UKRI
Science and Technology Facilities Council
Natural Environment Research Council

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

CEDA Annual Report

2020–2021

JASMIN

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Contents

Editors: Victoria Bennett, Poppy Townsend

Dr Victoria Bennett
Head of CEDA

Introduction

I am pleased to introduce this 2020–21 edition of the CEDA Annual Report. The past year has been dominated by the unexpected and, for many, stressful impacts of the global COVID-19 pandemic.

Nevertheless, under completely new ways of working, and interacting with our internal and external colleagues and collaborators, the CEDA team have achieved many exciting milestones, continued to lead and contribute to high profile projects, and developed tools and systems to support and improve our services.

In this report we present a representative selection of the year's activities, show off some of our highlight achievements, and report our usage and funding figures.

I look forward to another productive year ahead, taking the positives from what we learned during the last 18 months, but returning to more normal circumstances in general.

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

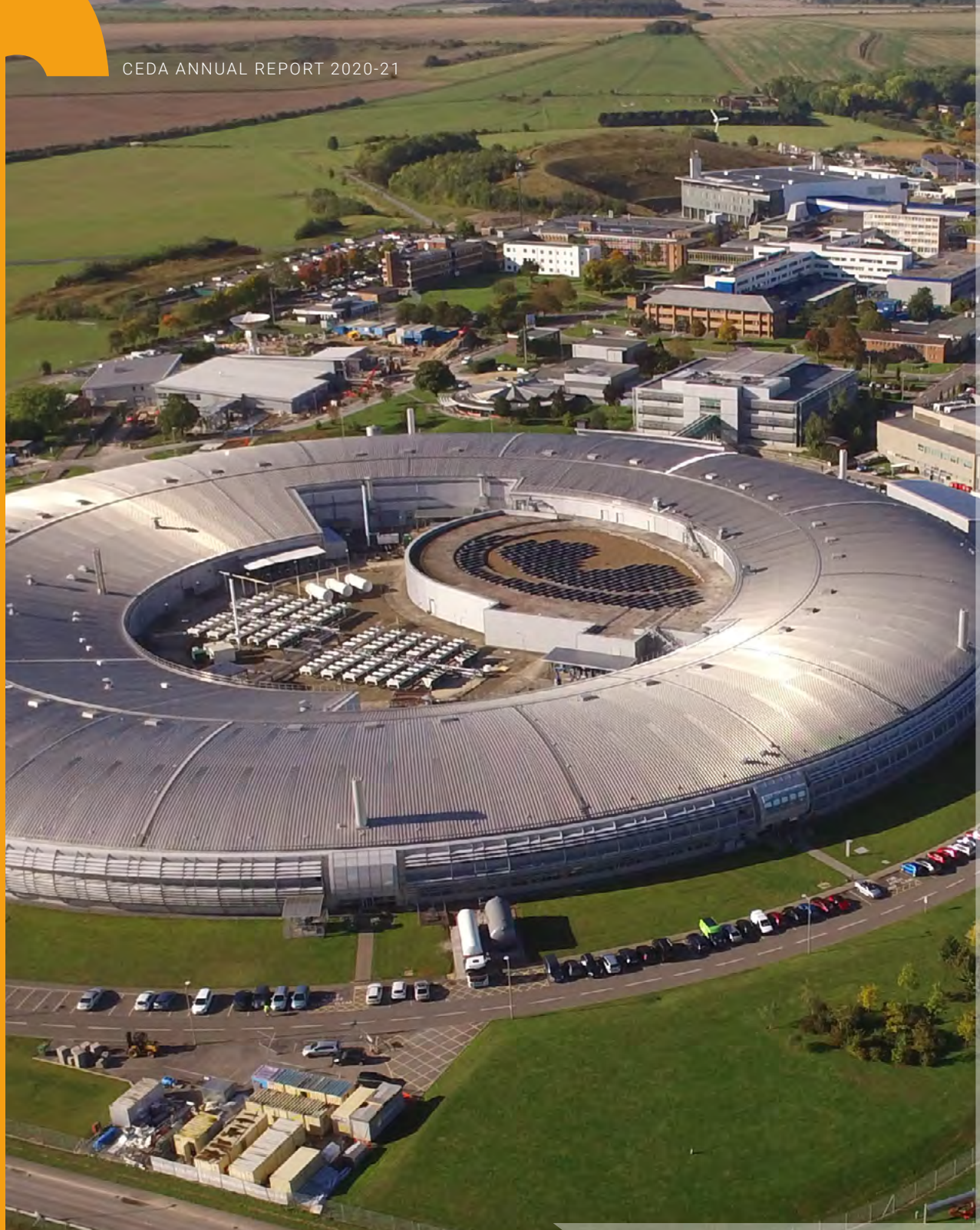
JASMIN

The Centre for Environmental Data Analysis (CEDA) is based in the Science and Technology Facilities Council (STFC)'s RAL Space department. CEDA operates data centres and delivers data infrastructure, primarily for the Natural Environment Research Council (NERC), and undertakes project work for a range of national and international funders. CEDA's mission is to provide data and information services for environmental science: this includes curation of scientifically important environmental data for the long term, and facilitation of the use of data by the environmental science community.

The role of CEDA staff continues to evolve to include services and support for users of increasingly large and complex datasets. Last year saw further investment in, and development of, JASMIN, enabling us to continue to grow and improve the capabilities offered to users. In addition, as in previous years, CEDA staff are involved in nearly all the major atmospheric science programmes underway in the UK, in many earth observation programmes, and in a wide range of informatics activities.

CEDA Archive
CEDA Archive was established in 2005, as a merged entity incorporating two NERC designated data centres: the British Atmospheric Data Centre, and the NERC Earth Observation Data Centre. Since April 2018, the CEDA Archive has been a component part of the NERC Environmental Data Service, which brings together the five NERC data centres into a single service commissioned by NERC as National Capability.

JASMIN
JASMIN is the data intensive supercomputer which provides the infrastructure upon which the CEDA archives and services are delivered. Increasingly, JASMIN provides flexible data analysis capabilities to a growing community, who benefit from high performance compute and a private cloud, co-located with petascale data storage.



Meet the team

Victoria Bennett
Head of the STFC Centre for Environmental Data Analysis (CEDA)

Jennifer Bulpett
Project Manager

Fatima Chami
JASMIN User Support

Esther Conway
Senior Data Scientist: Earth Observation

Steve Donegan
Senior Data Scientist: Earth Observation

Francesca Eggleton
Data Scientist

Wendy Garland
Senior Data Scientist: Aircraft

Hayley Gray
User Support and Metadata Services Operator

Andrew Harwood
Infrastructure Manager

Alan Iwi
Senior Software Engineer

Matt Jones
Senior DevOps Engineer

Martin Juckes
Head of CEDA-Atmosphere and deputy head of CEDA

Philip Kershaw
Technical Manager

Neil Massey
Senior Software Engineer

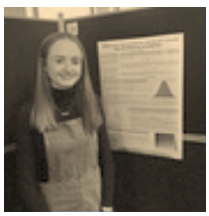
Alison Pamment
Data Scientist: metadata standards

Charlotte Pascoe
Senior Data Scientist: Models

Graham Parton
Senior Data Scientist: Observations

Sam Pepler
Head of Curation

Matt Pritchard
JASMIN Operations Manager



Elle Smith
Software Engineer

Richard Smith
Senior Software Engineer

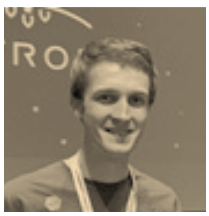
Ag Stephens
Head of Partnerships

Poppy Townsend
Communications Manager

William Tucker
Software Engineer

Alison Waterfall
Senior Data Scientist: Earth Observation

Matthew Wild
Senior Data Scientist: UKSSDC



Ed Williamson
Data Scientist: Earth Observation

Kate Winfield
Environmental Data Scientist

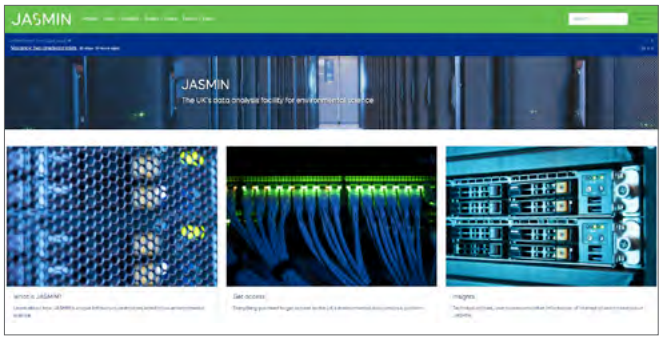
Highlights

Upgrading our websites

POPPY TOWNSEND, MATT PRITCHARD AND RICHARD SMITH

CEDA manages several websites that act as a brochure for our key services. These needed modernising to reflect and support the set of diverse services we currently offer. The CEDA, JASMIN and CEDA Archive websites have undergone a complete overhaul, in terms of design, content and structure. The new and improved sites were finalised this year.

Not only has the look and feel changed, but site navigation has also been improved. Using Google's programmable search, we have improved cross-site searching, specifying relevant search contexts for each domain. The addition of help beacons allows visitors to easily search documentation or contact the helpdesk.



Developing international and national strategies

MARTIN JUCKES

The serendipitous occurrence of several strategy discussions affecting CEDA within the same time frame gave us the opportunity to play a constructive role in shaping and aligning mid- and long-term objectives for a broad range of activities.

The three key areas of strategic discussions involved are:

- The UKRI Digital Research Infrastructure (DRI): this is a new initiative through which the nine UKRI Research Councils are working jointly to create a DRI which will include compute, software, people and the underpinning tools and networks needed to advance the work of the UK's researchers and innovators. The initial funding includes an allocation for JASMIN, recognising the facility as a national priority.
- The NERC Digital Strategy has been developed in parallel with the UKRI DRI Strategy, and will provide a foundation for the next phase of NERC Data Centre funding. The strategy emphasises the importance of consolidating JASMIN's role in supporting data-intensive computation for NERC.
- The IS-ENES (Infrastructure for the European Network for Earth System Modeling) Climate Data Infrastructure (CDI) Mission and Objectives has been developed, setting out priorities for the European collaborative services for climate model data as the network negotiates the transition between two EU funding cycles as well as initial steps to formalising the consortium as a legal entity.

Active engagement in these discussions reflects the high profile of CEDA among national and European service providers.

CEDA gains trusted repository status

GRAHAM PARTON

Researchers see value in the data they produce, which they want to be preserved in relevant, sustained, trustworthy repositories. However, with the increasing range of data services available, knowing which services meet these criteria can be hard to establish. Repositories, such as the CEDA Archive, can earn the Core Trust Seal as a recognition of their quality as a long-term archive for end-users. This is a peer reviewed process that examines all aspects of a repository from financial sustainability, policy and documented workflows, through to security and hardware choices. CEDA has successfully completed the thorough application process and final Core Trust Seal certification has now been approved.

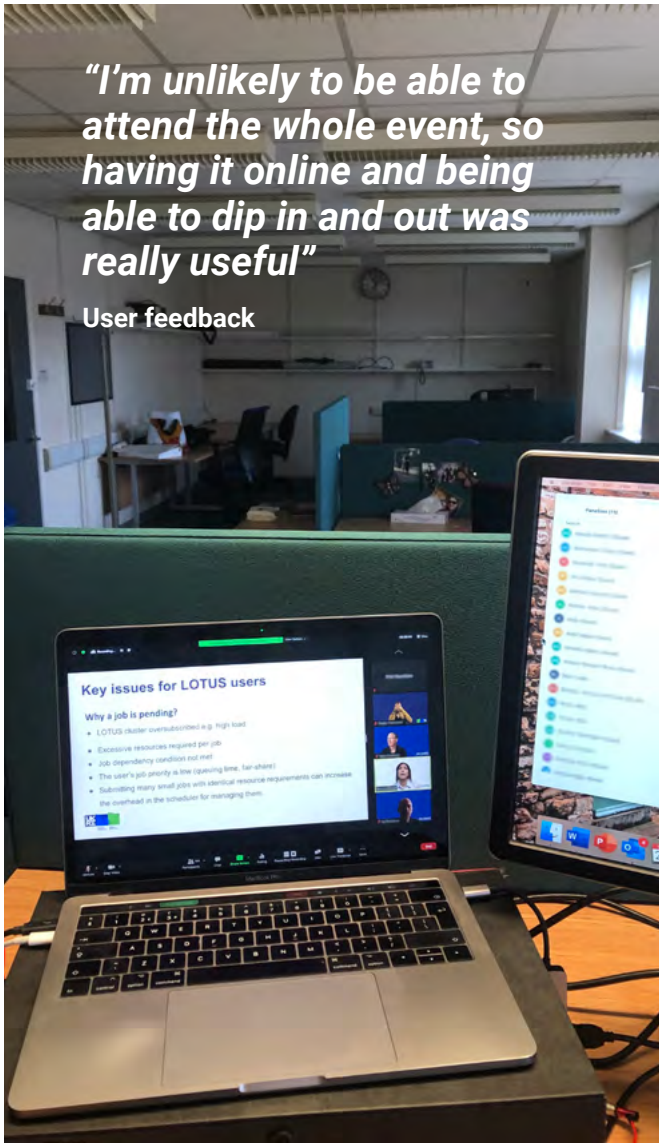
Engaging with users whilst working from home

POPPY TOWNSEND

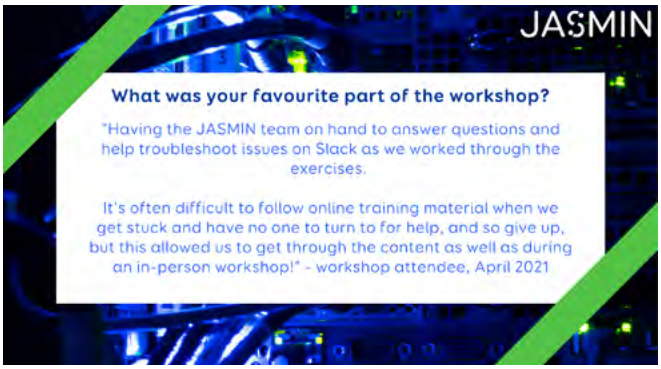
All staff members have been successfully working remotely during the coronavirus pandemic. Whilst the transition to home working wasn't too arduous for most work, the way we engaged with users for training and events required drastic changes. We adapted existing events and created new events where needed. Between 1st April 2020 and 31st March 2021, we delivered approximately 15 hours of virtual events, with over 600 attendees. Key events are described below.

JASMIN2020 Virtual Event

This replaced the user conference that was due to happen in June 2020. It was delivered via Zoom webinar on two consecutive mornings. It included update talks from JASMIN staff. Over 100 people attended each day and feedback was extremely positive. When asked 'would you prefer to have a virtual or face to face event next year?', over 80% of survey respondents said they'd prefer virtual or they didn't mind.



Presenting the JASMIN2020 virtual event to an empty office.



Feedback from an attendee at a virtual workshop.

User Seminar Series

Feedback from the JASMIN2020 virtual event suggested many users wanted to hear about other users' research. To facilitate this, a new monthly seminar series was developed. The aim is to encourage knowledge sharing and best practice between users – whilst showcasing the types of science enabled by JASMIN. The popular hour-long events include talks from two users with time for questions.

Hands-on workshops

Considerable work was undertaken to convert our in-person workshop into a virtual event. This included;

- re-writing exercises and recording videos – all resources are now freely available
- researching the best tools - we opted for a mixture of presenting live on Zoom and offering chat support via Slack
- changing the structure/format – exercises have now been split across a 'beginners' and 'advanced' workshop

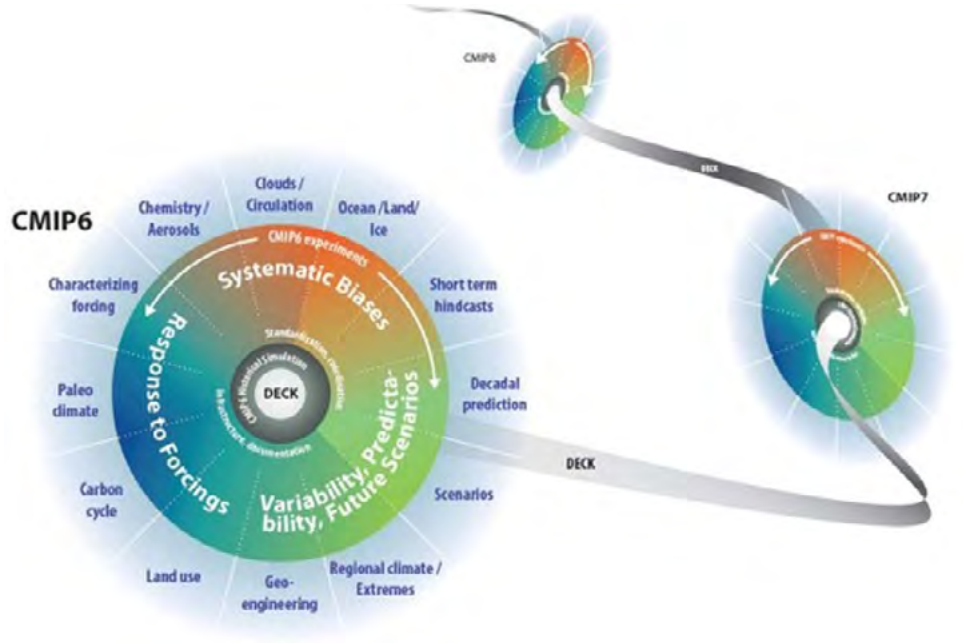
Future events

Over the last year, our events have been more accessible than ever before. We've reached the largest number of attendees ever and recordings have had over 2,500 views on Youtube. Less logistical effort was required by the CEDA team (e.g. no booking rooms, visitors through site security), however it can't be ignored that the events lacked networking interaction between ourselves and the user community. We will continue to listen to the user community's wishes when planning future events.

Disseminating essential climate data: CMIP and ESGF

RUTH PETRIE, ALAN IWI AND AG STEPHENS

The 6th Coupled Model Intercomparison Project (CMIP6) is an international coordinated effort to understand how the climate has changed in the past and may change in the future. It is the largest climate model intercomparison project to date and is coming to completion. Data produced by CMIP6 is essential to the 2021 assessment report by the Intergovernmental Panel on Climate Change (IPCC). CEDA supports CMIP6 by providing archival, management and publication of all received data and providing analysis capabilities through JASMIN.



This shows the continuity between CMIP6 to CMIP7 to CMIPx. Each evolution of the project runs a core set of experiments (the grey thread) and updates the other experiments as climate science evolves.

Credit: Eyring et al. Overview of the CMIP6 Experimental Design and Organization, GMD, 2016.

Currently, the CEDA Archive holds approximately three petabytes of CMIP6 data (of a total 20 Petabytes produced across the project). Researchers across the world use these important climate projections to explore topics such as the impact of climate change on hydroelectric power in Kenya, how future temperatures will affect human welfare, and many other important science areas.

Approximately half of the archived data is from the UK Met Office Hadley Centre, for which CEDA are the primary archive – this allows all international collaborators access to the Met Office model simulations data. The other half of the data is replicated from other contributing modelling centres from around the world. This replication of internationally produced data allows easy access and analysis for the UK research community.

Supporting international research

CMIP6 data are shared and published to a globally distributed software infrastructure called the Earth System Grid Federation (ESGF). This is an international collaboration between research institutions that allows the dissemination of outputs from important climate modelling efforts, such as CMIP6. CEDA is responsible for the UK's contributions to ESGF. We are part of a network of 33 nodes disseminating climate data around the world by providing tools and interfaces for data management, discovery and download.

CEDA staff are involved with ESGF in various ways, including; developing software components within the international collaboration, leading the design of future architecture, and maintaining operations for our data node as part of the network.

Automating processes

The sheer volume of data requires the ingestion into the CEDA Archive and subsequent publication to ESGF to be fully automated and flexible enough to work for all incoming data sources. To achieve this, CEDA have developed a dedicated tool called the CEDA REceive-to-Publish Pipeline (CREPP) tool. The tool has successfully been used to publish global climate simulation data of key interest to the climate science community from 30 of the world's leading climate research institutions.

Future architecture for ESGF

PHILIP KERSHAW

In 2019, CEDA led a programme of work to re-engineer the ESGF software system that was first developed over ten years ago. The work included the adoption of new technologies to facilitate cloud deployment. Working with US partners, a first system has been deployed on Amazon Web Services. A deployment by CEDA on JASMIN has been used in a new pilot system integrating tools and systems from other European partners. Work is also focussing on the development of new community standards for functions such as data discovery in order to facilitate interoperability and broader adoption of the systems developed.

Expanding JASMIN storage – new media for new times

SAM PEPLER AND MATT JONES

At the start of the JASMIN era, approximately 10 years ago, our services used just one main type of storage – fast parallel disk. This storage was reliable, provided parallel write capability and was fast; a great leap forward that suited the needs of both the JASMIN user community processing the data and the long-term CEDA Archive. However, needs change and technology progresses. This year, we finished a huge data migration effort to make use of more efficient storage systems.

We still use fast parallel disk (PFS) for fast, reliable storage, but we now have an increased array of storage types to accommodate other desired properties. As data volumes continue to grow, we need cheaper storage systems, such as a Scale out file system (SOF) – but this will sacrifice true parallel write capability. We also need systems that can handle many millions of files, which trades greater access speed for increased cost, also offering different interfaces than traditional file system access: object storage. Finally, we need storage which is performant for small file access (PFS and SOF aren't great for this): solid state drives (SSD).

This year the CEDA Archive hit a milestone: we moved the archive entirely off the parallel storage system – over 800 TB in a little over a month (this was alongside the group workspace migrations). Don't worry, this was not ten-year old kit, we had already migrated through several generations of this storage. The Archive is now principally on tape in a new tape library purchased last year and on SOF. It took a lot of planning and involved a dedicated cluster to do the migration of the data, but hopefully you did not notice and the transition was not too disruptive. The Archive was not the only thing to migrate; over 100 of the JASMIN Group workspaces moved too (totalling nearly 3 PB). This means there are now around 220 group workspaces on SOF, 22 on PFS, and 17 SSD volumes. The use of the object store is also increasing, with approximately 1.5PB of data across 17 group workspaces and the archive (total from June). The complete migration took around 6 months in total due to some complications and some hard-to-migrate group workspaces.

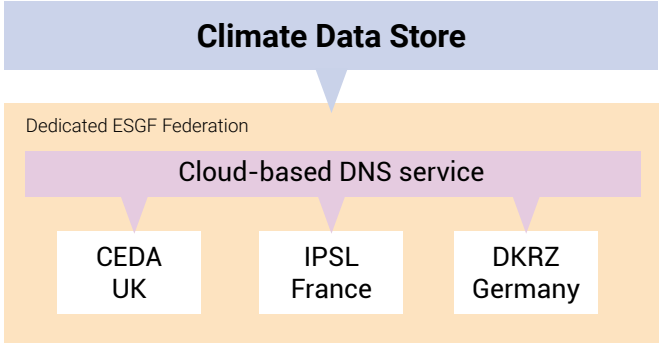
The CEDA Archive is a long-term effort; currently we can see a future that needs even more storage, and we will have to make better use of the cheaper tape. Twenty five years ago, we were moving data from a tape library onto disk, which was becoming very cheap. Whatever the future brings, we will have to adapt to new requirements and kit. This will take a massive effort, but we will endeavour to make sure we keep up.

Hardware in the JASMIN machine room.
Credit: STFC

Providing climate data to the Copernicus Climate Change Service

RUTH PETRIE, MARTIN JUCKES, AG STEPHENS AND PHIL KERSHAW

The Copernicus Climate Change Service (C3S) provides policymakers, businesses and researchers with a comprehensive climate data service through the C3S Climate Data Store (CDS). CEDA, with a consortium of European partners, provides the climate projections for the CDS. This includes the global climate model data from the Coupled Model Intercomparison Project – Phase 5 and 6 (CMIP5 & CMIP6) and the regional climate projections CORDEX data. As CMIP data are large in volume, a set of priority data variables were identified from CMIP5 from a broad range of models to become a core subset of data supplied to the CDS.



Schematic shows load-balanced data service providing data to the C3S Climate Data Store.

Data Quality Control

The data provided to the CDS are quality controlled so that the data provided meet a minimum level of data requirements. The most recent work done in 20/21 was to quality control the newly available CMIP6 data. The quality control consisted of a number of steps, including file-level metadata checks, adherence to CMIP6 data standards, interoperability standards and consistency checks required by C3S. These checks were performed at CEDA, Institute Pierre Simon Laplace (IPSL) and German Climate Computing Centre (DKRZ). The results were collated by CEDA and the data made available through the newly developed Web Processing Service for C3S.

Robust data access

Since C3S are running an operational-level service it is required that the data are available robustly. To deliver this, identical copies of the data are held at three partner sites CEDA (UK), IPSL (France) and DKRZ (Germany). The three sites are then load-balanced against user demand via a cloud based service. This will randomly pick one of the three sites, unless one of the sites is experiencing an outage (either planned or unplanned).

CEDA and their European collaborators have been funded by C3S for the last five years with the role to make these data available to the service, for uptake by a wide user community.

European Climate Data Explorer: an interactive tool for policy-makers

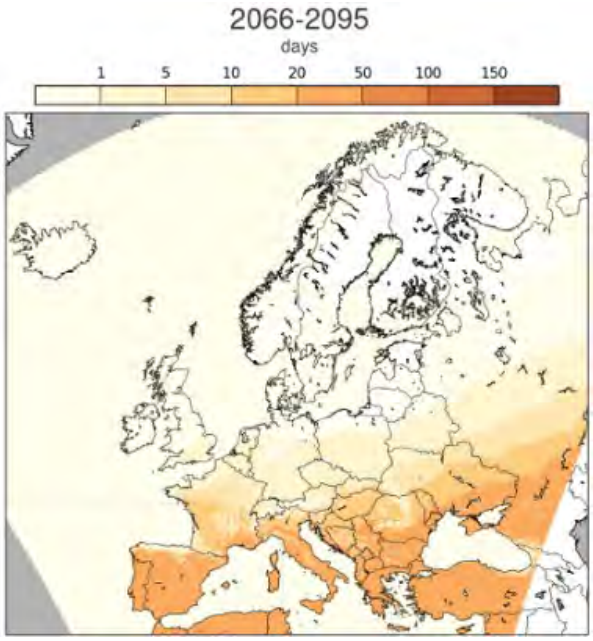
CHARLOTTE PASCOE, MARTIN JUCKES, JENNIFER BULPETT AND JONATHAN HAIGH

The European Climate Data Explorer (ECDE) is an interactive tool that provides access to climate impact data. The ECDE supports the European Green Deal, including the EU's new strategy on adaptation to climate change. The interactive data visualisation provided by the ECDE makes accessing climate impact data easier for policymakers and other stakeholders. It helps by removing otherwise complex data processing hurdles to data discovery and exploration.

CEDA was funded by the Copernicus Climate Change Service (C3S) to provide leadership and project management for the coordination and development of the ECDE. The ECDE provides novel interactive access to a diverse set of climate impact indices by combining live data from the C3S Climate Data Store (CDS) with visualisations that use the CDS Toolbox.

ECDE data is currently customised for climate indices in the health, agriculture, forestry, energy, tourism and coastal impact sectors. For example, the health sector has climate impact indices about thermal comfort and heatwaves as well as the suitability of Europe's climate for the tiger mosquito which can transmit diseases, such as dengue fever.

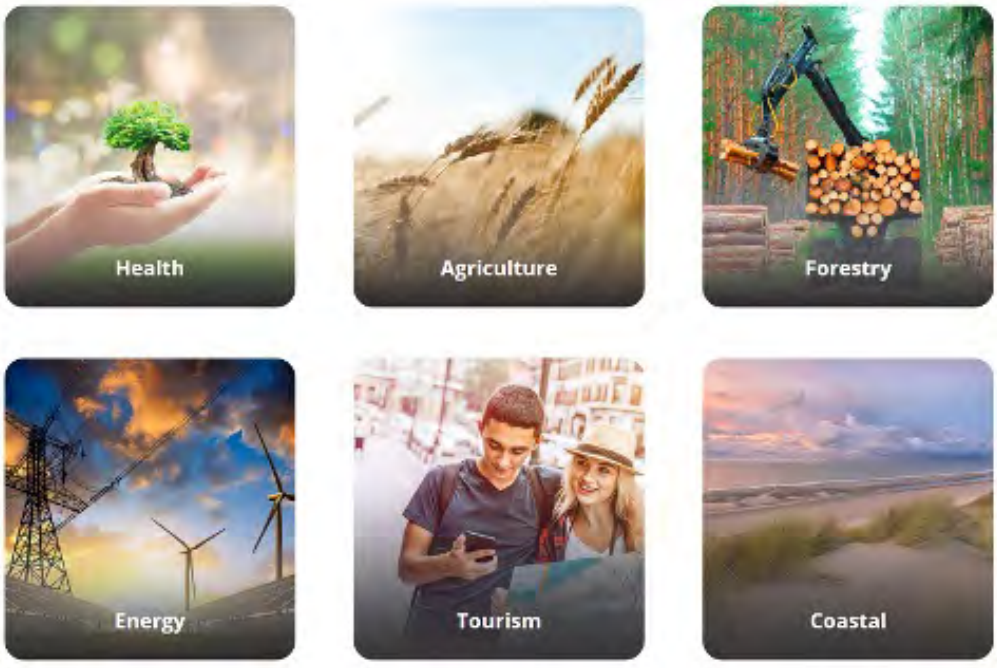
Each climate impact index has two ECDE pages: the first is a Europe-wide visualisation supported by text that provides context for how an index can be understood, the second allows users to explore an index further via an interactive map. The interactive map can be used to focus on smaller regions and to generate time series plots. The visualised data can be exported as images for use in reports and as data files for further analysis.



Map of a climate impact index from the ECDE's Forestry sector showing the projected change in the number of days per year with high fire danger under a high greenhouse gas emission scenario known as RCP8.5 for the period 2066-2095.

CEDA staff developed the editorial process for in-page documentation of climate impact indices which included a common information structure suited to the full range of ECDE indices. CEDA also provided the help resources for the ECDE.

The ECDE is hosted on Climate-ADAPT, a publicly accessible web portal managed by the European Environment Agency (EEA) in collaboration with the European Commission.



Screenshot from the European Climate Data Explorer, showing the six climate impact sectors addressed by the ECDE.

New technical developments

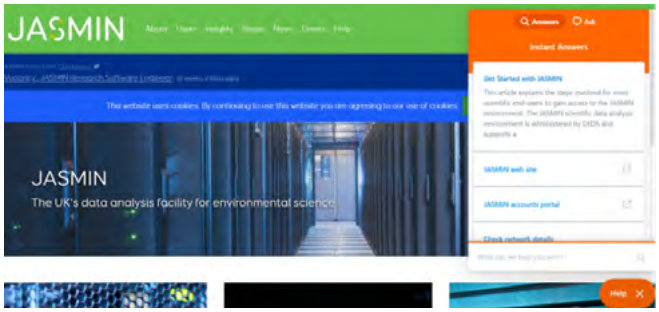
JASMIN Helpdesk

HAYLEY GRAY

Previously, all user queries went to a single helpdesk. In 2020, the helpdesk was separated into two (JASMIN and CEDA Archive). This split enabled more appropriate query categorisation.

As a result, there are now more accurate and detailed reporting metrics for each service - making common queries easier to identify. Help documents and saved replies are regularly updated to minimise the triage of queries so that more queries can be dealt with by the triage team. This also reduces the overall number of queries, allowing an efficient and less time consuming process. A new JASMIN specific help beacon allows users to self-triage their issues, accelerating helpdesk triage processes further.

These changes mean efforts can be focused on improving services for users, that will in turn, decrease helpdesk queries.

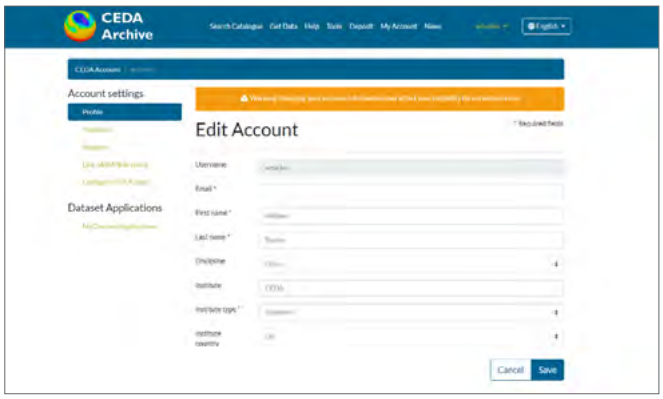


This shows the orange help beacon on the JASMIN website giving the user an option to search the documents or to ask a question to the helpdesk with dropdown options for self triage.

CEDA accounts portal

WILLIAM TUCKER AND ANDREW HARWOOD

We are developing a new user account management system for the CEDA Archive by adapting software developed for JASMIN. The system encompasses user registration, account management and authentication, and registration for access to restricted datasets. This will give us a more robust, maintainable and adaptable system with added functionality. For example, access to datasets requiring approval from external parties can be handled automatically, without requiring intervention from the helpdesk team. It will be deployed on our Kubernetes container orchestration platform, facilitating automated updates and software vulnerability checking.



An example of how the new accounts portal will look when a user accesses their account.

Metrics Service

MATTHEW JONES

Metrics enable the monitoring of systems and gathering numerical evidence to assist decision making. A system for collecting, storing, and displaying metrics about JASMIN is currently under development. Some example metrics include: storage usage, cloud and LOTUS utilisation, and user numbers over time.

The system has already been successfully used several times to easily provide metrics on request - such as for the CEDA-JASMIN Board reports. Soon this system will be production ready, with metrics dashboards designed for different audiences. Future plans will expand what is available through the metrics service – including additional JASMIN metrics, as well as various metrics for the CEDA Archive.

Example of dashboard which will be available with the JASMIN Metrics service, showing e.g. user numbers, and storage usage.

Cataloguing a large, fast-growing archive

GRAHAM PARTON

With over 340 million files in the CEDA Archive, most users will agree that it is pretty big... and it grows rapidly every day. Though CEDA uses tools specifically designed to index all these files at scale and speed, questions remain about how comprehensive CEDA's dataset catalogue coverage is of the Archive.

Comprehensive cataloguing is a vital element to curating data. CEDA has successfully deployed Elasticsearch, an industry standard no-SQL search and analytics engine, to provide a scalable, comprehensive and timely file-level index of the CEDA Archive which paves the way for future cataloguing and search tool developments. However, this technology is unable to provide users with a fuller, more detailed picture of what the data actually are and, more importantly, to aid dataset level discoverability. Details of who authored the data, why and how they were produced and their provenance, as well as basic information such as geographic and temporal coverage and parameters, are all useful contextual information essential to successful long-term re-use. While automatic harvesting from the file-level index can provide some of these elements for the dataset level, much of this content is still manually created by CEDA staff in collaboration with data providers.

Building on the file-level index for the archive and a range of tools developed in recent years, CEDA is now able to undertake timely, archive-wide audits of the data catalogue's coverage. These audits enable uncatalogued content to be identified and to monitor progress on achieving 100% coverage.

The audits utilise a directory tree Python package to build two complementary dataset-directory trees. The first represents the catalogue's existing dataset records and the parts of the CEDA Archive they cover. The second provides a top-down view of those parts of the archive lacking catalogue coverage, identified from the Elasticsearch file-based-index: some 3.4 million paths overall, but these can be grouped into around 1600 entries that are lacking catalogue coverage.

Comparisons between the two also identified a range of other issues that CEDA had been unaware of previously. Examples include: ghost entries remaining in the index from removed or moved archive content; empty directories set up historically for anticipated data that failed to materialise; and missing mappings to catalogue entries - an aspect of the index that is used in our download services.

Since the first run of the audit, over 130,000 entries have been cleaned up. Of the remaining 1600 archive paths requiring additional cataloguing or archive clean-up work, 20% of the resulting tasks have already been achieved.

Regular audits are planned to begin from summer 2021, with additional reporting of the percentage of total volume and number of files in the archive remaining to be catalogued, providing CEDA with the ability to set targets to achieve even comprehensive data catalogue coverage.

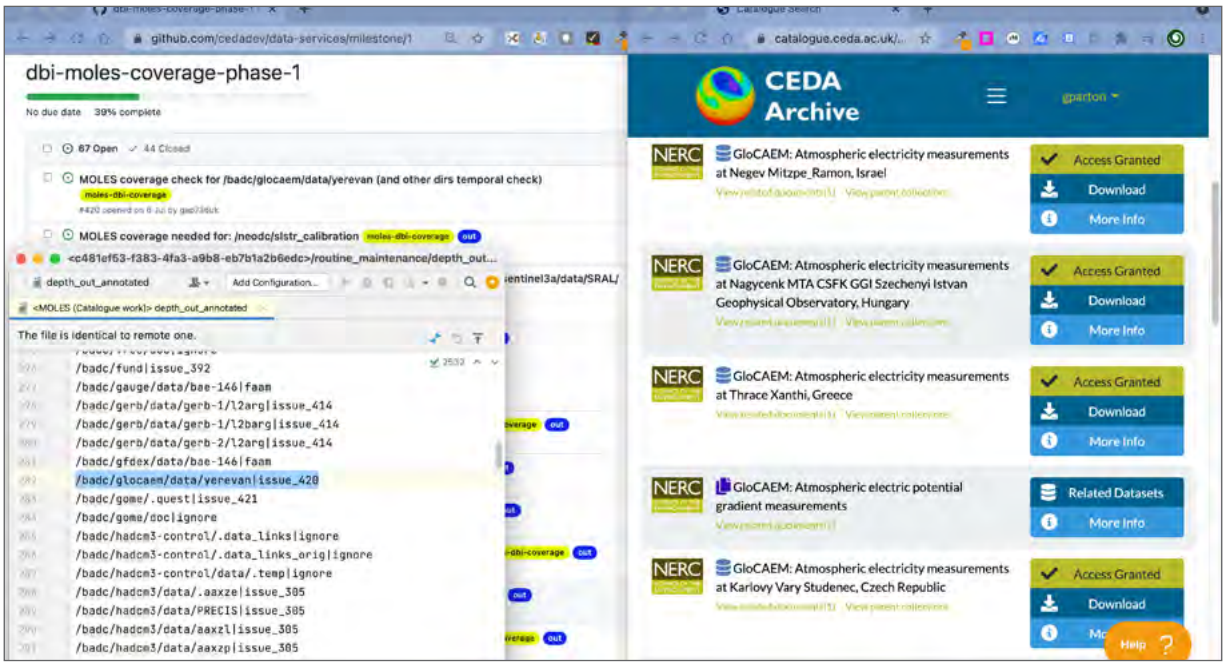


Image showing output of the coverage scan (bottom left). Here, the highlighted directory within the 'GloCAEM` dataset collection was lacking catalogue coverage (confirmed in the catalogue search - right window) leading to an internal task being created for the assigned staff member to resolve. These tasks are tracked with a dedicated Github milestone to help CEDA monitor progress towards full catalogue coverage (top left window).

Sentinel data: variety and volume

ED WILLIAMSON AND STEVE DONEGAN

Sentinel 2.
Credit: ESA

The CEDA Archive now holds over 10 petabytes of satellite data from the Sentinels – this increases in volume at over seven terabytes per day. We provide a mirror archive for Sentinel products, duplicating the ESA archives so that data products required by the Earth observation community are easily accessible. We update 13 Sentinel products daily with another eight being updated monthly – many of these products are essential for researchers studying global issues such as sea surface temperature, air quality, and land-use change.

Due to the ever-increasing volume of data, the team makes use of the differing types of archive storage to ensure data is shared in the most cost-effective and user-friendly way. Fast disk holds the latest data so users can access this instantly, older data is stored on tape as this has less user demand and is a cheaper storage solution. The Near Line Archive (NLA) allows users to pull back data from tape to access the data interactively (more information on the NLA in the 2017-18 Annual Report).

Providing useful data to the community

Many researchers were interested in investigating how COVID-19 lockdowns were changing countries emissions. Sentinel 5P products, which are used for air quality monitoring, were perfect for this. CEDA staff responded to this community need by updating and backfilling many of the relevant products so that comparisons could be made.

CEDA has aimed to archive the latest version of Sea and Land Surface Temperature Radiometer (SLSTR) data produced by Sentinel 3 satellites. Data released by ESA this year have been downloaded from ESA and NASA to populate the CEDA Archive. The latest version has provided users with access to the earlier products which hadn't previously been released as well as older products with the updated processing version.

Working with external data providers

CEDA also archives processed data products such as the Department for Environment, Food and Rural Affairs (Defra) and Joint Nature Conservation Committee (JNCC) Analysis Ready Data (ARD) for both Sentinel 1 and 2 over the UK. These new datasets take the initial processing task away from users – allowing researchers to immediately start tackling issues such as monitoring wildfires or detecting habitat change, rather than spending time preprocessing the data. These products have grown in size to almost 100 TB and were the largest volume downloaded via the web from the CEDA Archive for any Sentinel products during the last year. These data are produced by the Defra Earth Observation Data Service (EODS) on Microsoft Azure and JNCC Simple ARD Service hosted on JASMIN.

During the next year, a new satellite (Sentinel 6) will be launched with sea surface level data expected to be added to the CEDA Archive. The team will build on past experience to ensure these important data are archived in the most efficient way, allowing users access to the most relevant and up to date data.

Sentinel 2A image of Harwell Campus taken on 13 June 2021.
Processed by CEDA. Contains modified Copernicus data.

Providing air quality expertise to government

WENDY GARLAND AND MARTIN JUCKES

Air quality is a key priority area for the UK government. The Environment Agency (EA), on behalf of the Department for Environment, Food, and Rural Affairs (Defra), approached CEDA to participate in phase one of their air quality monitoring strategic review. The aim of this review is to look at future air quality monitoring needs in the UK and to generate a set of high level options for a next generation monitoring network to address those needs. CEDA's expertise in curating and archiving air quality data helped contribute to this review.

Current Air Quality (AQ) monitoring networks focus on collecting, processing and storing data for the purpose of fulfilling the statutory AQ reporting requirements. There are opportunities for improvements to be made in the data handling, storage and accessibility to enable wider reuse and accessibility.

CEDA's expertise and knowledge on data management topics contributed to Work Package 4 - A review of Data handling, Storage, Processing, Access, Tracking and Exposition. Wendy Garland was seconded part-time to EA for 3 months to complete this work with support from Martin Juckes.

The main components of the data value handling chain with reference to the AQ domain. A number of themes running through the whole process were also addressed: FAIR data principles, data and metadata standards, and identification of users and authentication of roles.

Observing systems

A comprehensive review was made of existing AQ data sources, including:

- Complex monitoring networks (both automatic and non-automatic) operated by or for EA
- Local authority AQ data and Highways England data
- NERC funded AQ data, such as observatories and supersites
- International programmes (including Earth Observation)
- Commercial projects such as the Breathe London Project

Data Services

The review also provides recommendations for good practice for data ingestion and curation, covering all relevant data types and associated metadata. Technology options for data access, including visualisation and transformation services were discussed - covering a wide scope of storage options and software environments with an analysis of the strengths and weaknesses of each. A selection of current viewing portals for similar geospatial environmental datasets was presented, highlighting desirable features and opportunities for a future AQ portal.

Finally, a review of similar systems with indicative costs at a range of scales, including our own CEDA-JASMIN set-up, showed how different levels of complexity can be achieved depending on effort and budget available. The WP4 report concluded with general data management recommendations for the Next Generation AQ Monitoring Network which will be compiled into the wider Strategic Review to be used by Defra to shape the next phase of their planning.

Supporting the STFC Air Quality network

WENDY GARLAND

CEDA is supporting the STFC Air Quality network (SAQN+) which brings together research, industry and policy to address air quality challenges - and introduces this community to STFC facilities. With two CEDA members on the local management committee, we have been involved in the setting up of the network, supporting events and providing information about STFC facilities - focusing on the CEDA Archive and JASMIN.

One of these events was the first SAQN+ collaboration workshop which was held online over 2 weeks in November 2020. This was very successful, engaging 30 participants, and led to several small projects being funded. CEDA's mentoring role helped participants to focus their proposals, advised on potential STFC facilities and input, and reviewed the end pitches to approve funding.

Early in 2021 we presented a webinar to the SAQN+ community (many of whom are not our usual CEDA audience) to explain about the CEDA Archive and JASMIN. We covered what air quality datasets we have available and what facilities JASMIN can offer them. We later contributed as panel members for a joint SAQN-social sciences workshop, commenting on several air quality projects running in the social science community.

Catalogue in the Cloud

GRAHAM PARTON AND WILIAM TUCKER

Building reliable, maintainable and resilient user facing services is a key aspect of CEDA's service delivery. As part of this effort, the CEDA data catalogue was successfully migrated to run on JASMIN's Community Cloud, a flexible environment supported by OpenStack — an open-source cloud computing technology. This has led to quicker, easier service deployments for service updates and reduced downtimes for users.

This web portal provides management facilities for virtual machines in the OpenStack cloud, such as the CEDA catalogue virtual machine

Delivering flexibility and resilience for CEDA Services

PHILIP KERSHAW, WILLIAM TUCKER AND RICHARD SMITH

In the 2019-2020 Annual Report, we wrote about the benefits of using Ansible to encapsulate our service deployments with scripts, significantly improving repeatability and saving development effort. These scripts were based on a set of shared recipes for building a service and have led to a more standardised set of applications. It's now easier and faster for the CEDA team to create new services and identify bugs in existing applications.

In addition, deploying the majority of our services in this way made a huge difference when carrying out operating system upgrades to our service hosts in 2019/2020 - saving time and effort compared with deploying hundreds of services each with their own dependencies. The improvements allowed us to migrate over 200+ virtual machines with ease.

Taking this to the next level, we're now using containerisation to further streamline our service deployment procedures. Containers are more lightweight than traditional virtual machines, making it easy to build complex services from small, modular parts. Our new JASMIN Notebook service, for example, uses a separate container for each user's Notebook. The advantages include resilience – individual processes are isolated from each other, one broken notebook does not affect others; and scaling - the number of concurrent users is only limited by the available computing resources.

Some of the technologies that make this possible include:

Docker - for building the containers used to run our own, or third-party, application code;

Kubernetes – a container orchestration platform;

GitLab CI/CD – which allows us to build Continuous Integration/Deployment pipelines (see Figure) to automate the entire deployment process from code commit through to final roll-out;

Helm charts – which replace Ansible playbooks as highly customisable descriptions of deployments; and

Rancher – providing a single interface for Kubernetes cluster monitoring/management.

We've built a suite of shared recipes using these tools which have already been used in the deployment of services such as the JASMIN notebook service and CCI data portal.

Here are some of the benefits we've been able to apply:

- automatic scaling based on demand – the amount of computing resources allocated to a given service can be configured to automatically shrink or expand according to how much it is being used;
- zero-downtime deployments – when new versions of a service are deployed, it initialises behind the scenes and takes over from the existing version only once it is confirmed they are up and running;
- easy rollbacks – the ability to bring failed deployments back to the last working state;
- failure recovery – an error in the application will automatically trigger a re-deployment.
- Using GitOps and Kubernetes also makes it easier to share production-ready deployments for internal review, helping to accelerate application development.

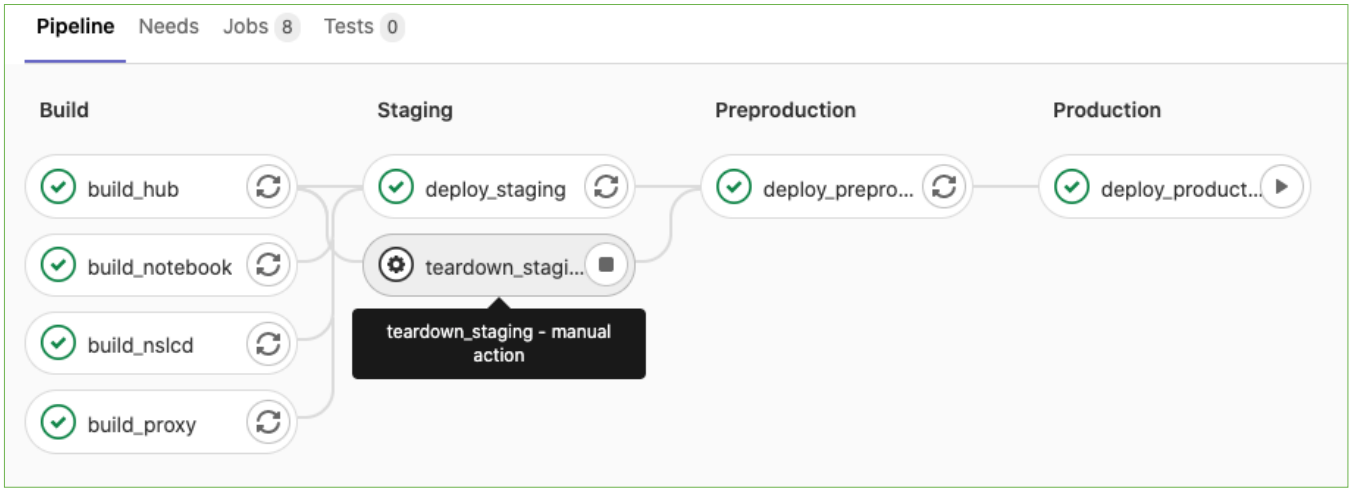


Figure: Continuous integration pipeline for JASMIN notebook service.

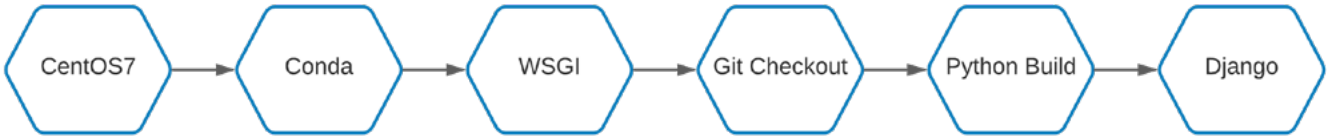


Figure: Different container images which make up a Django deployment.

JASMIN Cloud supports new data science techniques

PHILIP KERSHAW

JASMIN provides a platform giving users a lift-up with their work so that they have data, infrastructure, software tools and services ready-made for them, freeing up more of their time to dedicate to the research goals of their project. One example is the Cluster-as-a-Service system which enables users to make software environments for data analysis on JASMIN's Community Cloud from pre-assembled software building blocks. This is a system that has been in place for a few years now but in this article I want to explore how it has really come into its own with a recent project for the European Space Agency (ESA) Destination Earth Pre-cursors programme.

Working with colleagues from NCEO and industry, CEDA submitted a successful bid, led by Telespazio-VEGA, to build a climate explorer application specifically looking at modelling for soil moisture. The goal was to develop an Artificial Intelligence (AI) based emulator to a land surface modelling system. Building an emulator would enable it to run on a regular cloud computing system rather than the specialised High Performance Computing (HPC) environment needed for the model code. If we could achieve this, it could open up this application to a much wider audience.

JASMIN was uniquely placed, providing the project team with all the necessary components to build what was required: a HPC environment which could be employed to produce training data for the AI system and a cloud computing environment for hosting a data analysis environment to run the emulator. In the case of the latter, the Cluster-as-a-Service system was used to dynamically deploy an instance of Pangeo, a popular suite of software based on Jupyter Notebooks. The project team could quickly manage their own users and install all the required software. They were even able to build graphical web applications quickly using the Python Panel library.

One key challenge was how to share data outputs between JASMIN's HPC and the cloud environment. For this, JASMIN's object store was used together with the Python libraries xarray and zarr to make the data readily available. Model data was output as a series of spatial data slices one per time step. However, for analyses the project was primarily interested in time series data. Extracting these would be highly inefficient given the arrangement of the output data. CEDA staff supported the project team helping them using the xarray libraries to re-organise the data on object store as a set of slices oriented in the time dimension. This trivial re-organisation drastically cut data retrieval times making real-time analysis possible in an interactive map-based application (see Figure).

We are excited about the possibilities for future projects, with the ability to rapidly deploy interactive analysis environments and in addition, re-arrange data into an analysis-ready cache for the application in hand.

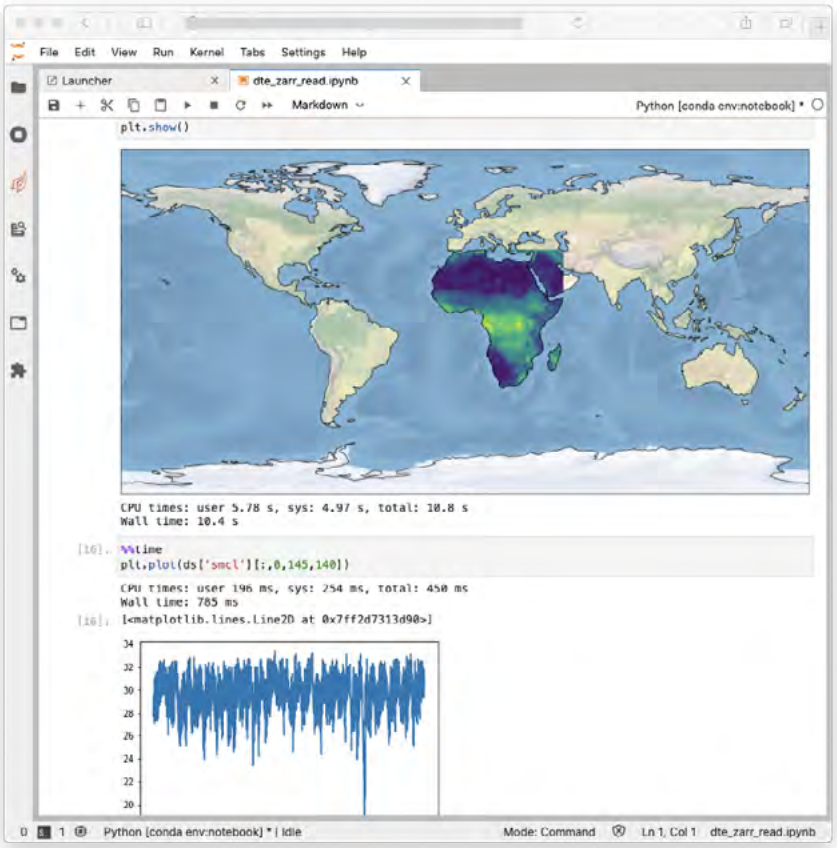


Figure: Soil moisture data generated through the project and accessed via Jupyter Notebook environment hosted on JASMIN.

The JASMIN Notebook Service: lowering technical barriers for scientists

AG STEPHENS, FRANCESCA EGGLETON, ESTHER CONWAY AND ED WILLIAMSON

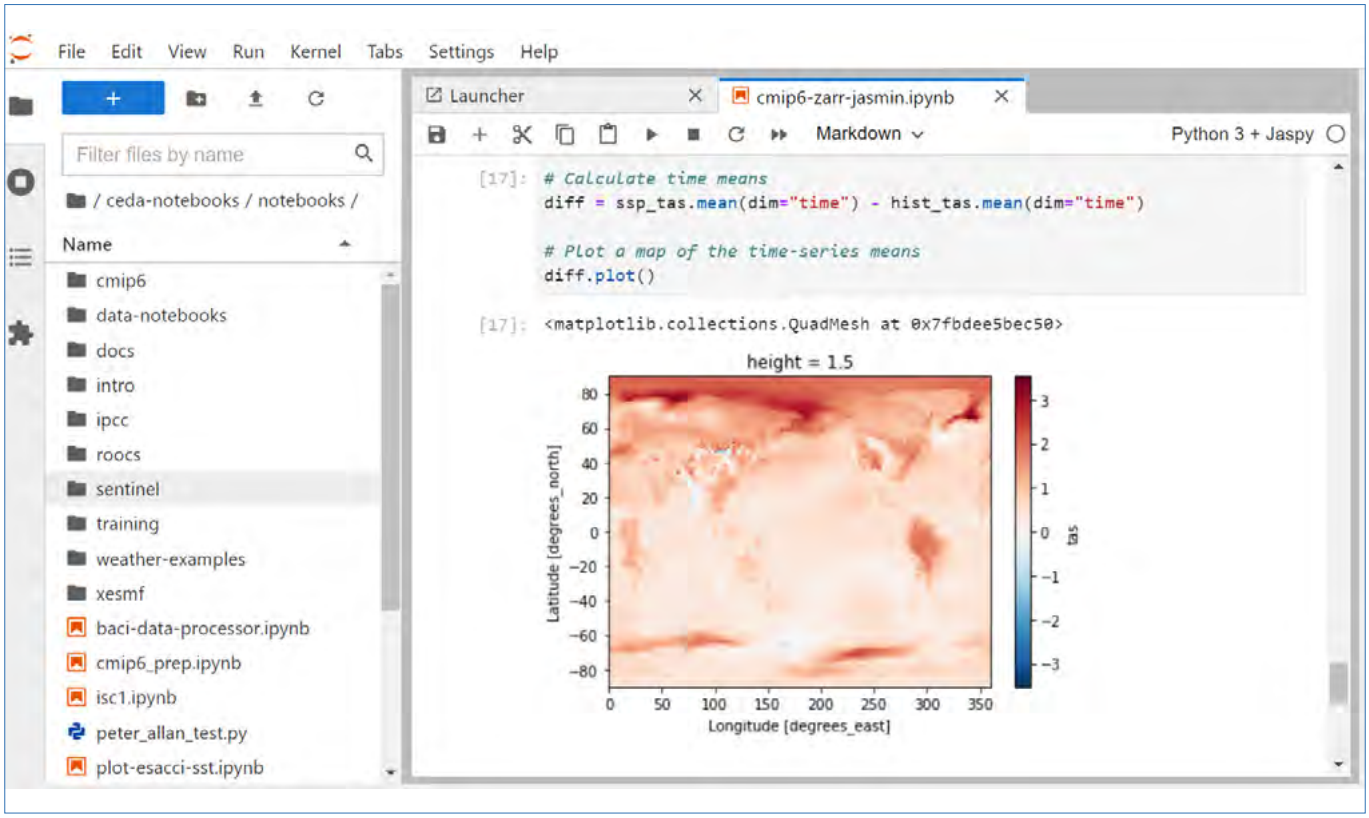
JupyterLab is a web-based interactive development environment for Jupyter notebooks, code, and data. Our JASMIN Notebook Service (JNS), launched in 2020, builds on JupyterLab to provide new functionality and opportunities for JASMIN users. The JNS gives the user access to a programming environment through a web-browser anywhere in the world. Each JNS session includes read access to the CEDA Archive, JASMIN Group Workspaces and Object Store – enabling analysis and visualisation of petabytes of environmental data. The service comes with a Python 3 environment with all the major open source data analysis packages pre-installed.

The JNS interface (see figure below) shows a navigation panel on the left which is directly accessing the JASMIN file systems. The main panel is where each notebook is displayed. Notebooks are separated into a series of cells, which can contain markup (i.e. documentation and notes) or code. Each code cell can be executed and outputs are displayed beneath them. In this example, a map of temperature change using CMIP6 data is created.

A suite of analysis software packages is pre-installed so that scientists can focus on the data processing. In cases where additional packages are needed, CEDA is developing notebooks to explain the installation procedure. These notebooks are also being developed to demonstrate use of certain libraries and how to work with data in the CEDA Archive.

The JNS is delivered through a stack of re-usable and cutting-edge technologies. The service is provisioned as a set of Docker containers which are managed using Kubernetes. This provides a dedicated share of the computational resource for each user session. Two-factor authentication is integrated with the JASMIN Accounts Portal so that the service is secure whilst re-using existing JASMIN login credentials.

Examples of analysis software packages that are pre-installed on the JASMIN Notebook Service.



The JASMIN Notebook Service interface, showing a map of global temperature change using CMIP6 data.

Helping our users

MIDAS data: improving user experience

HAYLEY GRAY

The Met Office Integrated Data Archive System (MIDAS) datasets hold some of the most popular weather data in the CEDA Archive. Over 5000 users, from a broad range of backgrounds, have accessed either the open access or the restricted access versions of these data over the last 12 months. However, using these data isn't always easy with many users encountering a range of challenges.

Work has been undertaken to address these challenges, including: clarifying the differences between the open and restricted datasets; production of a new user guide specifically for the MIDAS open dataset; and improvements to the station search tool to give clarity on which stations provide open data. Applying these changes has decreased the amount of time helpdesk staff spend on MIDAS queries - because the user community find the dataset much easier to use.

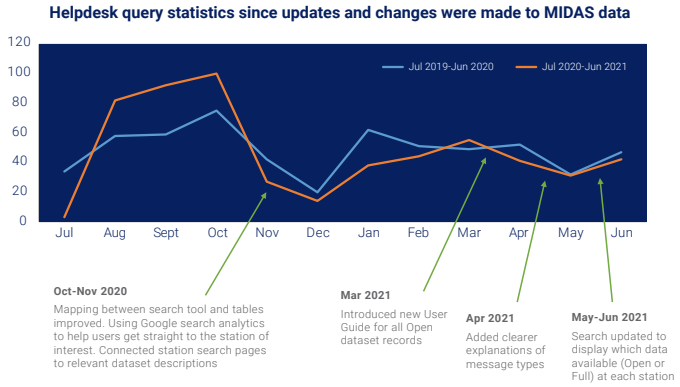


Figure: Shows an excel graph of the helpdesk query statistics compared to the previous reporting year since updates and changes were made to MIDAS data. The graph shows dips in the data correlating to the changes made as labelled meaning there are fewer queries coming into the helpdesk.

Evidencing our impact

POPPY TOWNSEND

CEDA is frequently required to evidence the impact it has on researchers and wider society. This was challenging without a formal collection process.

To understand how we could collect impact information, and to allow users to shape this process, over 500 users provided opinions, preferences and suggestions as to how to share impact, via focus groups and an online survey. We now ask users to provide impact information based on the findings.

Impact stories are regularly shared on our website and social media, and re-used by partner organisations if relevant.

A quote from a story about how CEDA services have been used to help predict seabird breeding abundance in a changing climate.

Supporting COVID-19 research

BRYAN LAWRENCE AND POPPY TOWNSEND

The coronavirus pandemic caused an unprecedented level of disruption to everyone's lives. The widespread health impacts required scientists across the world to work together to study the new respiratory virus. Researchers across UK Research and Innovation (UKRI) rose to the challenge and used complex computer models to better understand and predict the pandemic. This required access to increased compute resources – JASMIN is one of the infrastructures that was able to help. JASMIN supported interdisciplinary research teams who worked together to study the virus by providing collaborative workspaces for data storage and co-located processing. Access to our services increased the computing capacity available to COVID-19 related projects involving several UK universities and institutions.

JASMIN update 2021

18PB
New tape media

768
New computer cores

GPU facility
(More details available soon)

Increased support for small-file/ metadata-intensive workloads

Network improvements

30%
capacity increase for SOF storage

40%
capacity increase for object storage

125%
capacity increase for PFS storage

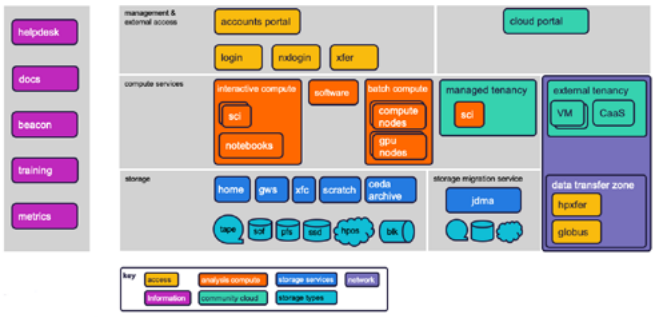
JASMIN continues to grow in all dimensions as the UK's environmental data analysis platform. Supporting over 240 science projects, it provides large-scale, collaborative compute and storage to facilitate the work of nearly 2,000 scientists across the UK and worldwide in addressing some of the most pressing issues of our time.

Running an infrastructure of JASMIN's size and complexity is challenging enough in a "normal" year, let alone during a global pandemic. With the hard work of the JASMIN team, within CEDA and in STFC's Scientific Computing Department, we kept services running: even introducing some new features to keep locked-down scientists occupied and able to continue their work.

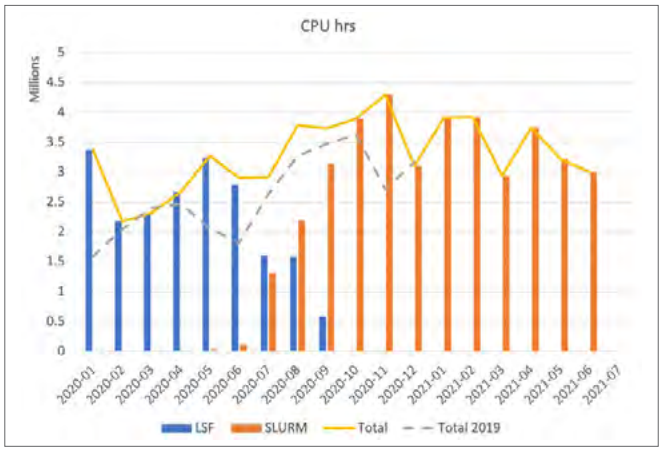
In 2020, users were able to benefit from:

- a new one Petabyte "scratch" storage area providing fast, temporary storage for users of JASMIN's batch compute cluster, LOTUS.
- special areas of solid-state disk to help projects store code and other small files more efficiently.
- improved diagrams used throughout documentation and training materials to help users orientate and navigate round the system
- a training course on how to use JASMIN, delivered as virtual workshop events but available online for self-service access.
- easier access from home networks to login servers and the graphical linux desktop service.
- building blocks for bespoke computing resources in the JASMIN Community Cloud. Cluster-as-a-Service enables projects to construct their own resources and workflows from pre-built components.
- early-adopter projects using new object storage in cloud-based workflows
- a small, proof-of-concept cluster of GPU processors, used to inform future expansion in this area to support machine learning and artificial intelligence workflows

Services



The service diagram introduced this year helps users orientate themselves and navigate through the system in training documentation.



Transition to SLURM as the scheduler for JASMIN's batch cluster, LOTUS. Vertical scale shows millions of CPU hours completed: blue bars for 2020 show LOTUS under the previous LSF scheduler and the transition to the SLURM scheduler during 2020/2021 (orange bars), with overall throughput increased and maintained. Grey dashed line shows the previous year (2019) for comparison. Some issues of high wait times due to overall load and seasonal peaks have been addressed with reconfiguration, but a wider review of demand management will follow.

Although delayed by side-effects of the pandemic, the same team that keeps everything running managed to ensure that new hardware and equipment for JASMIN's Phase 7 (2020-2021) upgrade was delivered, installed and is now being readied for use.

JASMIN is designed, integrated and operated by Science & Technology Facilities Council (STFC) on behalf of the Natural Environment Research Council (NERC). Architected jointly between STFC's Scientific Computing Department and the Centre for Environmental Data Analysis within RAL Space, it is operated and supported by a small but innovative team with expertise in computer science, research software engineering and environmental informatics.

Metrics and finance

CEDA exists to support the atmospheric, Earth Observation and near-Earth environment research communities in the UK and abroad through the provision of data management and access services. CEDA enhances this role through the development and maintenance of tools and services to aid data preservation, curation, discovery and visualisation; all of which add value for the world-wide user community.

The JASMIN data analysis facility provides petascale data-compute capabilities for the UK and wider environmental research communities. This section of the annual report presents summaries of CEDA Archive and JASMIN usage.

Usage of CEDA Data

CEDA delivers Data Archive services for the National Centre for Atmospheric Science (NCAS) and the National Centre for Earth Observation (NCEO). In addition, CEDA delivers the NERC/STFC funded UK Solar System Data Centre (UKSSDC) and the IPCC Data Distribution Centre for the Intergovernmental Panel on Climate Change (IPCC), as part of the NERC Environmental Data Service (EDS).

Annual CEDA archive usage: April 2020 – March 2021

Total number of users	32,000
Total data downloaded	650 TB
Total number of accesses	46,383,285
Total days activity	116,000

Note that a considerable amount of use of CEDA is by users on JASMIN, who would not be measured in most of these statistics because the data is directly available on the file system (and we are currently unable to gather these metrics).

We can break down the users accessing registered datasets by geographical origin and institute type. In total, there were over 12700 new user registrations. Note though, that many CEDA datasets are fully open so do not require user registration: those users are not captured here, unless they happen to have registered and logged in.

Yearly total of users by area

UK	8633
Europe	1368
Rest of the world	2411
Unknown	290

Yearly total of users by institute

University	9369
Government	1732
NERC	517
Other	787
Commercial	187
School	102

531 new datasets have been archived and made available via the CEDA catalogue – bringing the total number of datasets to 6782. The Archive now holds over 20 Petabytes of data.

Usage of JASMIN

JASMIN delivers analysis and compute capability for the environmental science community, predominantly for NCAS and NCEO, but also supports other NERC domains: oceanography, polar science, geology and earth sciences, ecology and hydrology.

As of March 2021 there were:
– 2410 JASMIN users
– 242 shared group workspaces, with nearly 20 Petabytes of storage allocated to them.

General overview

Supporting our users

Over 4650 queries were received by the helpdesk (covering both CEDA Archive and JASMIN services). These queries cover all aspects of data support except dataset/service applications or long term data management discussions. 1599 applications were processed for access to restricted datasets.

Collaborations

CEDA works closely with STFC's Scientific Computing Department to deliver the JASMIN infrastructure.
In 2020/2021, significant national and international collaborations have continued to support the international climate modelling community, EO and atmospheric research. On the national scale, CEDA itself reflects a collaboration between the earth observation community, the atmospheric sciences community (via NCEO and NCAS) and the space weather community.

Additionally, CEDA is:

1. Working closely with the other NERC Environmental Data Centres, as part of the NERC Environmental Data Service.
2. Operating and evolving the Earth System Grid Federation (ESGF) in partnership with the US Programme for Climate Model Diagnosis and Intercomparison and a range of global partners in support of the sixth Coupled Model Intercomparison Project (CMIP6).
4. Working with the wider UK atmospheric science and earth observation communities, via a range of projects, with NCAS and other NERC funding.
5. Working with the European Space Agency on projects such as the ESA Climate Change Initiative (CCI) Open Data Portal.
7. CEDA is part of the UK Collaborative Ground Segment for Sentinel data (with UKSA, Airbus, Satellite Applications Catapult) with the role to provide Sentinel data mirror archives and data processing capability for the UK academic community.

8. CEDA works with ECMWF to provide EO scientists with the high resolution atmospheric analyses they need to process satellite observations.
9. CEDA works with the UK Met Office to disseminate climate and weather data to the research community
10. With partners in Germany and the USA, CEDA provides data services on behalf of the IPCC (Intergovernmental Panel on Climate Change) through the Data Distribution Centre.
11. Supporting the Climate and Forecast Metadata (CF) Conventions with partners in University of Reading, UK Met Office, and multiple US research institutions.
12. With 20+ partners in the European Network for Earth System Modelling, CEDA is working to develop software and services for climate model data archives.
13. Working with academic partners in the UK Research and Innovation Cloud Working Group to share best practice, knowledge and strategy for use of cloud computing in the research domain.

Governance

The CEDA/JASMIN board reflects our funding arrangements: CEDA and JASMIN are NERC funded facilities based at STFC Rutherford Appleton Laboratory. NERC funding includes both central provision and support from NCAS and NCEO. Thus, all three, as well as the host organisation, are represented on the board.

Membership of the CEDA/JASMIN board (March 2021)

Bryan Lawrence, NCAS (Chair)
Poppy Townsend, CEDA, STFC (Secretary)
Victoria Bennett, Division Head CEDA, STFC
Chris Mutlow, Director RAL Space
Stephen Mobbs, Director NCAS
John Remedios, Director NCEO
Hartmut Boesch, Divisional Director, NCEO
Frances Collingborn, NERC
Michelle Odgers, UKSA
Tony Hey, SCD, STFC
Tom Griffin, SCD, STFC

The Board aims to meet at least annually, and up to semi-annually when necessary. This year, the CEDA-JASMIN board met in June and October 2020, and February 2021, with the next meeting planned in May 2021.

For CEDA's role in the NERC Environmental Data Service, there is additional governance and reporting, through the NERC Information Strategy Group and the NERC National Capability evaluation process.

Funding

In addition to supporting NCAS and NCEO, CEDA also delivers major projects with funding from a range of other bodies, including work for the European Space Agency (ESA), EC Copernicus Climate Change Service, BEIS, Defra and others, as well as participating and coordinating major European projects. Most of this funding comes to CEDA via a service level agreement (SLA) between NERC and STFC.

Annual total funding

Financial Year	13-14	14-15	15-16	16-17
NCAS income	829	829	808	808
NCEO income	392	390	393	393
Other NERC income	272	600	621	825
Other income	1486	1394	1505	1092
Total income	**2979**	**3213**	**3327**	**3118**

Financial Year	17-18	18-19	19-20	20-21
NCAS income	808	808	808	808
NCEO income	393	402	393	418
Other NERC income	816	733	883	784
Other income	1280	1458	1377	1476
Total income	**3297**	**3401**	**3461**	**3486**

Externally funded projects

The table below shows CEDA's Externally funded projects for 2020-2021 (non-core NERC) which were active during the reporting year.

Name	Description	Funder	Start date	End date	Value (£k)
ESA CCI Knowledge Exchange	Data archive for ESA Climate Change Initiative as part of wider activity including outreach and education	ESA	05/09/19	15/10/22	445
C3S_34a ESGF Data Node (CP4CDS)	Operational ESGF data node for C3S	C3S	01/09/16	31/05/20	1043
Pest Risk Modelling in Africa (PRISE)	JASMIN support for UKSA IPP project	UKSA	01/12/16	31/02/21	117.1
BEIS IPCC DDC	UK component of IPCC Data Distribution Centre	BEIS	26/09/18	31/03/21	262.2
C3S CORDEX4CDS	Regional Climate Projection data for C3S	C3S	01/05/17	01/10/21	184.6
C3S_34e CDS WPS Services	Designing an interface between CDS toolbox and remote processing using WPS	C3S	01/01/20	30/06/21	359.7
C3S_434 CDS to ClimateAdapt	Transferring information between Climate Data Store and European Environment Agency's portal	C3S	01/01/20	31/12/21	285
C3S_34g CMIP6	Including CMIP6 data in C3S Climate Data Store	C3S	01/02/20	30/04/21	168.3
MOHC Data Pipeline 18-21	Supporting CMIP climate model data movement from the Met Office to the CEDA archive and ESGF	BEIS	04/06/18	31/03/21	450
UKCP18 Services 20-21	To provide data services to support access to the next generation of climate projections for the UK	Met Office/ Defra	01/04/20	30/04/21	137.5
Support for Ensembles 20-21	CEDA support to multi-model climate ensemble archives	Met Office	01/04/20	31/03/21	50
UKSA DAP Support 20-21	Funding Esther Conway to attend ESA Data Access and Preservation WG for UKSA	UKSA	01/04/20	31/03/21	14
IS-ENES3	Phase 3 of the distributed e-infrastructure of the European Network for Earth System Modelling	H2020	01/01/19	31/12/22	748
C3S Oceans Data Archival	Data archival for C3S Oceans project (U. Reading)	CS3	01/01/19	30/06/21	19.5
ESA EPs Common Architecture	Consultancy for ESA Exploitation Platforms Common Architecture	ESA	06/11/18	28/02/21	97.5
JASMIN for ESA SST CCI+	JASMIN Support for ESA CCI+ Sea Surface Temperature processing	ESA	01/07/19	30/06/22	30
JASMIN for CCI WV	JASMIN Support for ESA CCI Water Vapour processing	ESA	01/01/19	31/12/20	10
ESA Digital Twin Earth Precursor	JASMIN support for development of an AI emulator for JULES land surface model	ESA	01/02/20	31/08/21	66.1
JASMIN for JNCC ARD Service	JASMIN Support for JNCC Sentinel ARD service provision	JNCC	01/03/20	31/03/22	33.7
JASMIN for JNCC Core	JASMIN Support for JNCC Sentinel ARD processing	JNCC	01/01/20	31/03/22	37
UKSA CDS Zone	JASMIN support for UK teams funded by C3S to produce Essential Climate Variable data for the C3S Climate Data Store	UKSA	01/04/20	31/03/21	14.8

Publications, posters and talks

Bennett, V., "Overview of CEDA and JASMIN", Presentation at NCAS / NCEO / MOHC Joint Forum on "Bringing together observations and models to improve our simulations of the coupled Earth System", 17th Nov 2020

Conway, E., "Jupyter Notebooks Survey results, Developing a Best Practice" CEOS WGISS-50, Virtual/Remote, September 22-24th, 2020

Conway, E. Heiko Balzer, Jupyter for CAPD – UK Overseas Development Aid, Agritech, etc., Presentation WGISS-49, April 21-23, 2020

Conway, E. "Jupyter Notebook Best Practice", Presentation at ESA Data Access and Preservation Working Group meeting, 22 March 2021

Moreno de Castro, M., Kulüke, M., Wachsmann, F., Kwee-Hinzmann, R., Kindermann, S., Nassisi, P., Levavasseur, G., Fiore, S., **Pascoe, C., Juckes, M**., Morellon, S., and Joussaume, S.: Skip high-volume data transfer and access free computing resources for your CMIP6 multi-model analyses, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-2513, https://doi.org/10.5194/egusphere-egu21-2513 , 2021.

Eggleton, F. and **Winfield, K**., "Open Data Challenges in Climate Science". Data Science Journal, 19(1), p.52. (2020), DOI: http://doi.org/10.5334/dsj-2020-052

Garland, W., Stephens, A., "Using CEDA for Air Quality Research", STFC Air Quality Network Webinar, 20 January 2020

Juckes, M., Taylor, K. E., Durack, P. J., Lawrence, B., Mizielinski, M. S., Pamment, A., Peterschmitt, J.-Y., Rixen, M., & Sénési, S. (2020). The CMIP6 Data Request (DREQ, Version 01.00.31). Geoscientific Model Development, 13(1), 201–224. https://doi.org/10.5194/gmd-13-201-2020

Juckes, M., Stockhause, M., Chen, R. S., and Xing, X.: IPCC Data Distribution Centre: FAIR data from Climate Research to Mitigation Policy, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-4737, https://doi.org/10.5194/egusphere-egu21-4737, 2021.

Kershaw, P., Halsall, K., Lawrence, B. N., **Bennett, V., Donegan, S., Iwi, A., Juckes, M**., Pechorro, E., **Petrie, R., Singleton, J., Stephens, A., Waterfall, A., Wilson, A**., & Wood, A. "Developing an Open Data Portal for the ESA Climate Change Initiative". Data Science Journal, 19, 16. (2020). https://doi.org/10.5334/dsj-2020-016

Kershaw, P., Matt Pryor JASMIN Jupyter Notebook Service Challenges and Opportunities , Presentation WGISS-49, April 21-23, 2020

Kershaw, P., A. Ghaleb, S. Ames, B. Evans, "ESGF Future Architecture Report", report published, July 2020, https://doi.org/10.5281/zenodo.3928222

Kershaw, P., "JASMIN, Hybrid Cloud and Bringing the Compute to the Data for ESGF", Session: Joint Panel Discussion 14 Sixth Symposium on US-International Partnerships - Update on CMIP-6 Data Delivery and Analysis Capabilities, AMS, February 2021

Kershaw, P., "JASMIN, Hybrid Cloud and bringing the Compute to the Data for ESGF", Session: TH073 - Update on CMIP-6 Data Status, Delivery and Analysis Capabilities, AGU, December 2020

Magagna, B., Buttigieg, P.L., **Pamment, A**., Moncoiffe, G., Stoica, M., 2020: Co-presented at I-ADOPT (InteroperAbility of Observable Property Terminologies) Working Group breakout session, RDA 16th Plenary Meeting (Costa Rica, virtual) November 9th - 12th 2020. Session notes: https://www.rd-alliance.org/plenaries/rda-16th-plenary-meeting-costa-rica-virtual/second-working-session-interoperable

Pascoe, C., Lawrence, B. N., Guilyardi, E., **Juckes, M**., & Taylor, K. E. "Documenting Numerical Experiments in Support of the Coupled Model Intercomparison Project Phase 6 (CMIP6)", Geoscientific Model Development, 13(5), 2149–2167 (2020) https://doi.org/10.5194/gmd-13-2149-2020

Popp, T., Michaela I. Hegglin, Rainer Hollmann, Fabrice Ardhuin, Annett Bartsch, Ana Bastos, **Victoria Bennett**, Jacqueline Boutin, Carsten Brockmann, Michael Buchwitz, Emilio Chuvieco, Philippe Ciais, Wouter Dorigo, Darren Ghent, Richard Jones, Thomas Lavergne, Christopher J. Merchant, Benoit Meyssignac, Frank Paul, Shaun Quegan, Shubha Sathyendranath, Tracy Scanlon, Marc Schröder, Stefan G. H. Simis, Ulrika Willén, "Consistency of satellite climate data records for Earth system monitoring", Bull. Amer. Meteor. Soc., Vol.101, Iss. 11, (2020), https://doi.org/10.1175/BAMS-D-19-0127.1

Pamment, A.: Presentation, "CF Standard Names", ESIP Marine Data Cluster meeting, 11 February 2021.

Pamment, A., "CF standard names", Presentation at 2020 CF workshop, 9 - 11 June 2020: http://cfconventions.org/Meetings/2020-workshop/CF_2020_standard_names_presentation.pdf

Pamment, A.: "CF Standard Names", Presentation, ESIP Summer Meeting 14 - 23 July, 2020.. Slides: https://doi.org/10.6084/m9.figshare.12736241.v1

Petrie, R., et al. including **Alan Iwi, Martin Juckes, Ag Stephens**, "Coordinating an operational data distribution network for CMIP6 data", Geosci. Model Dev., 14, 629–644, 2021, https://gmd.copernicus.org/articles/14/629/2021/gmd-14-629-2021.pdf

Sellar, A. et al. including **Ag Stephens**: Implementation of U.K. Earth System Models for CMIP6, Journal of Advances in Modelling Earth Systems, Vol. 12, Iss.4 (2020), https://doi.org/10.1029/2019MS001946

Smith, E.: "Reliable Operations on Climate Simulations" poster at STFC graduate conference, 12th March 2021

Stephens, A., Philip Kershaw, Alan Iwi, Matthew Jones, Bryan Lawrence, Neil Massey, Ruth Petrie, Matt Pryor, Eleanor Smith, "Cloud futures for CMIP data – evaluating object storage models and re-evaluating federation for data distribution", AGU, Session IN032-02, 14 December 2020,

Stephens A., Philip Kershaw, ESA Digital Twin Earth Climate Explorer Project on JASMIN, WGISS-51, Session on Data INTEROPERABILITY and USE, 21 April 2021

Stephens, A., Kershaw, P., Pritchard, M. "JASMIN: Infrastructure to support a diverse range of scientific use cases". Virtual workshop on requirements for a fast and scalable evaluation workflow - Preliminary, 19 May 2020.

Townsend, P. and C. Wilkinson: 'Gathering evidence of impact from research support services: Examining impact in the context of the Centre for Environmental Data Analysis', Research Evaluation, Dec 2020, rvaa031, https://doi.org/10.1093/reseval/rvaa031

Williamson, E., "JNCC Simple ARD Service", Webinar, July 2020: https://jncc.gov.uk/our-work/simple-ard-service-support-for-users/

Winfield, K., "Science and Engineering Careers Challenge", RAL Talk webinar, 7th August 2020

Winfield, K., Science and Engineering Careers workshop, Girlguiding groups. Various sessions throughout the year.

Winfield, K. "Early Careers Science and Engineering Careers Challenge", Appleton Space Conference 3rd December 2020