



## COMPUTING

### ECMWF contributes to Swiss supercomputing project in weather and climate science



EarthCARE satellite. Credit: ESA/ATG medialab

This article appeared in the Computing section of ECMWF Newsletter No. 183 – Spring 2025, pp. 32–35

## ECMWF contributes to Swiss supercomputing project in weather and climate science

Thomas Geenen, Ahmed Benallegue, James Hawkes, Simon Smart, Tiago Quintino, Umberto Modigliani (all ECMWF), Anurag Dipankar (ETH Zurich), Oliver Fuhrer (MeteoSwiss)

The SwissTwins project, led by the Swiss National Supercomputing Centre (CSCS), started in December 2022 and was established to support the Swiss weather and climate science community (<https://www.cscs.ch/about/collaborations/swisstwins>). One of the primary objectives is to ensure that Swiss research infrastructures remain well integrated in Europe. As part of this wider initiative, ECMWF has collaborated closely with CSCS and MeteoSwiss to establish efficient access to ECMWF's forecast products and meteorological archive for the training of artificial intelligence (AI)/machine learning (ML) models. These models have already shown great potential in weather forecasting and are increasingly being explored for climate modelling.

### Bridging data and computational power

Earth system sciences and the study of climate change require substantial computational power and access to large volumes of high-quality data. This is especially true for ML models and particularly for emerging trends in ML models trained on vast datasets, called foundation models. Compared to task-specific ML tools, foundation models need new training methods, and they require significantly more complex neural network architectures that are trained from massive datasets. Therefore, their development needs substantially more resources than that of task-specific models.

This understanding triggered the initial design concepts for the SwissTwins project and resulted in the design of the DataHypercube. The DataHypercube is a high-performance storage cluster that is deployed in ECMWF's data centre in Bologna (Italy), close to the datasets that are key to train both task-specific ML models as well as foundation models. ECMWF has one of the largest meteorological archives in the world and provides also other datasets widely used for ML model training, such as ERA5 meteorological reanalysis data. The DataHypercube is connected through an ultra-large bandwidth network connection to the Alps supercomputing system of the CSCS in Lugano (Switzerland). Alps hosts the MeteoSwiss weather prediction model and is also used by groups from ETH Zurich for Earth system sciences and climate change research. Leveraging ECMWF data handling software services, deployed on the DataHypercube, specific data can be extracted, pre-processed and staged for training purposes. The Alps supercomputer is one of the fastest AI supercomputers in Europe. Connecting the AI and ML capability of the Alps infrastructure with the unique data products and archive hosted by ECMWF in Bologna will open up unprecedented research opportunities. This project pilots a proof of concept, informing an outline methodology for possible future initiatives with other ECMWF Member States to leverage ECMWF data and services in a similar fashion.

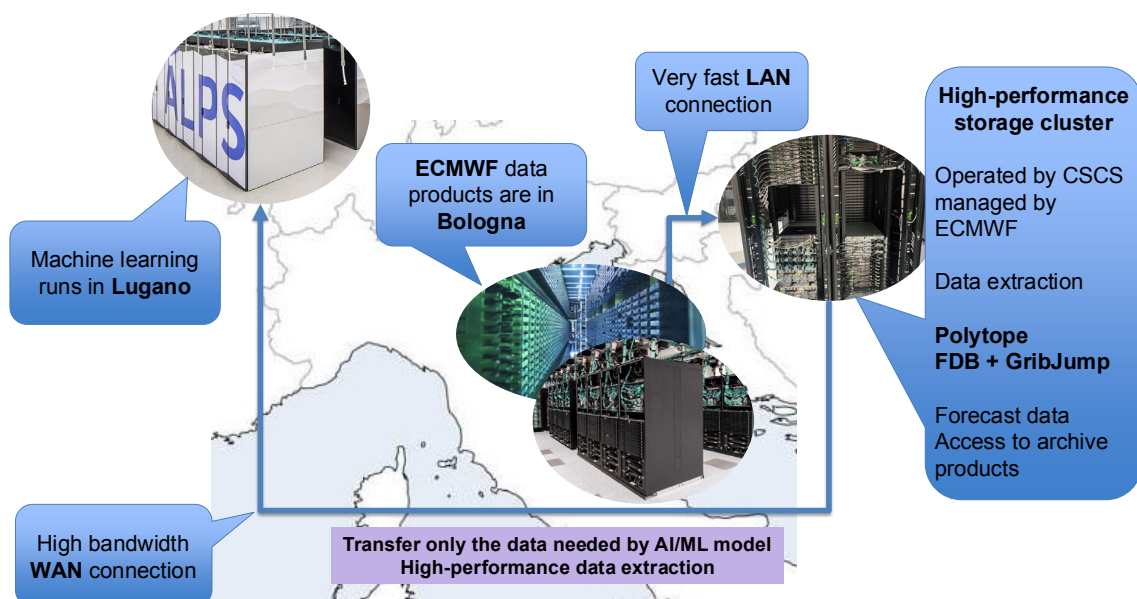
### Enabling AI/ML for weather and climate research

A significant part of the project is to set up the DataHypercube system in Bologna and to establish the high-performance network connection between the DataHypercube system and Alps in Lugano, but there is more. To allow for data access with low latency and the extraction of specific datasets to train AI/ML models for the benefits of the Swiss and international numerical weather and climate prediction community, the deployment of key ECMWF data handling services on this system is essential. ECMWF has been developing these tools and services over the last few years and further evolved them in the context of the EU's Destination Earth initiative (<https://destine.ecmwf.int>). The architecture of the software and services deployed on the DataHypercube will closely follow the developments of the Digital Twin Engine of Destination Earth and the architecture of the Destination Earth data bridges. In that sense, it is yet another example of using the ECMWF Software Engine (López Alós, 2024). In particular, it will leverage the Fields Database (FDB – Smart et al., 2017) for data storage and the Polytope service (Hawkes et al., 2020) for efficient data delivery. It uses the GribJump functionality to efficiently locate data in larger GRIB files (<https://github.com/ecmwf/gribjump>). FDB is ECMWF's semantic object store technology designed for efficient data handling of hierarchical, high-dimensional datacubes. FDB will be distributed across the cluster, ready to ingest data from ECMWF's forecasts. Its high-bandwidth communication protocols can be used directly for bulk consumers of this data, including by AI/ML models.

Polytope is ECMWF's data delivery service with the unique ability to efficiently extract features, such as point data (e.g. time-series, vertical profiles) or area data (e.g. polygon extractions), from the vast FDB data store without any intermediate copies or transposition of the data.

This will bring a new level of accessibility to the data, accelerating the work of Earth system and climate scientists. On Alps, an additional ECMWF data handling software component called earthkit will be deployed (Russell et al., 2024). This is to provide users with a rich API to extract and process the data from the DataHypercube on Alps.

The combination of a dedicated high-performance storage cluster with a high-performance network connection on Alps and the deployment of advanced data handling capabilities makes this a potent tool in the hands of weather and climate scientists (Figure 1). ETH Zurich's exascale simulation platform, EXCLAIM, will be benefiting from faster on-site access to ECMWF data through Alps. This improvement supports use cases on EXCLAIM, which aim to understand the evolving climate. These include kilometre-scale global climate simulations and limited-area large-eddy simulations.



**Figure 1** This diagram illustrates how the high-performance storage cluster and network will provide specific datasets from the ECMWF data holdings towards the Alps system in Lugano to be used by the Swiss weather and climate science community. (Copyright of the Alps photo: Marco Abram, CSCS)

The climate studies rely on various ECMWF products, such as reanalysis, daily operational analysis, and forecasts, to provide the necessary initial and boundary conditions for the modelling system used at ETH Zurich. These datasets are crucial for model validation and verification.

Enhancing the data cube on Alps with advanced tools like Polytope and earthkit is a big plus. These tools enable faster and more selective access to relevant data, which is expected to significantly improve user productivity.

### A dedicated multi-domain service

An end-to-end, ultra-large bandwidth network infrastructure is required to enable massive data transfers from ECMWF's meteorological data products and archives to CSCS supercomputing facilities. It is foreseen that during the ramp-up period and subsequent phases of the project, multiple AI/ML models in the weather and climate domain will be consuming data from the DataHypercube simultaneously, requiring a further increase of the total aggregate throughput as the number of users grows.

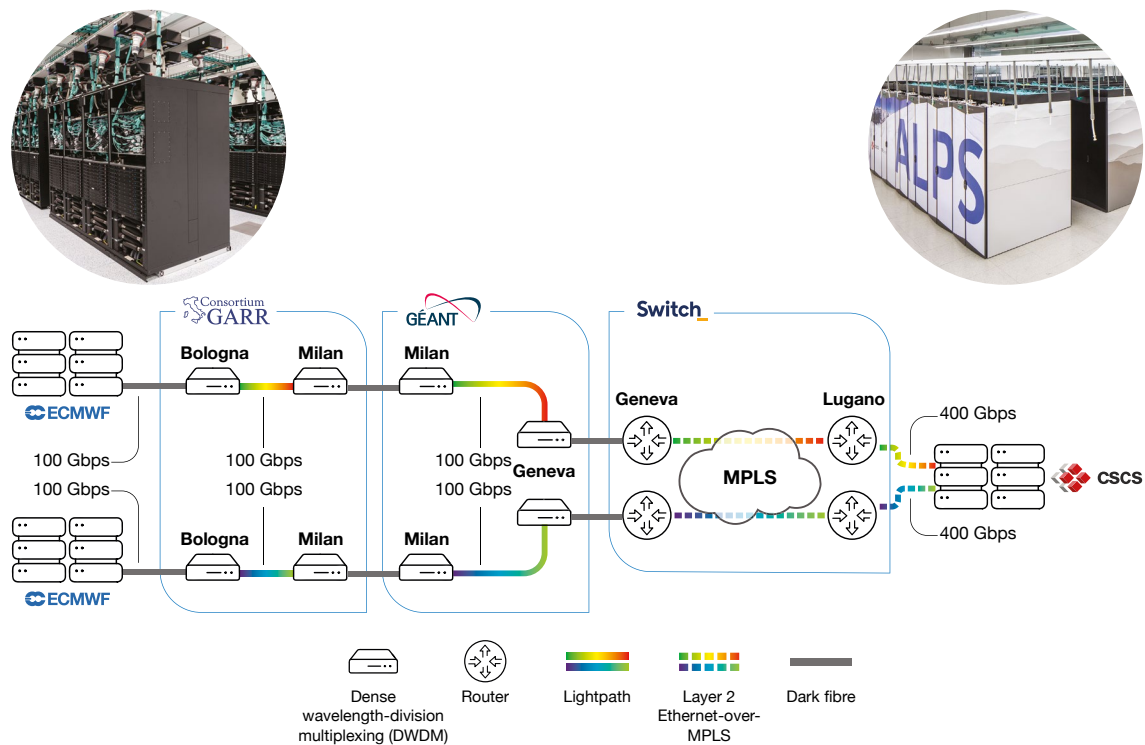
Thanks to collaboration between GARR (Gruppo per l'Armonizzazione delle Reti della Ricerca, the Italian national computer network for research and education), SWITCH (the Swiss national research and education network) and GÉANT (Gigabit European Academic Network, the pan-European data network for the research and education community), the two centres are now interconnected by a state-of-the-art, ultra-large bandwidth, dedicated connection that seamlessly traverses the three network domains.

In designing this link, the requirements included not only extremely high capacity and scalability but also significant resilience, reliability, security, and direct user access to the optical infrastructure.

This dedicated connection comprises two separate 100 Gbps links, independently managed by the three operators. To ensure the security and reliability of the service, new resources have been deployed. At the ECMWF site in Bologna, two new fibre links have been installed by Lepida, GARR's local partner in the Emilia-Romagna region, to access the backbone. These complement the existing GARR fibre connections that are already used for operational forecast product dissemination at ECMWF.

From there, two optical light paths have been configured to reach GARR PoPs (Points of Presence) in Milan. Here, the Italian network connects with the GÉANT network, which carries the traffic to its PoP located at CERN in Geneva. From Geneva, SWITCH takes over and delivers the data to its final destination at CSCS in Lugano.

The implementation methods differ between the networks: GARR and GÉANT have chosen an optical domain approach, while SWITCH uses a Layer 2 Ethernet-over-MPLS (Multi-Protocol Label Switching) service (Figure 2).



**Figure 2** A resilient, fast, dedicated, secure connection has been created between ECMWF and CSCS. This diagram shows the details of the high-performance network connection between the ECMWF data holdings in Bologna and CSCS Alps in Lugano. See <https://www.garrnews.it/caffe-scientifico/tre-reti-al-servizio-della-terra>. (Copyright of the Alps photo: Marco Abram, CSCS)

“This type of service needs a high level of collaboration and trust among operators and would not be possible outside the research networks ecosystem,” said Sabrina Tomassini (Head of Network Department at GARR). When it comes to implementation, management, and monitoring, each network is responsible for its domain, therefore a high level of communication is necessary. National Research and Education Networks (NRENs), such as GARR, GÉANT and SWITCH, offer multi-domain services, transcending their domains to harmonise services across networks. This capability is unique to NRENs, as commercial operators are confined to their domains of competence.

To ensure a high level of security and resilience, a deep technical analysis was conducted, and diversification, shelters, and robust pathways were employed to identify and address potential faults, thereby ensuring the highest possible resilience of the service.

### In Bologna as in Lugano

“This service is one of the first to leverage the capabilities of the new next-generation GARR-T network for optical infrastructure access,” stated Sabrina Tomassini. “The advantage is that the entire 100 Gbps channel capacity is wholly dedicated to the user, without statistical multiplexing.”

“Moreover, a peculiarity of this interconnection is that the sites are linked at the layer 2 level. This enables CSCS to have full visibility of the DataHypercube machines located at ECMWF, as if they were located in its local network in Lugano. This is a feature enabled by the new GARR-T network,” said Alessandro Inzerilli (head of the Network Operations Centre and Operations groups at GARR).

The DataHypercube platform itself also plays an important role in enhancing the data-proximity experience for the users on Alps. The interactive, user-facing data handling services that serve time-critical data are hosted from a Kubernetes cluster (a three-node cluster) with access to a fast flash storage pool of 500 TiB (11 storage nodes), which hosts the FDB services. In addition, the DataHypercube contains a second tier of capacity storage (2 PiB ClusterStor Lustre appliance) for specific historical datasets that can be pushed to the system based on user request. The DataHypercube hardware specifications are provided in Table 1.

Hardware component	Number	Specifications
Compute and login nodes	3	2x AMD EPYC 7713 64c 512 GB RAM 30 TB local SSD
Storage/compute nodes	11	1x AMD EPYC 9534 64c 384 GB RAM 61 TB SSD
ClusterStor appliance (Lustre)	1	HPE Cray E1000 2 PB usable space
High speed network switch	6	200 Gbit/s HPE Cray Slingshot-11
External network switch	2	100 Gbit/s Aruba router

**Table 1** The hardware specification of the DataHypercube installed in the ECMWF data centre in Bologna.

### Going forward

The project, currently planned for two years, serves as a pilot. If successful, it could lead to further developments, including the possibility of increasing capacity to 400 Gbps in the future. One of the advantages of the optical link with GARR is that any potential upgrade to 400 Gbps could be performed relatively quickly with minimal technical effort from ECMWF.

This high-performance connection, together with the deployment of a high-performance storage cluster and data handling services both on Alps and in the Bologna data centre, opens new opportunities for research in meteorology and scientific computing, enabling closer interaction between ECMWF and the Swiss numerical weather and climate prediction community. It facilitates the sharing of large volumes of data and a tight integration of analysis software using earthkit on Alps and backend services like polytope and FDB, which are deployed on the DataHypercube in Bologna. The project potentially accelerates progress in climate modelling and weather forecasting and is an example of close collaboration with our Member States and international organisations to generate value through innovation.

**Further reading**

**López Alós, A.**, 2024: The modernisation of the Data Stores at ECMWF, *ECMWF Newsletter No. 181*, 9–10. <https://www.ecmwf.int/en/newsletter/181/news/modernisation-data-stores-ecmwf>

**Smart, S.D., T. Quintino & B. Raoult**, 2017: A Scalable Object Store for Meteorological and Climate Data, *PASC '17*, **13**, 1–8. <https://doi.org/10.1145/3093172.3093238>

**Hawkes, J., N. Manubens, E. Danovaro, J. Hanley, S. Siemen, B. Raoult et al.**, 2020: Polytope: Serving ECMWFs Big Weather Data, *EGU General Assembly 2020 session entry*. <https://doi.org/10.5194/EGUSPHERE-EGU2020-15048>

**Russell, I., T. Quintino, B. Raoult, S. Kertész, P. Maciel, J. Vardell et al.**, 2024: Introducing earthkit, *ECMWF Newsletter No. 179*, 48–53. <https://www.ecmwf.int/en/newsletter/179/computing/introducing-earthkit>

© Copyright 2025

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

The content of this document, excluding images representing individuals, is available for use under a Creative Commons Attribution 4.0 International Public License. See the terms at <https://creativecommons.org/licenses/by/4.0/>. To request permission to use images representing individuals, please contact [pressoffice@ecmwf.int](mailto:pressoffice@ecmwf.int).

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error or omission or for loss or damage arising from its use.