

# Early Adopter: ECMWF

**Early Adopter's Name:**

*ECMWF - European Centre for Medium-Range Weather Forecasts*

**Organisation type:**

International Organisation

**Organisation size:**

**Large Enterprise**

**Organisation Research Field(s):**

Natural Sciences – Earth and related environmental sciences (1.5)

**Organisation Profile:**

ECMWF is both a research institute and a 24/7 operational service, producing global numerical weather predictions and other data for our Member and Co-operating States and the broader community. The Centre has one of the largest supercomputer facilities and meteorological data archives in the world. Other strategic activities include delivering advanced training and assisting the WMO in implementing its programmes.

In addition, we operate two services from the EU's Copernicus Earth observation programme, the Copernicus Atmosphere Monitoring Service (CAMS) and the Copernicus Climate Change Service (C3S). We also contribute to the Copernicus Emergency Management Service (CEMS).

**Organisation website URL:**

<http://www.ecmwf.int>

**Suggested Use case title:**

Real Time Dissemination of NWP Open Data

**Problem definition:**

ECMWF provides its users with Numerical Weather Prediction (NWP) products by running a time-critical Global Atmospheric model which runs four times per day on a strict operational schedule. The output of the forecast model (IFS) is then post-processed and about 420 millions of user tailored products are computed and disseminated to our Member States, Cooperating States and, and commercial users on a daily basis. These products total to a volume of about 42 TIB that are sent out distributed over the 4 forecast times on a day.



We aim at providing a fraction of this data (see description below), using high-bandwidth transfers to make the data available as soon as possible after being generated and also aim at retaining the data as long-term storage on ArchiveR / EOSC.

### Is this use case new for your organisation?

Using external cloud providers for dissemination of real-time open data is a new use case at ECMWF.

The intended target timeline is the following:

- **2022:** Set up preliminary test implementation (duration: 1 year since platform availability)
  - Definition of a subset of data to make available
  - Definition of the metadata required to use the data.
  - Test data accessibility by end-users.
  - Develop end-user tools to use the data.
- **2023 and latter:** Consolidation of the test implementation (duration: approximative)
  - Turn the test pipeline into a reliable production pipeline
  - Assess the most relevant format for the data (GRIB, NetCDF, zarr, parquet, other)
  - Assess data discoverability.

### Data and metadata Characteristics:

#### **ECMWF Open Data (real-time):**

The dataset contains all products that the UN World Meteorological Organization has endorsed to be available to all National Hydrological and Meteorological Services around the world, including all [WMO Essential and Additional datasets as described on ECMWF's website](#).

The single **ECMWF Open Data (real-time)** dataset will be released with a CC-BY-4.0 open license which means it can be re-distributed and used commercially.

ECMWF expects to upload every day 928 files, with sizes between 2.4 MB and 2.5 GiB for a total volume of approximately 700 GiB. In addition a varying number of files characterizing Tropical cyclones will be uploaded.

Access to the data is expected to follow two distinct patterns:

1. Recent data download: Part of our users will download data that have been uploaded during the latest day, most likely as soon as the data is available.
2. Random download: Part of our users will download a randomly selection of our data, that have been uploaded at any date and time.

File layout may differ to allow flexibility to use the most appropriate format when the data becomes available, i.e. we may split files differently or merge them if appropriate.



Domain specific metadata will need to be provided along with the data itself. The metadata follows international standards for meteorology and climate. The complete implementation of this metadata is not fully defined yet.

**Cost requirements:**

Due to the large data volumes ECMWF requires the cost per GB to be as low as possible, ideally matching the cost of archiving on tape.

**Benefits and expected impact:**

*How would the community that you are addressing benefit from the ARCHIVER solutions? Please describe which would be the foreseen impact of this activity*

Weather forecast products are known to have a significant socioeconomic impact. The specific WMO datasets proposed in this use-case

include Tropical Cyclones forecasts as well as 10 day weather forecasts with fields for pressure, temperature, and wind speed based on both the HRES (High Resolution) and ENS(Ensembles) models of ECMWF.

These datasets are expected to be of great interest for the wider European scientific community as well as industry

**Additional technical details:**

A list of functionalities to be tested:

- Upload of daily data (approximately 700GB per day)
- Verification of uploaded data: list of the data. Checksum of the data.
- Deletion of the data we uploaded (in case of corruption, and for cleanup).
- Download of the uploaded data (for the latest 2 months)
- Access to random part of the data (HTTP range request or similar)
- Statistic gathering usage analytics (e.g. number of users, number of accesses, location, etc.)

The required storage capacity for the testing needed will be 43 TB (700GB per day x 2 months).

The needed time to access the platforms for the tests is 3 months. However, if at all possible, a more permanent test platform would be very welcome to enable testing the upload of data in a longer, sustained semi-operational mode.

30 Days is the timeframe in advance it's needed to be informed of the testing slot in order to prepare the tests.

**Contact person & details:**

Florian Pappenberger

Director of Forecasts @ ECMWF

