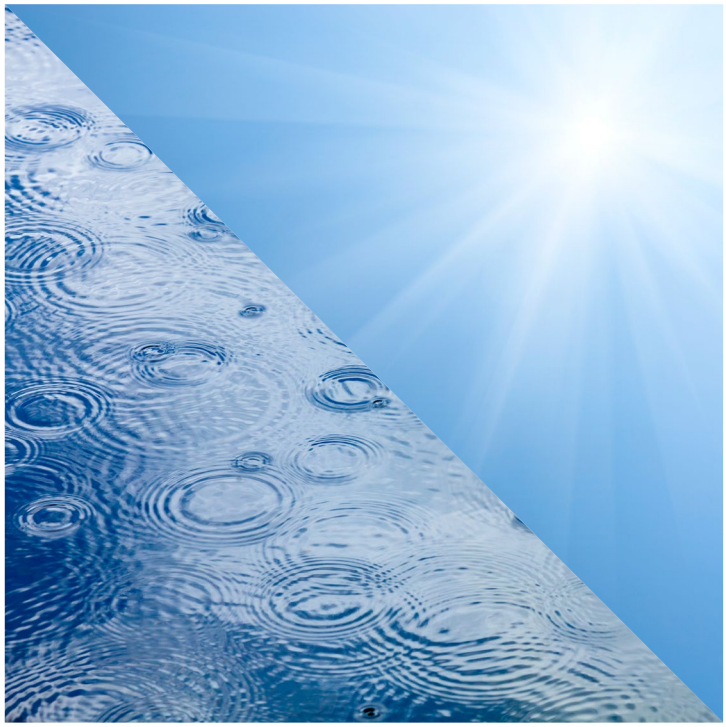
ECMWF Feature article

from Newsletter Number 184 – Summer 2025

COMPUTING

ECMWF's improving data services in the era of cloud computing and machine learning



Cover photos: naumoid & Andrey_Ivanov/iStock/Getty Images Plus

www.ecmwf.int/en/about/media-centre/media-resources

doi: 10.21957/73e53hn4ss

This article appeared in the Computing section of ECMWF Newsletter No. 184 – Summer 2025, pp. 37–40

ECMWF's improving data services in the era of cloud computing and machine learning

Stephan Siemen, Umberto Modigliani

For decades, ECMWF has played a pivotal role in providing comprehensive and reliable meteorological and climate data to its Member States and the broader global community. From offering real-time and historical weather forecasts to delivering sophisticated model outputs and observations, ECMWF has supported a wide array of applications, ranging from everyday weather prediction to advanced research and disaster risk management. These services, accessible through diverse platforms and APIs (Application Programming Interfaces), have empowered users to obtain precise and tailored datasets, thereby bolstering research and services across numerous sectors.

From its inception, ECMWF's core mission, as enshrined in its Convention, included the "collection and storage of appropriate meteorological data" and ensuring that these data were "available to the meteorological offices of the Member States". Consequently, the development, maintenance, and operation of data provision services has been fundamental to ECMWF's identity since its creation.

However, the landscape of data services is constantly evolving. ECMWF has witnessed a substantial surge in data volumes, driven by advances in weather forecasting technology and an escalating demand for more accurate and granular predictions. The introduction of higher-resolution models and the integration of more observational data from diverse sources, such as satellites and radars, has led to an exponential increase in the data processed and stored by ECMWF. Simultaneously, the number of users who integrate forecast data into their systems and services has expanded, placing greater demands on the services delivering this data. The variety of datasets handled by ECMWF has also grown, thanks to numerous collaborations that integrate data from other centres.

This article will explore the evolution of ECMWF's data provisioning services in response to these growing demands, the increasing variety of data sources, and the potential offered by the introduction of cloud computing and machine learning (ML). It will also present the various ways in which these developments aim to serve Member States and the wider community, highlighting the positive impact of ECMWF's many collaboration activities on these developments.

The continuous growth of data volumes

The exponential growth of data volumes at ECMWF presents significant challenges. The organisation must manage the logistical and computational demands of processing, storing, and disseminating vast amounts of information in an effective way. Robust data management processes, including data modelling and the management of data holding catalogues, are essential. This necessitates continuous upgrades to computing capabilities and the optimisation of data handling strategies to ensure the timely and reliable delivery of forecasts. The increase in data volumes is both a driver of progress and a challenge that demands continuous innovation and investment.

ECMWF's day-to-day data volumes have grown dramatically over the years. The increasing availability of satellite data has been a key driver on the observation side. For forecast data, increases in resolution, frequency of runs, the number of ensemble members, and the number of parameters created have all contributed to a steady increase in volume. While the majority of data represents forecast and reanalysis data from operational or research and collaboration activities, observations collected across the globe are a vital component and are essential for many research activities at ECMWF and its Member States.

ECMWF has undertaken various efforts to mitigate the growth of its data archive, including stricter governance and deletion of obsolete research data, as well as the implementation of better compression and efficient storage techniques. These activities have significantly influenced the growth curve, damping the exponential trend it represents (see Figure 1).

Currently, ECMWF's overall production corresponds to 360 TB of forecast data per day. Projections indicate that this daily production will exceed 1 PB by 2027. This remarkable amount highlights a growing gap between what is produced and what can be effectively handled by users. While it is challenging for users to handle the increasing volume of data, ECMWF is actively working to offer solutions that enable users to realise more of the benefits that fully utilising this data can bring.

 doi: 10.21957/73e53hn4ss

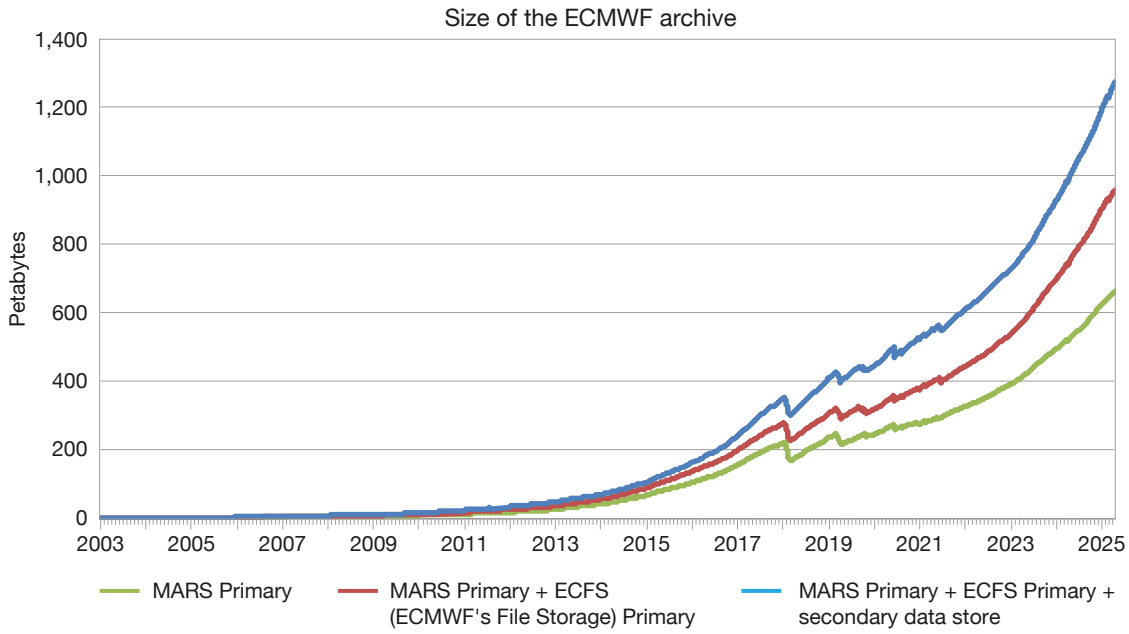


Figure 1 ECMWF's archive has substantially grown over the last two decades despite some reductions, and further growth is projected.

Embracing open data

ECMWF forecasts have always been provided at no extra charge to our Member and Co-operating States, and provisions under the umbrella of the World Meteorological Organization (WMO) were in place to support national meteorological and hydrological services (NMHSs) and international collaborations worldwide. Fees for other users were set in coordination with Member States.

In December 2020, the ECMWF Council unanimously adopted a move towards a more open data policy. This decision aimed to make some of ECMWF's model output available under an open data licence. During this transition, a tiered approach has been designed to provide balanced services for all users. The free and open data tier provides a subset of the full ECMWF Catalogue, available with no charges under the CC-BY-4 open licence.

The implementation period, which ends in the autumn of 2025, has enabled ECMWF to prepare for the change. While the provision of open data will offer many benefits and opportunities, it will also present some challenges to implementing good and efficient management of user demand. When users take up certain services, they need to be encouraged to use relatively tailored and small requests rather than large-scale ones, which can stress the system.

As the open data offering expands, ECMWF will also remove barriers for WMO members by removing all data and service charges for full-resolution data in support of UN initiatives such as Early Warnings for All (EW4ALL). Between now and 2027, the quality and accessibility of data will be improved for all WMO members, first focusing on less developed nations in the WMO's Regional Association for Europe (WMO-RAVI) and the countries supported by the Systematic Observations Financing Facility (SOFF) initiative. SOFF aims to support some countries in generating and exchanging data compliant with the Global Basic Observing Network (GBON), which is critical for improved weather forecasts and climate services.

doi: 10.21957/73e53hn4ss

ECMWF's data provision services

Data provision by ECMWF takes place through two avenues: the operational dissemination system and the Meteorological Archival and Retrieval System.

Dissemination system

ECMWF's most mature and time-critical method of data provision is the operational dissemination system, which is part of the ECMWF Production Data Store (ECPDS). This advanced platform is designed to efficiently collect and distribute vast amounts of meteorological data to users, including Member States, research institutions, and commercial entities. The ECPDS is also part of a collaborative effort with Member and Co-operating States through an Optional Programme (see Gougeon, 2019).

Meteorological Archival and Retrieval System

The Meteorological Archival and Retrieval System (MARS) is the backbone of ECMWF's data services, providing a centralised repository for a vast collection of meteorological data. MARS allows users to retrieve specific datasets based on various criteria, such as time, location, and parameters.

Over time, ECMWF has continuously evolved data access methods to improve the user experience and cater to diverse needs:

- **Web-based access to data and APIs:** ECMWF has developed web-based interfaces and APIs that allow users to access data programmatically. These tools provide a user-friendly way to explore the data holdings and retrieve the desired information.
- **ecCharts:** ecCharts serves as a 'window' to ECMWF's forecasts, providing interactive visualisations and analysis tools. This platform allows users to explore forecast data in a graphical format, enabling them to quickly identify trends and patterns.
- **OGC standards:** ECMWF has adopted Open Geospatial Consortium (OGC) standards to ensure interoperability and compatibility with other geospatial data systems. This allows users to seamlessly integrate ECMWF data into their existing workflows and applications.

All these developments were conducted in parallel with new services across ECMWF, such as the Copernicus Climate Change and Atmosphere Data Stores, which ensured harmonisation and benefited from each other's experiences.

Processing close to the data: the emergence of cloud services

The increasing volume of data and the growing demand for customised processing have led to the emergence of cloud-based services. These services bring the processing capabilities closer to the data, reducing the need for users to download large datasets and enabling them to perform complex analyses in a scalable and efficient manner.

- **The European Weather Cloud (EWC):** The EWC is a collaborative initiative with the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), aimed at providing a cloud-based platform for meteorological research and development. It enables users to access ECMWF data and computing resources in a secure and scalable environment, fostering collaboration and innovation (Figure 2). The EWC was able to benefit from the experiences gained from building the cloud-based EU Copernicus reference service WEkEO.
- **Interactive compute platforms – Jupyter notebooks:** ECMWF has embraced Jupyter notebooks as an interactive compute platform, allowing users to explore and analyse data using popular programming languages like Python. Jupyter notebooks provide a flexible and reproducible environment for data analysis, making it easier for users to share their work and collaborate with others.

 doi: 10.21957/73e53hn4ss

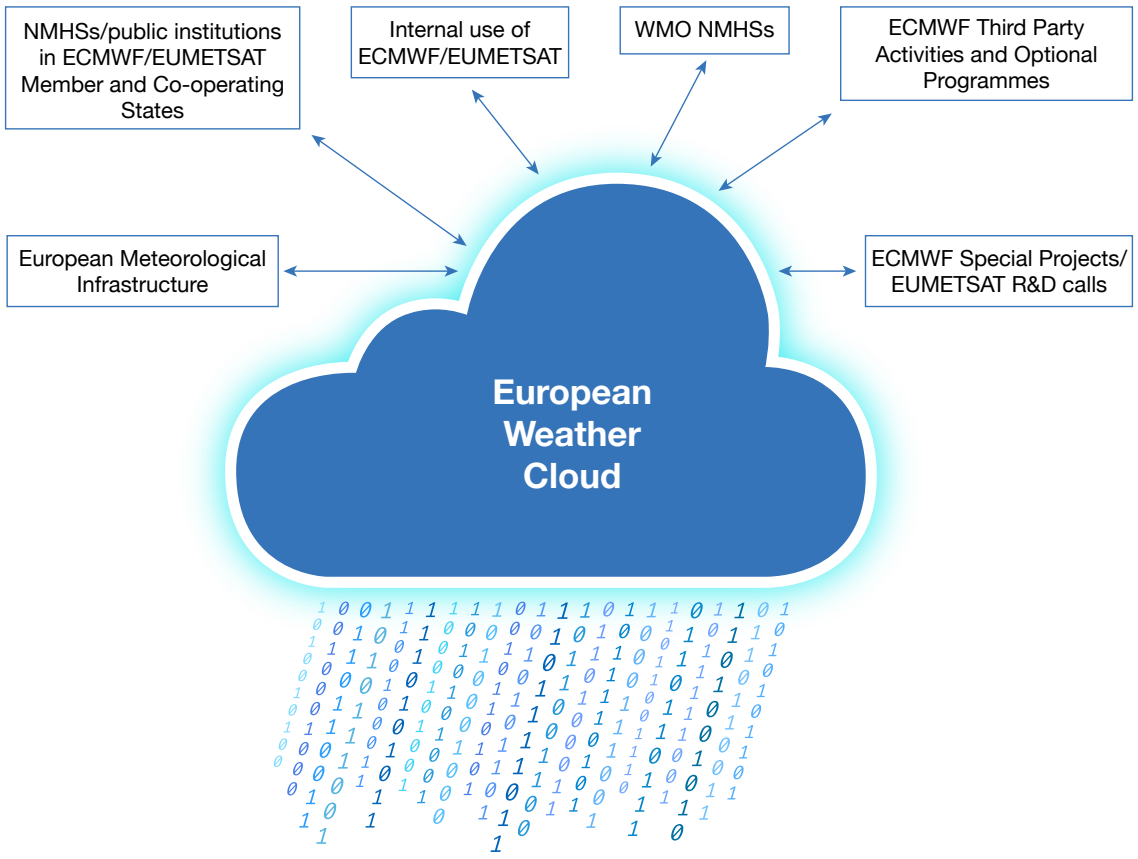


Figure 2 The EWC has a range of users, including European public institutions and NMHSs that are members of the WMO.

New opportunities and challenges through machine learning

Machine learning (ML) is revolutionising many fields, and meteorology is no exception. ECMWF recognises the immense potential of ML to improve weather forecasting, climate modelling, and data analysis. The organisation is actively exploring ways to integrate ML into its data services, offering new opportunities for users to leverage this powerful technology. ECMWF successfully operationalised its Artificial Intelligence Forecasting System (AIFS) in February and July 2025.

ECMWF is experimenting with chatbots as a new way of interacting with forecasts. These chatbots can provide users with quick and easy access to information, answering questions about weather conditions, forecast accuracy, and other topics.

ML can also be used to support the intercomparison and verification of different weather models. By training ML algorithms on historical data, it is possible to identify biases and errors in different models, leading to improved forecast accuracy.

ECMWF's data holdings, especially in the MARS archive, are an invaluable source for many ML applications, and a concentrated effort with Member States is under way to make this data accessible within popular ML frameworks. This work will open ECMWF's data to an even wider and faster-developing user community, benefiting all users overall.

Supporting the community

ECMWF is committed to supporting its user community by providing comprehensive resources and tools.

- **User support:** ECMWF offers dedicated user support to assist users with any questions or issues they may encounter. The user support team provides guidance on data access, software tools, and other topics.
- **Tools and software:** ECMWF develops and maintains a range of software tools to facilitate data processing and analysis. These tools are freely available to users and are designed to be user friendly and efficient.

doi: 10.21957/73e53hn4ss

- **Supporting community projects:** ECMWF actively supports community projects aimed at advancing meteorological research and development. This support includes providing data access, computing resources, and technical expertise.

The need for data spaces

European data spaces have attracted increasing attention since their inception in 2020 by the European Commission. A data space is an evolution from traditional data services and cloud systems to offer a decentralised infrastructure for trustworthy data sharing and exchange in data ecosystems, based on commonly agreed principles. It brings together relevant data infrastructures and governance frameworks to facilitate data pooling and sharing needed for new challenges, such as cross-discipline ML.

These data spaces aim to create a single European market for data, enhancing competitiveness and data sovereignty. They have progressed from conceptual frameworks to operational systems, and a common access layer, called SIMPL, is being developed. For the weather community, these data spaces present a unique opportunity to integrate meteorological data with other sectors, such as agriculture, energy, and mobility. This integration could lead to more uptake and impact of our community's data, and enhanced decision-making for weather-dependent industries, ultimately contributing to the goals of the European Green Deal. The European Open Science Cloud (EOSC) emerged as one of the first data spaces focusing on supporting the research community, and it would benefit from weather and climate data. Other sectorial data spaces, such as the Green Deal and Health ones, will also require access to substantial amounts of weather and climate data. In parallel, the EU's Copernicus Programme has started its own data space for Earth observation data, which could incorporate the Atmosphere Data Store and the Climate Data Store run by ECMWF. With these data spaces developing, ECMWF is working with its Member States and partners, EUMETSAT and EUMETNET, to form a community-wide approach to connect all these data spaces in a coordinated and harmonised way. This work has started with a series of workshops, with the latest in June 2025 bringing together key stakeholders at a side event to a meeting of the RODEO project (*https://rodeo-project.eu/*), which aims to make meteorological high-value datasets easily available. ECMWF is contributing its wide-reaching experience in providing data services for forecast data. It works with the community to develop a roadmap for a common approach.

Conclusion

ECMWF's data provision services have undergone a remarkable evolution, driven by the increasing volume and variety of data, the emergence of cloud computing, and the transformative potential of ML. By embracing these advances, ECMWF is empowering its Member States and the wider community to make better use of meteorological data, leading to improved weather forecasts, climate models, and decision-making across a wide range of sectors. ECMWF remains committed to continuous innovation and collaboration, ensuring that its data services continue to meet the evolving needs of its users in the years to come.

Further reading

Gougeon, **L.**, 2019: The ECMWF Production Data Store. *ECMWF Newsletter* **No. 159**, 35–40. *https://doi.org/10.21957/83deq5lgc0*

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK