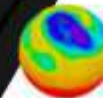


DIGITAL RESEARCH INFRASTRUCTURE WEBINAR SERIES

GETTING TO KNOW JASMIN AND CEDA

10 June 2026

ORGANISED BY:



Centre for Environmental
Data Analysis

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

JASMIN



Guest Speakers:

Alison Pamment, Senior Data Scientist, Centre for Environmental Data Analysis (CEDA)

Matt Pritchard, JASMIN User Services and Operations Manager

Hosted by: Digital Curation Centre (DCC)

A bit about the Digital Curation Centre

A centre of expertise in digital information curation, research data management, and Open Science since 2004.

INTERNATIONAL JOURNAL OF DIGITAL CURATION

<https://www.dcc.ac.uk/publications/ijdc>



<https://dmponline.dcc.ac.uk/>



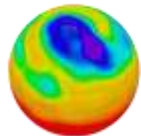
<https://dcc.ac.uk/events/idcc26>



<https://www.dcc.ac.uk/projects>

Training | Events | Tools | Advocacy | Consultancy | Guidance | Publications | Projects

A bit about you



**Centre for Environmental
Data Analysis**

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Getting to know CEDA and JASMIN

Alison Pamment and Matt Pritchard
June 2026



**National Centre for
Atmospheric Science**

NATURAL ENVIRONMENT RESEARCH COUNCIL

Overview of CEDA



CEDA and what we do

Mission: Deliver innovative Atmospheric and Earth Observation data services for the benefit of science and society



EDS CEDA Archive

- Long term data archive
- Over 20 Petabytes of atmospheric and Earth observation data



JASMIN

- Data intensive supercomputer
- Funded by NERC
- Jointly operated by STFC Scientific Computing & CEDA

Organisation relationships



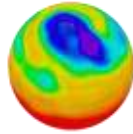
The NERC Environmental Data Service

We provide a focal point for scientific data and information spanning all environmental science domains. We are a network of distributed data centres, with domain specific expertise:



British
Oceanographic
Data Centre

British Oceanographic Data Centre
Discipline: Marine



**Centre for Environmental
Data Analysis**

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Centre for Environmental Data Analysis
Discipline: Atmospheric, Earth
Observation, and Solar and Space
Physics



UKCEH

**Environmental
Information
Data Centre**

Environmental Information Data Centre
Discipline: Terrestrial and Freshwater



NGDC

National Geoscience Data Centre
Discipline: Geoscience



**British
Antarctic Survey**

UK Polar Data Centre

UK Polar Data Centre
Discipline: Polar and Cryosphere

The NERC Environmental Data Service (EDS)

- The Environmental Data Service (EDS) is there to support and advise you on all aspects of Data Management, including data submission
- The NERC EDS is responsible for curating research data and making them *freely available* to all users with an interest in re-using them *for any purpose* (including commercial gain)

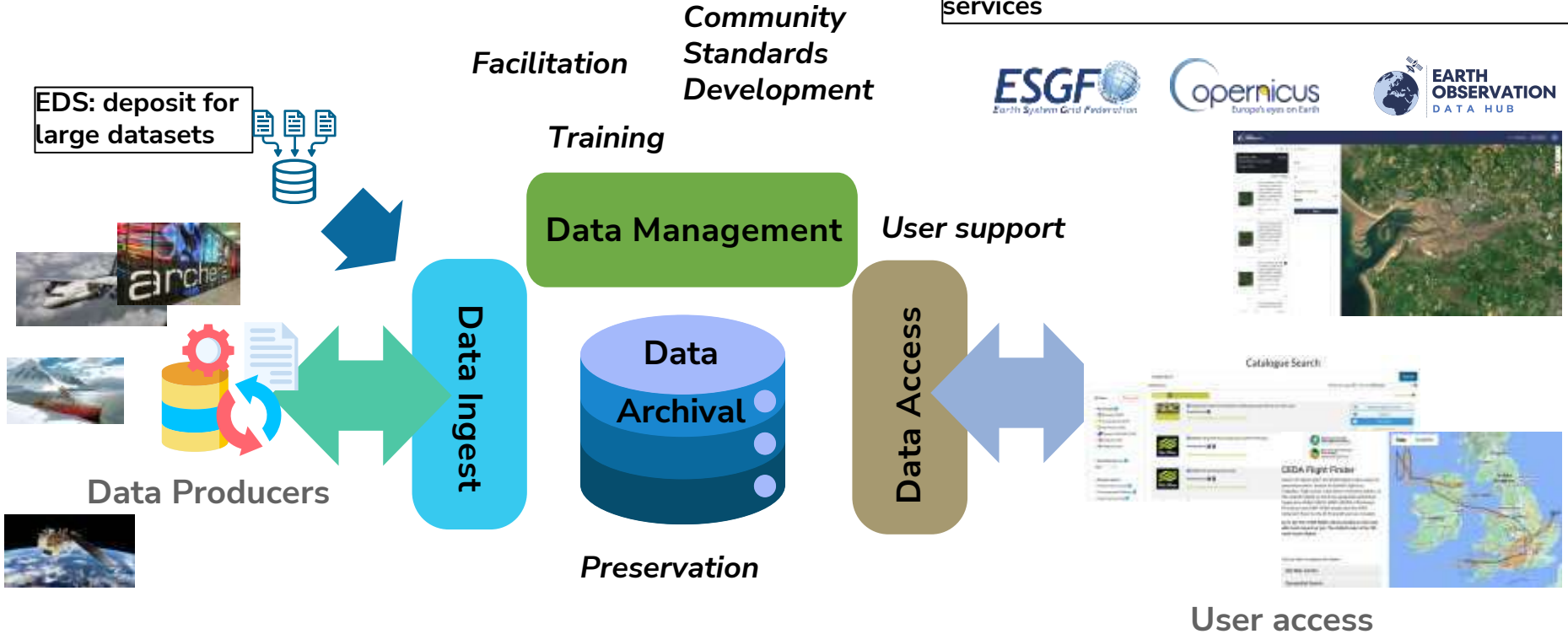


The CEDA Archive

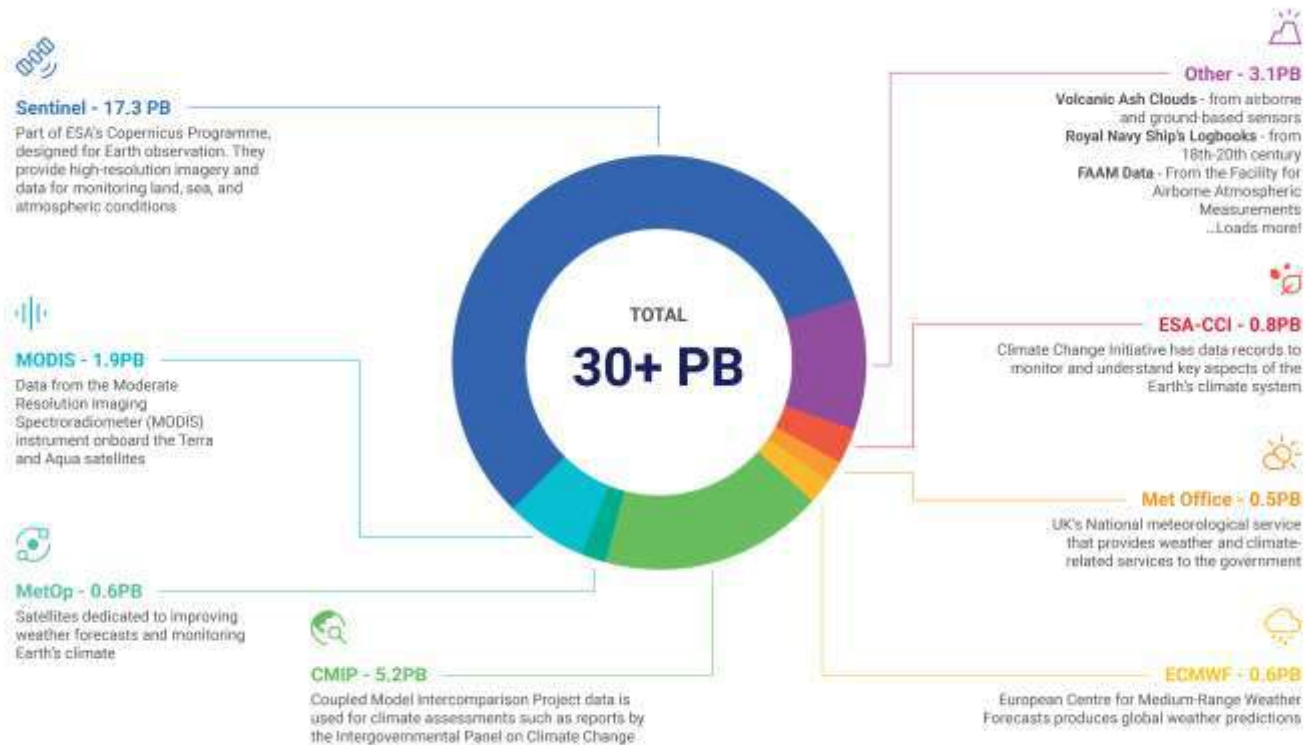


Our bread and butter...

Adding value – developing software tools and services

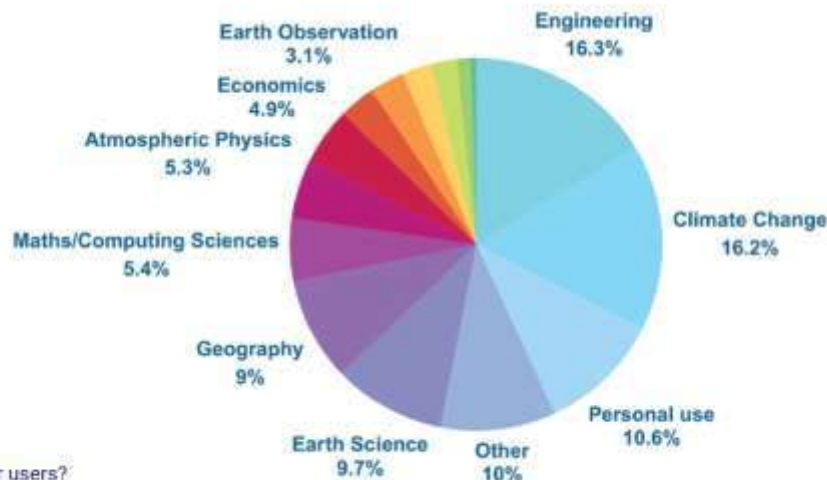


Data in the CEDA archive



1 Petabyte = 1,000,000,000,000,000 bytes i.e. 1×10^{15} bytes
1 PB \approx 20,000 Blu-ray disks, 30 PB \approx 600,000 Blu-ray disks

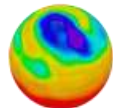
Who uses CEDA data?



Who are our users?
A pie chart showing CEDA
Archive users by domain.

In 2024-25:

- Data accessed almost 140 million times, with over 1.9 PB downloaded
- 330 active NERC projects
- Over 26 million files deposited into archive
- 263 new datasets published, bringing total to 9283

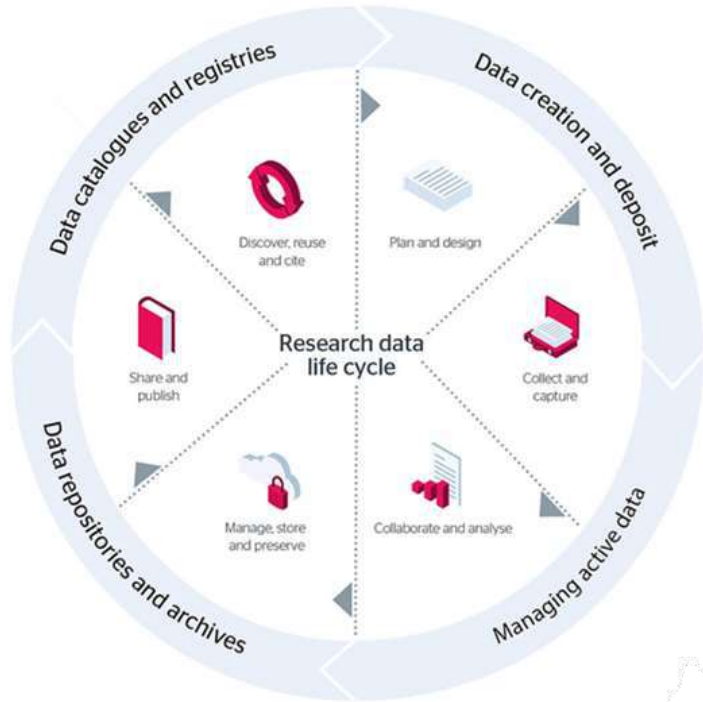


**Centre for Environmental
Data Analysis**

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL



Research data lifecycle



The data life cycle is the sequence of stages that a particular unit of data goes through from its initial generation or capture to its eventual archival and/or deletion at the end of its useful life.

- 1) Data Management planning
- 2) Data collection and capture, collection and analysis
- 3) Data storage and archiving, sharing and publishing
- 4) Data cataloguing, discovery and reuse



Data Management Plans (DMPs)

NERC Grant Requirements:

- DMP created at start of project
- Reviewed annually

The grant is assigned to a datacentre, who provide support to create and use the DMP

Project PI creates a DMP using the FAIR **Data Stewardship Wizard**, a questionnaire style knowledge model

Meant to be a living document that is can updated throughout the project



DSW

Less time and energy

Collaborative effort for all PIs

Standardised and structured

Machine-readable

Consensus across EDS

Easy updates

Extensive help resources



CEDA core supported file formats

<https://help.ceda.ac.uk/article/104-file-formats>

Gridded data

- NetCDF
- HDF
- GRIB
- PP Binary format

CSV/text data

- BADC CSV
- NASA Ames
- HITRAN
- JCAMP-DX

Image data

- JPEG/JPEG2000
- TIFF
- GeoTIFF

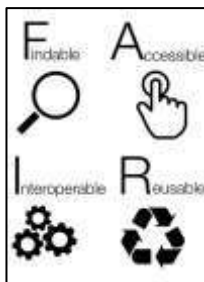
Metadata standards

CF Metadata conventions

CF + NetCDF = CFNetCDF

Controlled vocabularies

CF standard names for parameters



https://catalogue.ceda.ac.uk



precipitation Search

2028 Results Sort by Relevance ▾

« previous 1 2 3 4 ... 99 100 101 102 next »

Filter by record type ⓘ

- Datasets (1898)
- Dataset Collections (33)
- Projects (31)
- Instruments (26)
- Platforms (20)
- Computations (20)



  UKCP09: Met Office gridded land surface climate observations - precipitation and temperature indices at 5km resolution

[View parent collections](#)

[Register/Login for access](#)

[Explore](#)



[More Info](#)

  Global Precipitation Measurements (GPM) Integrated Multi-satellite Retrievals (IMERG) L3 Half Hourly 0.1 degree x 0.1 degree

[Register/Login for access](#)

[Explore](#)

[More Info](#)

  CMORPH 0.25 degree daily precipitation estimates

[View parent collections](#)

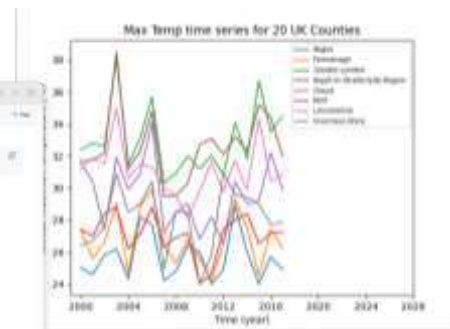
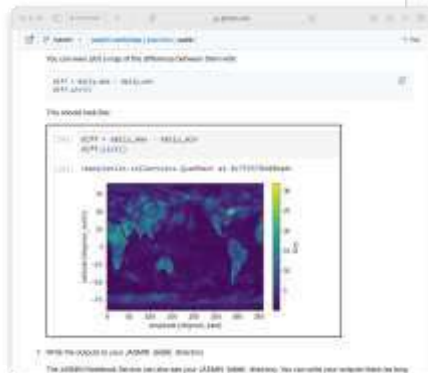
[Register/Login for access](#)

[Explore](#)

[More Info](#)

Training and skills development

Introduction to Scientific Computing Course



JASMIN Beginners workshops



EDS data management training



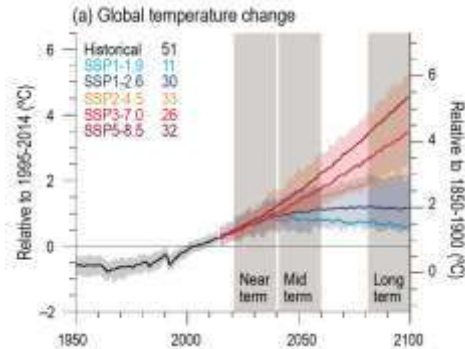
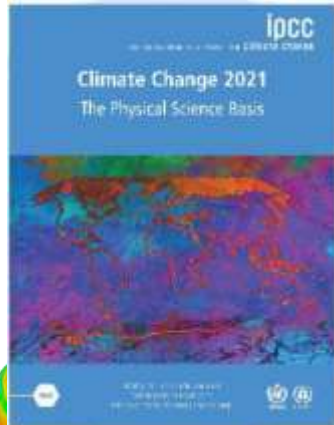
Projects at CEDA



CEDA support for CMIP7



- Data Request Task Team
- Controlled Vocabularies Task Team
- Climate Data Citation Task Team
- Climate model documentation team (Charlotte Pascoe co-lead)
- Climate data access Task Team
- ESGF development
- CEDA archive
- Direct access to data in IPCC report Figures and **complex citations** to provide the metadata (i.e. which datasets contributed to the final diagram)



IPCC Working Group I, 6th Assessment Report, Chapter 4, Figure 4.2a.



EO DataHub

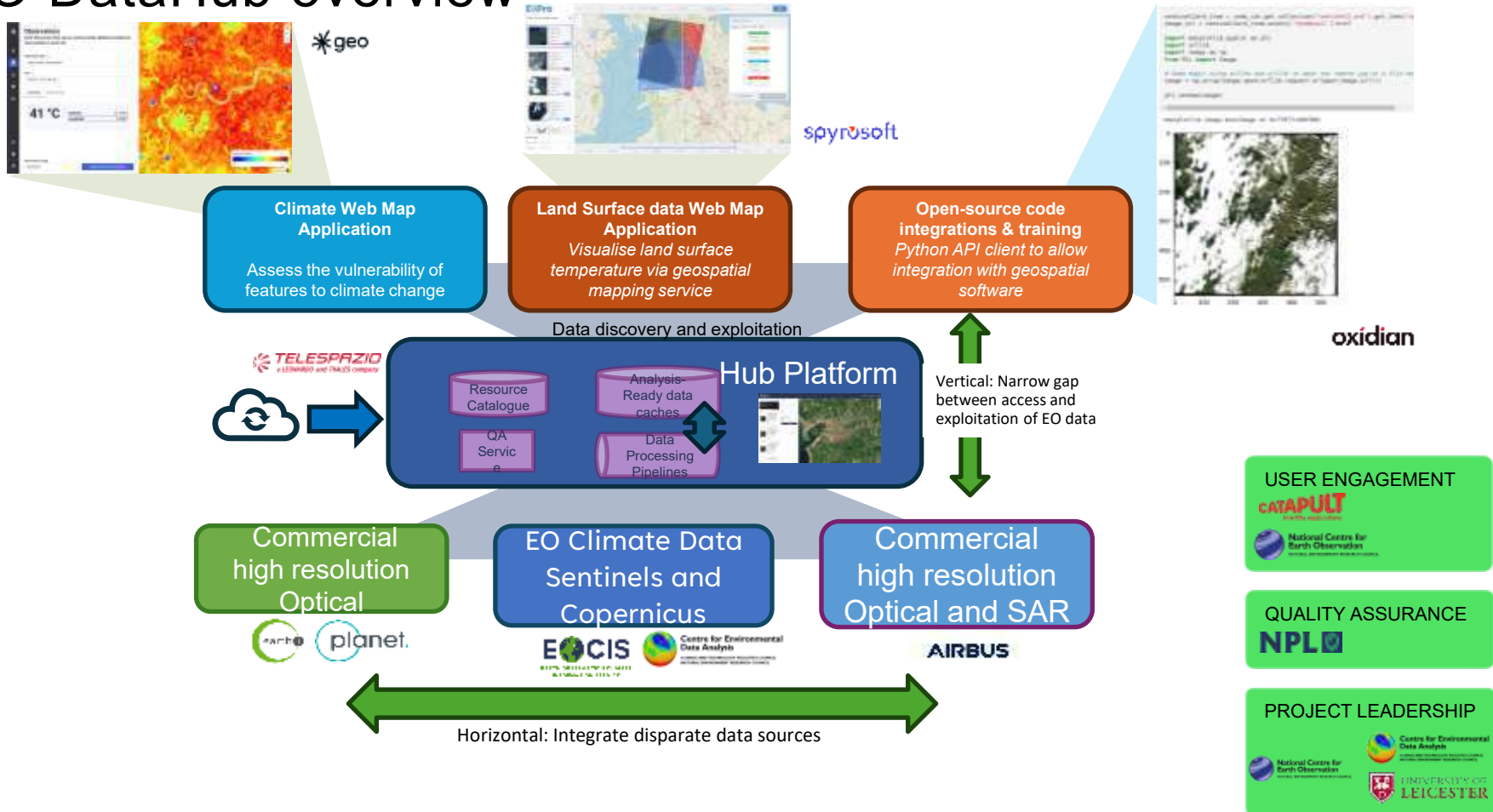
A UK Pathfinder project aiming to deliver easy access to EO data for effective decision-making across academia, government and business

Vision: is to deliver an innovative data and analysis platform that brings value across the breadth of UK EO sector users.

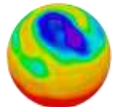
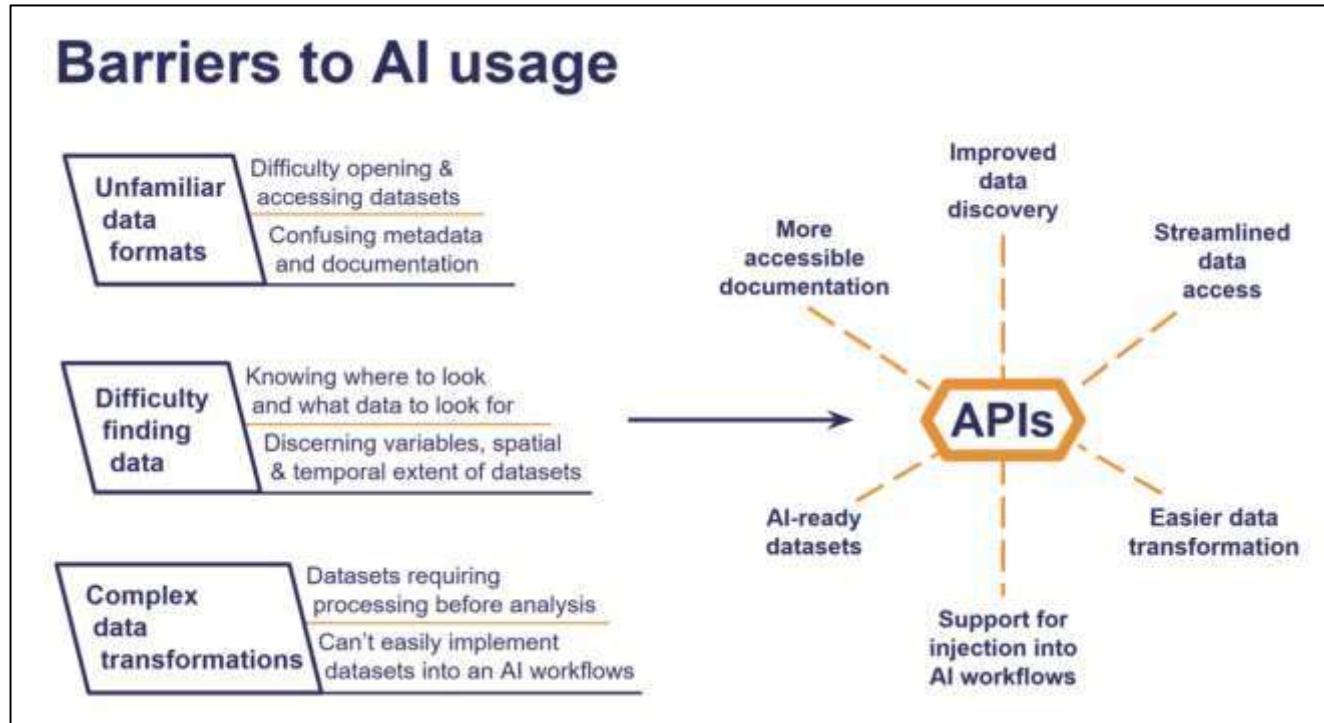
- 2-year project, started Feb '23, ~£10m investment funded from DSIT's Earth Observation Investment Package
- Goal – to address key barriers to effective exploitation of EO data: access and discovery, interoperability, transparency, and trustworthiness



EO DataHub overview



High-five for AI



JASMIN



JASMIN

The UK's data analysis platform for environmental science



Science and
Technology
Facilities Council



Natural
Environment
Research Council

Dr Matt Pritchard

Digital Curation Centre Webinar

June 2026

History

- Early 2012
 - CEDA Archive was outgrowing its hosting environment
 - Increasing need to support scientific workflows for analysing data, as well as “just” downloading data
 - To tackle this, JASMIN was created
- Design considerations
 - Avoid the need to move large volumes of data around
 - Enable collaboration around data
 - Provide appropriate balance between flexibility & performance

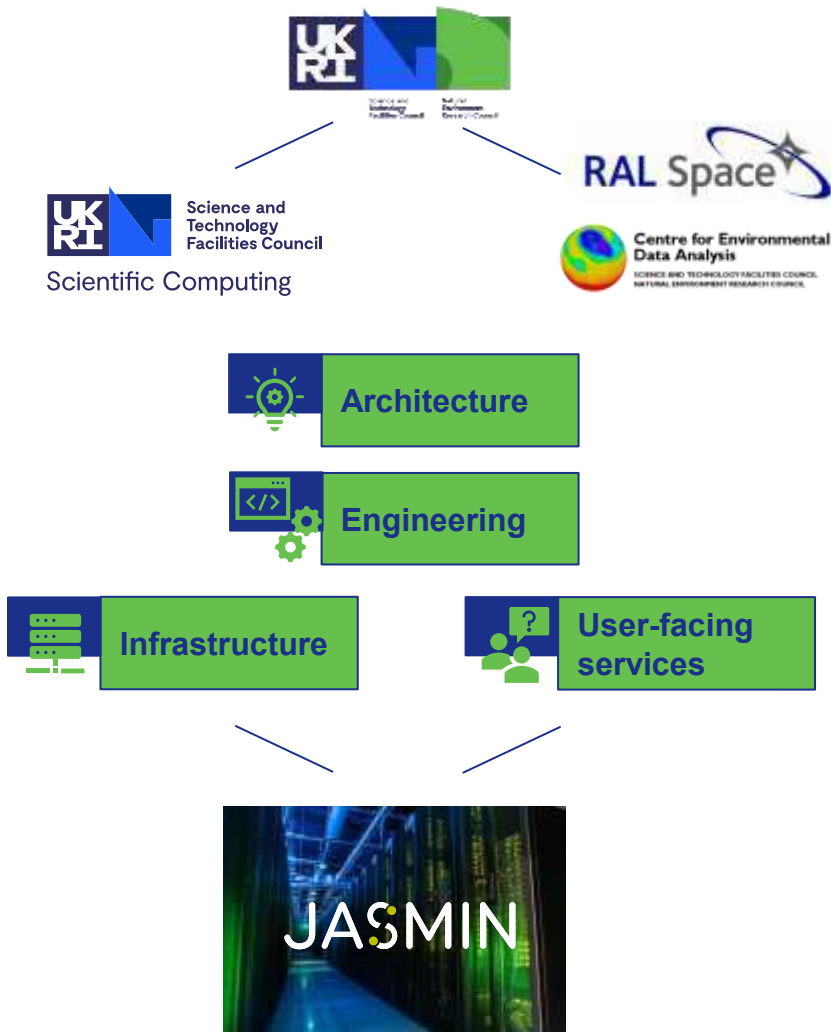


Organisation

- Funded primarily by NERC to support the environmental science community
- Infrastructure operated by STFC Scientific Computing
- User-facing services run by CEDA

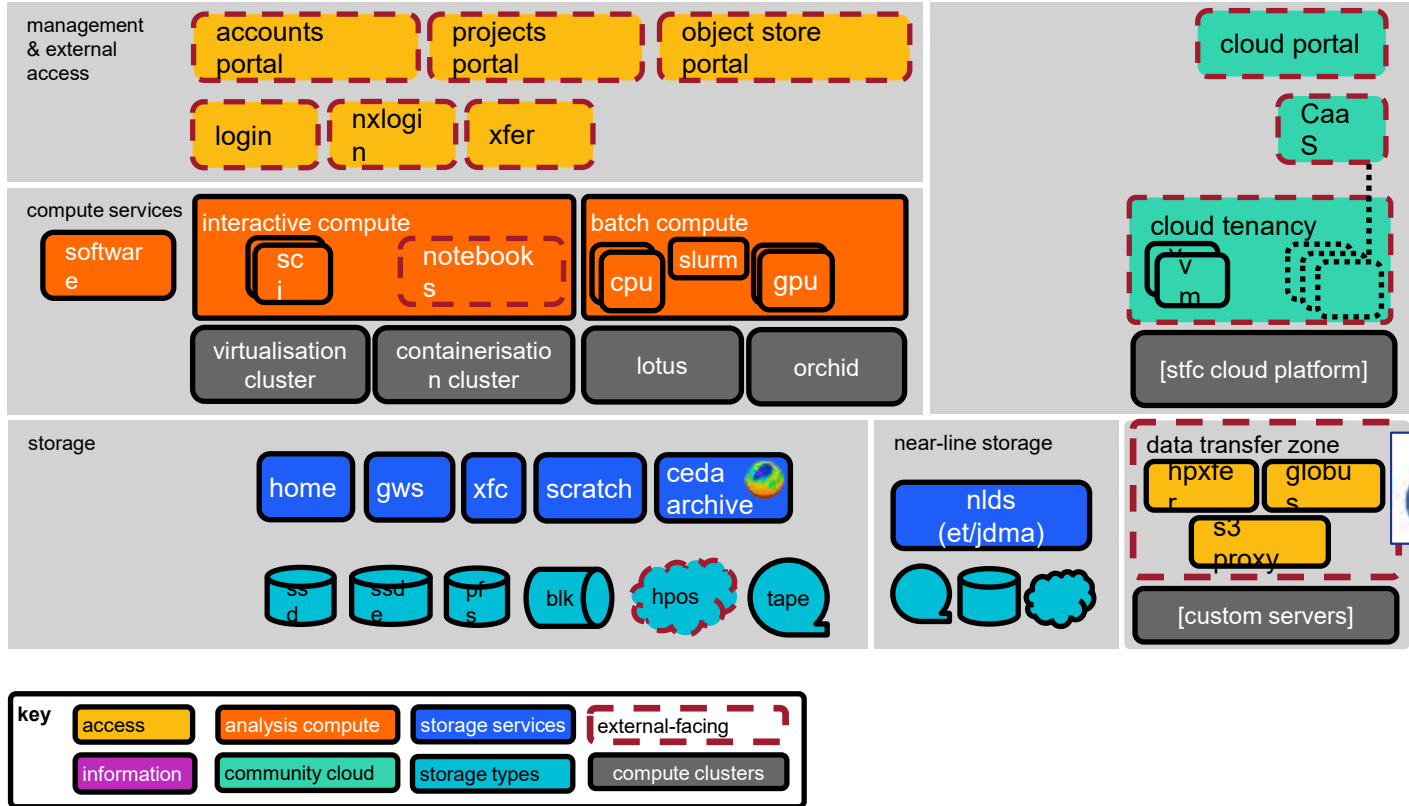
Result:

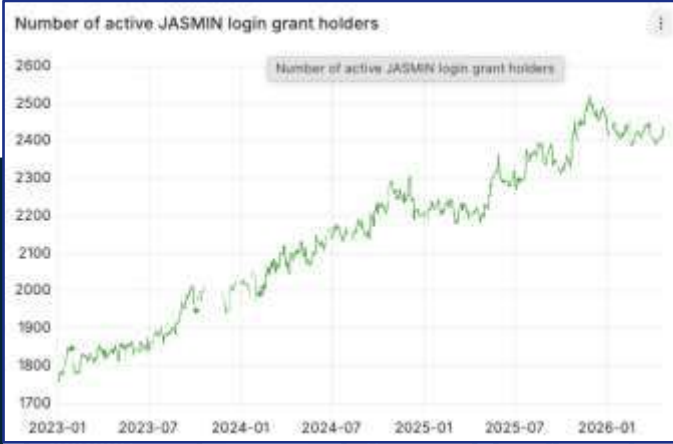
- Hosting environment for CEDA Archive & related services
- Co-located HPC data analysis platform
- Support from specialist teams



Services

- helpdesk
- docs
- beacon
- training
- dashboard
- status
- discussions





2400 users
400+ projects

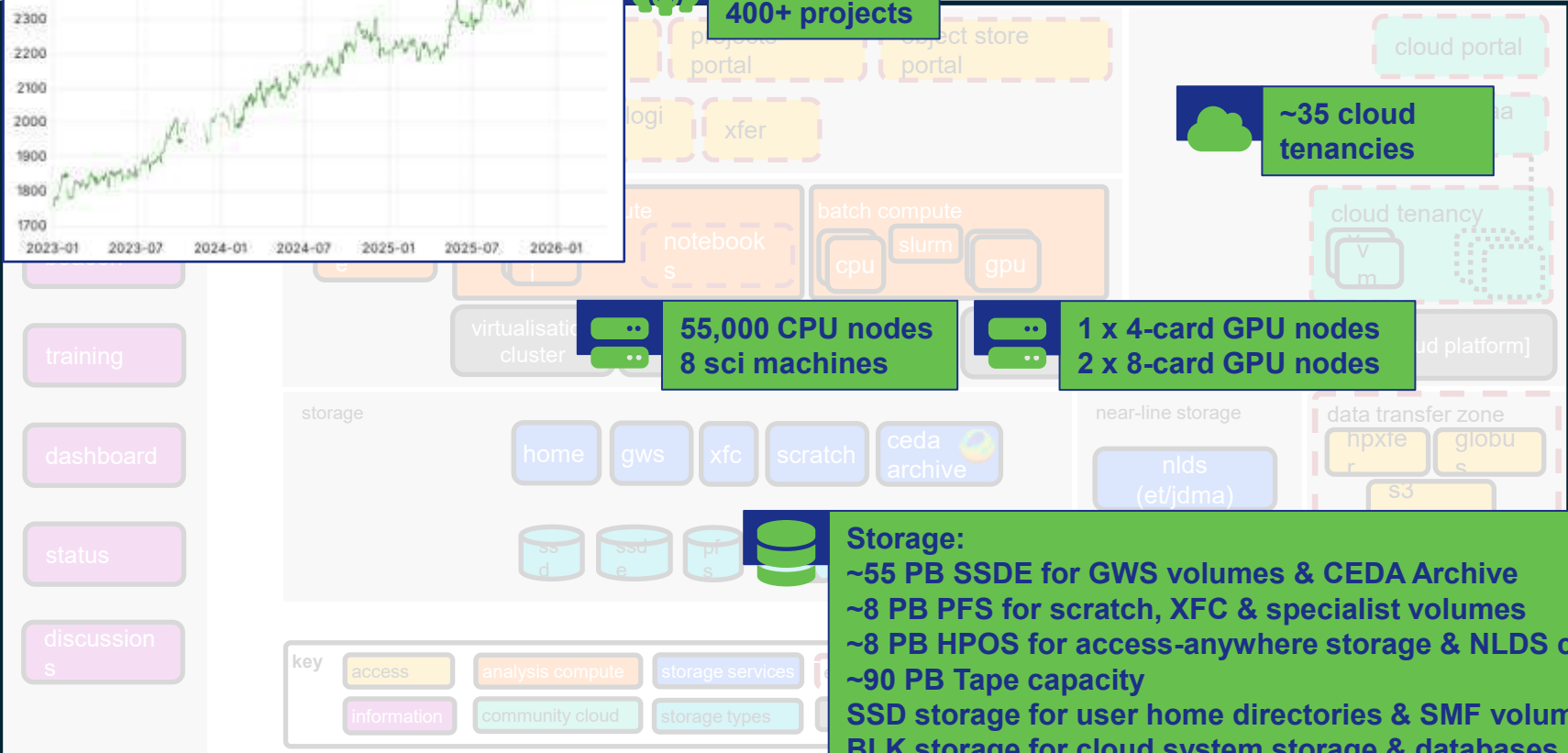
~350 kW
Power consumption

~35 cloud tenancies

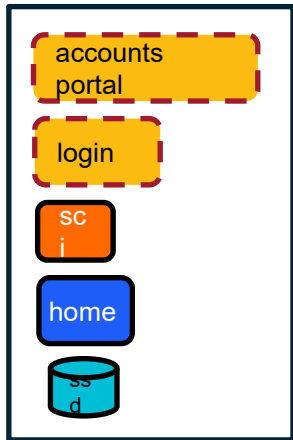
55,000 CPU nodes
8 sci machines

1 x 4-card GPU nodes
2 x 8-card GPU nodes

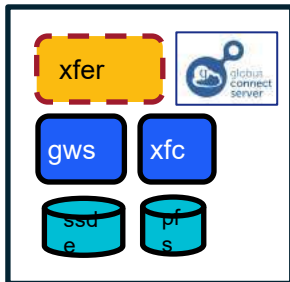
Storage:
 ~55 PB SSDE for GWS volumes & CEDA Archive
 ~8 PB PFS for scratch, XFC & specialist volumes
 ~8 PB HPOS for access-anywhere storage & NLDS cache
 ~90 PB Tape capacity
 SSD storage for user home directories & SMF volumes
 BLK storage for cloud system storage & databases



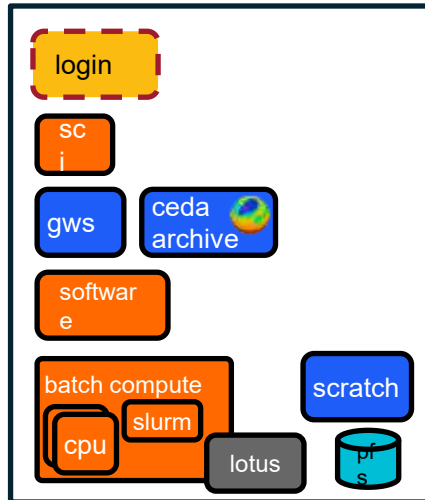
Different ways of using JASMIN



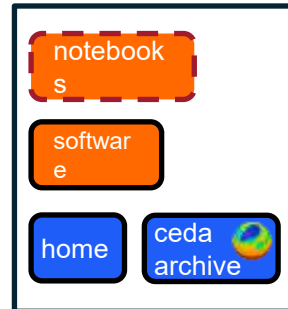
Log in & access home directory.



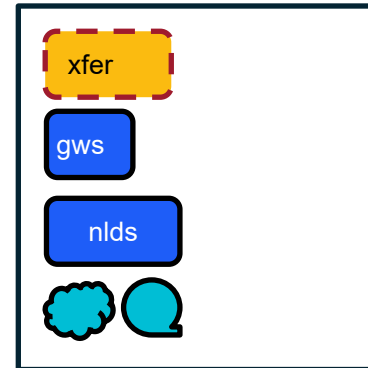
Transfer data, share with collaborators



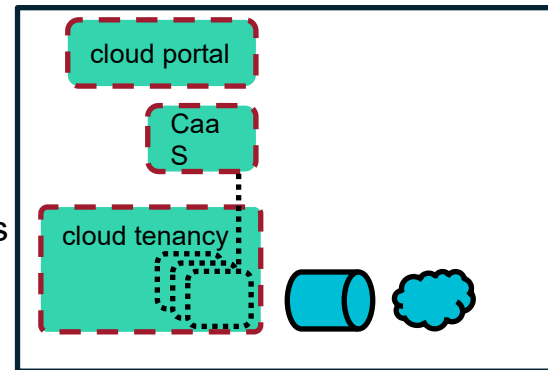
Develop a process on a sci server.
Process at scale on lotus cluster.



Visualise CEDA Archive data in a Jupyter Notebook.



Move data to/from near-line storage.



Develop cloud-native workflows for operational deployment elsewhere.

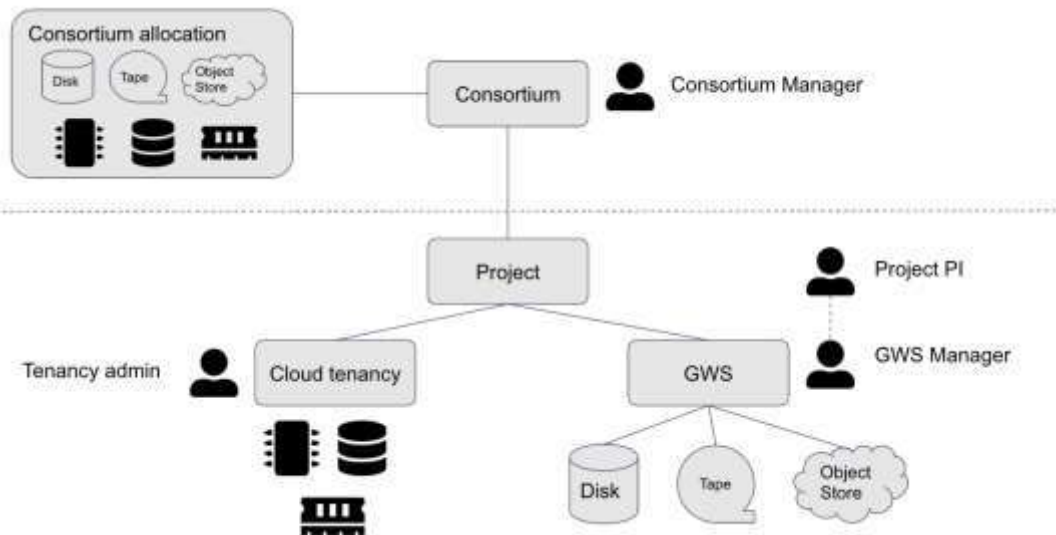
Resources for projects

The screenshot shows the JASMIN website home page. The top navigation bar includes 'Data', 'Discussions', and a search icon. A left-hand navigation menu lists various services: getting started, interactive computing, batch computing, software on journal, data transfer, short term project storage, long term archive storage, email, for cloud, tutorials, workflow management, and uncategorized. The main content area features the heading 'Requesting resources' with sub-sections for 'Requesting resources' and 'How projects and resources are managed'. The 'Requesting resources' section includes filters for 'Quota', 'Requirement', 'Ours', 'Workspace', and 'Tenancy', and a 'Share via' button. The 'How projects and resources are managed' section explains that resources are allocated to science communities through consortia and that project requirements should be discussed with Consortium Managers.

The screenshot shows the 'Projects List' page on the JASMIN website. The page title is 'Projects List' and it includes a sub-header 'View of all projects, including filters for tags and consortia. Some columns can be sorted.' Below this, there is a note about earliest and latest end dates for projects. The page features two filter buttons: 'aria -' and 'All Consortia -'. A table displays project details with columns for 'vm', 'Tags', 'Earliest End Date', 'Latest End Date', 'QuoBytes', 'Panasas', 'Cerning Object Store', 'PURE S50', 'Tape Storage', 'BLK storage for VIO7', 'RAM for VIO7', 'VCPU for VIO7', 'ext IP addresses', 'managed IP addresses', and 'PURE SSDE'. Two projects are listed: 'ARIA_GWANT' and 'PROMOTE'. The footer contains 'Useful links', 'Contact us' (with links to Documentation, Helpdesk, Twitter, and YouTube), and information about JASMIN's operation by the Science and Technology Facilities Council on behalf of the Natural Environment Research Council. Logos for STFC and NERC are present, along with a 'Help ?' button.

vm	Tags	Earliest End Date	Latest End Date	QuoBytes	Panasas	Cerning Object Store	PURE S50	Tape Storage	BLK storage for VIO7	RAM for VIO7	VCPU for VIO7	ext IP addresses	managed IP addresses	PURE SSDE
aria		2008-07-04	2008-07-04	20000	-	-	-	-	-	-	-	-	-	-
atmos		2001-03-31	2001-03-31	-	78000	-	-	440000	-	-	-	-	-	-

Community management of resources



JASMIN Consortia

Abbreviation	Name	Manager
atmos	Atmospheric Science	David Hassell, University of Reading
director	Director's cross-cutting activities	Adrian Hines, CEDA
ecohydro	Ecology & Hydrology	Mike Brown, UKCEH
eo clim	Earth Observation & Climate	Federica Moscato & Alison Waterfall, CEDA
genomics	Genomics	TBD
geology	Geology	TBD
npg	National Public Good	Adrian Hines, CEDA
oceanss	Ocean & Shelf Seas	Andrew Coward, NOC
poisci	Polar Science	TBD
semp	Solid Earth & Mineral Physics	John Brodholt, UCL
ukncsp	UK National Climate Science Partnership (was JWRCP)	Bryan Lawrence, University of Reading

Welcome to CEDA!

Status
Status of CEDA services

URL	Status	Summary	Address
2022-01-21 09:00	Operational	Interactive computing on the JASMIN Notebooks Service	2022-01-21 09:00
2022-01-21 09:00	Operational	Batch computing on the JASMIN Notebooks Service	2022-01-21 09:00
2022-01-21 09:00	Operational	Data transfer from JASMIN	2022-01-21 09:00
2022-01-21 09:00	Operational	Software on JASMIN	2022-01-21 09:00

News

- Update to JASMIN desktop
- REMOVED: JASMIN desktop software
- Use of the console carefully

Information

- CEDA News
- Status page
- JASMIN-USERS email list
- JASMIN Help docs
- Discussion fora on GitHub



JASMIN

Documentation sections

- About this site**: Information about this documentation site itself.
- Getting Started**: Key things to know to get your first batch job.
- Interactive computing**: How to use interactive computing resources.
- Batch Computing**: How to use the CORUS CPUs and DRACO GPUs clusters for batch processing.
- Software on JASMIN**: Information about what software is available and how to use it.
- Data transfer**: How to move data around from JASMIN.

JASMIN

JASMIN Notebooks Service

Getting started

Interactive computing

Batch computing

On this page

What is a Jupyter Notebook?

A Jupyter Notebook is an interactive document containing live code and visualizations that can be viewed and modified in a web browser. These documents can be shared, often using GitHub, and many projects distribute reusable code as Jupyter Notebooks. Users interact with their notebooks using the open-source Jupyter Notebook server application.

Training

Training events

- JASMIN Workshop
- Hackathons
- Tailored training
- Support for other events



Training accounts

Preconfigured with required access
Used for our training events
Can be “borrowed” for yours

JASMIN

JASMIN training accounts

In this article:

- Overview
- For event organisers
- For event attendees

Overview

What are training accounts?

JASMIN training accounts are TEMPORARY accounts that provide a partner with access to JASMIN for a specific event (such as a training workshop) organised in advance.

The benefits of using these accounts are:

- permissions/services are consistent across all training accounts

Current priorities (Spring 2026)

- Complete storage replacement
- ORCHID GPU cluster: develop case for replacement
- Environmental sustainability
- Exploring resilience options
- Increase support to Hackathons & 3rd-party training
- Developing closer links with other DRIs
- Working with other communities
- Further integration into UK Digital Research Infrastructure

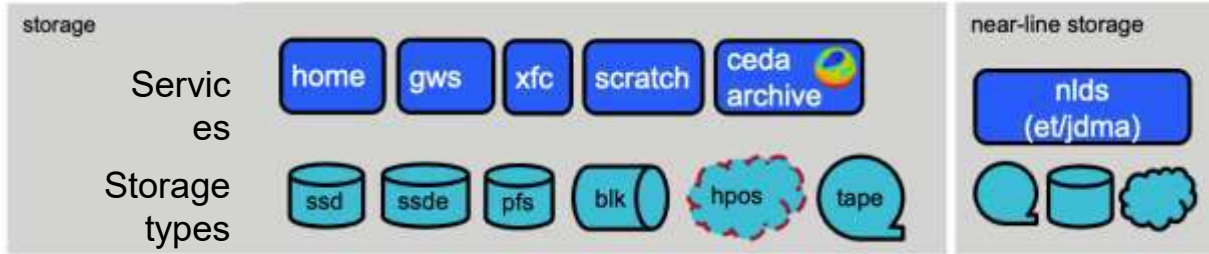


In more detail...

- Storage
- Compute
- Software
- Cloud
- Data transfer
- Your JASMIN account
- Expectations



Storage











- Mix of storage types to support several services

















Storage features



	Features				
Disk (file system)	Parallel Read	 ✓	 ✓	 ✓	 ✓
	Parallel Write	 OK except parallel netCDF	 OK except parallel netCDF	✓	✓
	Small files (<8Mb)	✓	✓		
	Other				Phasing out during 2026
Near-line	NLDS: JASMIN Group workspace data Also used for CEDA Archive backup & staging				
Object Store	Interface like Amazon S3. Disk-based, but simpler with no file system: just buckets & objects. Access from anywhere. Tape-like storage for GWS & cloud-based workflows.				
Block	For cloud tenancy storage and other internal uses (virtualisation, databases)				



Storage deployments

		Path(s)	Mounts						Backup	Other
			login	(hp)xfer	cloud	sci	nb	lotus orchid		
		/home/users/<username>	✓	✓	✓	✓	✓	✓	✓	100GB quota
	  	/gws/ssde/j25a/<gwsname> /gws/ssde/j25b/<gwsname> /gws/pw/j07/<gwsname> /gws/smf/j07/<gwsname> /gws/nopw/j04/<gwsname>		✓	✓	✓	✓	✓	✗	One or more volumes per project. Sized to project requirements (typically 1Tb – 1Pb) “SMF” vols on SSD were for small files, may not be needed now most GWSs moving to SSDE.
	 	/work/scratch-pw4 /work/scratch-pw5 /work/scratch-nopw2			✓	✓	✓	✓	✗	For use only as part of LOTUS/ORCHID cluster jobs. Auto-deletion of files older than 28 days.
		/work/xfc/vo1N/user_cache/<username>		✓	✓	✓	✓	✓	✗	For transient use in transfer tasks. User clean-up expected!
	 	/badc /neodc ..dc		✓	✓	✓	✓	✓	✓	Read-only access throughout JASMIN, if CEDA account linked. No need to copy archive data: process in-place.

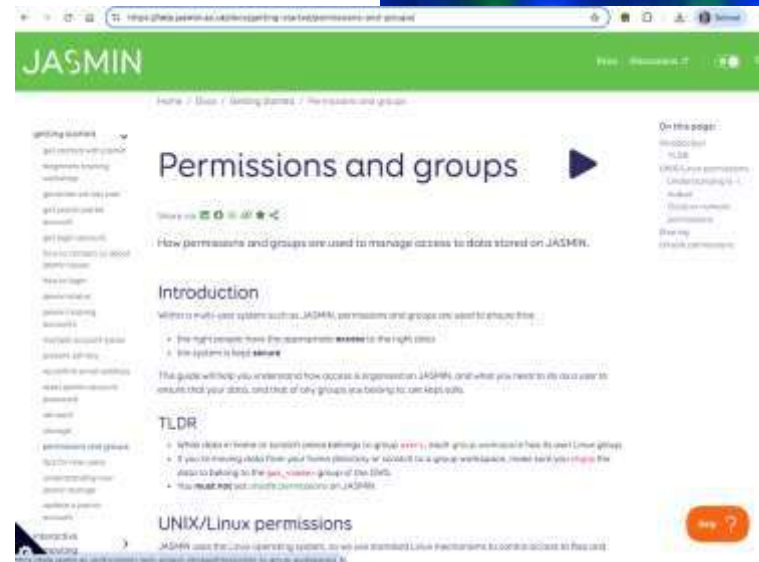
Other storage resources

- Object storage (S3)
 - Simpler than file system: buckets + put/get
 - Access-anywhere
- Near-Line Data Store (NLDS)
 - Command-line interface + S3 cache + Tape storage + catalogue
 - Interface to put / get data from local disk, manage holdings
 - Encouraged vs storing all data online
 - lower energy
 - lower cost
 - more easily scalable



Group Workspaces

- GWS Manager controls membership
 - You apply, manager can approve/deny
 - GWS Manager can also just “grant”
 - GWS has own unix group
- GWSs are NOT BACKED UP
- For science data only
- Not permanent
- Consortium manager



The screenshot shows the JASMIN website's 'Permissions and groups' page. The page title is 'Permissions and groups' and it includes a navigation breadcrumb: 'Home / Docs / Getting started / Permissions and groups'. The main content area is titled 'Permissions and groups' and contains an introduction section. The introduction states: 'Within a multi-user system such as JASMIN, permissions and groups are used to ensure that...'. It lists two bullet points: '• The right people have the maximum **access** to the right data' and '• The system is kept **secure**'. Below this, it says 'This guide will help you understand how access is organised on JASMIN, and what you need to do as a user to ensure that your data, and that of other groups you belong to, are kept safe.' There is also a 'TLDR' section with three bullet points: '• While data is in use or backup please always do `git pull --all` until git pulls everything back to user's local copy', '• If you're moving data from your home directory or `~/data` to a group workspace, make sure you **copy** the data to belong to the `jas_min` group of the LDRS', and '• You **must not** set chmod permissions on JASMIN'. At the bottom, there is a section for 'UNIX/Linux permissions' which states 'JASMIN uses the Linux operating system, so we use standard Linux mechanisms to control access to files and...'. On the right side of the page, there is a sidebar titled 'On this page:' with links to 'Introduction', 'Understanding...', 'Usernames and...', 'Groups and...', 'Permissions and...', and 'Share my...'. There is also a search icon in the bottom right corner.



Compute

- Sci servers
 - interactive processing
 - Compilation
 - Testing, estimation of resources needed
 - Job submission to LOTUS, monitoring
 - Common software stack installed
 - Not for huge processing!
- LOTUS batch cluster
 - for tasks which don't need to be run interactively
 - Job scheduler: **Slurm**, applies fair-share system
 - Same software as interactive compute
 - Same storage resources
 - Enables compilation on specific hardware

Sci machines for general use:

5 virtual machines

`sci-vm-0[1-5].jasmin.ac.uk`

2 physical machines

`sci-ph-0[1-3].jasmin.ac.uk`

- Over 55,000 cores (277 nodes)
- AMD Architecture
- Scheduler partitions (queues) with different attributes (QoS)
- MPI parallel processing possible

Training exercises:

<https://github.com/cedadev/jasmin-workshop>

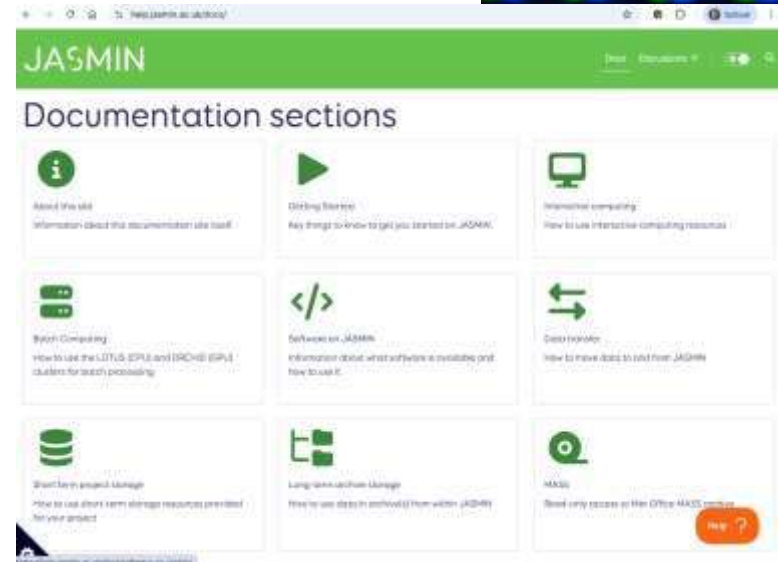
Notebooks service

- Use JASMIN interactively
 - Jupyter Notebooks: live code & output in web interface
 - Great for quick visualisations
 - Scalable
 - Now possible to use GPU too



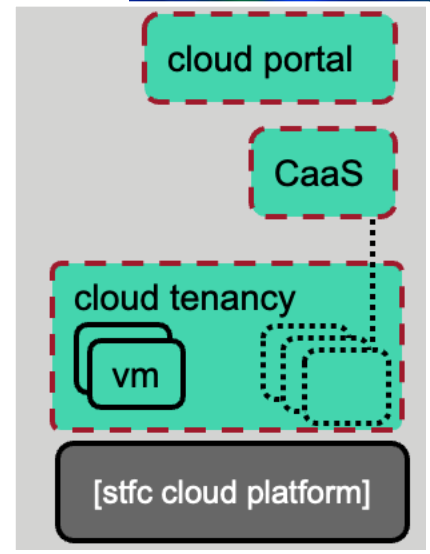
Software

- Analysis software common to batch & interactive compute
 - Stack developed with user community
 - **jaspy**, **jasr**, **jasmin-sci**
 - Additional tools for compiling and building software
 - Restricted software
 - Data movement software
-
- Same software available across
 - sci / notebooks / lotus



Cloud

- Separate part of infrastructure
 - STFC Cloud Platform: Openstack
- Create bespoke compute setup
 - VMs
 - Cluster-as-a-Service : use “building blocks”
 - Managed via “Azimuth” cloud portal
 - <https://github.com/azimuth-cloud/azimuth>



Your JASMIN account

- Each user has a JASMIN account
 - Profile to attach privileges for JASMIN services
 - **jasmin-login** = needed for SSH login access & Notebooks service
 - GWS membership(s)
 - Apply for services at JASMIN Accounts Portal
 - GWS membership is approved by GWS manager (not CEDA/JASMIN team)
 - Different to your CEDA account, if you have one
 - Needs to be linked for CEDA Archive access



The screenshot displays the JASMIN Accounts Portal interface. At the top, a green navigation bar contains the JASMIN logo and the text 'JASMIN Services'. Below this, the main content area is titled 'Login Services'. A sidebar on the left lists various service categories: 'My Services', 'Discover Services', 'Login Services', 'Sci Analysis VMs', 'Data Workspaces', 'Project VMs', 'Cloud Resources', 'Additional Services', and 'Cloud Store'. The 'Login Services' section is highlighted, showing a card for 'jasmin-login'. The card text reads: 'Access to the JASMIN shared services, i.e. login, transfer, scientific analysis servers, Jupyter notebook and LOTUS'. To the right of the card is a blue button labeled 'More information'.



Further information

- Help docs <https://help.jasmin.ac.uk>
- Workshop materials online <https://github.com/cedadev/jasmin-workshop>
- JASMIN website <https://www.jasmin.ac.uk>
- Status page <https://www.ceda.ac.uk/status> & message-of-the-day (MOTD) in terminal
- Discussion forum <https://github.com/cedadev/jasmin-community/discussions>
- Emails lists e.g. JASMIN-USERS
- Helpdesk support@jasmin.ac.uk (or use Help Beacon)
- Events www.ceda.ac.uk/events
 - User seminar series
 - Webinars
 - Training
 - Hackathons
 - Presence at community events (NCAS, NCEO, NERC-DG)



Expectations (yours and ours!)

- **JASMIN is a research infrastructure**

- At cutting edge of data-intensive environmental informatics
- Not a 24x7x365 enterprise
- Small team, no out-of-hours cover

- **More flexible than a supercomputer, but...**

- Relies on you being sensible
- You need to know what you're doing
- Test stuff before scaling up!
- Build failure-tolerance into your workflows

- **Please acknowledge JASMIN**

- <https://help.jasmin.ac.uk/docs/uncategorized/acknowledging-jasmin/>
- Tell your funding body what you've done with JASMIN
- ...And tell us! (impact stories)
- Helps ensure continued funding
- Makes the case for improving the service further



JASMIN



Useful Links and Contacts

CEDA

Documentation: <https://help.ceda.ac.uk/>



Email: support@ceda.ac.uk

JASMIN

Documentation: <https://help.jasmin.ac.uk/>



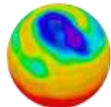
Email: support@jasmin.ac.uk

NERC EDS

Website: <https://eds.ukri.org/>



Email: data@nerc.ukri.org



**Centre for Environmental
Data Analysis**

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL



Thank you for listening!



**DIGITAL RESEARCH
INFRASTRUCTURE
WEBINAR SERIES**

**GETTING TO KNOW
ELIXIR UK**

17 June 2026

ORGANISED BY:



D | C | C

Thank you for joining!

Subscribe to our monthly newsletter Pipeline with RDM related news and events from around the globe

<https://www.dcc.ac.uk/form/dcc-associates>

 <https://twitter.com/digitalcuration>

 <https://www.linkedin.com/company/digital-curation-centre/>

 <https://bsky.app/profile/digitalcuration.bsky.social>