

OpenEO: an API that unifies access to Earth Observation data processing platforms



ifgi

Institute for Geoinformatics
University of Münster

Edzer Pebesma

52°North Geospatial Sensor Webs Conference 2018, Sept 3-5 2018



Open **Innovation**
Open **Science**
Open to the **World**

– a vision for Europe





EUROPEAN COMMISSION

DIRECTORATES-GENERAL FOR RESEARCH AND INNOVATION (RTD) AND
COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY (CONNECT)

BACKGROUND DOCUMENT

PUBLIC CONSULTATION

'SCIENCE 2.0': SCIENCE IN TRANSITION

0. OVERVIEW OF THE CONSULTATION PROCESS

This background paper to the online public consultation gives a short description of the term 'Science 2.0' as used in this consultation. The goal of the consultation is to better understand the full potential of 'Science 2.0' as well as the desirability of any possible policy action.

European Commission - Press release

European Cloud Initiative to give Europe a global lead in the data-driven economy

Brussels, 19 April 2016

The Commission today presented its blueprint for cloud-based services and world-class data infrastructure to ensure science, business and public services reap benefits of big data revolution.

Europe is the largest producer of scientific data in the world, but insufficient and fragmented infrastructure means this 'big data' is not being exploited to its full potential. By bolstering and interconnecting existing research infrastructure, the Commission plans to create a new **European Open Science Cloud** that will offer Europe's 1.7 million researchers and 70 million science and technology professionals a virtual environment to store, share and re-use their data across disciplines and borders. This will be underpinned by the **European Data Infrastructure**, deploying the high-bandwidth networks, large scale storage facilities and super-computer capacity necessary to effectively access and process large datasets stored in the cloud. This world-class infrastructure will ensure Europe participates in the global race for high performance computing in line with its economic and knowledge potential.

Focusing initially on the scientific community - in Europe and among its global partners -, the user base will over time be enlarged to the public sector and to industry. This initiative is part of a package of measures to strengthen Europe's position in data-driven innovation, to improve competitiveness and cohesion and to help create a [Digital Single Market](#) in Europe ([press release](#)).

Carlos **Moedas**, Commissioner for Research, Science and Innovation, said: *"Our goal is to create a European Open Science Cloud to make science more efficient and productive and let millions of researchers share and analyse research data in a trusted environment across technologies, disciplines and borders. We listened to the scientific community's plea for an infrastructure for Open Science and with this comprehensive plan we can get down to work. The benefits of open data for Europe's science, economy and society will be enormous."*

Günther H. **Oettinger**, Commissioner for the Digital Economy and Society, said: *"The European Cloud Initiative will unlock the value of big data by providing world-class supercomputing capability, high-speed connectivity and leading-edge data and software services for science, industry and the public sector. With this initiative, our ambition is to be in the global top-three in high performance computing by 2020. We will also be looking into the potential of quantum technologies which hold the promise to solve computational problems beyond current supercomputers."*

NATURE | NEWS FEATURE

[E-alert](#) | [RSS](#) | [Facebook](#) | [Twitter](#)

1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016



PDF

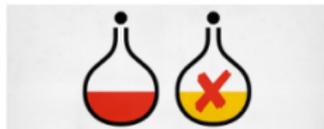


Rights & Permissions



More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature's* survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research.

Crisis talks



1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.



Like

Recent

Read

Commented

1. [US advised to stick with troubled fusion reactor ITER](#)
Nature | 26 May 2016
2. [Two-hundred-terabyte maths proof is largest ever](#)
Nature | 26 May 2016
3. [Sights and sounds of seabed cities](#)
Nature | 26 May 2016

Newsletter

The best science news from *Nature* and beyond, direct to your inbox every day.

Email*

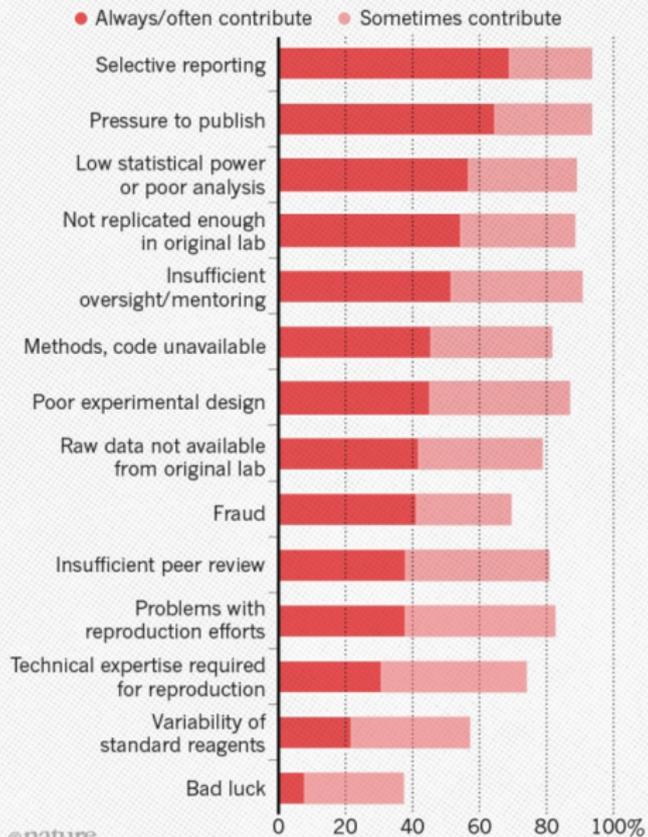
Submit

IS THERE A REPRODUCIBILITY CRISIS?



WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

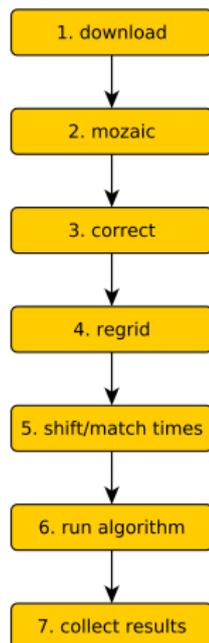
Many top-rated factors relate to intense competition and time pressure.



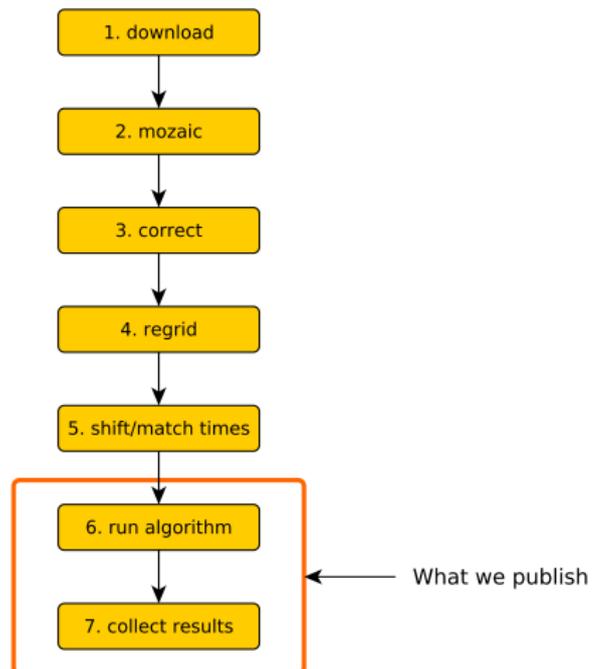
What is data science?

- ▶ partly business driven: "making money out of numbers"
- ▶ "science begins with a question, data science begins with data"
- ▶ combines: domain knowledge, understanding data analysis, computational skills
- ▶ *share* (software, scripts) how we do stuff

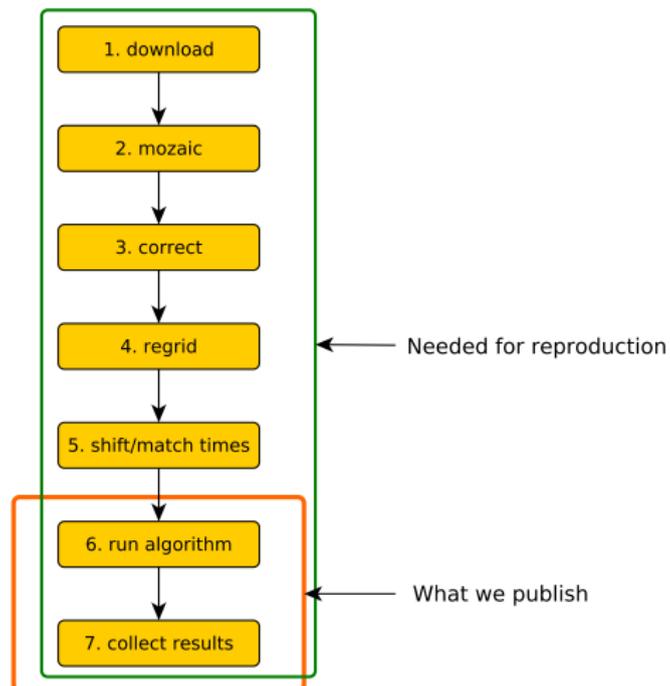
Current Earth Observation Research:



Current Earth Observation Research:



Current Earth Observation Research:



Science domains that do share workflows

- ▶ statistics, bioinformatics, geoinformatics (R, python)
- ▶ astrophysics
- ▶ high energy physics (WLCG: shared usage of 170 data centers)

Google Earth Engine (GEE): only feasible offering?

Yes, it allows you to

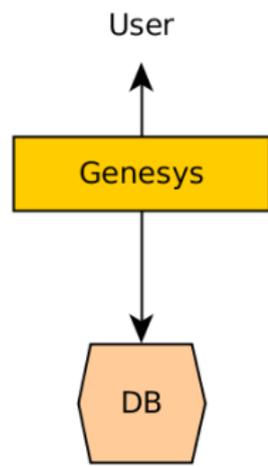
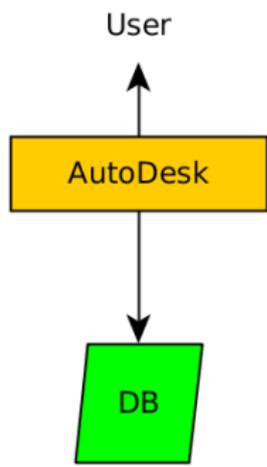
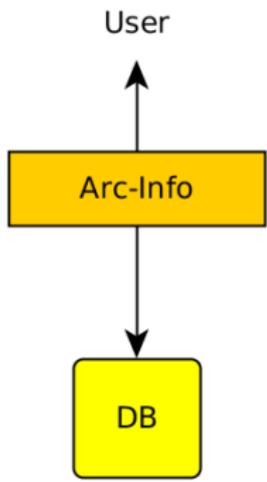
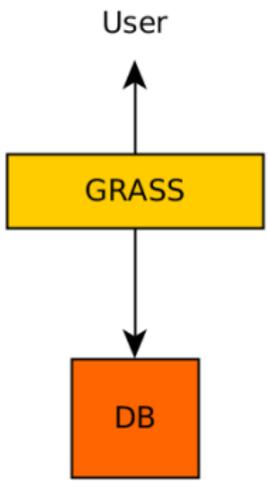
- ▶ process and combine practically all available sat imagery
- ▶ scale up to continents and large time periods
- ▶ on-the fly resampling/mosaic
- ▶ not worry about tiles, but work on *image collections*

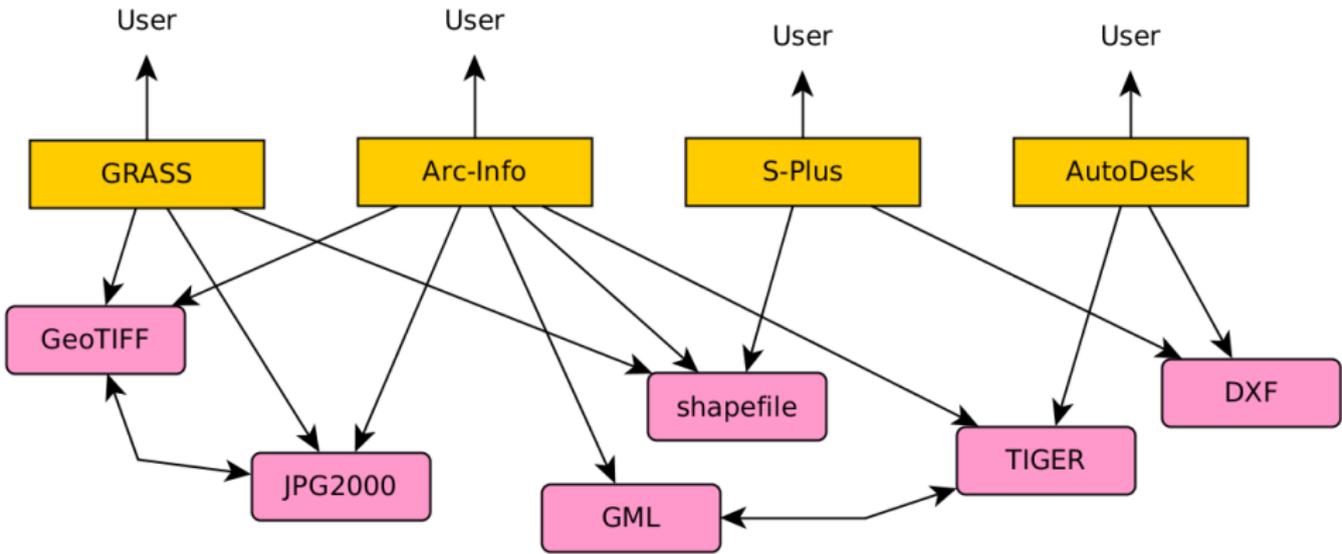
It also

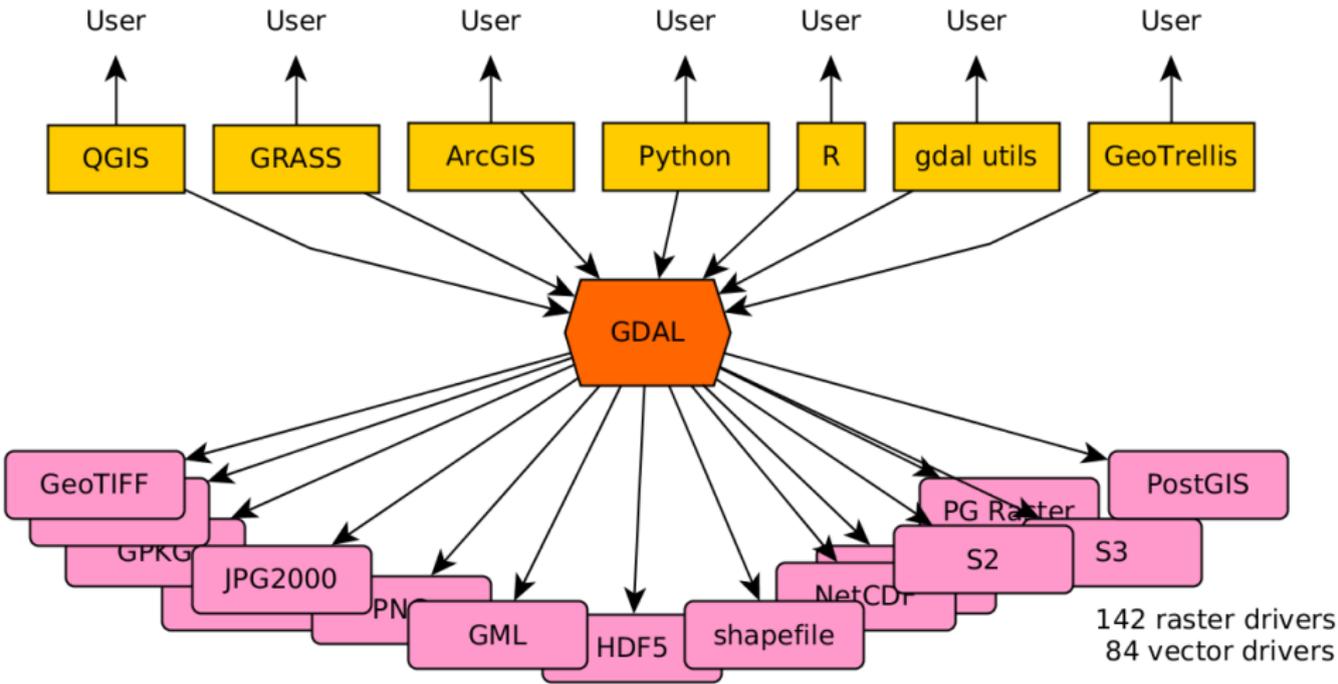
- ▶ has an EULA
- ▶ must have limited capacity
- ▶ has no SLA (or under development)
- ▶ is in a public cloud
- ▶ doesn't run arbitrary, user-defined functions
- ▶ is difficult to validate (and who is going to do this?)
- ▶ ... is not open source

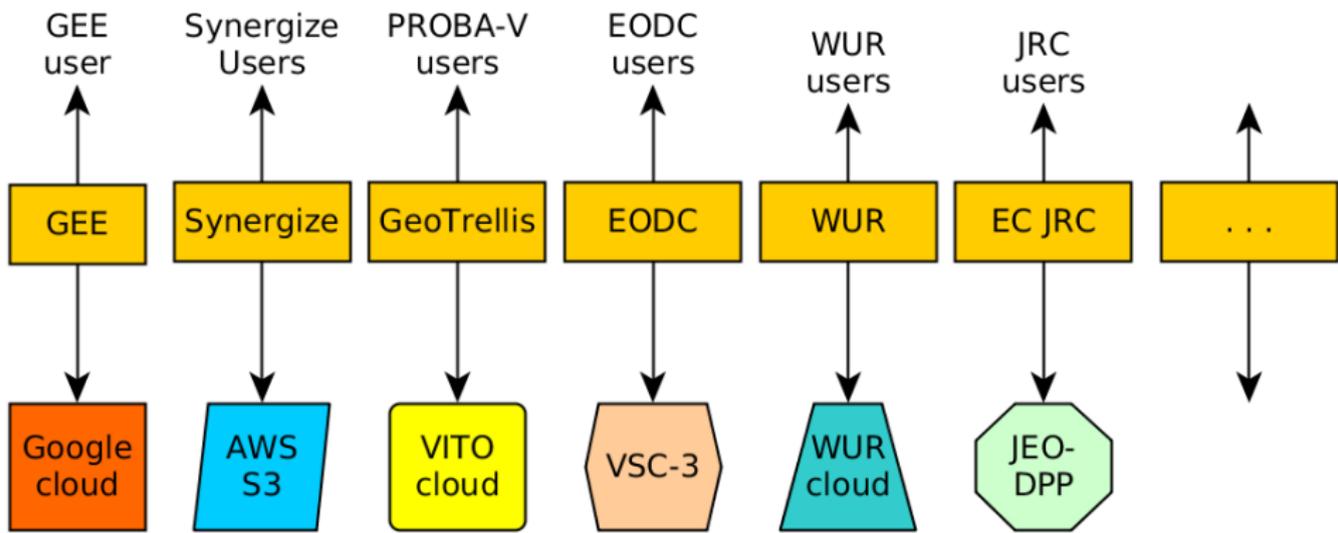
Other cloud platforms for satellite image processing

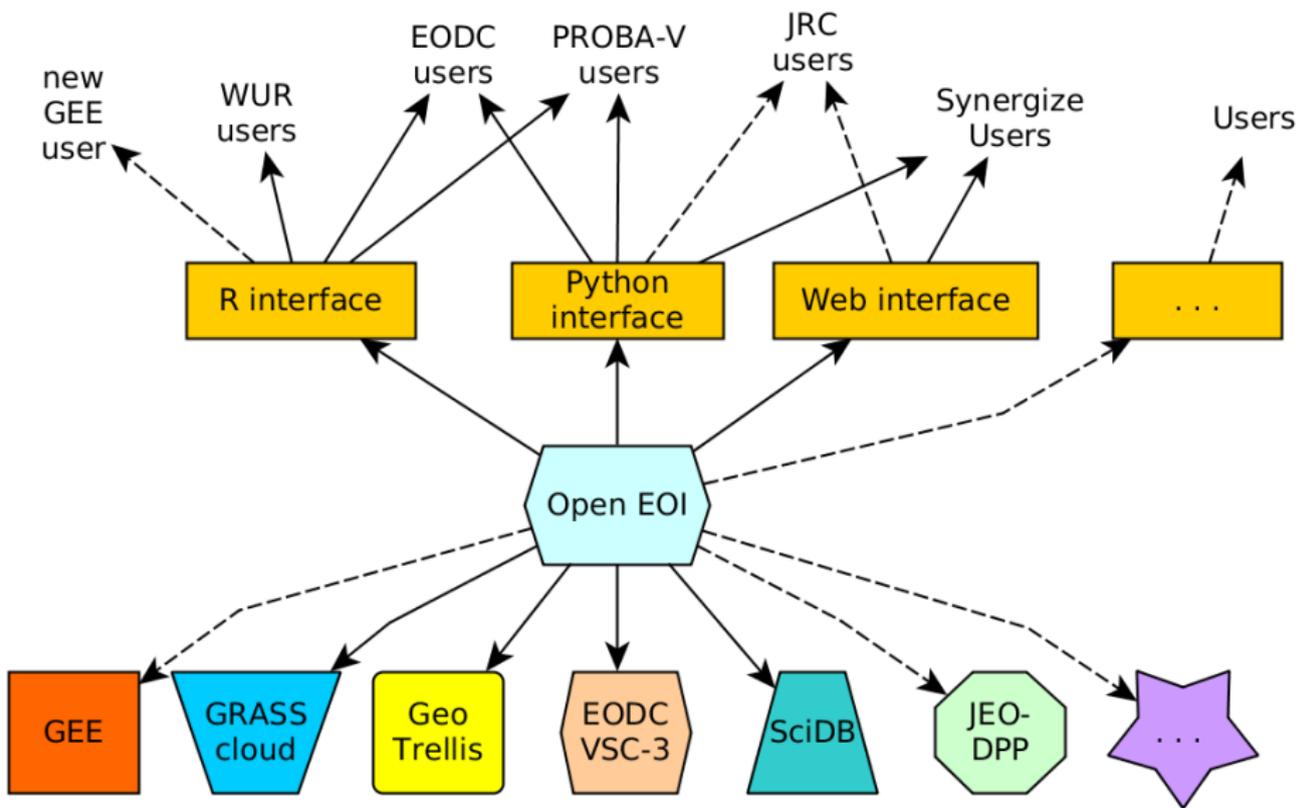
- ▶ ESA Data and information access service (DIAS; 5)
- ▶ Thematic Exploitation Platforms (TEPs)
- ▶ Meteorological and Environmental Earth Observation (MEEO), Brockmann
- ▶ ArcGIS online (?)
- ▶ JRC's EO-DPP

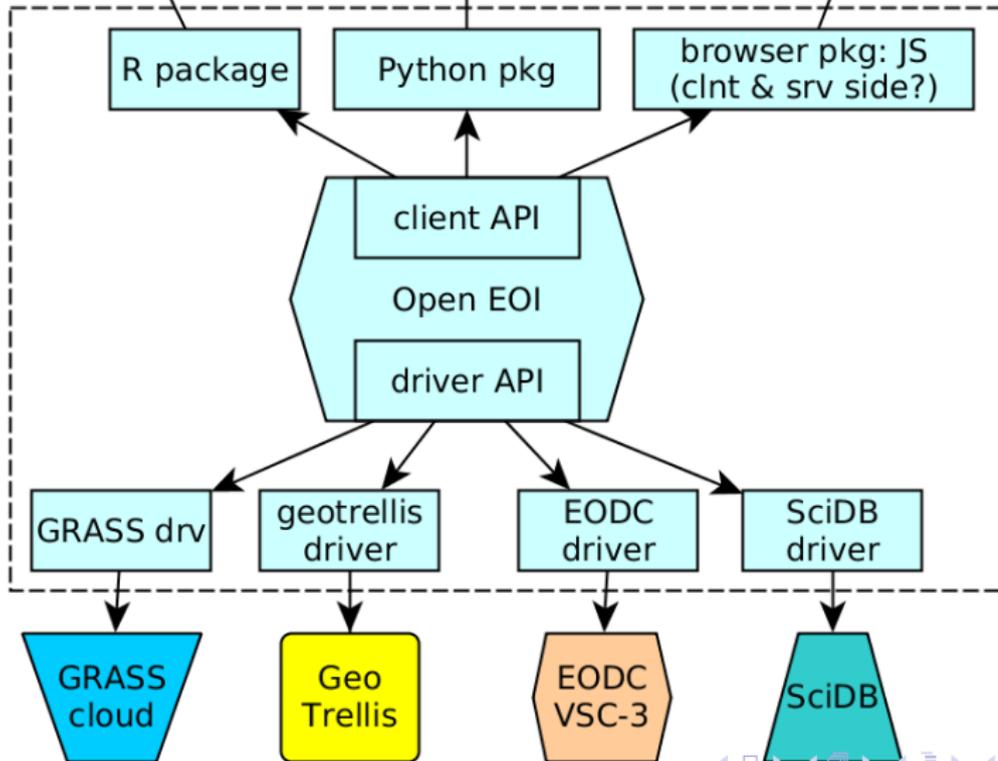




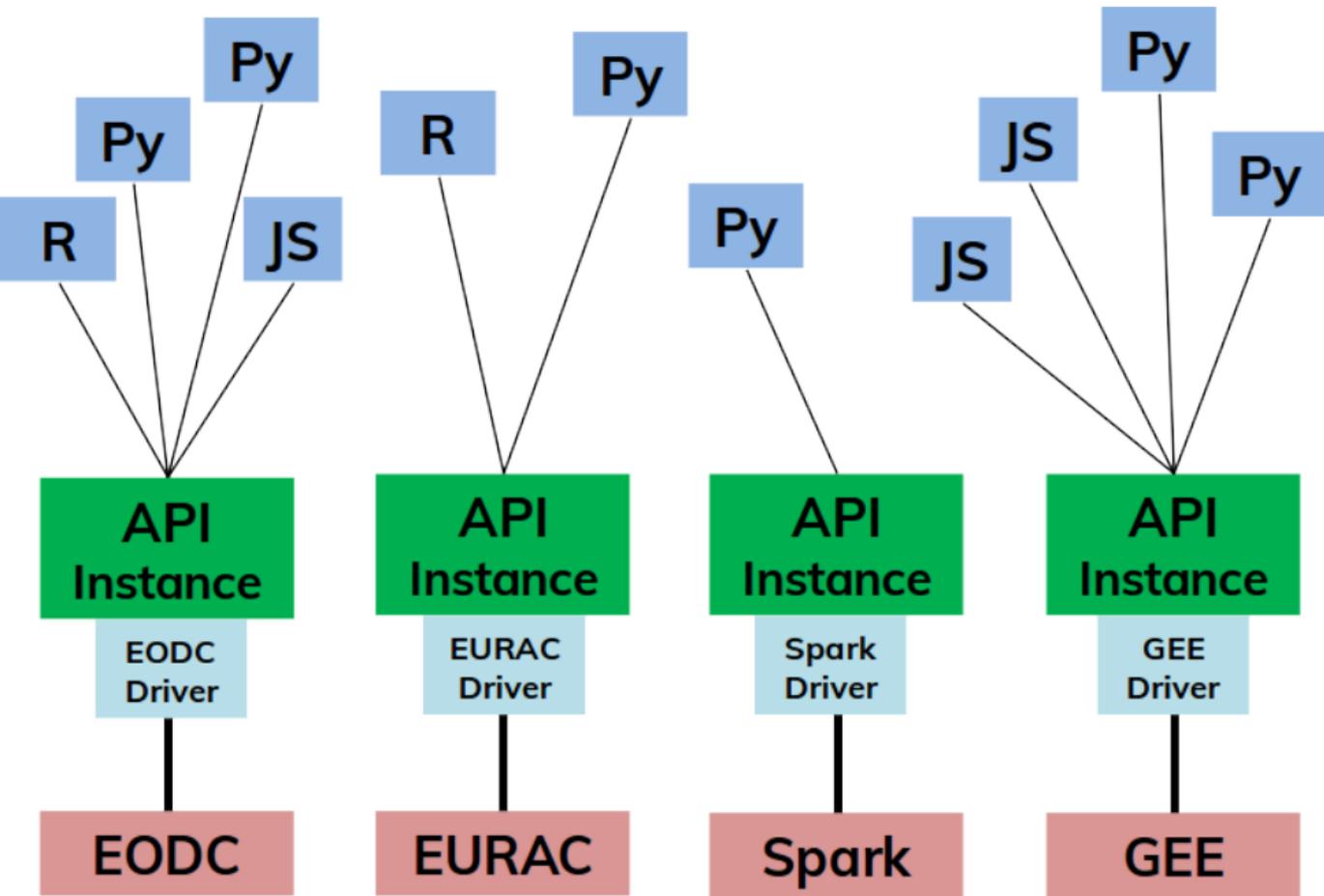








- ▶ H2020, Oct 2017-2020,
- ▶ <http://openeo.org/>
- ▶ openEO develops an open API to connect R, python and javascript clients to big Earth observation cloud back-ends in a simple and unified way.



Server: <http://127.0.0.1:8080>

Data: ASTERIAST_L1T_003

Processes: filter_bbox

Visualizations: None

Script:

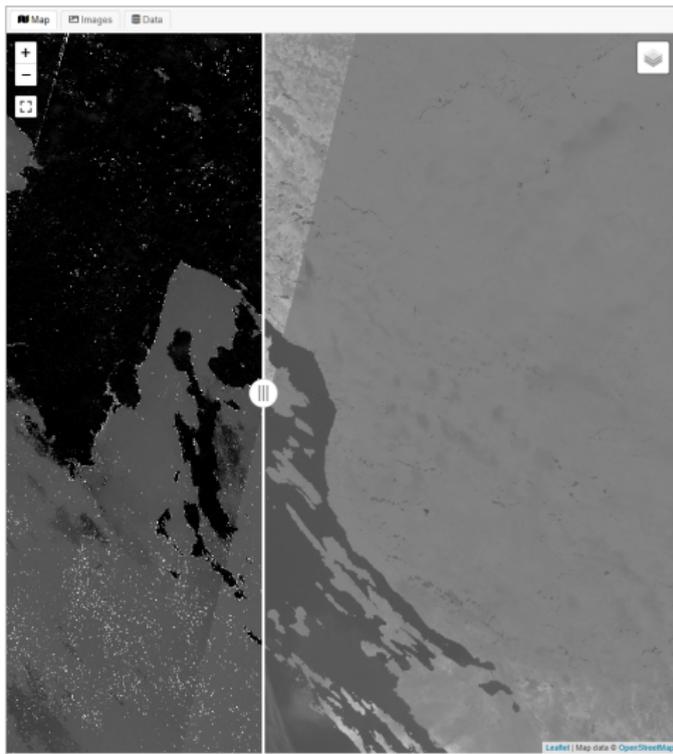
```

1 OpenEO.Editor.ProcessGraph = OpenEO.ImageCollection.create("COPERNICUS/S2")
2   .filter_daterange("2017-01-01", "2017-01-31")
3   .NDVI("B4", "B8")
4   .min_time()
5   .process("stretch_colors", {min: -1, max: 1}, "imagery");

```

Jobs Services Process Graphs Account

ID	Status	Submitted	Last update	Costs	Actions
53NG2BxxlaoervMm	canceled	2018-03-20 21:27:09	2018-03-20 21:27:09	0	
EvAAEyaVlcdHqx3G	canceled	2018-03-20 21:31:18	2018-03-20 21:31:18	0	
UrfZXzzKGdE9Go7	submitted	2018-03-20 21:39:39	2018-03-20 21:39:39	0	
XtaMa5UhnvrqEGZ	submitted	2018-03-21 15:28:25	2018-03-21 15:28:25	0	
rRrw1mikM6HJnQ9S	submitted	2018-03-28 09:25:04	2018-03-28 09:25:04	0	



```
1 { "process_id": "min_time",
2   "args": {
3     "imagery": {
4       "process_id": "/user/custom_ndvi",
5       "args": {
6         "imagery": {
7           "process_id": "filter_daterange",
8           "args": {
9             "imagery": {
10              "process_id": "filter_bbox",
11              "args": {
12                "imagery": {
13                  "product_id": "S2_L2A_T32TPS_20M"
14                },
15                "left": 652000,
16                "right": 672000,
17                "top": 5161000,
18                "bottom": 5181000,
19                "srs": "EPSG:32632"
20              }
21            },
22            "from": "2017-01-01",
23            "to": "2017-01-31"
24          }
25        },
26        "red": "B04",
27        "nir": "B8A"
28      }
29    }
30  }
31 }
```

Cube view

File-agnostic access to EO imagery through a **data cube view** boosts usability of EO data.

In openEO:

- ▶ spatial dimensions are complemented with other dimensions such as the **temporal or spectral dimensions**
- ▶ researchers can directly filter, aggregate, or map functions over dimensions of a **user-defined cube** without being concerned about how the data in the processing platform is organised (granules, collections, coverages, ...)
- ▶ **raster** and **vector data cubes** are integrated.

Proof of Concept

The Month 6 (April 2018) proof of concept involved:

- ▶ **coupling 3 clients** (Python, R, JavaScript web-editor: figure left) **to 7 back-ends** (Sentinel Hub, GRASS GIS, EODC OpenStack, WCPS, Python GeoPySpark / GeoTrellis, Google Earth Engine, R) **for**
- ▶ **3 use-cases** with band indexes, time series, aggregation over polygons, and user-defined (Python) functions
- ▶ **source code and API docs** on GitHub
- ▶ P.o.C. **demo videos** on the project web site

Why don't we build upon existing standards?

- ▶ which standards?
- ▶ many are too generic, and not expressive enough for this problem
- ▶ WCPS, for example, has only a limited set of hard baked-processes, and doesn't integrate vector operations
- ▶ WCS doesn't want to consider an image collection as a coverage (OGC WCS Interface Standard - Earth Observation Application Profile, version 1.0.0; 10-140r1. Open Geospatial Consortium, 2014)
- ▶ lots of issues are not addressed (user management, accounts, rights, where to put results etc)
- ▶ no standards exist for describing (discovering, processing, publishing) image collection / dataset series (but STAC is coming!).

Working software is more useful than unimplemented standards.

Why don't we build upon existing standards?

- ▶ which standards?
- ▶ many are too generic, and not expressive enough for this problem
- ▶ WCPS, for example, has only a limited set of hard baked-processes, and doesn't integrate vector operations
- ▶ WCS doesn't want to consider an image collection as a coverage (OGC WCS Interface Standard - Earth Observation Application Profile, version 1.0.0; 10-140r1. Open Geospatial Consortium, 2014)
- ▶ lots of issues are not addressed (user management, accounts, rights, where to put results etc)
- ▶ no standards exist for describing (discovering, processing, publishing) image collection / dataset series (but STAC is coming!).

Working software is more useful than unimplemented standards.

Upcoming challenges

- ▶ MANY!!
- ▶ A big one: UDFs (user-defined functions): how can I have my back-end execute my arbitrary (python, R) function on selected imagery?
- ▶ validating (verifying) back-ends against each other
- ▶ combining several back-ends
- ▶ User adoption: how/when will users start to adopt this (clients AND servers need to work, be useable, and be affordable!)