# Polytope: Feature Extraction for Petabyte-Scale Datacubes

Mathilde Leuridan[1*], James Hawkes[1] , Tiago Quintino[1] , Simon Smart[1], Emanuele Danovaro[1]

(1) ECMWF; (*) mathilde.leuridan@ecmwf.int

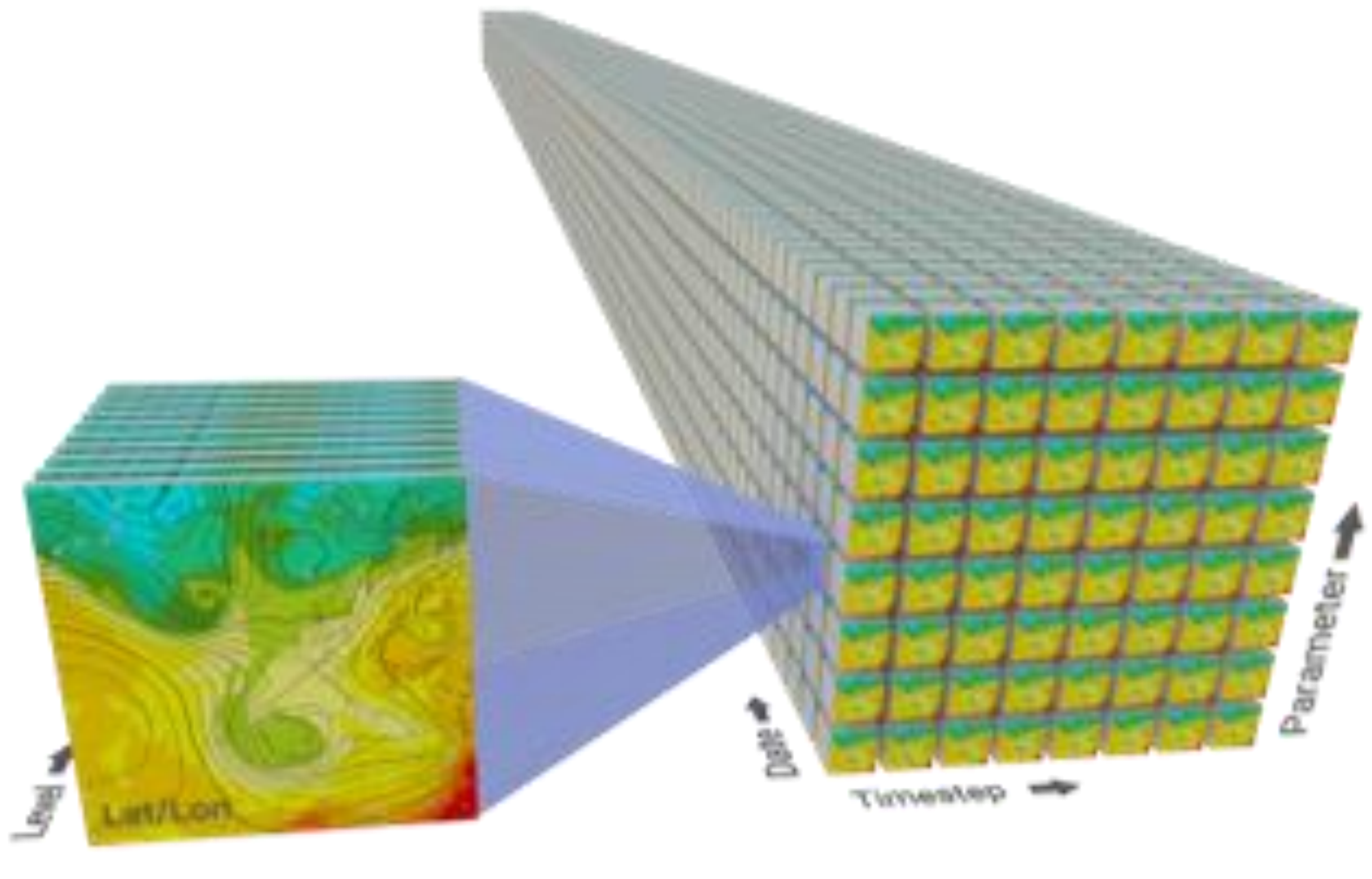
## 1. Introduction

ECMWF generates roughly 120 TiB of daily raw weather data, expected to surpass a petabyte in the coming years due to model advancements and higher resolution forecasts.

These improvements will, in principle, help scientists better forecast weather events, but distributing such vast amounts of data efficiently is increasingly difficult using current data access mechanisms. ECMWF is thus developing "Polytope", a feature extraction concept utilising higher-dimensional computational geometry, which aims to facilitate more efficient data access, and provide an overall data usability improvement.

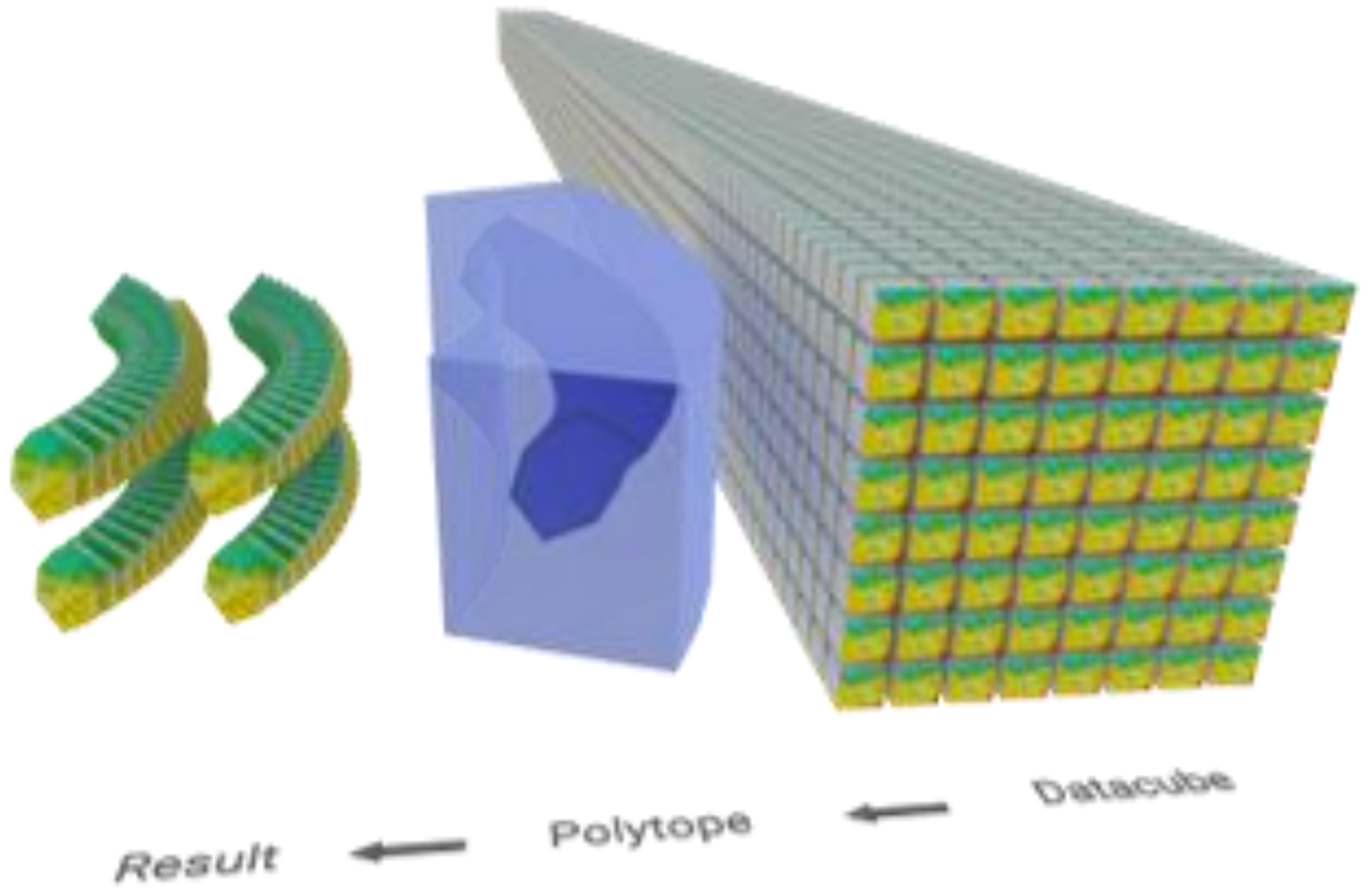
## 2. Motivation

Weather data can be represented as a high-dimensional datacube (usually 6D or 7D). The current data access mechanism on such datacubes can only return "bounding boxes" of data, which does not capture a wide range of possible user requests. Moreover the current data extraction mechanism has to read complete global fields from the I/O system, which is very inefficient. This quickly becomes a bottleneck to scalability.



ECMWF's NWP Datacube

To tackle these scalability issues, ECMWF is taking a new approach to data and feature extraction. The novel Polytope concept revolves around retrieving n-dimensional polygons, or so-called "polytopes", instead of bounding boxes from the datacube.

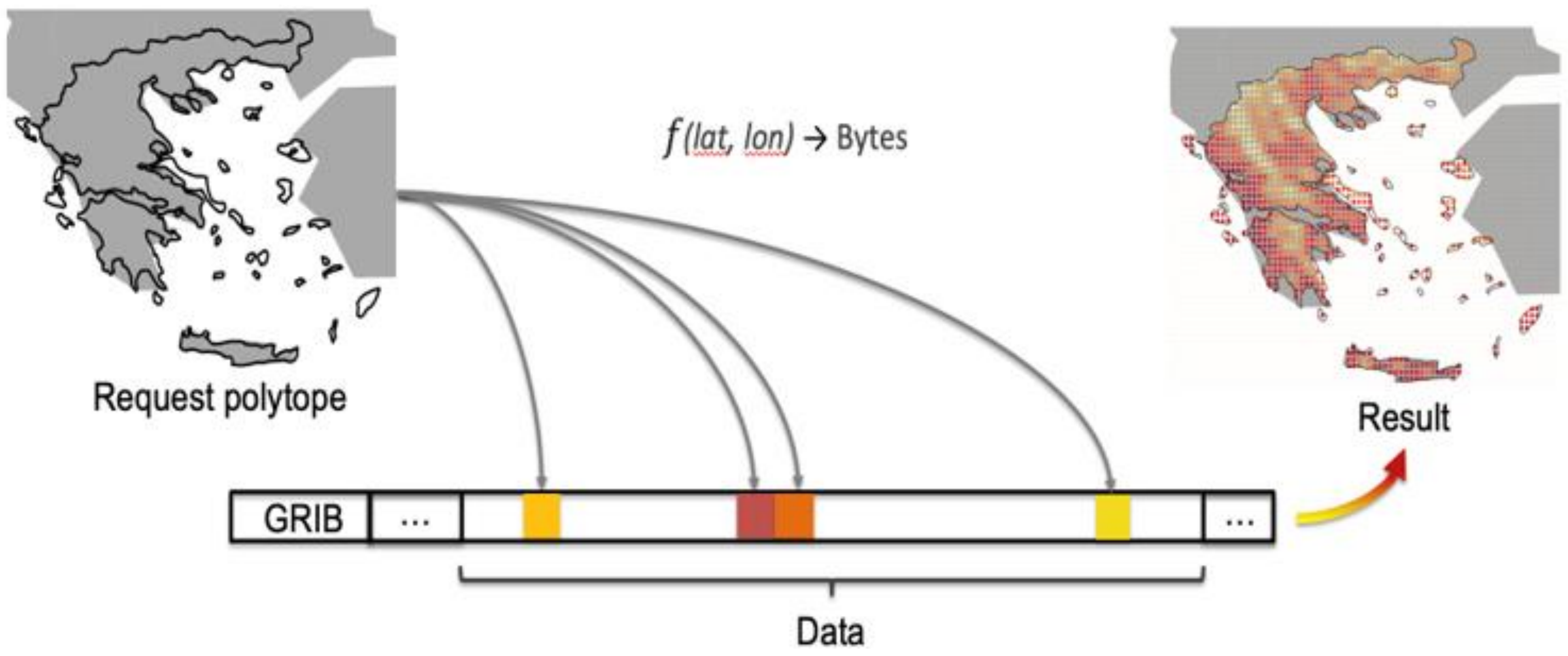


6D Polytope Data Extraction

## 3. Mechanism

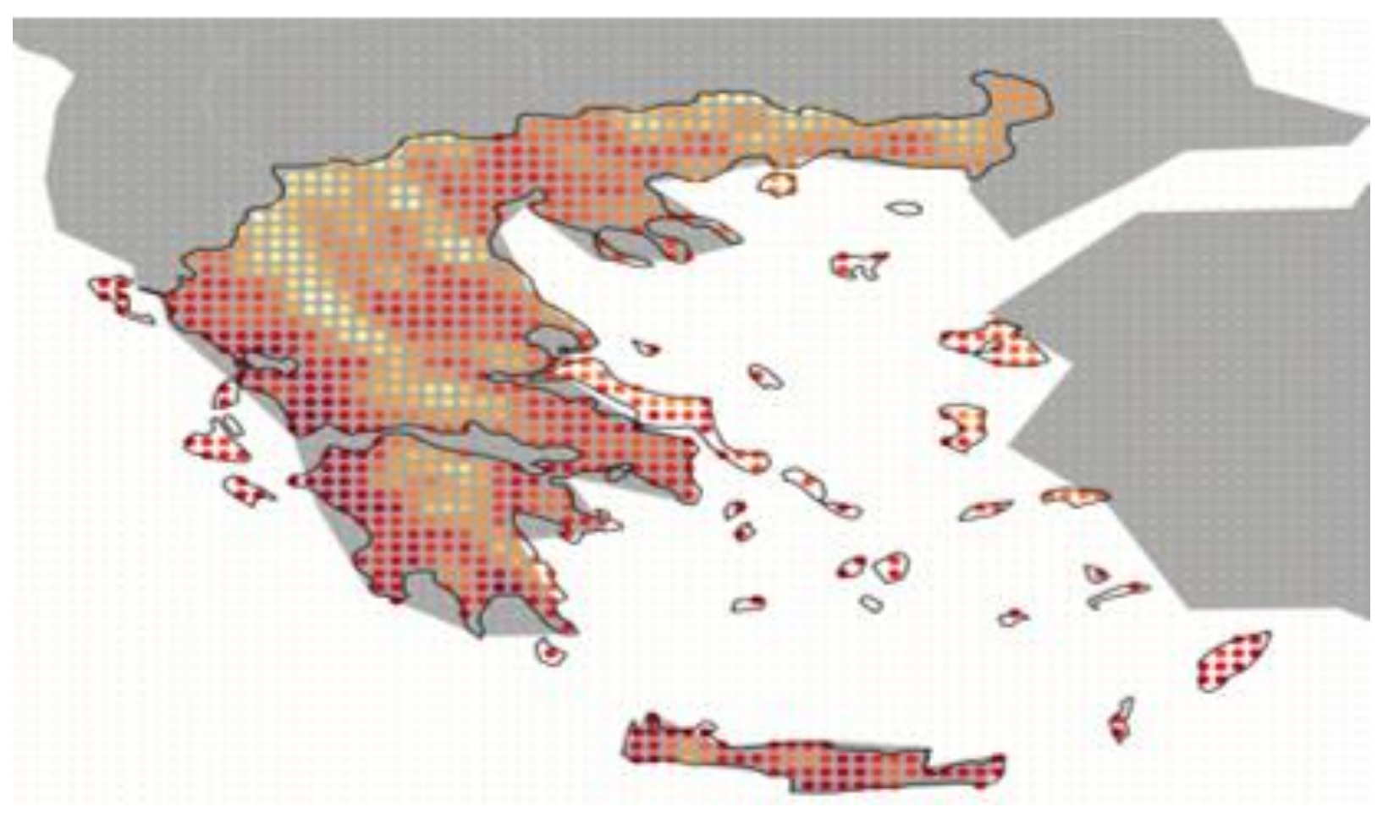
Polytope allows users to request arbitrary complex shapes. However, the Polytope algorithm is much more than a simple post-processing technique used to cut intricate shapes.

Indeed, the Polytope algorithm can be thought of as a function which calculates the precise bytes to retrieve from the datacube. Importantly, this means that only those specific bytes will be accessed on the I/O system, instead of whole fields. This is much more efficient than traditional techniques.



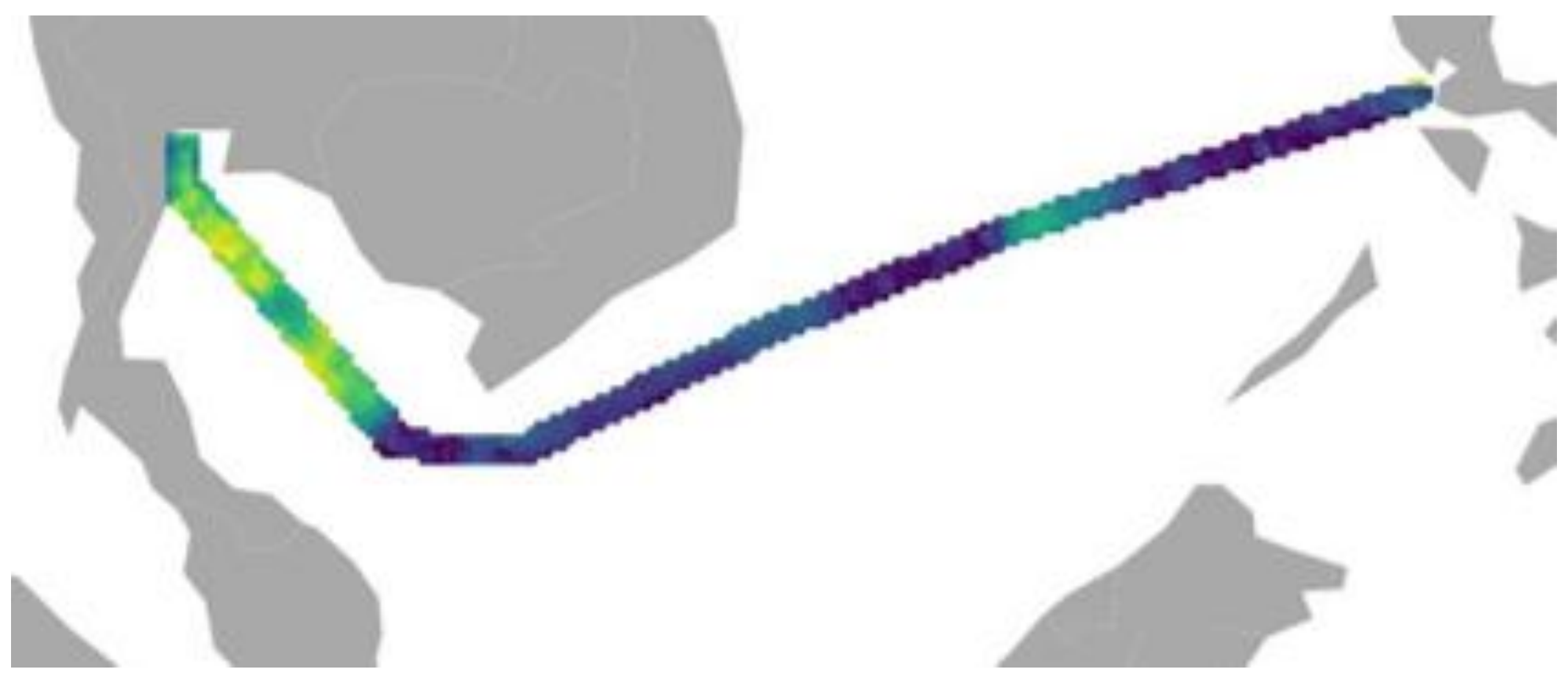
## 4. Examples

**User A** wants to extract temperature data over Greece to assess the risk of a fire happening. Using Polytope, she can extract exactly the relevant data points from the datacube.

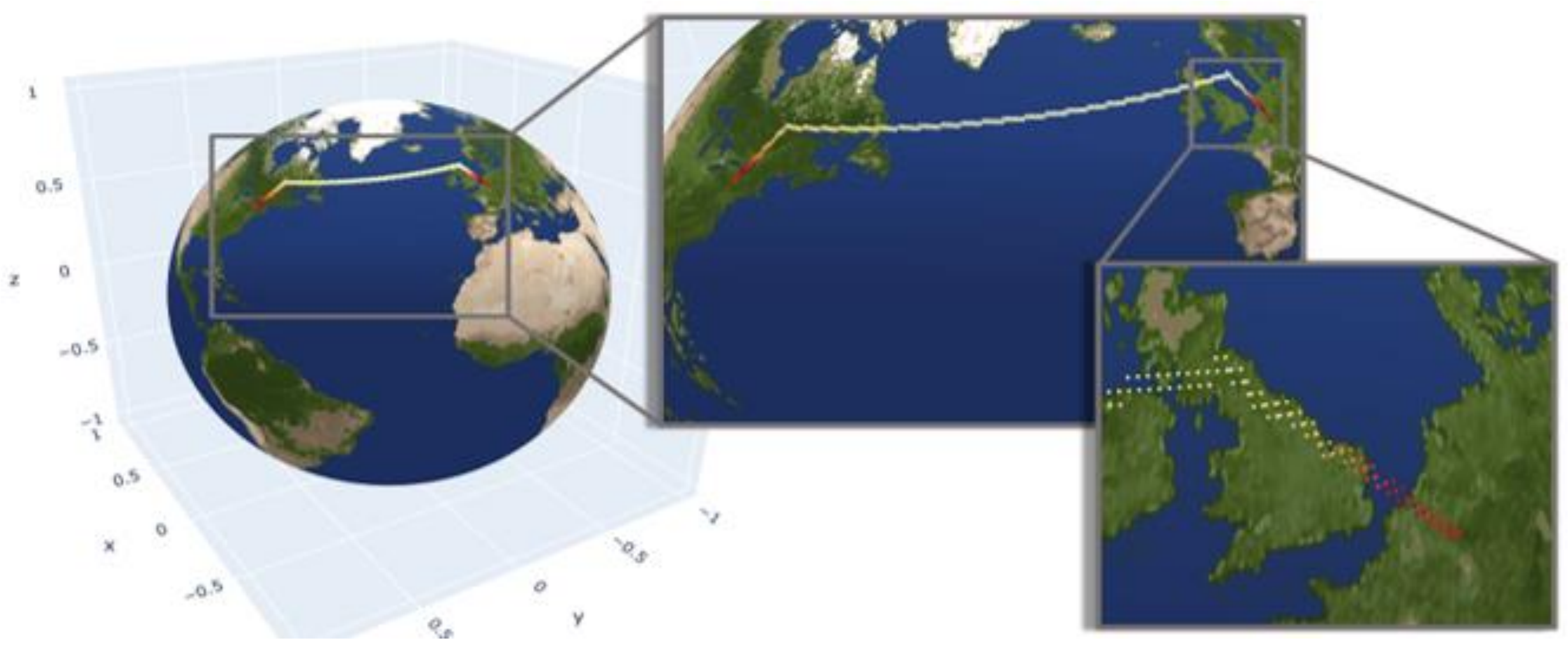


**User B** wants to extract temperature data over Italy for the past 20 years to assess the local effects of climate change. Polytope returns the coloured points below.

**User C** wants to extract wind speed data over a shipping route in the Indian Ocean. Polytope returns only the points corresponding to the ships' location in space and time.



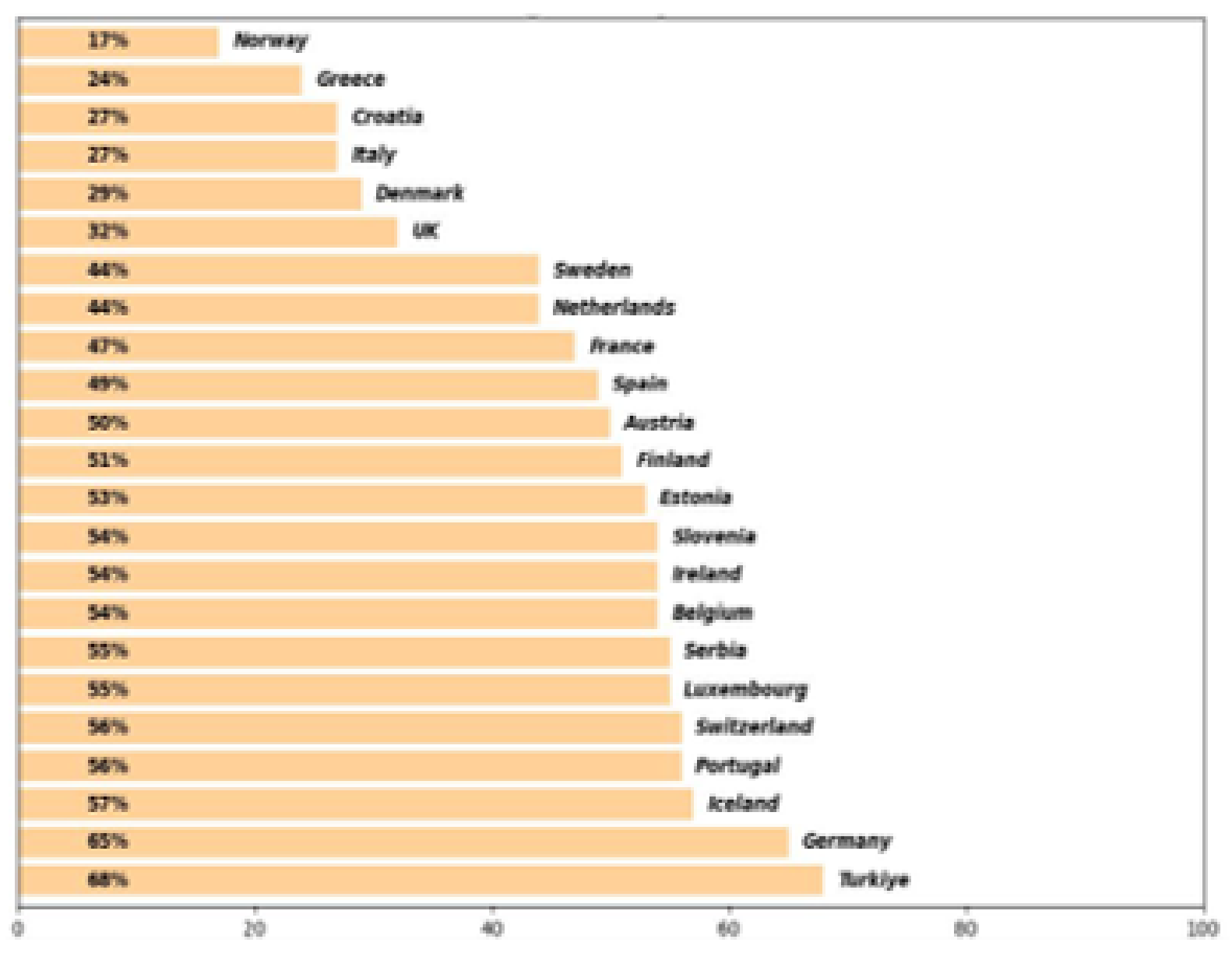
**User D** wants to extract wind speed data over a flight path from Paris to New York. Only the exact spatio-temporal points corresponding to the flight are extracted.



These examples clearly highlight how Polytope reduces the post-processing burden on users once the data is retrieved. Indeed, in the examples above, we see that the users get back exactly the data they asked for and thus do not need to take any additional processing steps after extraction.

## 5. Data Reduction

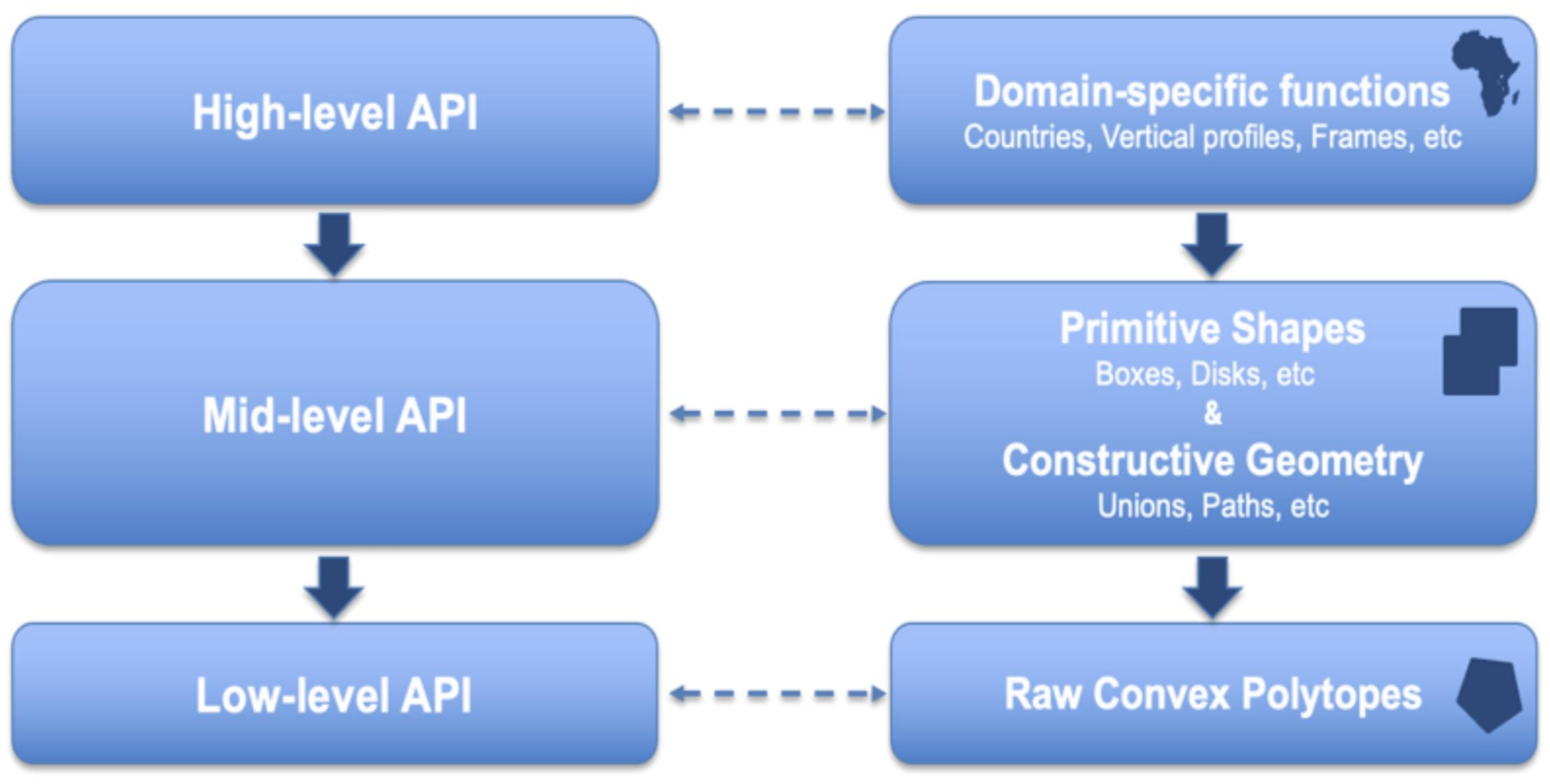
Already for country extraction, we observe a 30-70% data reduction compared to the bounding box approach. This is a significant improvement, especially when considering only these extracted bytes are read from the I/O system instead of whole fields. Data reductions exceed 99.99% for higher-dimensional request shapes such as timeseries, vertical profiles or trajectories.



Country Data Reduction (%)

## 6. APIs

To accommodate for a range of users and possible requests, Polytope has 3 different API levels: the low-, mid- and high-level APIs, summarised in the figure below.


High-level API

Domain-specific functions
Countries, Vertical profiles, Frames, etc

Mid-level API

Primitive Shapes
Boxes, Disks, etc
&
Constructive Geometry
Unions, Paths, etc

Low-level API

Raw Convex Polytopes

Funded by the European Union

Destination Earth implemented by ECMWF esa EUMETSAT