

Challenges and
Needs of

REPRODUCIBLE WORKFLOWS

of **Open Big Weather** and
Climate Data

#repwork19

Reading, 14 Oct 2019

Julia Wagemann

PhD candidate at University of Marburg
Visiting Scientist at ECMWF

@JuliaWagemann

Basic **NEEDS** for reproducible workflows:

❖ **Open data**

Basic **NEEDS** for reproducible workflows:

❖ **Open data**

❖ **Open-source software
/ FOSS4G**

Basic **NEEDS** for reproducible workflows:

❖ **Open data**

❖ **Open-source software
/ FOSS4G**



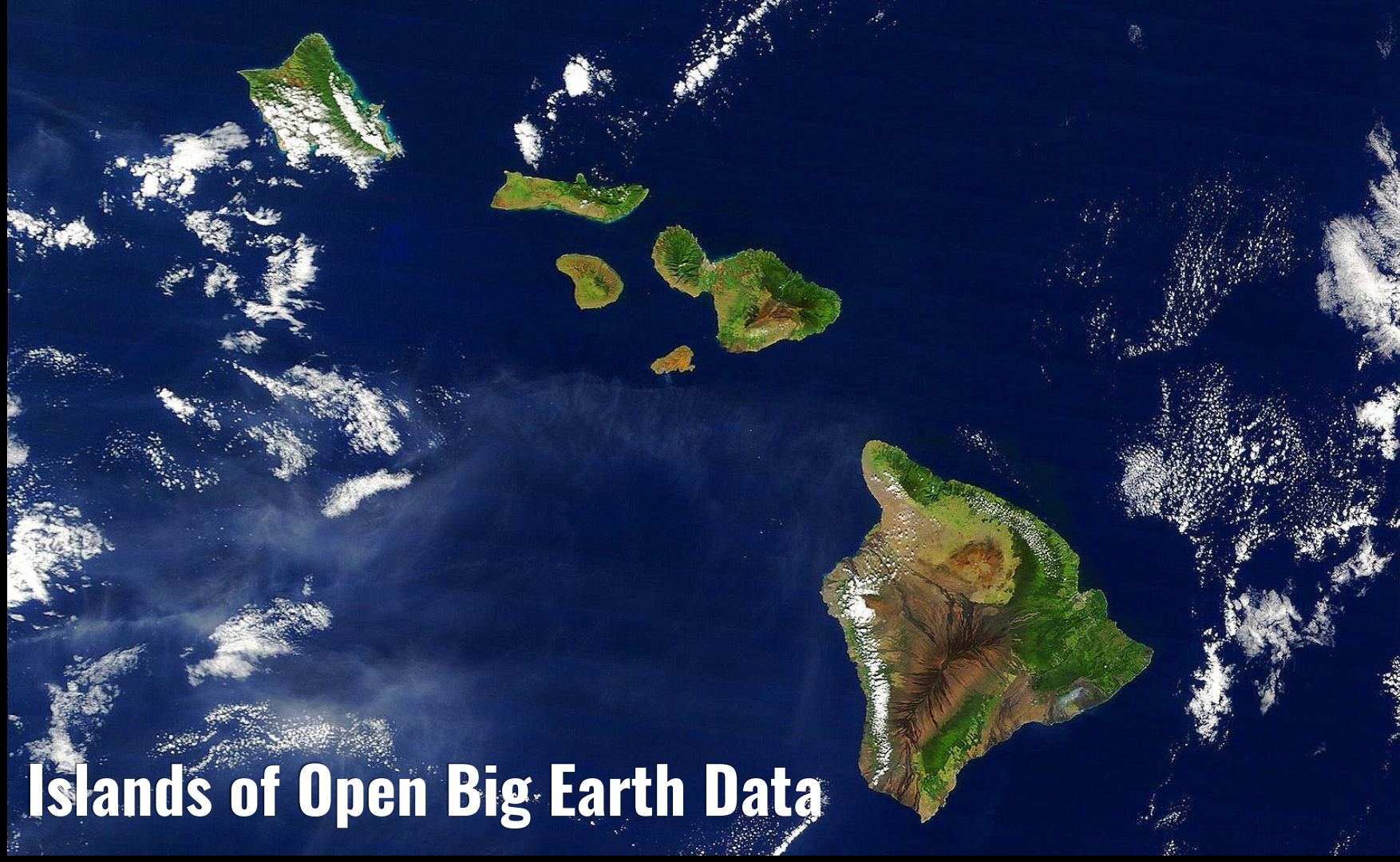
Basic **NEEDS** for reproducible workflows:

❖ Open data



❖ Open-source software / FOSS4G





Islands of Open Big Earth Data

A satellite image of the Hawaiian Islands, viewed from space. The islands are shown in green and brown, indicating vegetation and topography. The surrounding ocean is deep blue, and white clouds are scattered across the scene. Two yellow text boxes are overlaid on the image: one in the top left corner and one in the bottom right corner. The title 'Islands of Open Big Earth Data' is at the bottom left.

**Earth Observation
community**

**Meteorological /
climate community**

Islands of Open Big Earth Data

A satellite view of the Hawaiian Islands, showing the main islands and surrounding waters. The islands are green and brown, contrasting with the deep blue of the ocean and the white of the clouds.

Earth Observation community

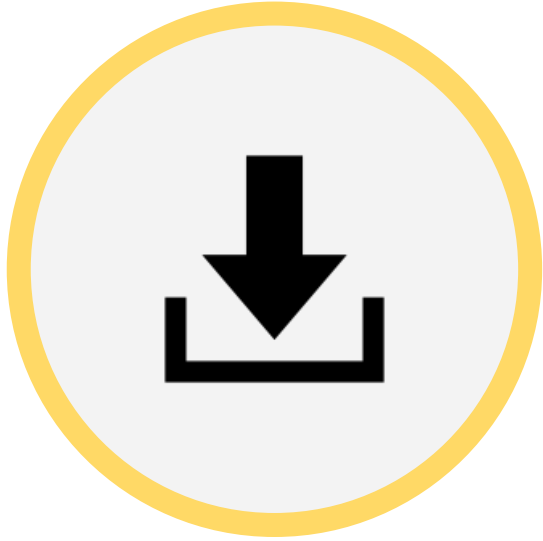
- Google Earth Engine
- GeoTiff, JPEG2000

- Copernicus Climate Data Store
- GRIB, NetCDF

**Meteorological /
climate community**

Islands of Open Big Earth Data

Principle components of a Big Earth Data workflow



Access



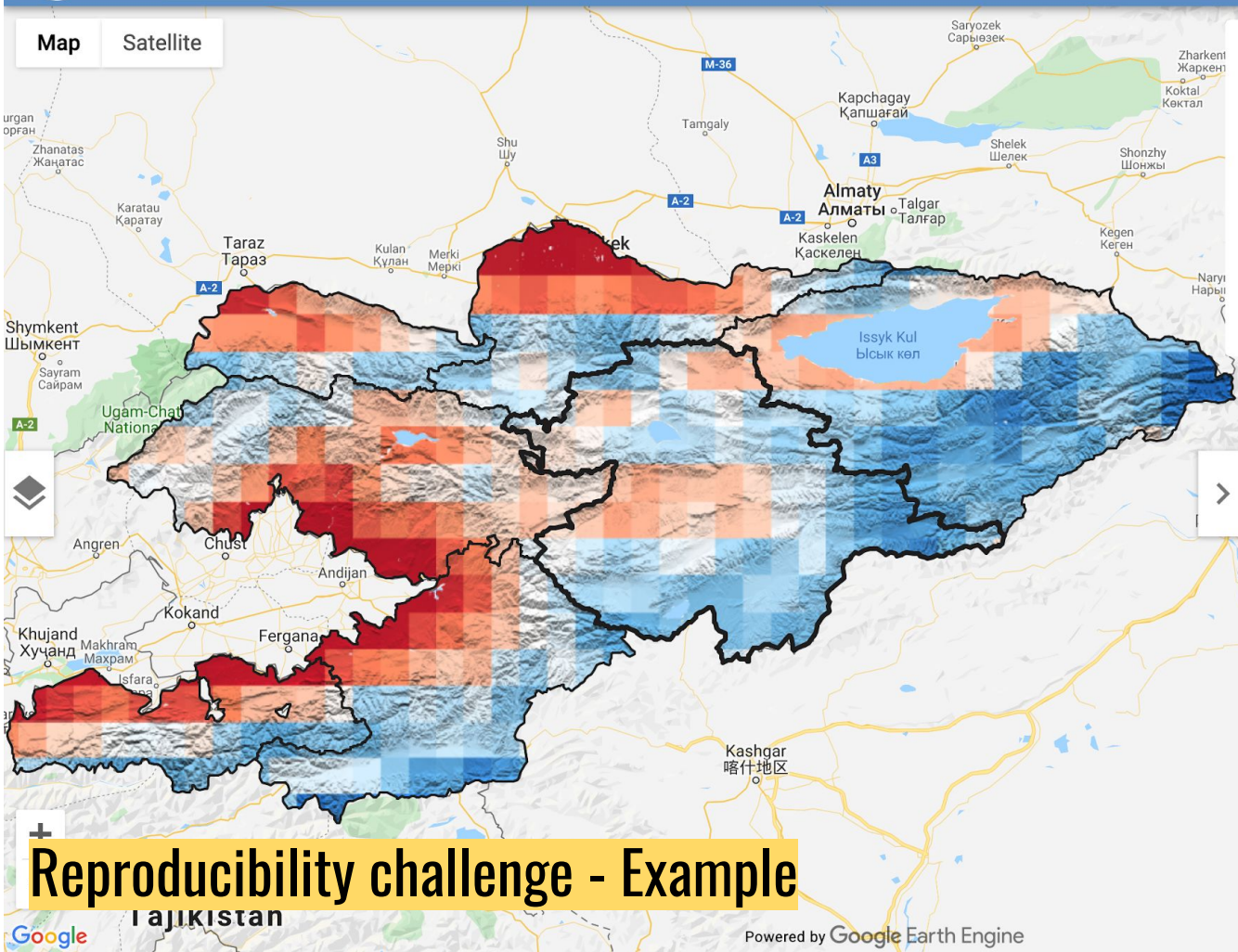
Processing



Visualization

**EMERGENCY COMPARE****Map**

Satellite



Reproducibility challenge - Example

Powered by Google Earth Engine

Naryn

Metrics

Precipitation

Temporal aggregation

Monthly time series

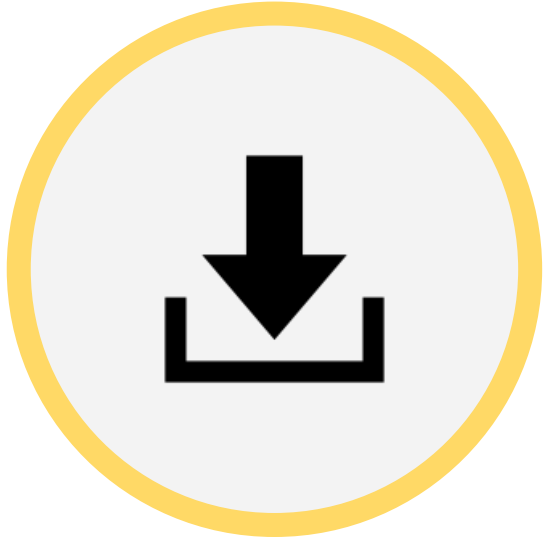
🔄 2018 - 2018

PROCESS**Naryn**

Precipitation

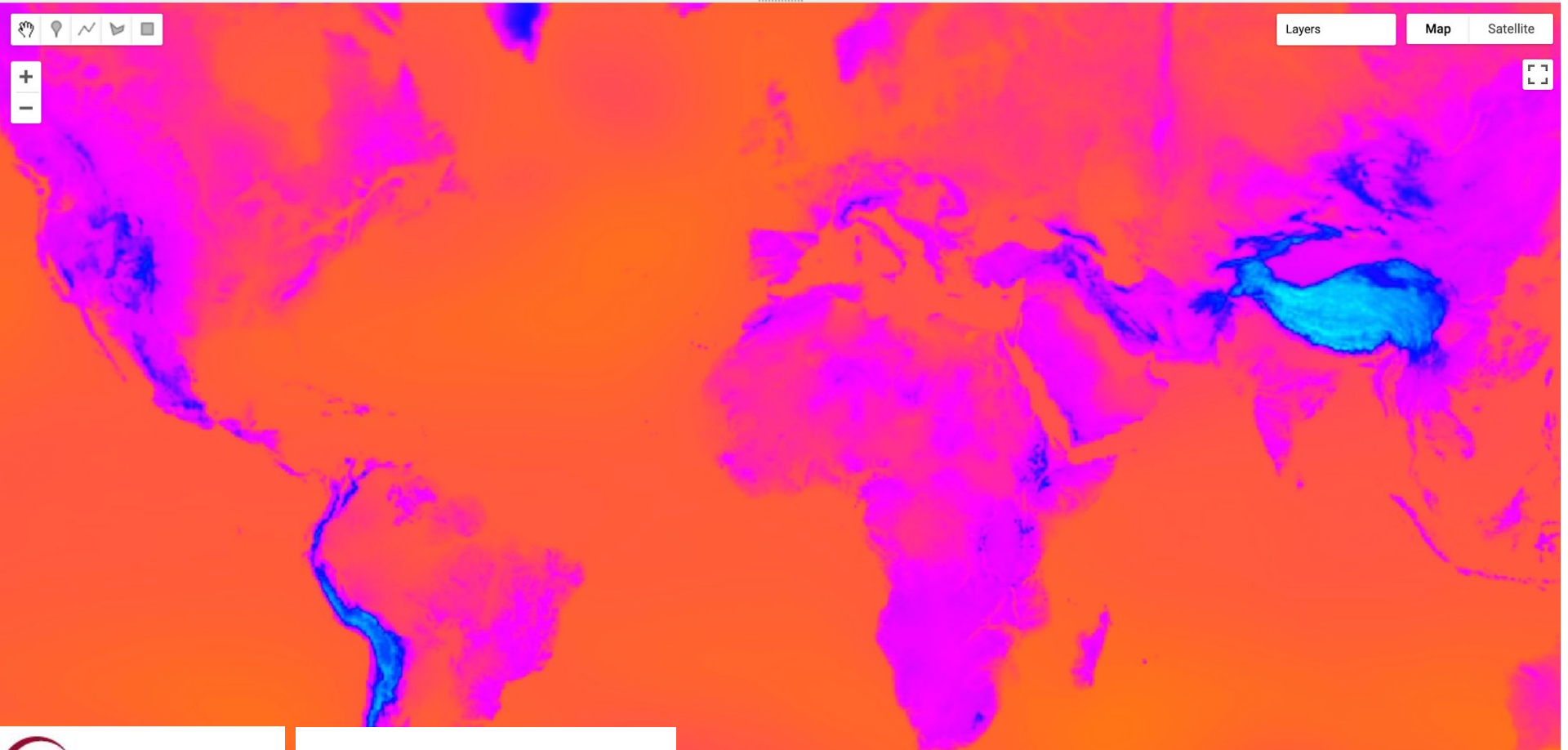


Reproducibility challenge - DATA ACCESSIBILITY



Access

- different data are accessible via **different data access systems**
- it is **still about downloading data**
- **community-specific data formats**
(GRIB, NetCDF, GeoTiff)
- **data structure and complexity**
(analyses vs forecast, multiple dimensions)



Layers

Map

Satellite



Reproducibility challenge - DATA ACCESSIBILITY - Question

**If a process involves data
download that might take
several months...**

Do we call it reproducible?

One great example of a reproducible workflow...



Matteo De Felice
@matteodefelice

Can we use a #Python only workflow to calculate the skill scores on @CopernicusECMWF seasonal forecasts? Yes, and the whole process is *unbelievably* faster than a few years ago. The @GoogleColab notebook is accessible here:

[colab.research.google.com/drive/1wWHz_SM ...](https://colab.research.google.com/drive/1wWHz_SM...) 1/4



8:32 AM - 12 Oct 2019

13 Retweets 62 Likes



4 13 62

cdsapi

xarray

cfrib



xESMF

cartopy

https://colab.research.google.com/drive/1wWHz_SMCHNuos5fxWRUJTcB6wgkTJQCR



Tom Kralidis

@tomkralidis

The older I get, the more I find myself quoting Cliff Kottman: "Interoperability doesn't happen by accident". Amen!

@opengeospatial

opengeospatial.org/ogc/honors/kot...

6:42 AM - 9 Oct 2019

10 Retweets 27 Likes



↻ 10

♥ 27



Interoperability does not
happen by accident. (Cliff Kottman)

REPRODUCIBILITY

does not happen
either by accident

Reproducibility is *'going the extra mile'*



Additional NEEDS for reproducible workflows:

- ❖ Open data
- ❖ Open-source software / FOSS4G

Additional NEEDS for reproducible workflows:

- ❖ Open data
- ❖ Open-source software / FOSS4G



Reproducibility
as personal code of conduct



Additional **NEEDS** for reproducible workflows:

- ❖ Open data
- ❖ Open-source software / FOSS4G



Reproducibility
as personal code of conduct



Prioritise
Interoperability
of data systems



**“Problems cannot be solved
by the same level of thinking
that created them!”**

(Albert Einstein)

#repwork19

Reading, 14 Oct 2019

Thank you!

Questions?

Julia Wagemann

PhD candidate at University of Marburg
Visiting Scientist at ECMWF

@JuliaWagemann