines). Dots denote significant anomalies.

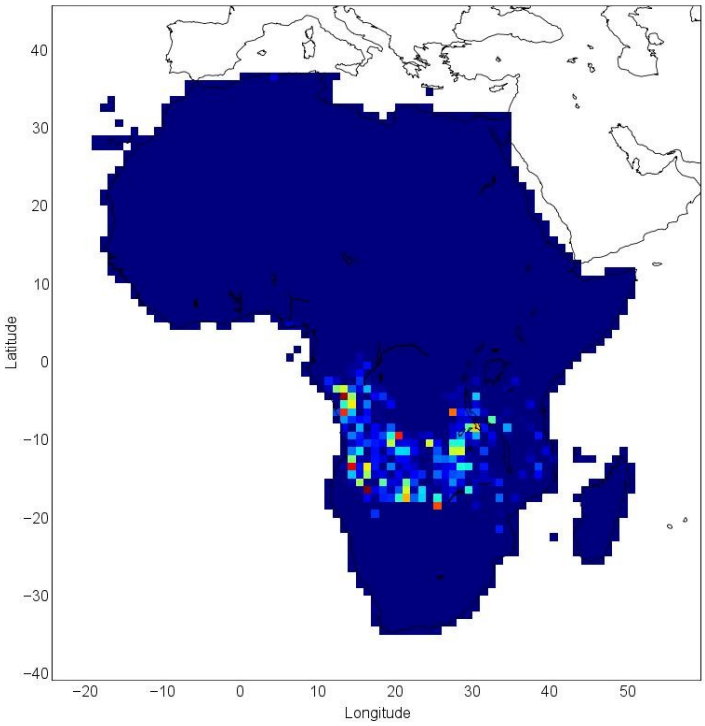
Highlight: Curating long-term Earth Observation Datasets

Victoria Bennett

The data generated from the activities within the National Centre for Earth Observation (NCEO) are a valuable resource, and as such require effective management so they remain an asset in years to come. In particular, the long term global data sets produced using Earth Observation are key to understanding and monitoring global change. The professional curation of this data ensures that the impact of NCEO's activities reaches a wider audience, enabling knowledge exchange between different disciplines, sectors and organisations.

Datasets already available though the NEODC and BADC data centres from NCEO Science Themes include products relating to global plankton and primary productivity, air-sea gas exchange, atmospheric profiles of ozone and other constituents, clouds and fire radiative power. These products have been derived from satellite datasets with algorithms developed by NCEO scientists, using data from a range of satellite instruments (SeaWiFS, TOPEX, GOME, MIPAS, SEVIRI, and the (A)ATSR series) on board ESA, NASA and Eumetsat satellites.

More are expected in the coming years, including long-term datasets of ice extent and ocean topography at high latitudes, global sea and land surface temperature, methane, volcanic emissions, aerosol and surface reflectance data. Model output from climate, numerical weather prediction and chemical transport models is also stored and disseminated to users in NCEO and NCAS.



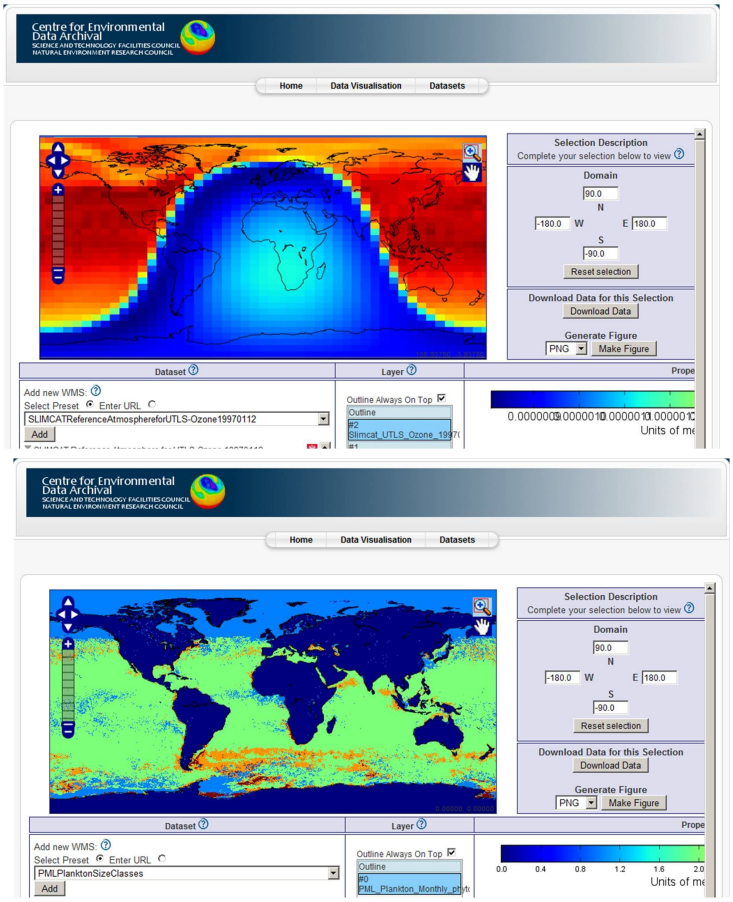
Fire Radiative Power data over Africa (1st August 2004) produced by NCEO researchers at King's College London from Meteosat Second Generation SEVIRI data. Plot produced using CEDA's web-based NCEO visualisation service.

The data and documentation can be discovered, searched and browsed via the data catalogue and, given appropriate data types and formats, can feed into visualisation and analysis tools. We have worked with NCEO data providers to ensure EO data products are provided in well-documented standardised formats (CF-compliant netCDF).

Highlight: Visualising Earth Observation data products

Victoria Bennett, Dominic Lowe, Phil Kershaw, Stephen Pascoe, Graham Parton & Peter Norton

Development work carried out at CEDA has allowed datasets produced by scientists in the National Centre for Earth Observation (NCEO) to be made available for web-based visualisation and download through standards-compliant web services.



Web-based interactive visualisation of NCEO datasets: Model data from the SLIMCAT/TOMCAT models produced at University of Leeds, and plankton products derived from satellite data (SeaWiFS) by researchers at Plymouth Marine Laboratory

The visualisation portal allows users to view datasets in the CEDA archives and overlay different views, e.g. different time steps or parameters in the same data, or different datasets. The graphics interface lets the user choose colour scales, data ranges, geographical area, contour style, etc, and produce and save an image for use in reports or publications. The data values themselves can also be downloaded directly via the interface. The tools have been developed to handle any geospatial data which is provided in the right formats, so can be scaled to handle many more datasets as they become available.

Producing the visualisation service for NCEO data has involved:

a) liaising with the data providers to ensure the datasets are provided to the data centres in the correct formats with appropriate metadata (CF-compliant netCDF)
b) designing and setting up a new web-based visualisation portal (http://ceda.ac.uk/nceo)
c) enabling security to be applied to restricted datasets (restricting access to registered users only), and linking this to the access control already in use in the BADC and NEODC data centres
d) procuring new hardware, for deployment of the system, and
e) writing software tools to automatically extract metadata from the files to feed into the visualisation layer.

The datasets are deployed behind OGC (Open Geospatial Consortium) standards compliant web services, which can feed into other Web Map Service (WMS) visualisation systems as required.

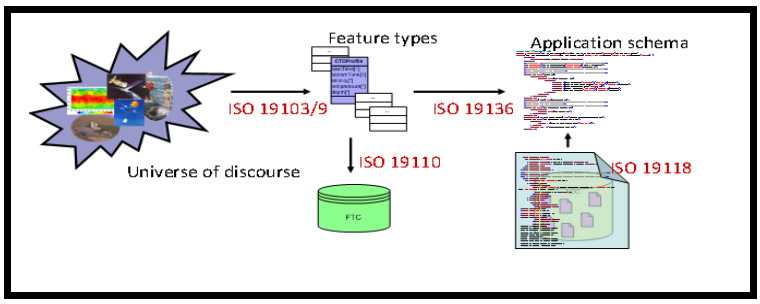
The NCEO visualisation service has also been used to display and make available the ERA-interim re-analysis dataset from the European Centre for Medium Range Weather Forecasting (ECMWF), to registered NCAS and NCEO users. The ERA-interim dataset, which is updated in the archive on a monthly basis, is now routinely scanned on ingestion to extract the metadata required for the visualisation service.

Highlight: Working with the European Space Agency to extend Earth Observation metadata standards - *The Heterogeneous Missions Accessibility Follow On (HMA-FO) project.*

Dominic Lowe, Victoria Bennett & Andrew Woolf (STFC e-Science).

The European Space Agency (ESA) Heterogeneous Missions Accessibility[6] (HMA) programme was developed to involve space-data stakeholders in a ground segment harmonisation and standardisation process for satellite data products. HMA aims to create an efficient information-based environment based on the concepts of data and service interoperability.

In the early stages of the HMA initiative ESA together with other GMES[7] (Global Monitoring for Environment and Security) participating agencies modelled the metadata of Earth Observation products as geographic features encoded in the OGC[8] Geography Markup Language (GML). CEDA in collaboration with STFC e-Science centre is now taking part in a next stage project, HMA Follow-On, in a consortium with two Belgian space data/GIS companies; Erdas and GIM.



Model Driven Architecture

Our role in this project is to build on our expertise in current metadata standards and data modelling (gained through NERC initiatives such as MOLES[9] and CSML[10]) along with our experience in data management to:

- Review and revise the existing GML EO metadata schema based on current international best practice such as the ISO Observations & Measurements[11] standard.
- Move to a Model Driven Architecture approach (such as that prescribed by INSPIRE[12]) for further schema development.
- Extend the current Earth Observation metadata to describe additional data products:
  - Radar Altimeter data
  - Limb Sounding data
- Co-author a new OGC standard for a GML profile for Earth Observation.

At the halfway stage of this project we have recently (May 2010) delivered a draft data model to ESA and the other stakeholders and are beginning work on the new OGC standard.

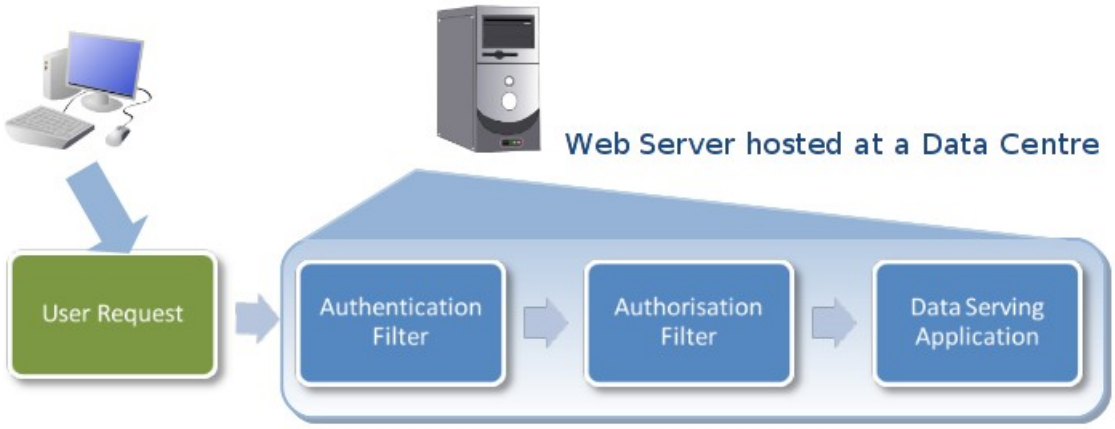
[6]HMA: http://wiki.services.eoportal.org/tiki-index.php?page=HMA-FO
[7]GMES: http://www.gmes.info/
[8]OGC: Open Geospatial Consortium, http://www.opengeospatial.org/
[9]MOLES: Metadata Objects for Linking Environmental Sciences, http://proj.badc.rl.ac.uk/moles
[10]CSML: Climate Science Modelling Language, http://csml.badc.rl.ac.uk/
[11]O&M: http://www.opengeospatial.org/standards/om
[12]INSPIRE: http://inspire.jrc.ec.europa.eu/

Highlight: Deploying NERC DataGrid Security for the CMIP5

Phil Kershaw

The NERC Data Grid was a federated data infrastructure that delivered a variety of data-related services With that ability to provide widespread access to data and services there alsocame a paradoxical need to restrict access to registered users, to support the licensing arrangements associated with third party data, to protect finite computing resources, to keep users up to date with changes to data and services, and to feedback usage statistics to data owners and sponsors. In a federated environment, this came down to a requirement to support "single-sign-on": that is the ability to use one username and password (preferably from a home institution) across other organisations in the federation.

The NDG security system was developed to meet these needs and this year has seen a number of developments and deployments. The software has been re-engineered into a modular architecture of pluggable components or filters which can be arranged into the desired configuration to filter access to any given data web-based data access service. This is illustrated in the diagram below, a user makes a request from a browser or other program on their desktop computer to a web application serving environmental data hosted at a data centre. The application at the data centre is configured with a set of security filters to intercept requests for data, allowing only authorised requests to reach through to the data service .



Data Provider protects a data service with NDG Security filters

Primary Applications in 2009/2010:

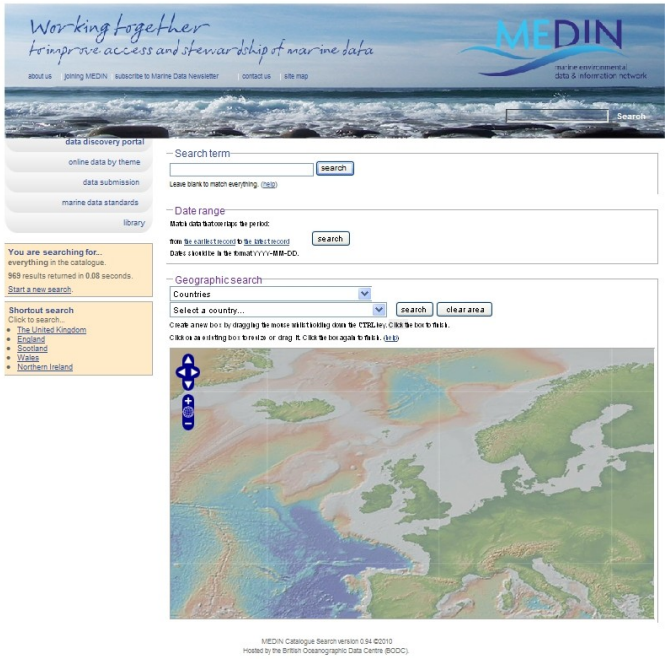
1. **Within CEDA:** Systems to support NDG security were deployed within the standard CEDA infrastructure, and in a number of CEDA applications, including the CEDA OGC Web Services (COWS) – a major application library.
2. **CMIP5:** A major collaboration with the US Argnonne National Laboratory and National Centre for Atmospheric Research (NCAR) under the auspices of the "Earth System Grid Federation" to support CMIP5, has resulted in deployment of the NDG filter approach for CMIP5. Initial tests with NCAR have shown the first step of interoperation: authentication interoperability, and that authentication is already built into the live CMIP5 metadata questionnaire.
3. **NetCDF:** Code to provide a client to NDG/ESGF security has been supplied to Unidata to absorb into their NetCDF libraries.
4. **OpenDAP:** This architecture has also been applied to OPeNDAP (Open-source Project for a Network Data Access Protocol) services. OPeNDAP is a data access service in widespread use across the atmospheric science and oceanography communities, and the PyDAP implementation has been instrumented with NDG security. Discussions with other implementers are under way.

Highlight: Generation of an updated Discovery Service API for the MEDIN portal

Steve Donnegan, Maurizio Nagni & Matt Pritchard

The Marine Environmental Data Information Network (MEDIN) wished to upgrade their Data Discovery Portal for users to find datasets related to data in the marine domain held by a number of MEDIN partner data centres. MEDIN previously hosted their own Discovery portal using the same database, web service interface and API as the current NERC Data Discovery Service. This was based on information extracted from older format Marine Data Information Partnership records (MDIP, precursor to MEDIN).
MEDIN have recently defined a new profile of the ISO19115/19139 metadata standard tailored towards MEDIN's explicit data requirements and they required that their new Discovery portal should be based on information extracted from the new metadata format. In addition to this MEDIN also required that the new portal should allow more intensive and "intelligent" searching of the database than that offered by the existing NERC Discovery Service API. This updated portal should allow searches to be layered and targeted towards specific fields within the database as well as allowing the user to construct advanced searches.

CEDA successfully bid for the work to develop and deploy an updated Discovery Service API to support the new MEDIN portal (the portal itself was to be developed by the GeoData Institute at the University of Southampton). CEDA defined the behaviour of the new API, not only ensuring that it matched the MEDIN requirements but also that it accommodated likely future requirements of the NERC Discovery Service. CEDA worked with the GeoData developers to ensure that the API used by the MEDIN portal met the requirements, that all software bugs had been fixed and that the new API operations met the original MEDIN specifications. This API is undergoing the transition to operational service for the MEDIN portal.



Screenshot of the MEDIN portal

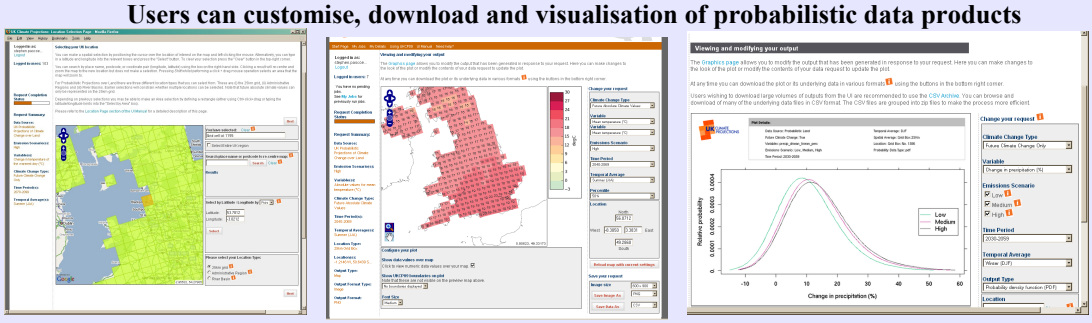
The new API and the associated metadata ingestion system, capable of taking not only the MEDIN ISO19139 but also original GCMD DIF9.4 (and easily other profiles of ISO19139), will form the basis of a "revitalised" NERC Data Discovery Service. This will ensure that the NERC service will also benefit from the advanced search capabilities now offered by the MEDIN Discovery Portal.

CEDA has successfully demonstrated that not only can it quickly respond to the requirements of groups such as MEDIN but also that it can deliver, deploy and run the software operationally as well as working with an external development team to ensure that the specifications are met.

Highlight: UK Climate Projections (UKCP09) User Interface

Stephen Pascoe, Ag Stephens, &Alan Iwi

Users can customise, download and visualisation of probabilistic data products



In response to the clear evidence that anthropogenic emissions are a major contributor to climate change[13] the climate science community is being called upon to provide high-resolution projections of climate change that can be used by decision-makers to plan future strategy. For instance national and local government need the best projections of regional and local climate to feed into National Climate Change Risk Assessment[14] cross-departmental and local strategies.

The UK Climate Projections (UKCP09, launched in June 2009[15]) provide both land and marine products for a range of key meteorological variables over various temporal and spatial scales. The products were developed by the Met Office Handley Centre (MOHC). The real complexity lies in the probabilistic nature of the dataset. For each combination of parameters the user is presented with a set of plausible projections expressed as probability density functions.

This extremely innovative data product is delivered through a customised interface (*http://ukclimateprojections-ui.defra.gov.uk*) developed and operated by CEDA as an operational service. Selected projections can be customised, visualised and downloaded through the highly dynamic user interface. From these projections the user may derive simulated hourly time-series of future weather simulated for a given location using the UKCP09 Weather Generator developed at the Newcastle University and the University of East Anglia.

The system has been designed from the beginning to be highly resilient and scalable. Using the Service Orientated Architecture approach (SOA) the web application communicates with a cluster of back-end services via standard Open Geospatial Consortium protocols: Web Map Service and Web Processing Service. The SOA design provides resilience against server failure and allowed the system to expand across additional hardware to scale to the anticipated demand of 1000 simultaneous users during launch period.

UKCP09 UI registered users by sector

Sector	Users
Leisure and Tourism	39
Industry	26
Multi-sector	415
Non-sector specific	688
Planning	407
Retail	7
Services	100
Telecoms	12
Transport	112
Waste management	27
Water resources	281
Agriculture	145
Nature Conservation	287
Buildings	197
Defence and security	103
Emergency planning	58
Energy	199
Financial services	39
Fisheries	26
Flood management	370
Forestry	33
Healthcare	57
Heritage	30
Total	**3658**

[13] http://www.ipcc.ch/ipccreports/ar4-wg1.htm

[14] http://www.defra.gov.uk/environment/climate/adaptation/ccra/index.htm#assessment

[15] http://ukclimateprojections.defra.gov.uk/

Statistics: CEDA Help desk and associated services

Anabelle Guillory & Graham Parton.

The CEDA Helpdesk (BADC and NEODC user support) includes responding to user queries and handling of electronic application forms for access to restricted data. Helpdesk software is used to distribute, track and analyse queries more efficiently. 93% of user queries are handled by 2 CEDA Data Scientists while the 7% remaining are covered by other CEDA team members as necessary. CEDA continues to provide prompt and effective support services to the user community at a high priority level.

Statistics for period 1st April 2009 to 31st March 2010

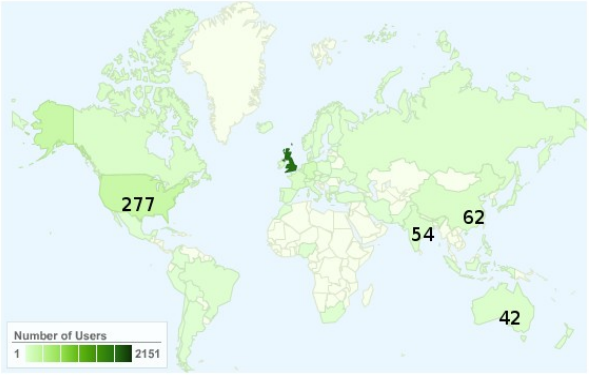
CEDA Queries closed	4086 - 91% are BADC queries
Total CEDA registered users (to 21/05/2010)	15579 - includes 2964 new users in period
NERC funded active users (to 21/05/2010) *(An active users has access to one or more restricted datasets)*	5% of Total CEDA registered users - or 23% of all active users
Identifiable users actively downloading	2819 - up 29.5% from last year
Total download volume	124TB - in 17 million files

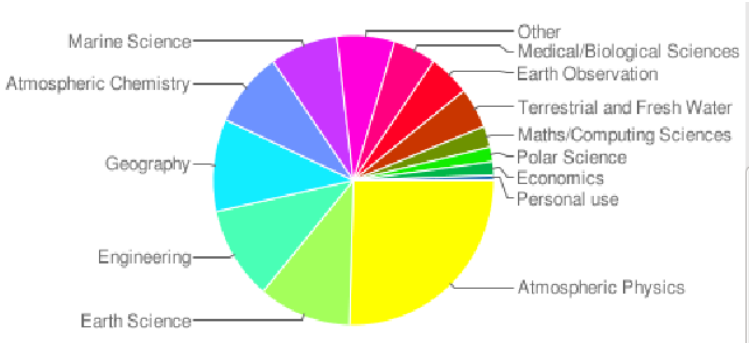
While all new queries receive an initial response within one working day, one BADC query takes on average 5 business days to close.

The quotes below (from a survey carried out as part of a Research Information Network study in January 2010) highlight the continuing good performance of the CEDA helpdesk service, despite a steady rise in the number of users:

> **"These services have enabled research to be undertaken that would otherwise have been impossible. Please keep up this excellent work."**
>
> **"Helpful and knowledgeable staff"**
>
> **"Support have been very helpful with queries."**
>
> **"Helpdesk has been extremely helpful in directing me to data I need, and dealing with access issues."**

While 65% of CEDA users are based in the UK (mostly universities), the outreach of CEDA services goes well beyond the UK borders and atmospheric physics as shown below.





Registered users with at least one active licence (1 June, 2010) by geographic and discipline origin (source: CEDA user database). In excess of 2100 in the UK, 130 in Germany, 60 in France, 50 in Italy, 40 in Spain (European numbers not explicitly shown on map).

Statistics: CEDA Infrastructure

Dan Hagon, Andrew Harwood & Sam Pepler

CEDA has considerable physical infrastructure which supports the data centres and projects. The physical infrastructure itself consumes significant effort to maintain and develop. With around two hundred computer systems, and omitting consideration of systems support, migration of data and services alone require two full time staff (Dan Hagon who is responsible for managing data migration, and Andrew Harwood, who is responsible for software migration). The major elements of the infrastructure are:

1. **Server rooms.** Two server rooms in separate buildings are used in order to provide robust service. The server rooms have the expected server room infrastructure, including air-conditioning, raised floors and industrial power supplies.
2. **Networking.** In 2009/2010 the networking was upgraded to 10Gbit/s links. As well as access at 10Gbit/s to the RAL site access router, the two server rooms are themselves linked with a 10Gbit/s connection so that they function as a single unit.
3. **Primary Storage.** The primary storage is commodity Network Attached Storage (NAS) which provides a cost effective and scalable means to accommodate large volumes of data. In order to minimise the risk of data loss, multiple copies of data are kept, and CEDA endeavours to ensure no NAS system is older than four years Disk storage has increased by a factor of two during 2009/2010, with new systems procured for CMIP5.
4. **Fast Cache Storage.** The expected imminent advent of CMIP5 presented CEDA with a new challenge: supporting massive volumes of data on ingest, and serving multiple clients simultaneously. While NAS is suitable for most of the CEDA load, which tends to have infrequent demand for simultaneous large downloads, the situation for CMIP5 is forecast to be quite different. To that end, in 2009/2010 CEDA has procured a Lustre file system (initially 180 TB), which will provide a scalable solution for high-volume parallel input/output. The system is now running and is undergoing performance tests and configuration before it is deployed. It is hoped that other tasks requiring fast, parallel access to data may also benefit from the lustre system.
5. **Tertiary Storage.** Backup is provided by the Atlas Data Store who provide a "Data Migration Facility" which provides CEDA with 500 TB of physical storage backed by a 5 PB tape robot.
6. **Application Servers.** CEDA uses a mix of native and virtualised servers to run a plethora of applications beyond simple data access, ranging from ingestion, to metadata systems and data visualisations. During 2009/2010 and in the future, CEDA will continue to deploy more applications on virtualised systems where performance permits.

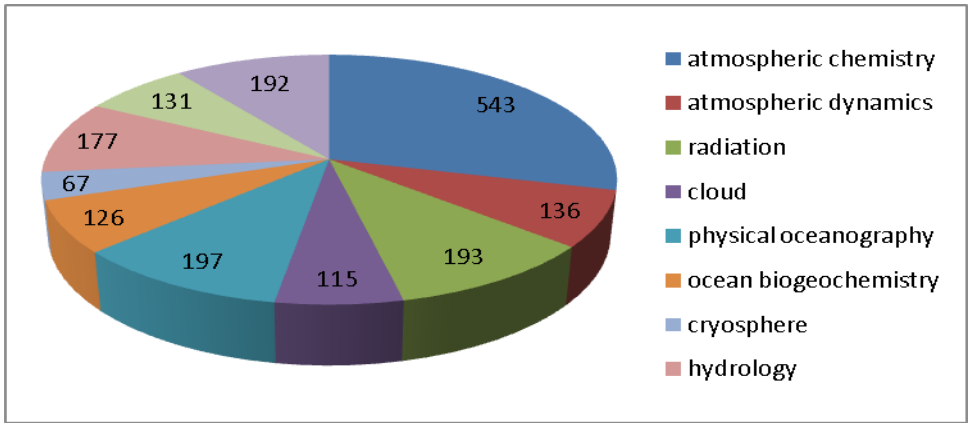
CEDA Storage	
Primary Storage Capacity (Usage) (Note the excess capability is in anticipation of CMIP5)	928 TB (534 TB)
Fast Cache Storage	180 TB
Tertiary Storage (Available Disk for Migration)	500 TB Tape (500 TB)
Total number of primary data files (BADC only, NEDOC statistics not currently available)	93 Million
CEDA Computer Systems	
Storage Servers	33
Hypervisors (Virtual Server Hosts)	14
Virtual Application servers	37
Other Physical Servers	92

Statistics: Standard Vocabularies to Facilitate Sharing of Climate Model Results

Alison Pamment & Bryan Lawrence.

The NCAS-British Atmospheric Data Centre plays an important role in supporting the development of the CF (Climate-Forecast) metadata conventions. The CF conventions are a way of describing climate and weather datasets to an agreed standard thus facilitating easier exchange of data between researchers. The conventions are particularly well suited to describing model data and developments are under way to better adapt them for the description of instrument data. CF metadata record information about the data, such as which model produced them, the time and geographical domain to which the data apply and, importantly, the geophysical variables (e.g. atmospheric pressure, sea surface temperature, wind speed) that are represented by the data. At the BADC we are responsible for developing and maintaining the part of the CF conventions known as the "standard names" – this is a vocabulary of carefully agreed names to describe the thousands of geophysical variables that can be output from models or observed by instruments. Labelling data with standard names provides an important means of searching for them within an archive. However, CF standard names aren't just a label, as each name is also associated with appropriate units and a definition. This means that a researcher comparing data having the same standard name but derived from a number of different models can be confident that he or she is indeed examining the same parameter in every case. The standard names are published on the CF website and are also available via the NERC vocabulary server maintained at the British Oceanographic Data Centre.

The latest generation of earth system models being used for CMIP5 (Coupled Model Intercomparison Project phase 5) contain representations of many physical, chemical and biological processes that have not previously been included in climate models while many other processes are being modelled in more detail than ever before. The results of CMIP5 will form the basis of the next IPCC (Intergovernmental Panel on Climate Change) scientific assessment report. It is a requirement of the CMIP5 project that every model output variable must have a CF standard name and this has resulted in the agreement of over 800 new standard names in the last year, covering such diverse areas of climate science as cloud-climate feedbacks, physical oceanography, atmospheric chemistry, radiative effects of atmospheric aerosol, sea ice dynamics and ocean biogeochemistry. The total number of standard names is now in excess of 2000 and it is anticipated that the list will continue to grow for the foreseeable future. The BADC can provide advice and guidance to researchers who are new to the CF community and who would like to contribute either to the list of standard names or to the process of developing other aspects of the metadata conventions.



The number of standard names, arranged by science domain, in version 14 of the CF standard name table (May 2010).

Statistics: CEDA Software Distributions

Phil Kershaw, Dominic Lowe, Stephen Pascoe, Steve Donegan & Matt Pritchard

CEDA has a considerable software infrastructure which supports the data centres and projects. While much of the software is customised for internal use, CEDA also releases a considerable amount of software as open source. There are three broad grouping to the software CEDA users and makes public for reuse:

1. Security software which provides implementations of key standards necessary to support federation authentication and authorization (so that CEDA internal systems can be used for federated as well as local applications).
2. Discovery systems software to support the NERC Data Discovery Services (since these are common problems).
3. Data manipulation and visualisation packages (used internally, and made available for reuse elsewhere).

Package	Description	Available at
1. NDG-SAML Version (Python)	CEDA implementation of SAML (Security Assertion Mark-up Language) – needed for the Earth System Grid Federation (ESGF) and CMIP5.	http://pypi.python.org/pypi/ndg-saml/
1. NDG-XACML Version (Python)	CEDA implementation of XACML (eXtensible Access Control Mark-up Language). Enables the expression of access control policies to determine who or what has the rights to access a given dataset or other resource. Also for ESGF and CMIP5.	Available on ndg subversion repository (public, but not yet "released".) http://ndg.nerc.ac.uk/dist/
1. MyProxyClient Version 1.1.1 (Python)	Lightweight python based client to the MyProxy package developed by the US National Center for Supercomputing Applications. It enables users to manage their personal identity tokens using remote token repositories.	http://pypi.python.org/pypi/MyProxyClient/
1.MyProxyWebService Version 0.1 (Python)	Enhances the MyProxy service software by adding a HTTP based interface to the server side software enabling any simple Web based client to access it and obtain identity tokens.	Available on ndg subversion repository (public, but not yet "released".) http://ndg.nerc.ac.uk/dist/
1. NDG-Security Version 1.5.5 (Python)	A complete toolkit to manage access control in a federated infrastructure compliant with the system developed for the ESGF and CMIP5. It includes an implementation of the single sign on technology OpenID and features pluggable components for securing any given Web based application.	Available on ndg subversion repository (public, but not yet "released".) http://ndg.nerc.ac.uk/dist/
2. MILK Server Version 3 (Python)	Provides a metadata graphical user interface for the discovery service and other metadata services (ATOM metadata editor & eXist db client and MOLES V2).	Available on ndg subversion repository (public, but not yet "released".) http://ndg.nerc.ac.uk/dist/
2. OAI Document Ingester Version 3.0 (Python)	Code to insert GCMD DIF metadata sourced from an OAI provider or ATOM feed into a postgres database used as the data source for the NERC (v3) Discovery Web Service.	Egg available on ndg subversion repository (public, but not yet "released".) http://ndg.nerc.ac.uk/dist/
2. OAI Document Ingester Version 4.2	Code to insert GCMD DIF and ISO19115 (Gemini/MEDIN) metadata sourced from an OAI provider into a postgres database used as the data source for the MEDIN (v4) Discovery Web Service.	Available on ndg subversion repository (public, but not yet "released".) http://ndg.nerc.ac.uk/dist/
2. OAI Info Editor v1	Web client that provides a public interface allowing users to initiate an OAI harvest and ingest sequence into a Discovery	Available on ndg subversion repository (public, but not yet

(Python)	metadata database. Provides admin role for editing.	"released".) http://ndg.nerc.ac.uk/dist/
2. Discovery WS Version 3 (Java)	Provides a SOAP based interface to allow querying of the Discovery database according to terms and settings submitted. Version 3 is intended to work with OAI_Document_Ingester v3 and MILK server v3.	Public, but not yet "released".
2. Discovery WS Version 4 (Java)	Provides a SOAP interface to the Discovery Database based on ISO ingestion from OAI_Document_Ingester v4 intended to support the MEDIN portal.	Public, but not yet "released".
3. COWS-Server Version 1.4.1 (Python)	Provides OGC web services (Web Map Service, Web Feature Service, Web Coverage Service) for climate data from a variety of data sources.	Available on ndg subversion repository (public, but not yet "released".) http://ndg.nerc.ac.uk/dist/
3. COWS-Client Version 1.4.0 (Python)	Provides a graphical user interface to the COWS server interfaces which can be accessed from a browser. Can also be used to access Web Map Services (WMS) from other data providers.	Available on ndg subversion repository (public, but not yet "released".) http://ndg.nerc.ac.uk/dist/
3. COWS-WPS Version (Python)	An implementation of the OGC Web Processing service that supports synchronous and asynchronous process execution on grid and cluster resources. (COWS-WPS is the unifying technology behind the UKCP09 UserInterface.)	Available on ndg subversion repository (public, but not yet "released".) http://ndg.nerc.ac.uk/dist/
3. OWSLIB Version 0.3.1 (Python)	OWSLib is a community-based open source project which provides a python API for accessing OGC services and making requests for maps/data/features. CEDA is one of the main contributors to this project.	http://pypi.python.org/pypi/OWSLib/0.3.1
3. NAPPY Version (Python)	Python input/output package for handling NASA Ames files, including conversion to/from NetCDF.	Available on ndg subversion repository (public, but not yet "released".) http://ndg.nerc.ac.uk/dist/
3. CSML Version 2.7.13 (Python)	The Climate Science Modelling Language (CSML) package provides a set of python modules for reading and writing CSML documents and interfacing CSML with climate data formats such as NetCDF.	Available on csml subversion respository (public, but not yet "released"). http://ndg.nerc.ac.uk/dist/
3. CDAT-lite Version 5.2-1 (Python)	CDAT-Lite is a package for manipulating climate science data. It is a subset of the CDAT tools developed at Lawrence Livermore National Laboratory which focusses on data management and analysis distributed in a compact package.	http://pypi.python.org/pypi/cdat-lite/5.2

Statistics: Publications

- Callaghan, S; Pepler, S; Hewer, F; Hardaker, P; Gadian, AM; How to publish data using Overlay Journals: The OJIMS Project. Ariadne Oct 2009
- Callaghan, S; Hewer, F; Pepler, S; Hardaker, P; Gadian, AM; Overlay journals and data publishing in the meteorological sciences. Ariadne Jul 2009
- Haines, Keith; Hermanson, L; Liu, Chunlei; Putt, Debbie; Sutton, R. T.; Iwi, A. M.; Smith, Doug; Decadal climate prediction (project GCEP). Philosophical Transactions Of The Royal Society A: Mathematical, Physical And Engineering Sciences, 367 (1890) , 925--937, 2009
- Juckes, Martin; Lawrence, B. N.; Inferred variables in data assimilation: quantifying sensitivity to inaccurate error statistics. Tellus A, 61 (1) , 129--143, 2009
- Latham, S. E; Cramer, R; Grant, M; Kershaw, P; Lawrence, B. N.; Lowry, R; Lowe, D.; O'Neill, K; Miller, P; Pascoe, S.; Pritchard, M; Snaith, H; Woolf, A.; The NERC DataGrid services. Philosophical Transactions Of The Royal Society A: Mathematical, Physical And Engineering Sciences, 367 (1890) , 1015--1019, Mar 2009
- Lawrence, B. N.; Lowry, R; Miller, P; Snaith, H; Woolf, A.; Information in environmental data grids. Philosophical Transactions Of The Royal Society A: Mathematical, Physical And Engineering Sciences, 367 (1890) , 1003--1014, Mar 2009
- Lowe, D.; Woolf, A.; Lawrence, B. N.; Pascoe, S.; Integrating the Climate Science Modelling Language with geospatial software and services. International Journal Of Digital Earth, 2 (1 supp 1) , 29--29, 2009
- Parton, G. A.; Vaughan, G; Norton, E. G.; Browning, K. A.; Clark, P. A.; Wind profiler observations of a sting jet. Quarterly Journal Of The Royal Meteorological Society, 135 (640) , 663--680, 2009
- Shaffrey, L.C.; Stevens, I.; Norton, Warwick; Roberts, M. J.; Vidale, P-L; Harle, J. D.; Jrrar, A.; Stevens, D. P.; Woodage, Margaret Jean; Demory, M-E; Donners, John; Clark, D. B.; Clayton, A.; Cole, J.; Wilson, Simon Spencer; Connolley, W. M.; Davies, T. M.; Iwi, A. M.; Johns, T. C.; King, J. C.; New, A. L.; Slingo, Julia Mary; Slingo, Anthony; Steenman-Clark, L.; Martin, G. M.; U.K. HiGEM: The New U.K. High-Resolution Global Environment Model - Model Description and Basic Evaluation. Journal Of Climate, 22 (8) , 1861--1896, 2009
- Thomas, G. E.; Poulsen, C. A.; Sayer, A. M.; Marsh, S. H.; Dean, S. M.; Carboni, E.; Siddans, R.; Grainger, R. G.; Lawrence, B. N.; The GRAPE aerosol retrieval algorithm. Atmospheric Measurement Techniques, 2 (2) , 679--701, Nov 2009

CEDA Funding 2009/2010

CEDA is funded by a wide range of sources, through direct funding via service level agreements and on a project basis.

In 2009/2010, at CEDA:

- no new NERC grants were begun.
- one new EC project began: IS-ENES (InfraStructure for a European Network for Earth Simulation)

Financial Summary:

	NCAS	NCEO	Other NC	Data RP	Data RM	TOTAL SLA	Other
Carry-In	-236	-164	-255	-174	-40	-869	39
Income	-866	-389	-22	-430	-29	-1736	-710
Spend	1208.2	428	241	424	20	2321.2	596
Carry-Out	106.2	-125	-36	-180	-49	-283.8	-75

CEDA financial summary for 2009/2010.

Most of the funding to CEDA comes from a service level agreement (SLA) between the Natural Environment Research Council (NERC) and the Science and Technology Facilities Council (STFC).

Many of the programmes funded by the SLA are multi-year programmes, with funds being allocated in one year, but not spent until some years later. Funds are generally deferred by a combination of three mechanisms: simple accounting carry over from one year to the next, or formal deferment of milestones at either STFC or NERC (in which case the funds remain outside of the CEDA account). Because there are very large fluctuations in income from one year to the next, because large item spends come from accumulating capital, and because staffing is relatively static, there can be considerable carry overs from one year to the next.

The table above tells us for 2009/2010 that:

- There was a significant overspend in direct NCAS funding.
- There was a significant underspend in direct NCEO funding.
- Other NERC national capability funding slightly underspent.
- Data management funding associated with NERC research programmes underspent considerably (this is normal, as generally the money arrives before the work).
- Data management funding associated with NERC research mode programmes (generally consortium grants) underspent slightly.
- Overall, we see that for SLA activities, CEDA spent a considerable amount of the accumulation: primarily on hardware to support CMIP5.
- We also see that of the total spend, roughly 20% was from non-SLA activities (primarily EC grants and government contracts).

Targets for 2010-2011

<table>
<tr><th>Project Topic</th><th>National Capability within the project</th></tr>
<tr><td>T5.1 Acquire, ingest, and catalogue, appropriate data from a range of sources including the Met Office and NERC-funded projects.</td><td rowspan="3">• METAFOR and IS-ENES FP7 projects
• NERC SIB projects
• UK Climate Predictions (UKCP09)
• DECC support for CMIP5
• DECC support for www.ippc-data.org
• Met Office support for CMIP5
• BADC and NEODC
• Development of the ISIC visualisation centre.</td></tr>
<tr><td>T5.2 Maintain and upgrade computing systems and networks to support data holdings and data access.</td></tr>
<tr><td>T5.3 Develop, maintain and upgrade necessary software and information systems to support data curation and data access. (Including support for CF-netCDF.)</td></tr>
<tr><td rowspan="2">T5.4 Curate existing information according to best practice principles: create, delete, migrate data and information as necessary.</td><td>Research Programme Activities</td></tr>
<tr><td rowspan="2">Data support for:
• HIRDLS
• APPRAISE
• QESDI
• Rapid Watch</td></tr>
<tr><td>T5.5 Provide prompt and effective user support. Provide additional services to users such as a meeting registration service and distributed document management service.</td></tr>
<tr><td rowspan="2">T5.6 Provide support for CMIP5 by providing a UK data node, a replicated copy of the global core archive, and appropriate interfaces (software, hardware, and networks). Deploy and maintain any additional necessary services (e.g. the CMIP5 questionnaire.)</td><td>Research Mode Activities</td></tr>
<tr><td>Data support for:
• CASCADE
• ABACUS
• COBRA</td></tr>
<tr><td>T5.7 Support the UK earth observation community by continuing to provide high speed UK cache archives for ESA, NASA (and other high volume remote data).</td><td rowspan="6"></td></tr>
<tr><td>T5.8 Provide data management services (data management plans and formal archives) for NCAS and NCEO themselves (including FAAM, ARSF, UFAM and the NCEO scientific themes).</td></tr>
<tr><td>T5.9 Contribute to the implementation of the NERC Science Information Strategy Programme, in particular leading and/or managing the projects dealing with architecture, data centre metrics, data citation and publishing, data value check lists and grants information for data management.</td></tr>
<tr><td>T5.10 Collaborate with the British Library, formulate and implement an automatic method of assigning digital object identifiers (DOIs) to datasets to allow data to be easily cited.</td></tr>
<tr><td>T5.11 Work with commercial and academic partners within the NCEO and STFC communities to deliver the ISIC (International space innovation centre) scientific visualisation services. Contribute toe the development of a plan for ISIC phase 2.</td></tr>
<tr><td>T5.12 Plan and begin the implementation of a merger of aspects of the UK Solar System Data Centre into CEDA.</td></tr>
</table>

Appendix: CEDA Organisational Structure

The following diagram indicates who delivers the major functional activities within CEDA (note that some people deliver functions in more than one facet of the CEDA operation)