EGU24-3937, updated on 09 Nov 2024
https://doi.org/10.5194/egusphere-egu24-3937
EGU General Assembly 2024

New horizons for the Data Store Infrastructure at ECMWF

Angel Lopez Alos, Baudouin raoult, Ricardo correa, Andre obregon, Chris stewart, James varndell, **Edward comyn-platt**, Eduardo damasio da-costa, and Marcus Zanacchi
European Centre for Medium Range Weather Forecast (ECMWF), Reading, United Kingdom

Since its official launch in 2018 supporting the implementation of the Copernicus Climate Change Service (C3S), the Climate Data Store (CDS) software infrastructure has evolved in many ways driven by an expanding catalogue of resources, a growing user community and the evolution of technologies and standards. On 2020 a twin instance, the Atmosphere Data Store (ADS), as support of the Atmosphere Monitoring Service (CAMS) was released. Since then, Infrastructure was renamed as Climate and Atmosphere Data Store (CADS). Combined, CDS and ADS, provide nowadays service to more than 270k registered users, delivering over 130 TBs of data on daily average in the form of more than 700k processed requests.

In 2024, a modernized CADS will take over. A configurable framework built on cloud oriented and state-of-the-art technologies providing more scalable, wider, and open access to data and services which will foster the engagement with a broader user community and will facilitate interaction with different platforms in the future EU Green Deal Data Space.

Despite changes, CADS foundational principles of simplicity and consistency remains along with FAIR. A rigorous content management methodology is at the core of the system, supported by automatic deployment tools and configuration files that range from web portal content to metadata, interactive forms, dynamic constraints, documentation, adaptors, and quality control. This versatile mechanism provides huge flexibility for adaptation to different standards and FAIR principles.

In addition to improved capabilities for discovery, search and retrieve, the modernized system brings new or re-engineered components aiming to improve the usability of resources, such as compliant OGC APIs, integrated and interactive Evaluation and Quality Control (EQC) function, open-source expert python packages (earthkit) for climate and meteorological purposes able to deploy & run anywhere, or Serverless Analysis-Ready Cloud Optimized (ARCO) Data and Metadata Services supporting responsive WMS/WMTS interfaces.

Modernization also involves the underlaying Cloud Infrastructure which aligned with the ECMWF's Strategy for a Common Cloud Infrastructure (CCI) brings extended compute and storage resources and more importantly, closer, and efficient access to ECMWF resources, data, and services.

All new capabilities combined power a new generation of interactive user applications, training material, EQCs functions, and efficient access mechanisms to large data volumes driven among

others by ML/AI requirements.

Here we describe the new horizons that the modernized Data Store infrastructure open to users, introduce the broad spectrum of functionalities, open-source code, and material currently available and we open for debate the expectations and requirements that will foster the future evolution of the different components of the infrastructure.