



New horizons for the Data Store Infrastructure at ECMWF

Edward Comyn-Platt, Angel Lopez Alos, James Varndell, Eduardo Damasio Da-Costa, Marcus Zanicchi, Gionata Biavati, Petrut Corbazan, Iryna Rozum, Baudouin Raoult, Ricardo Correa, Andre Obregon, and Chris Stewart



The CDS (and ADS) history

What are the CDS and ADS?

The CDS and ADS are single points of access to the wide range of quality-assured datasets produced by or in collaboration with the Copernicus Climate Change Service (C3S) or the Copernicus Atmospheric Monitoring Service (CAMS), respectively.

Objectives of the CDS/ADS



Make data discovery simple and relevant



Provide easy-to-use applications to explore data



Provide online capabilities to process the data



Enable reproducible and traceable research

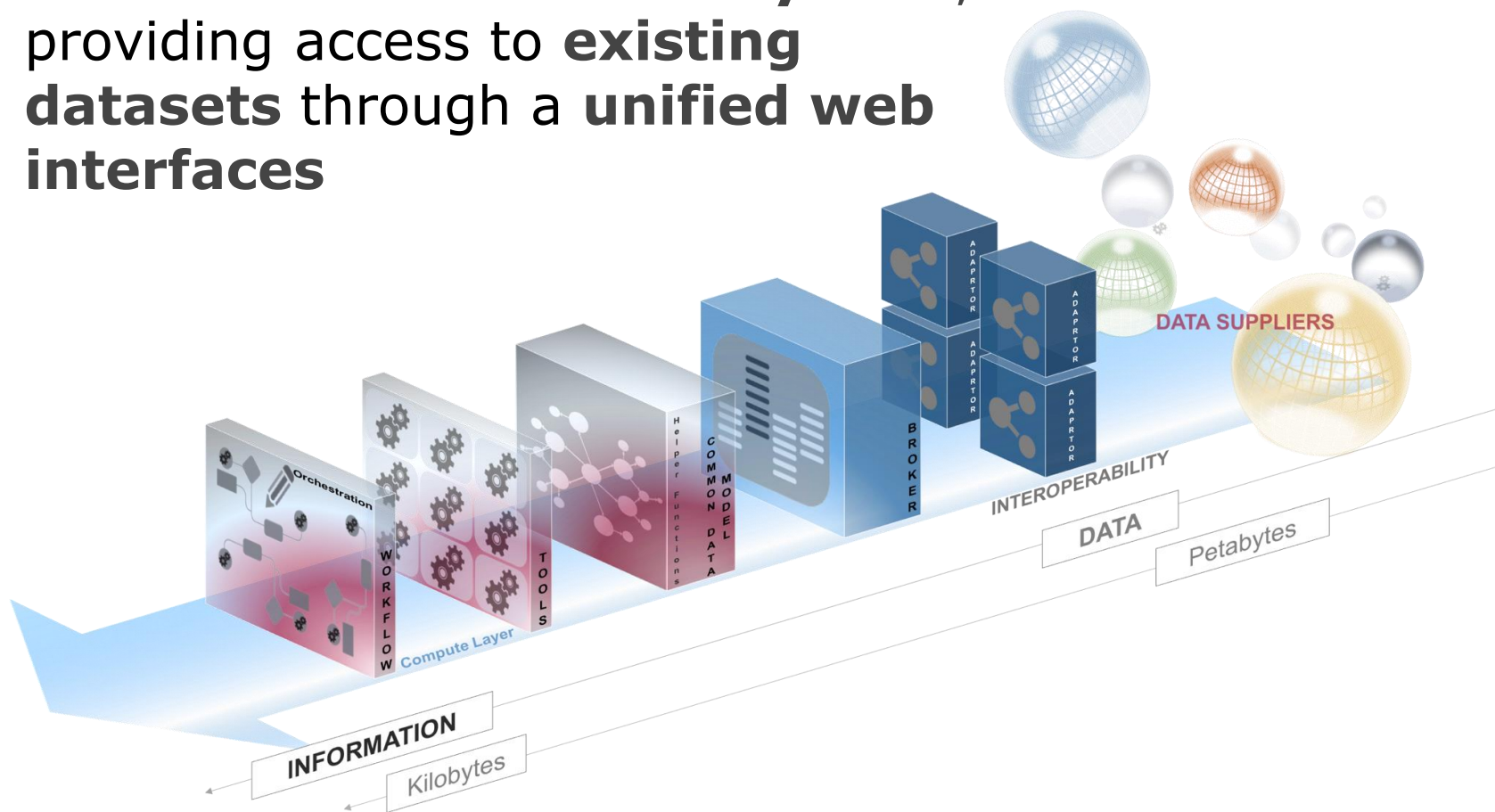


Users spend less time handling data

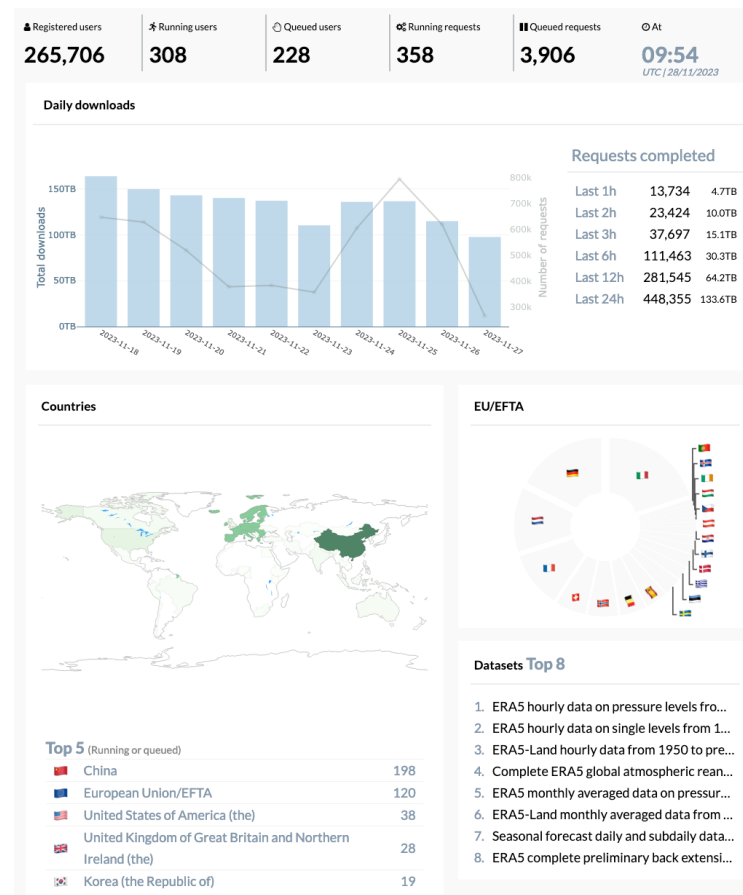


Conceptual design of the CDS

The CDS is a **distributed system**, providing access to **existing datasets** through a **unified web interfaces**



Typical performance





Modernisation of the system

Objectives of modernisation



Capitalize **experience, feedback and lessons learned** from 5 years of operation



Engage with a **broader user community** and make services more **accessible**



Ensure compatibility with **state-of-the-art solutions**



Embrace an open-development approach for **traceability** and **collaboration**



Strengthen synergies with related platforms and projects





Modernisation of the system

Flexible deployment

The Kubernetes deployment means that the system can be easily redeployed on other infrastructure for other projects.

The system is highly scalable and the object store approach downloads provides the fidelity required to serve the growing user base.

JupyterHub and earthkit for online processing

Jupyter notebooks have become a popular development environment for python users and there are a range of training resources available.

earthkit ensures synergy in public facing software developments of C3S and other ECMWF activities.



A FAIR Catalogue

Following the FAIR principles and adopting standardised webAPIs will increase uptake of the catalogue(s) and be compatible with machine learning methods



earthkit is the namespace for ECMWF open-source python code

Objectives of earthkit



Lower the barrier to entry



Make common tasks simple and efficient



Provide quality assurance to computation

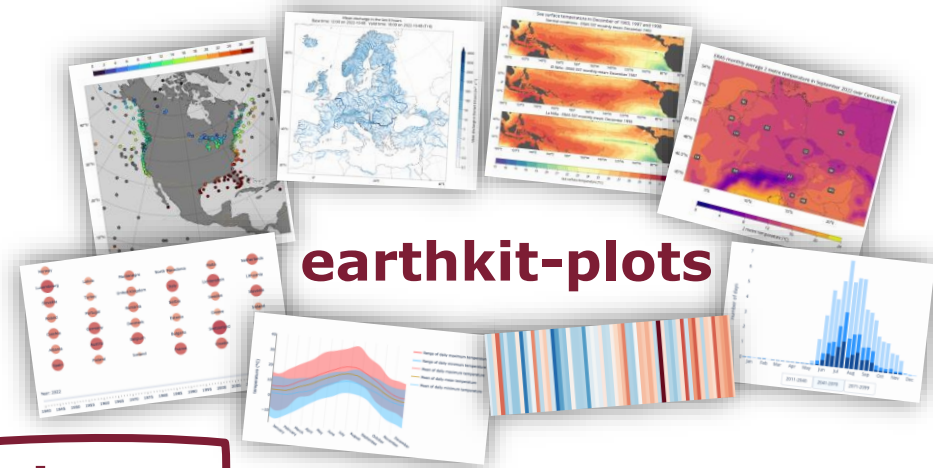


Interoperable with data and tools from many sources



Embrace open-source approach for traceability and collaboration

<https://earthkit.readthedocs.io>



earthkit-data
A polymorphic data object for handling many data formats

earthkit-aggregate
Quality assured methods for aggregating data in time and space

earthkit-regrid
A package for regriding data using high-performance algorithms

earthkit-meteo
Quality assured meteorological and hydrological calculations

earthkit-TBC
There are more earthkit packages in the pipeline for 2024

Documentation, Notebooks and Training Material



Demonstrate how to access and use C3S and CAMS data effectively



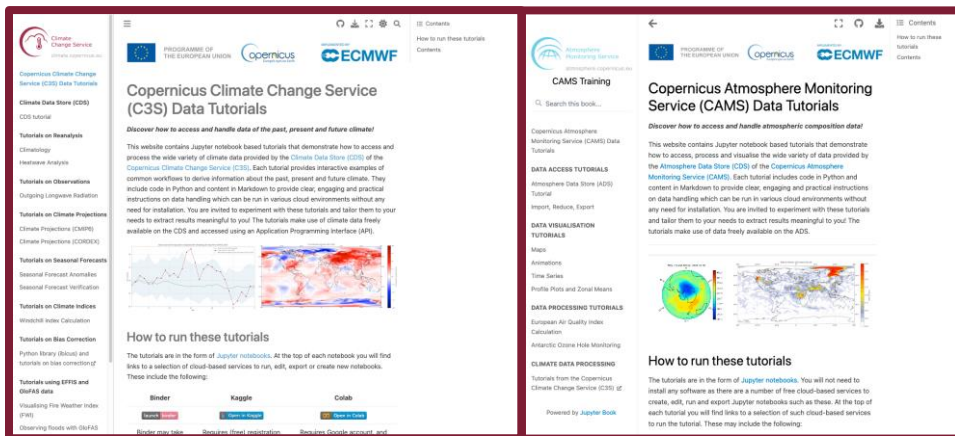
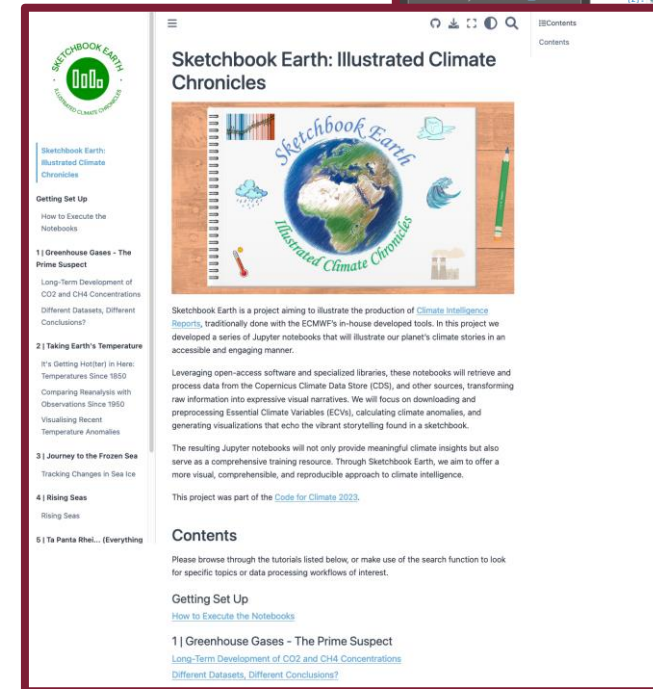
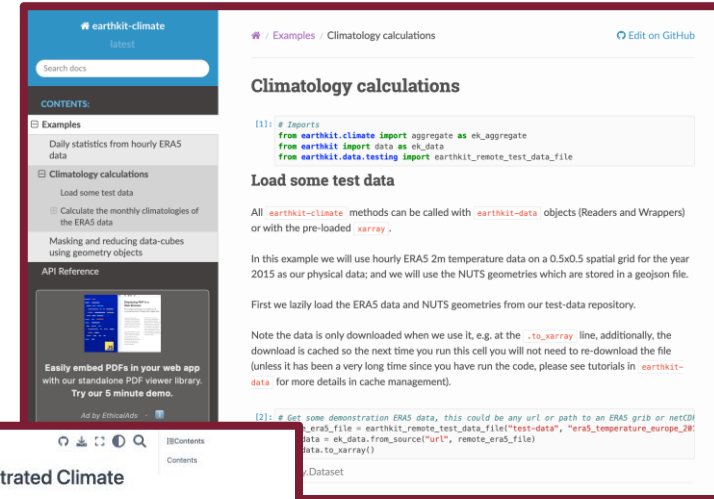
Highlight the important features and applications of data products



Document public facing software packages with in-context examples



Provide traceability for published reports and documents



Evaluation and Quality Control

The EQC component will have a more prominent role in the modernised web-portal. The information provided has been simplified and consolidated into three distinct components.



Quality assurance

Based on explicit requirements
Simple checklist with details on demand



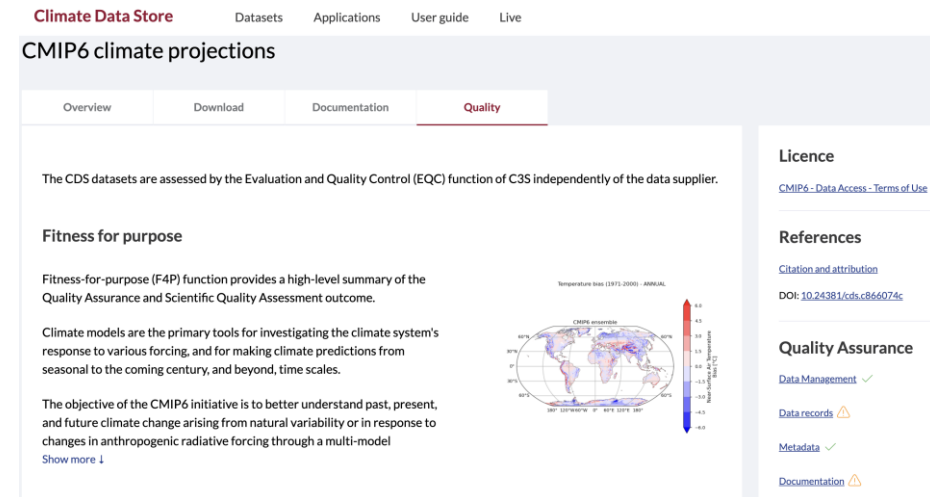
Quality assessment

Addressing specific user questions
Providing scientific expertise



Fitness for purpose

Succinct overview with implications for best practice



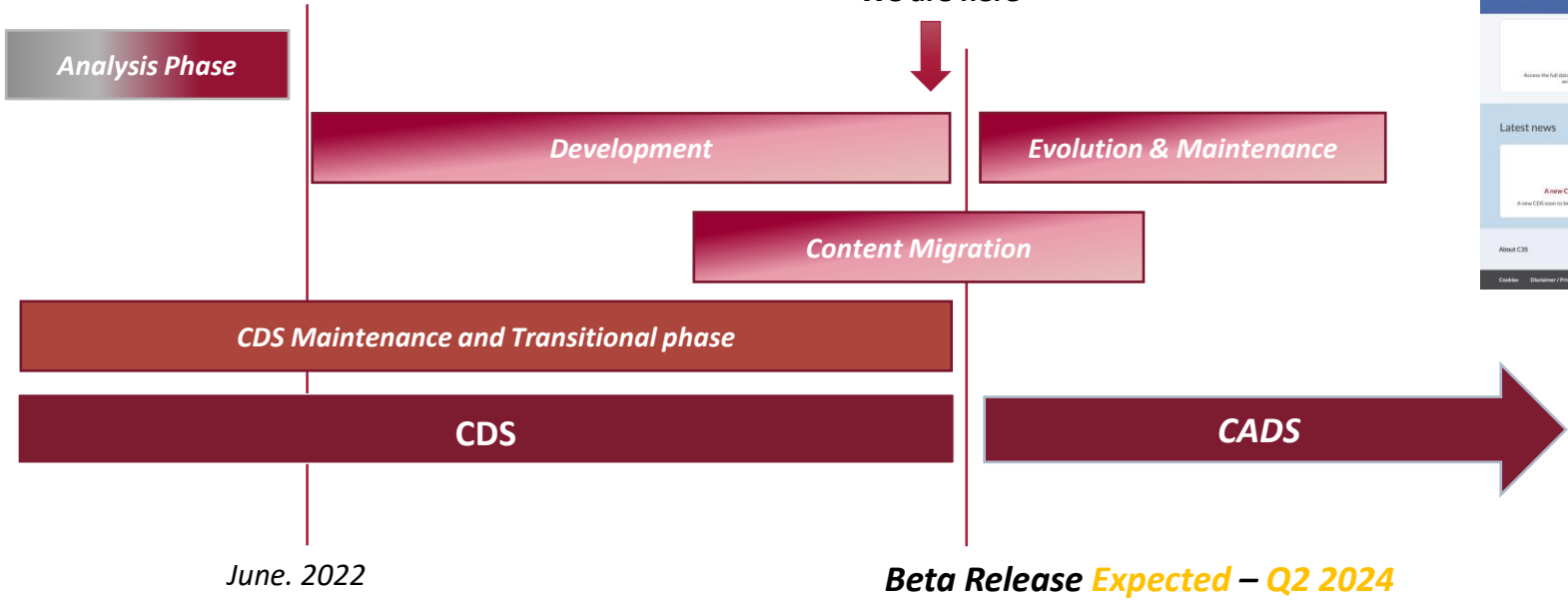
“How, and how well, can I use these data for my purpose?”



Climate Change

Any questions?

Roadmap



The image shows two screenshots of the Climate Data Store website. The top screenshot displays search results for 'Crop productivity and evapotranspiration indicators from 2000 to present derived from satellite observations'. The bottom screenshot shows the details for 'ERAS hourly data on single levels from 1940 to present', including an overview, download options, and a license section.