EMS Annual Meeting 2021 | 7 September 2021

Making ECMWF Open Data more easily accessible via cloud-based services

Julia Wagemann, Umberto Modigliani, Stephan Siemen, Vasileios Baousis, Florian Pappenberger

ECMWF

ECMWF Open Data (real-time) – Overview and roadmap

- **Aim:** to gradually (until 2025) open up a subset of the ECMWF model outputs (forecast data)
  - Currently only available to ECMWF Member and Co-operating states and licensed entities

- **License:** CC-BY-4.0

- **Format**: GRIB, but could be transformed to netCDF

- **Total data volume (per day):** 710 GB

- **Domain**: global

- **Horizontal resolution**: 0.4 degrees

- **Steps**: up to step 240 (by 3 or 6)

- **Output frequency**: 6 hours (00, 06, 12, 18 UTC)

- **Levels**: single and pressure levels (~ 9 levels incl. 1000, 925, 850, 700, 500, 300, 250, 200, 50)
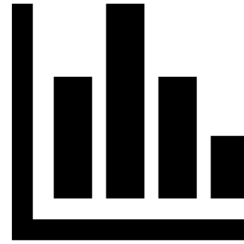
Open Data - Implications

Diversification of users

Spikes in data access

FAIR principles

Shift towards non-expert users

Especially **forecast data is of value as soon as the forecast is issued**

Open data policy is important pre-requisite

Better understanding of users and their requirements

Expected spike in data requests every 6 hours

But **FAIR principles have to be fulfilled** to unleash the full potential of data uptake

Survey on user requirements of open Big Earth data – Overview

- Nov 2018 to May 2019
- 32 questions
- 231 respondents

- Majority from Europe and USA / Canada
- 70% between 30-50 years
- Around half indicated to work in research, followed by public and private sector

Analysis of the **current state**

Wagemann et al. (2021): Users of open Big Earth data – An analysis of the current state. *Computers and Geosciences 2021. (in press)*

| Data use | Data challenges | Data handling |

Future requirements

Wagemann et al. (2021): A user perspective on future cloud-based services for Big Earth data. *International Journal of Digital Earth. (under review)*
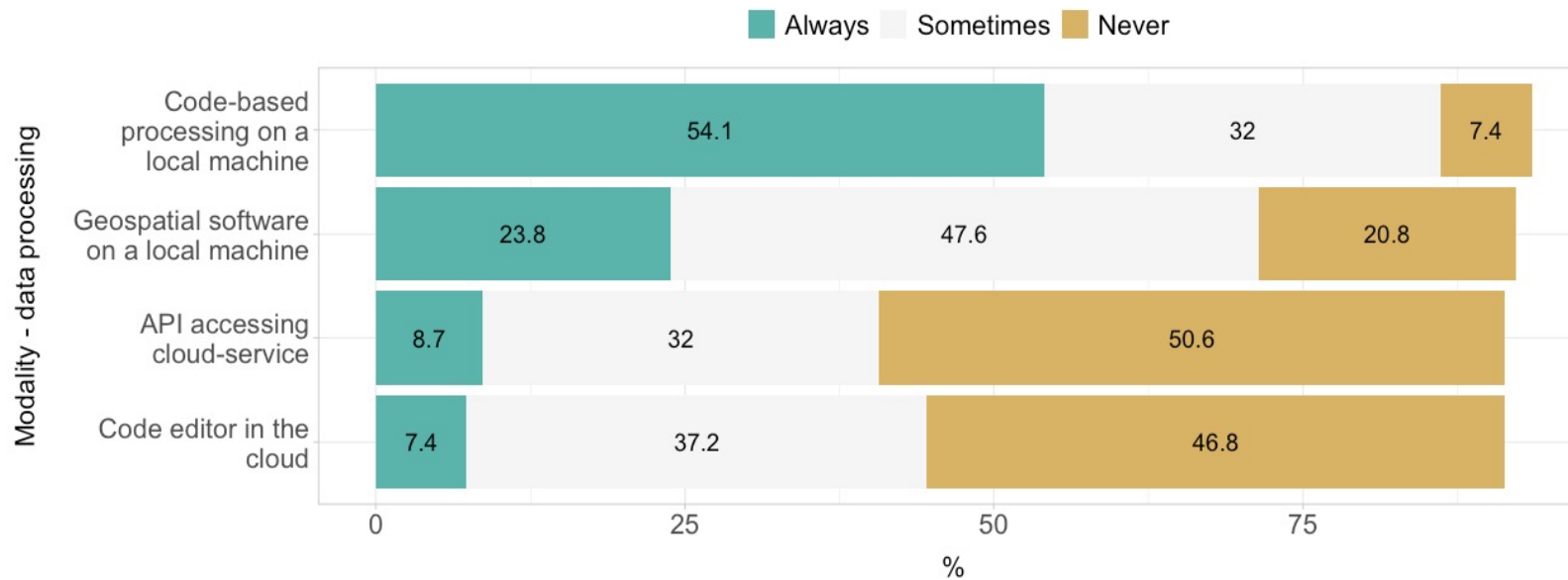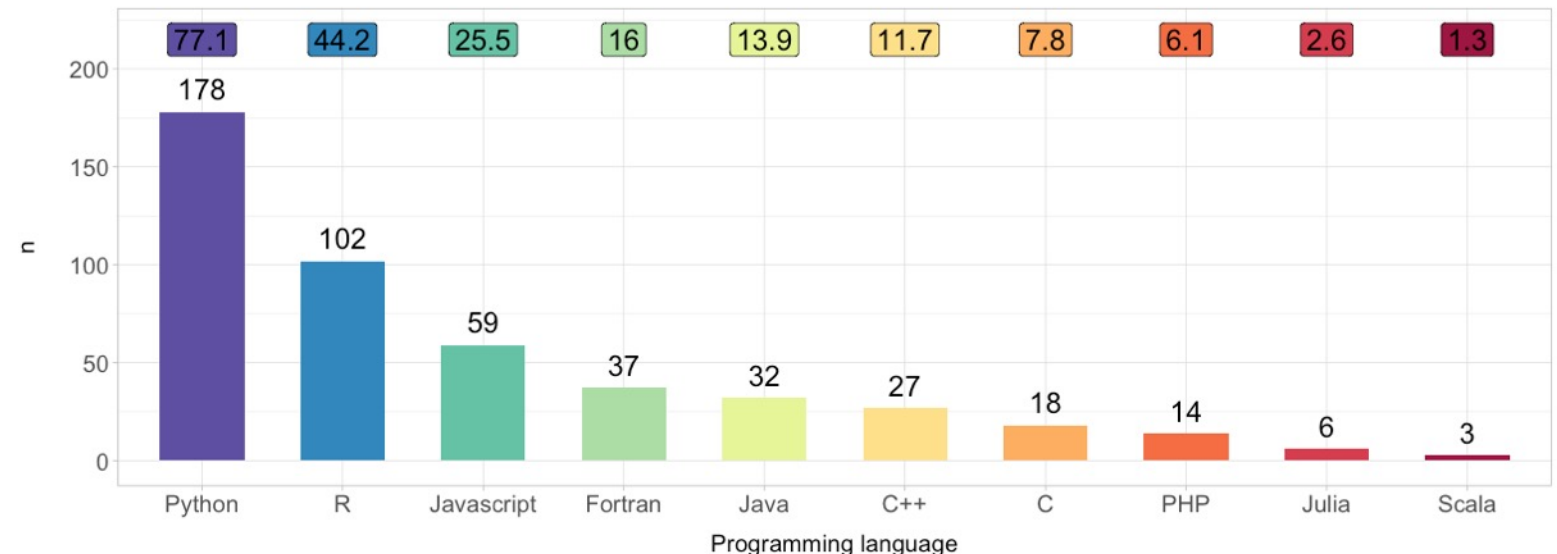
| Future data services |

Survey on user requirements of open Big Earth data – Main outcomes
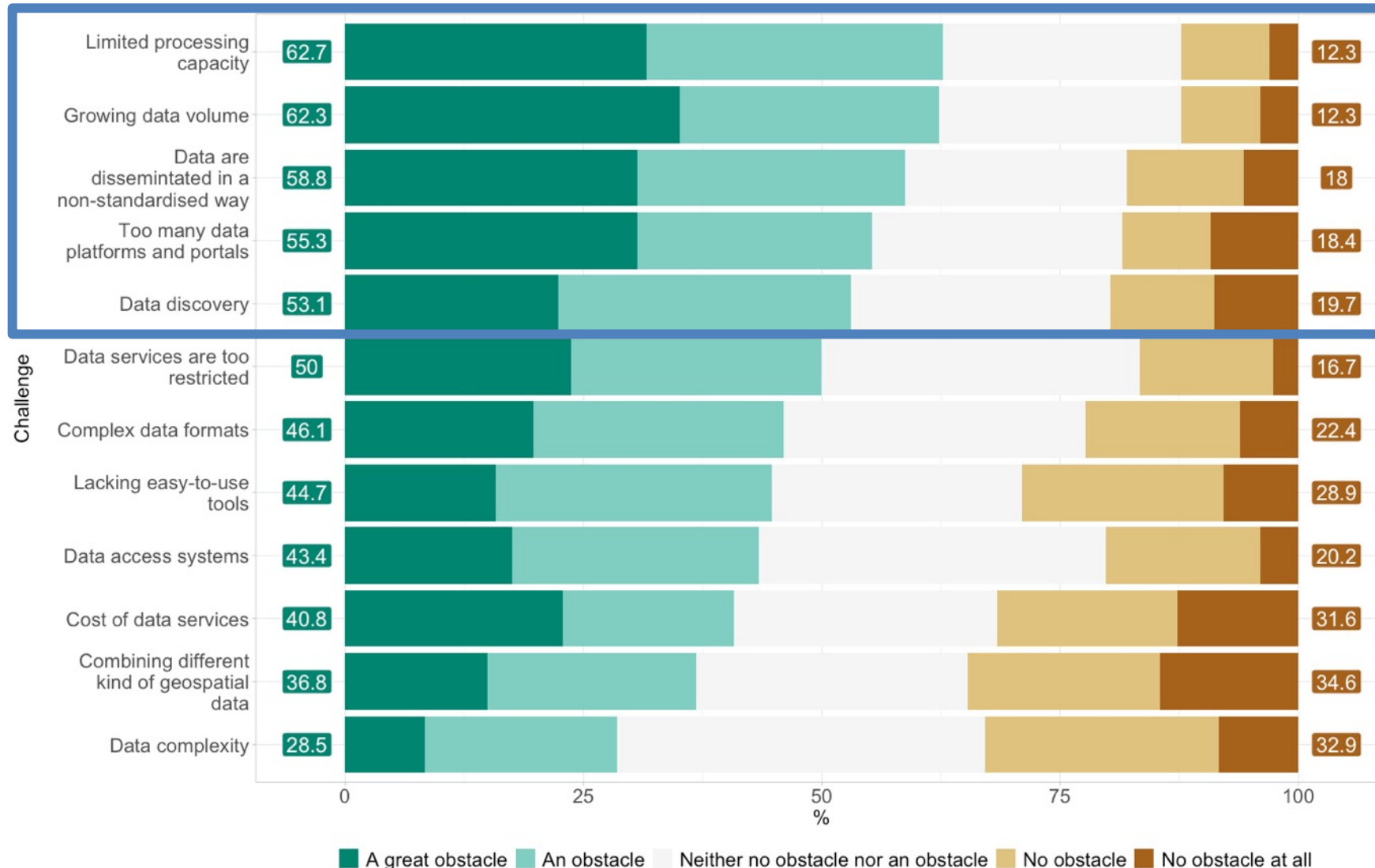

- Majority processes data locally with a code-based processing routine

- Python and R are the most used programming languages

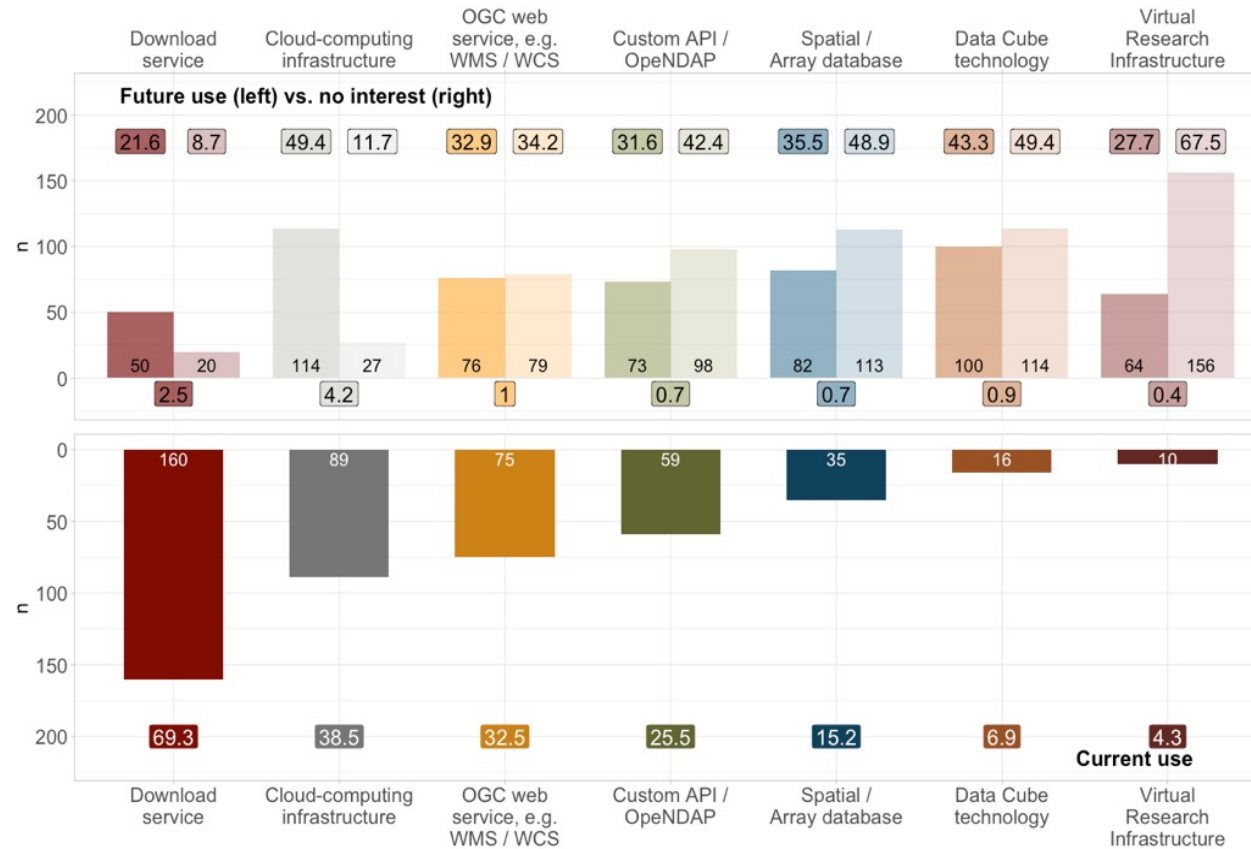- Python twice as much as R used for meteorological and climate data

ECMWF

Survey on user requirements of open Big Earth data – Main outcomes


Top 5 challenges related to:

- growing data volume

- limited processing capacity, and

- heterogeneity of data and data systems

# Survey on user requirements of open Big Earth data – Main outcomes

- ~ 70% are using a Download service to access data

- High interest in future use of cloud-based services

- Continued future interest in download services

70% are motivated to migrate to cloud-based services in the future
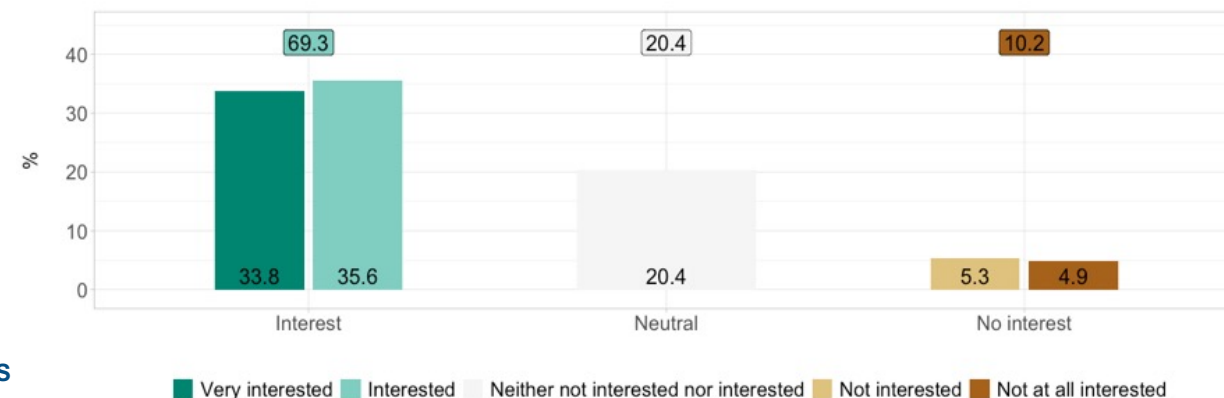
Survey on user requirements of open Big Earth data – Main outcomes

- Preference of publicly funded cloud services stronger in Europe and in publicly-funded work places (e.g. university, government)

- Only 1 out of 4 is able to estimate technical requirements for data storage and processing in the cloud

- 2 out of 3 rated all security aspects as risk or major risk

Preference of cloud service policy
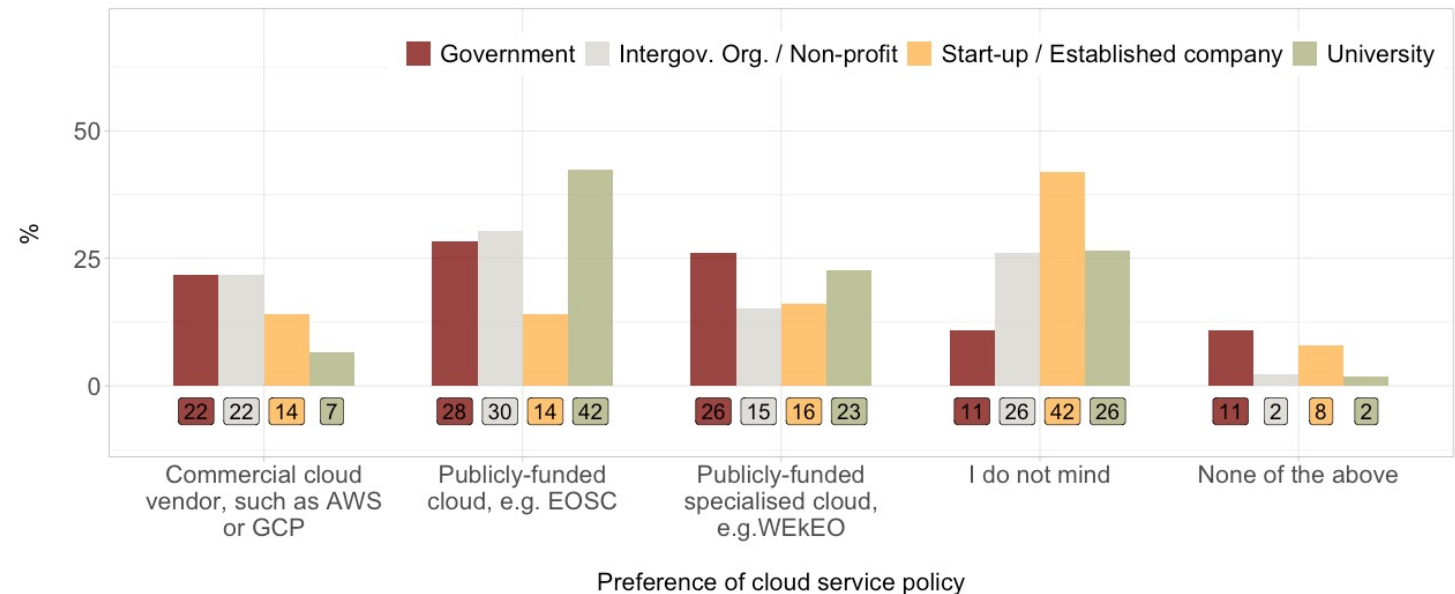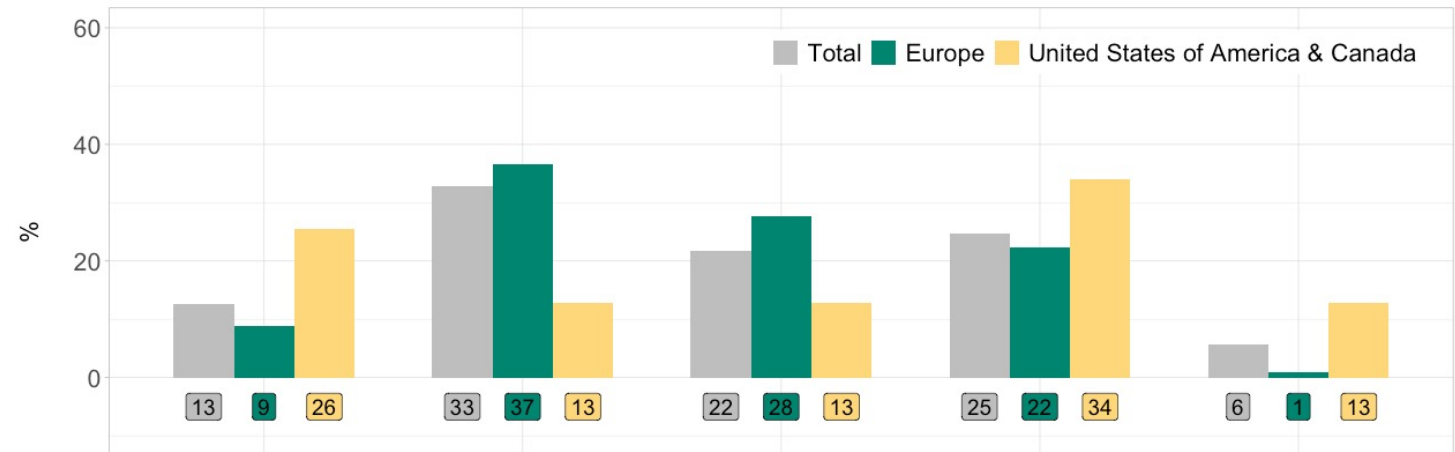
ERA5 on Google Earth Engine - an interoperability trial

ERA5 Daily aggregates – Latest climate reanalysis produced by ECMWF / Copernicus Climate Change Service
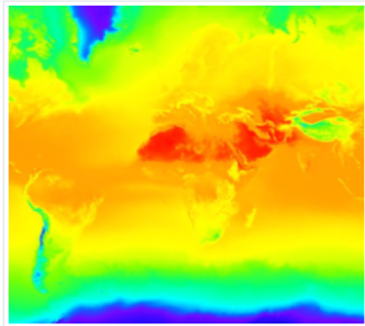
Dataset Availability

1979-01-02T00:00:00 - Present

Dataset Provider

ECMWF / Copernicus Climate Change Service

Earth Engine Snippet

`ee.ImageCollection("ECMWF/ERA5/DAILY")` ☒

Tags

wind   temperature   dewpoint   surface   pressure

precipitation   climate   ecmwf   copernicus   era5

reanalysis

https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_DAILY

https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_MONTHLY

Description	Bands	Image Properties	Terms of Use	Citations

ERA5 is the fifth generation ECMWF atmospheric reanalysis of the global climate. Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset. ERA5 replaces its predecessor, the ERA-Interim reanalysis.

ERA5 DAILY provides aggregated values for each day for seven ERA5 climate reanalysis parameters: 2m air temperature, 2m dewpoint temperature, total precipitation, mean sea level pressure, surface pressure, 10m u-component of wind and 10m v-component of wind. Additionally, daily minimum and maximum air temperature at 2m has been calculated based on the hourly 2m air temperature data. Daily total precipitation values are given as daily sums. All other parameters are provided as daily averages.

ERA5 data is available from 1979 to three months from real-time. More information and more ERA5 atmospheric parameters can be found at the Copernicus Climate Data Store.

Provider's Note: Daily aggregates have been calculated based on the ERA5 hourly values of each parameter.

ERA5 on Google Earth Engine Overview

7 TBs

3 IMAGE COLLECTIONS

HOURLY | DAILY | MONTHLY

104,000

243,113

9 VARIABLES

- 2m air temperature (min, mean, max)
- 2m dewpoint temperature
- Total precipitation

- Surface Pressure
- Mean sea-level pressure
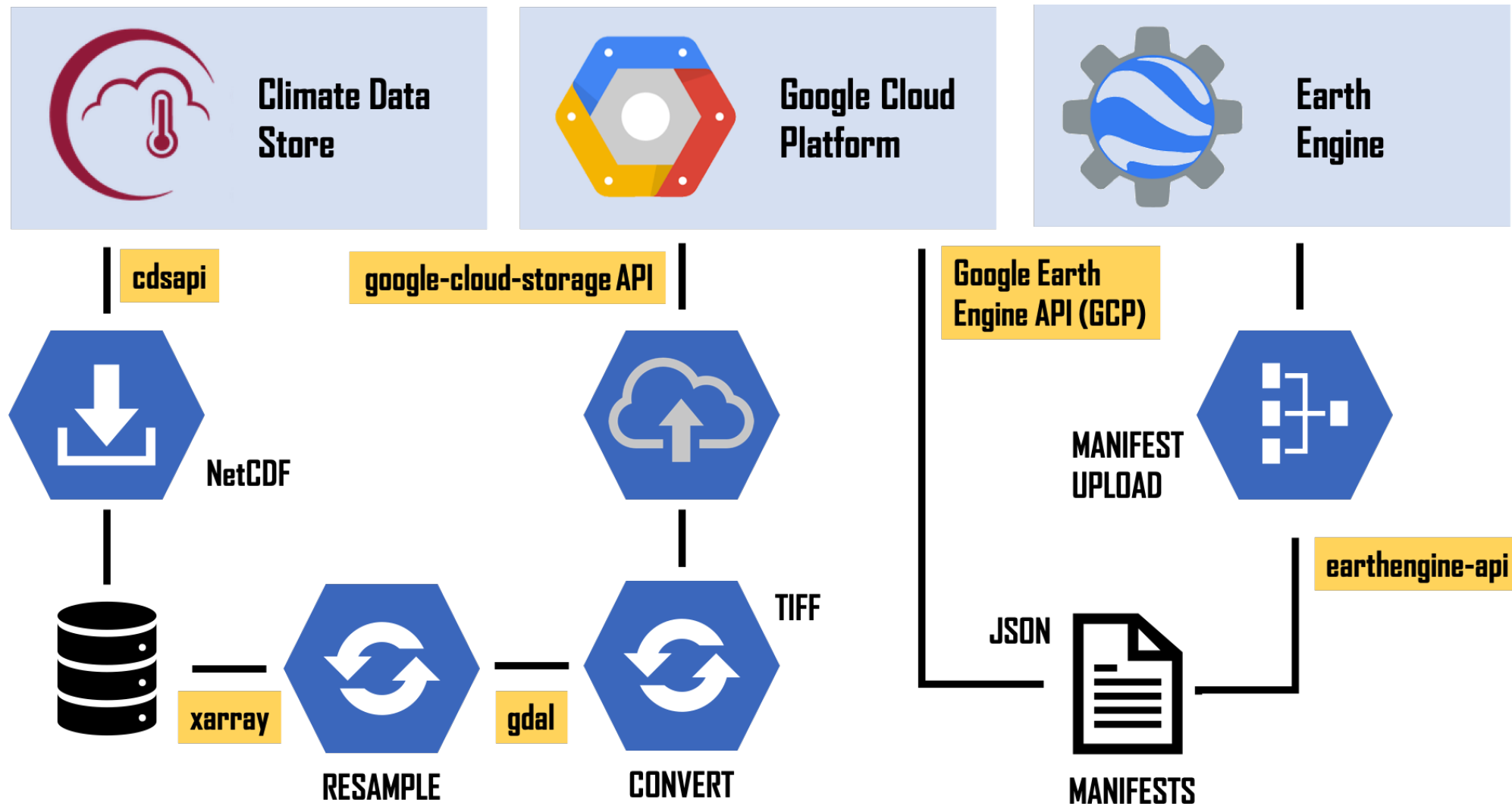- 10m u- and v-component of wind

ERA5 on Google Earth Engine – Interoperability challenge

Future plans

Currently in discussions with commercial cloud service provider to make ECMWF Open Data (real-time) available via a public datasets program

Conclusion

Offering Open Data via a cloud-based service improves accessibility of open data, but interoperability of cloud-based systems remains a key challenge

Users are motivated to migrate to cloud-based services, but show an insufficient literacy in cloud services and a lack of trust in terms of security and emerging costs

Substantial investment efforts needed for capacity-building and training

To make open data more FAIR in the cloud, an internationally coordinated effort among data providers of Big Earth data is required

Thank you!

More information:

Wagemann, J., Siemen, S., Seeger, B. and J. Bendix (2021):
[Users of open Big Earth data – An analysis of the current state](#).
Computers and Geosciences (2021). doi:
10.1016/j.cageo.2021.104916

[ERA5 in Google Earth Engine Github repository](#)

[ERA5 in Google Earth Engine Daily Aggregates](#)
[ERA5 in Google Earth Engine Monthly Aggregates](#)

ECMWF