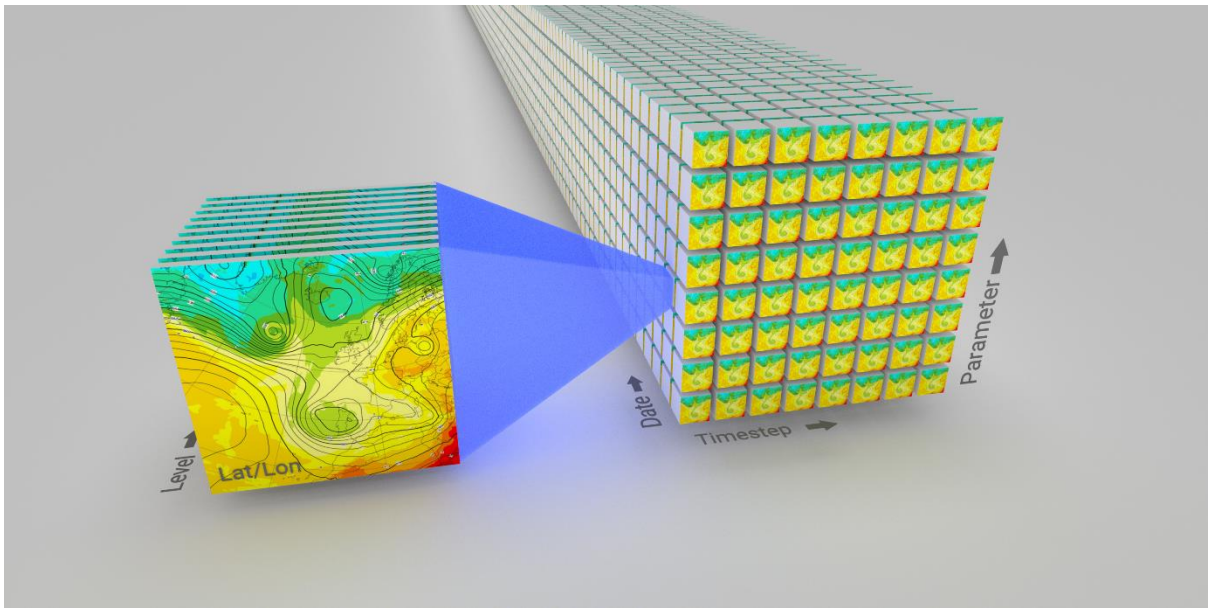
Polytope: Serving ECMWFs Big Weather Data

James Hawkes, Nicolau Manubens, Emanuele Danovaro, John Hanley, Stephan Siemen, Baudouin Raoult, Tiago Quintino

ECMWF, Reading, United Kingdom
james.hawkes@ecmwf.int

Every day, ECMWF produces approximately 120TiB of raw weather data. This data is used to produce approximately 30TiB of user-defined products, which are disseminated worldwide. The raw data is also stored in the world's largest meteorological archive (MARS), currently holding over 300 PiB of primary data -- which is also served around the world on demand.



ECMWF's data can be represented as a 6-dimensional data cube, with axes for latitude, longitude, vertical level, parameter, forecast date and timestep. For ensemble forecasts there is an additional dimension representing each ensemble member.

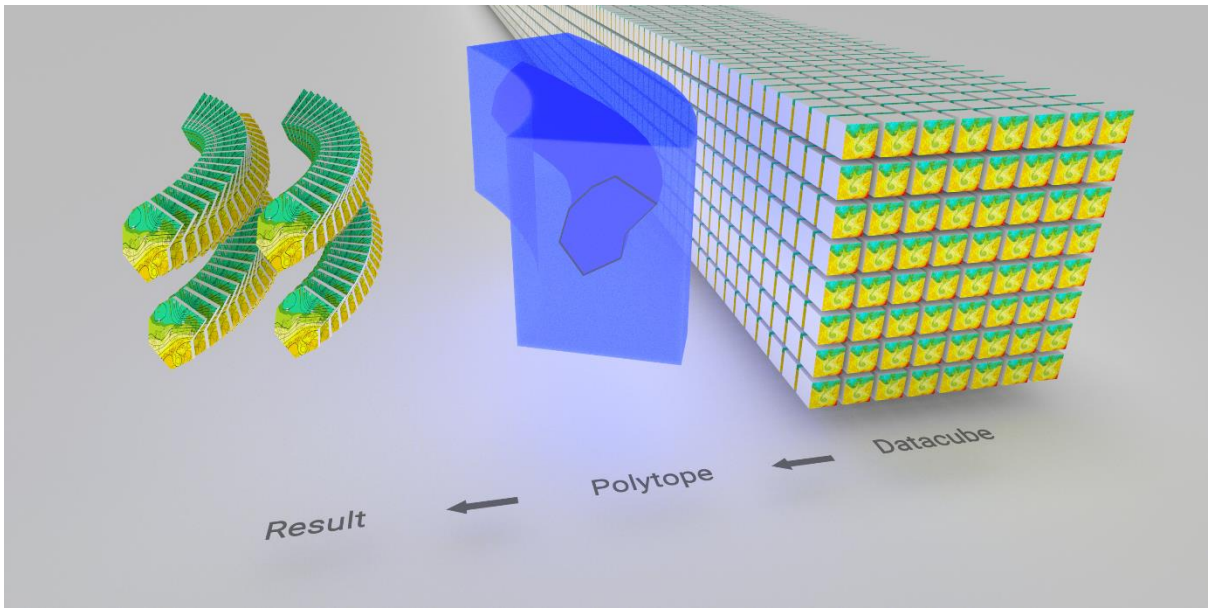
As the resolution of ECMWFs global weather models increase over the next few years, the amount of raw data produced per day will increase into the petabytes, and the distribution of products and archived data will become impossible.

Currently, the smallest discrete unit of data is a two-dimensional global field, so users who require global data already have quite efficient data access. However, there are many uses of our data where entire global fields are not required, and transmitting the entire fields takes up valuable bandwidth. Common examples include:

- Computing time-series data for a single geographical point or region, for multiple parameters
- Retrieving vertical profiles for a single geographical point or region, for multiple parameters
- Evaluating spatiotemporal paths (routes in time), for example, predicting weather along a flight-path

Users performing these kinds of analyses must download entire global data fields and perform additional processing on that data. This is not only expensive, but a cognitive burden on the user, because they must learn to use various meteorological data extraction tools.

To this end, we are developing a service which will provide in-situ, on-the-fly data extraction and processing, which will improve data access efficiency and ease of use. It provides a simple REST API, so users can make requests easily and without complex, domain-specific clients. The core principle is that users will be able to request any subset of ECMWF's datacube, expressing their request as an n-dimensional shape – a Polytope – giving its name to the service. This polytope may be defined as simple numeric ranges or as custom shapes, or some combination (to be determined).



The Polytope service will allow users to define an n-dimensional 'polytope', which describes the data they require. For a flight path, the polytope may be a box-shaped domain following a flight over time, perhaps repeating every day for a month. Polytope will efficiently extract the required data in-situ, returning just the data the user needs.

Polytope has domain-specific knowledge of the data at ECMWF, and is able to do byte-addressability within each field, meaning that it can extract just the relevant bytes from a global field – reducing the amount of data transferred to the user by orders of magnitude. It will connect to ECMWF's meteorological archive (MARS) as well as having access to real-time data directly from the HPC.

Polytope is being partly developed under LEXIS, an EU-funded Horizon 2020 project which focuses on large-scale HPC & cloud workflows. Polytope forms a part of ECMWF's strategy for the convergence of HPC & cloud, and will provide a means to efficiently serve Big Weather Data to many cloud-based systems, both at ECMWF and across Europe.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825532