Open Data at ECMWF: Progress and future plans

Emma Pidduck, Victoria Bennett, Xiaobo Yang, Maartje Kuilman, Ilaria Parodi and Baudouin Raoult

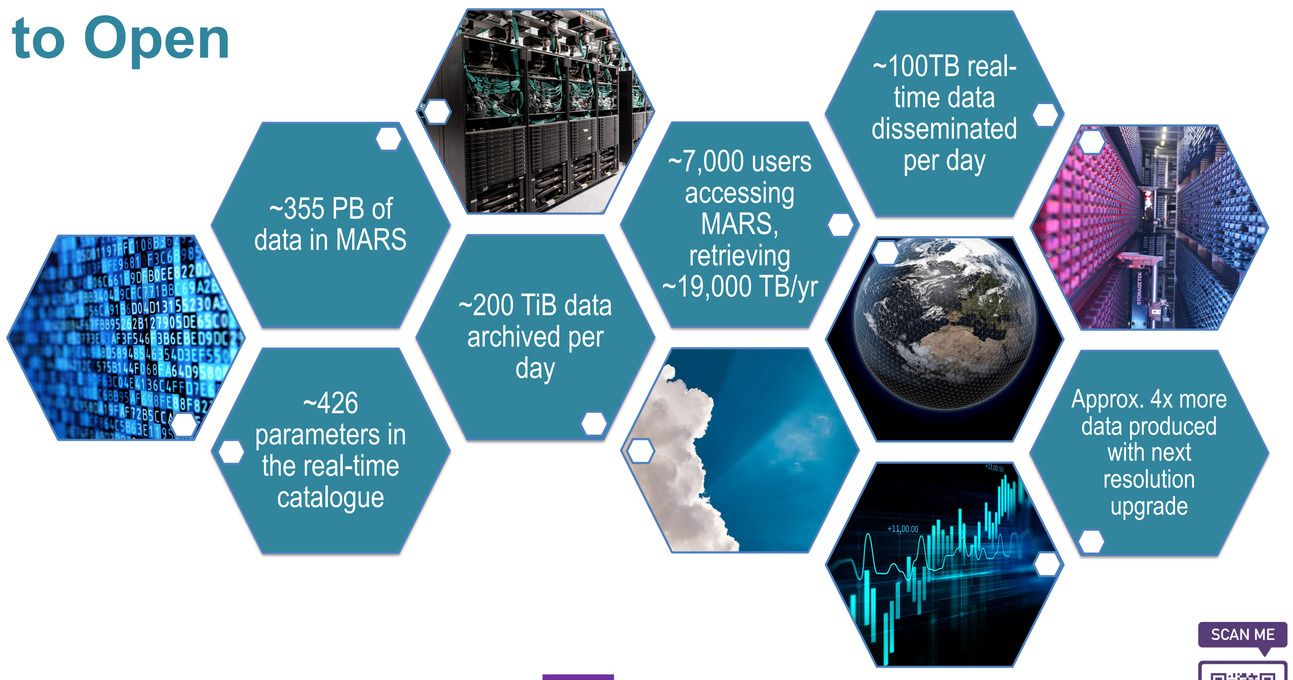
ECMWF, Forecast Department, Bonn (Germany), Reading (United Kingdom) and Bologna (Italy)

Challenges of Transitioning from 'Closed' to Open

Transition from 'closed' to Open causes a decrease in revenue associated which needs to be covered by alternative sources – in most cases, this means the Member and Co-operating States pay more.

Transition also requires significant upgrades to the infrastructure and services, which requires additional funding and resources.

During the transition, a subset of the full catalogue will be governed by an open licence, while the rest remains closed. This means users are governed by two licences for selected portions of the same data catalogue.



View the recorded presentation from #UEF2022:

SCAN ME

Datasets Available – Open Licence

Real-time Open Data – CC-BY-4.0

Approximately 700 GB per day of HRES (10-day) and ENS (15-day) forecast products at 0.4 resolution (~36km) with 3h timesteps to 144 and 6-hourly timesteps to full forecast range. Four forecasts released per day via ECMWF HTTPS service and Microsoft Azure.

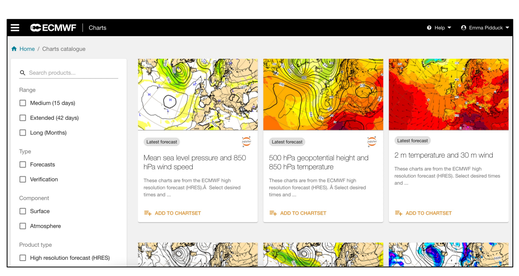
Recently released!

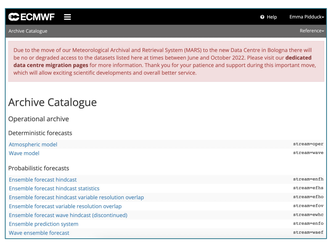
SCAN ME

OpenCharts – CC-BY-4.0

Around 250 charts available freely and openly on ECMWF OpenCharts application. Custom chartset capability, Jupyter Notebook and integration with ECMWF Real-time Open Data.

SCAN ME





Historical Forecasts – CC-BY-4.0

Past forecasts from the atmospheric, wave, ensemble and seasonal models, are accessible via ECMWF WebAPI from the MARS Archive. Approximately 355 PB of data dating back to 1975.

SCAN ME

Research datasets

TIGGE Dataset

TIGGE dataset consists of ensemble forecast data from 10 global NWP centres. Dataset comprises 6-hourly model outputs for a range of parameters.

CC BY NC

S2S Dataset

CC BY

Subseasonal to seasonal (S2S) prediction fills the gap between weather and seasonal-decadal climate prediction. Dataset comprises historical data.

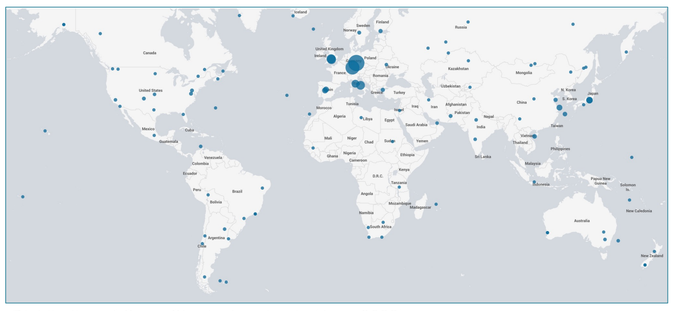
YOPP Dataset

CC BY

Dataset from the Year of Polar Prediction programme. Targeted to enable a significant improvement in environmental prediction capabilities for the polar regions and beyond.

Data Usage

Since the release of Open Data, ECMWF has seen a 2.5x increase for the volume of data delivered as part of the available Open Data, while visits to the supporting documentation are approximately 650 users per month with 2,000 page views per month. The ECMWF Python package is downloaded ~100 times per day.



Distribution of Open Charts downloads (June 2022)

ECMWF Real-time Open Data Delivery Statistics

2021	2022
2.8 TB/day	7.0 TB/day
38,000 requests/day	800,000 requests/day

Delivery via Microsoft Azure

Reported 23TB in one week and 800k user retrievals (June 2022)

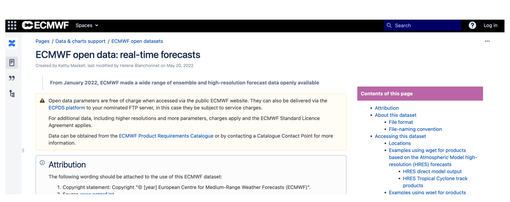
"FAIR" Data Access

Access to the Open and Free data has been designed according to the "FAIR" principles of data distribution. Alternative access methods, including interpolation, area-subsetting, compression, 24/7 operational support, are subject to delivery fees, for which the fees can be waived for research projects (subject to management approval).

"FAIR" (Findable, Accessible, Interoperable and Reusable):

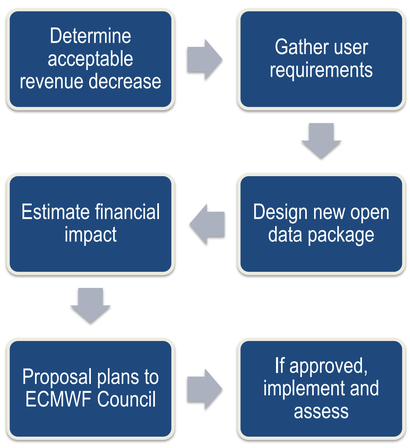
- ✓ Design of an API to easily download the data
- ✓ Development of open-source Python libraries to process and visualise the data
- ✓ Development of Jupyter notebooks that showcase the use of libraries and the data
- ✓ Updated and maintained user documentation to support increasing number of active users





The Process

As part of ECMWF's 10-year strategy, the goal is to have a fully open data catalogue by 2026 with a subset of the catalogue available for free. Between 2022-2026, ECMWF will undergo a phased transition of increasing open data and decreasing the revenue from data sales, with all phases approved by ECMWF Council on an annual basis.



During this transition, user requirements combined with input from Member and Co-operating States are collected to inform future decisions relating to open datasets.

User feedback is important at all stages and we welcome input via the ECMWF Support Portal.

What happens next?

CAMS Meteorological Parameters Released in Real-time!

Meteorological fields from the Copernicus Atmospheric Monitoring Services (CAMS), which are released 6-days after production will become available in real-time from the Atmospheric Data Store (ADS). Available data will be sufficient to initialise Limited Areas Models and includes surface, pressure and model level data at 36km resolution.

C3S Seasonal Parameters Released in Real-time!

Seasonal data produced by ECMWF and available via the Copernicus Climate Change Service (C3S) is currently available with a 4-day delay via the Climate Data Store. In 2023, the fields will be released in real-time at the current temporal and spatial resolution.

Conference Poll

For Open Data, which would you prefer…?

High-frequency time steps
OR
Higher resolution data

Scan here to submit a vote and see live results

Useful Links

ECMWF Real-time Catalogue: https://apps.ecmwf.int/shopping-cart/
ECMWF Archive Catalogue: https://apps.ecmwf.int/archive-catalogue/
Support for WMO: https://www.ecmwf.int/en/about/what-we-do/serving-meteorology/working-wmo
ECMWF Support Portal: https://www.ecmwf.int/en/support