



JupyterLab for Earth Observation applications with HTCondor scaling and Voilà dashboarding

Earth Observation big data analytics and visualisation

davide.de-marchi@ec.europa.eu
armin.burger@ec.europa.eu
pierre.soille@ec.europa.eu

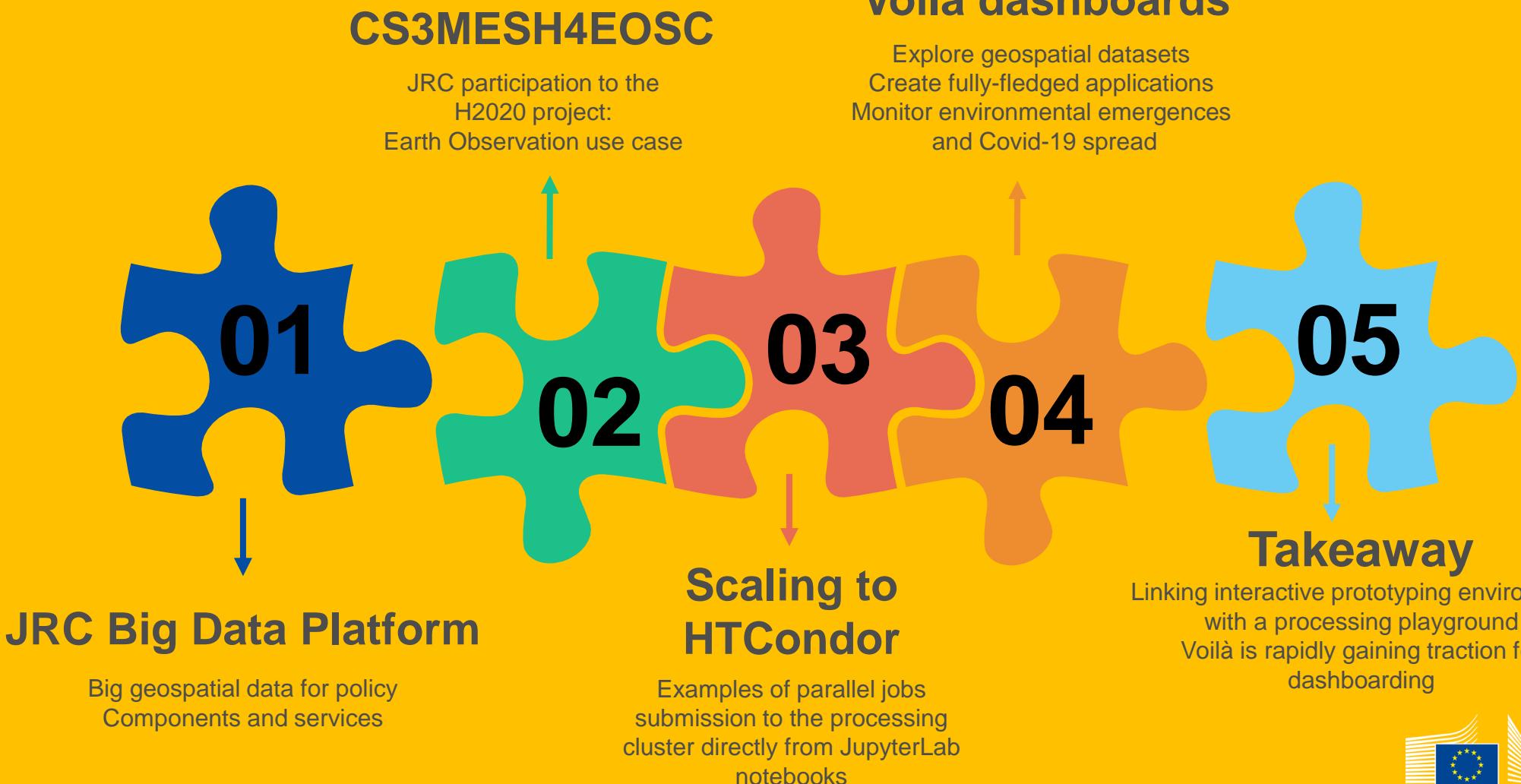
Davide De Marchi, Armin Burger and Pierre Soille

European Commission, Joint Research Centre (JRC)

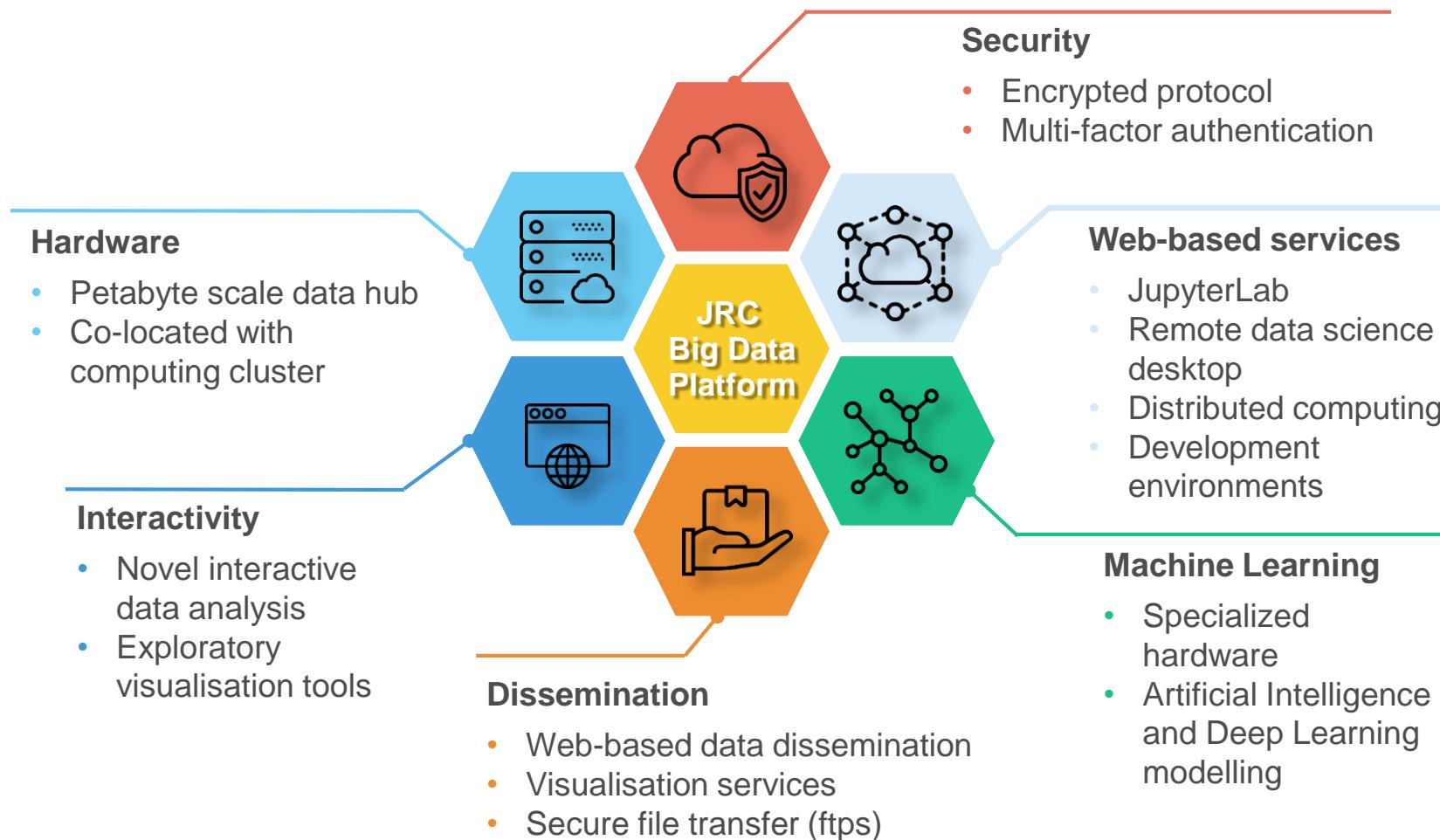
CS3 2021 - Workshop on Cloud Storage
Services for Synchronization and Sharing

26 January 2021

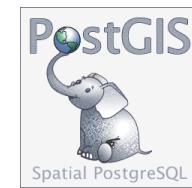
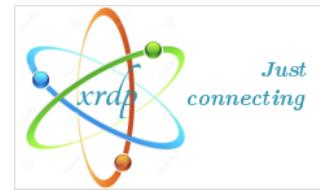
Outline



JRC Big Data Platform components



Main software stack



PostgreSQL



Facts and Figures



**~160
servers**

Installed in the
**JRC Data
Center**

To manage the
storage JBODS,
the processing jobs
and all the services



**~2,500
cores**

For JEO-
batch/desk/lab
and other
services

12-19 GBs of RAM
per core



**9 GPU
servers**

28 Nvidia GPU's
in total

Accessible from
JEO-lab Jupyter
notebooks for
machine learning
and deep learning
applications

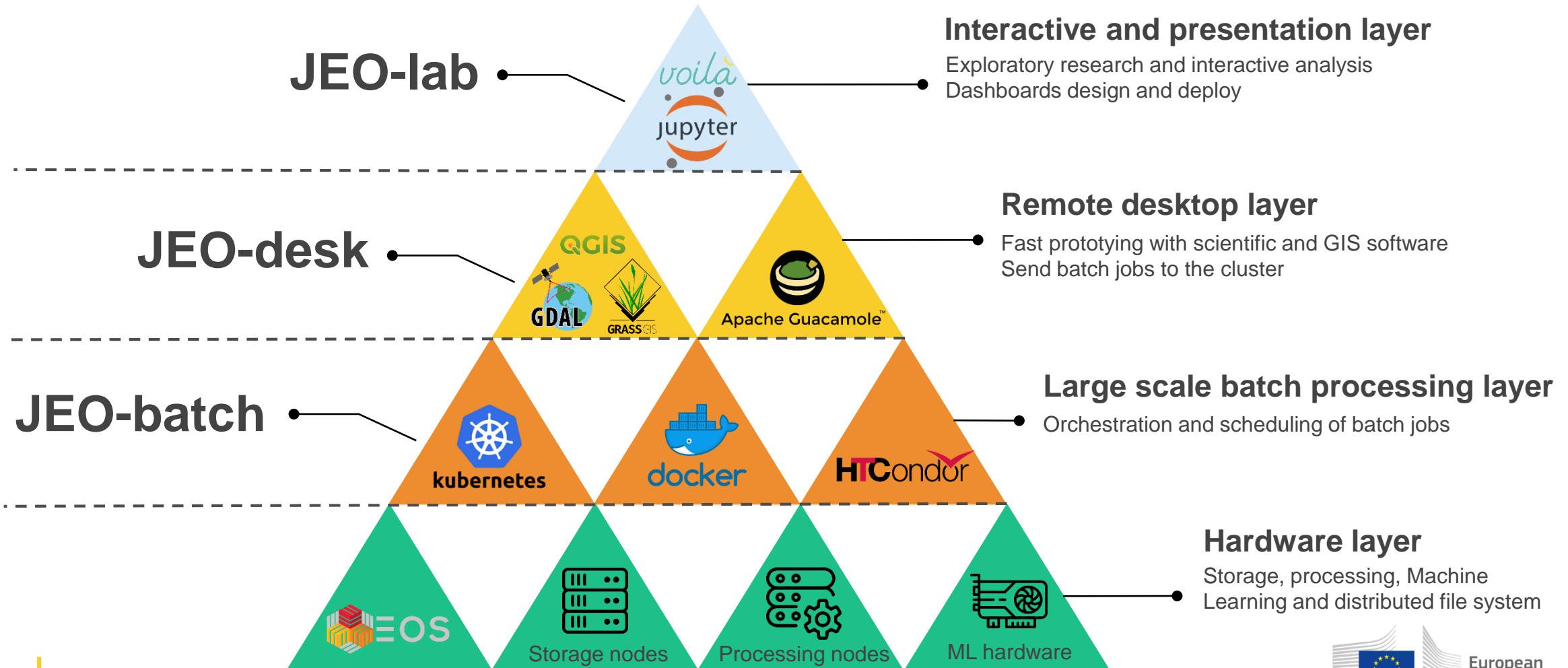


**21.4 PiB
gross
capacity**

10.7 PiB net
capacity (using
replica 2)

Currently used: 7.2
PiB net

JRC Big Data Platform main services



Participation to the CS3MESH4EOSC project



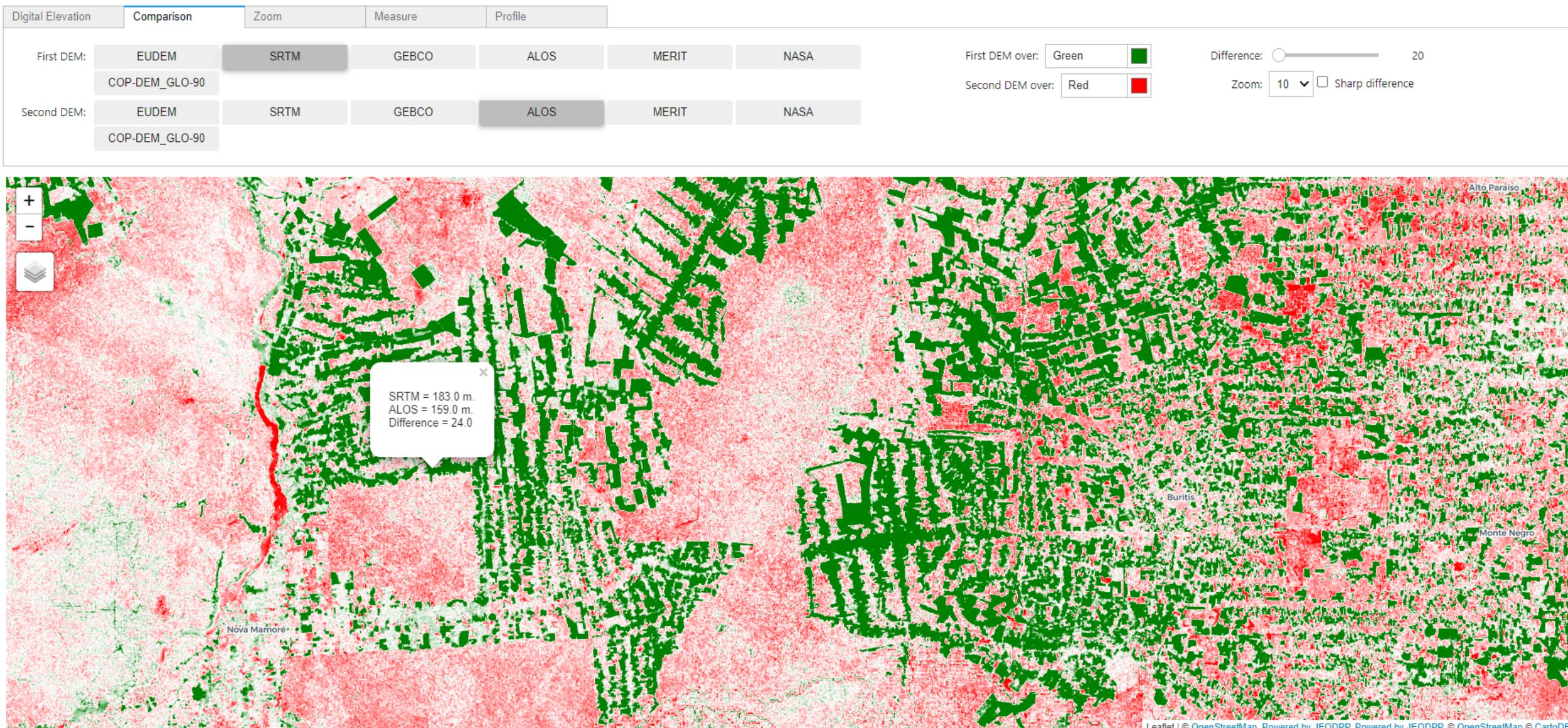
- Earth Observation use case in task 4.1 Data Science Environments
- Scaling of processing chains to the HPC cluster
- Voilà dashboarding
- Integration of CERN Swan and Kubernetes
- Assessing possible future federation using ScienceMesh IOP

Scaling to HTCondor



- Python bindings to submit jobs to HPC cluster directly from JupyterLab
- Analysis workflow linking multiple data sources to generate new insights
- Custom python function transforming input n-D NumPy array into desired output
- On-the-fly definition of resources (CPU and RAM)
- GUI for monitoring job execution based on ipywidgets

Example: calculate % of deforestation in areas of Amazon



Example: calculate % of deforestation in areas of Amazon

The screenshot shows a Jupyter Notebook interface with the following details:

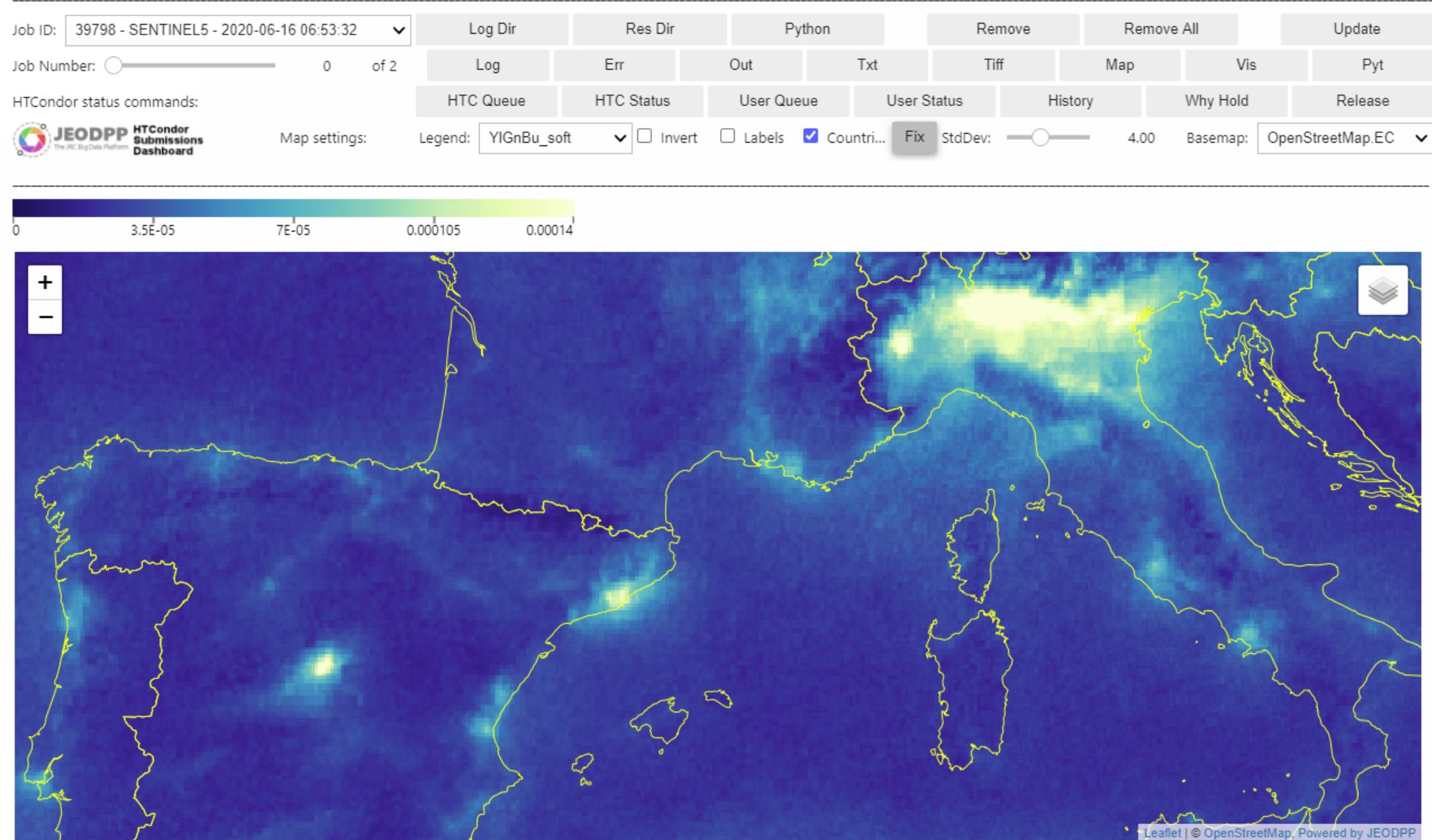
- Toolbar:** Includes icons for file operations (New, Open, Save, etc.), a search bar, and tabs for "Launcher" and "Deforestation.ipynb".
- Code Cell [100]:** Displays the command `p.scaleGraph()`. The output shows a hierarchical structure of three jobs under a cluster:
 - Cluster:** Contains three **Job#0**, **Job#1**, and **Job#2**.
 - Job#0:** Contains three images (`img[0]`, `img[1]`, `img[2]`) with their respective coordinates and file names.
 - Job#1:** Contains three images (`img[0]`, `img[1]`, `img[2]`) with their respective coordinates and file names.
 - Job#2:** Contains three images (`img[0]`, `img[1]`, `img[2]`) with their respective coordinates and file names.
- Text Cell [100]:** Displays the output `Out[100]:`
- Code Cell [101]:** Displays the definition of the `Deforestation` function:

```
def Deforestation(img):
    polygon, dem1, dem2 = img[0], img[1], img[2] # Access the numpy arrays for polygon, dem1, dem2

    difference = dem2 - dem1 # Difference in height between the two DEMs
    deforested = np.logical_and(polygon>0, difference>10) # Deforested if inside the polygon and difference is greater than 10 meters
    numdeforest = np.count_nonzero(deforested) # Count number of deforested pixels
    numtotal = np.count_nonzero(polygon) # Count total number of pixels inside the polygon
    if numtotal > 0:
        print("Percentage deforestation = %.2f%%" % (100.0 * numdeforest / numtotal))

    return deforested.astype(np.uint8) # Return the array of deforested pixels converted in Byte
```
- Text Cell [102]:** Displays the command `p.scaleSubmit("DEFORESTATION", Deforestation, 3, '6GB')` and its output `Out[102]: 42729`.
- Text Cell [103]:** Displays the command `scaleDashboard()`.

Example: effects of Covid19 lockdown measures on air quality

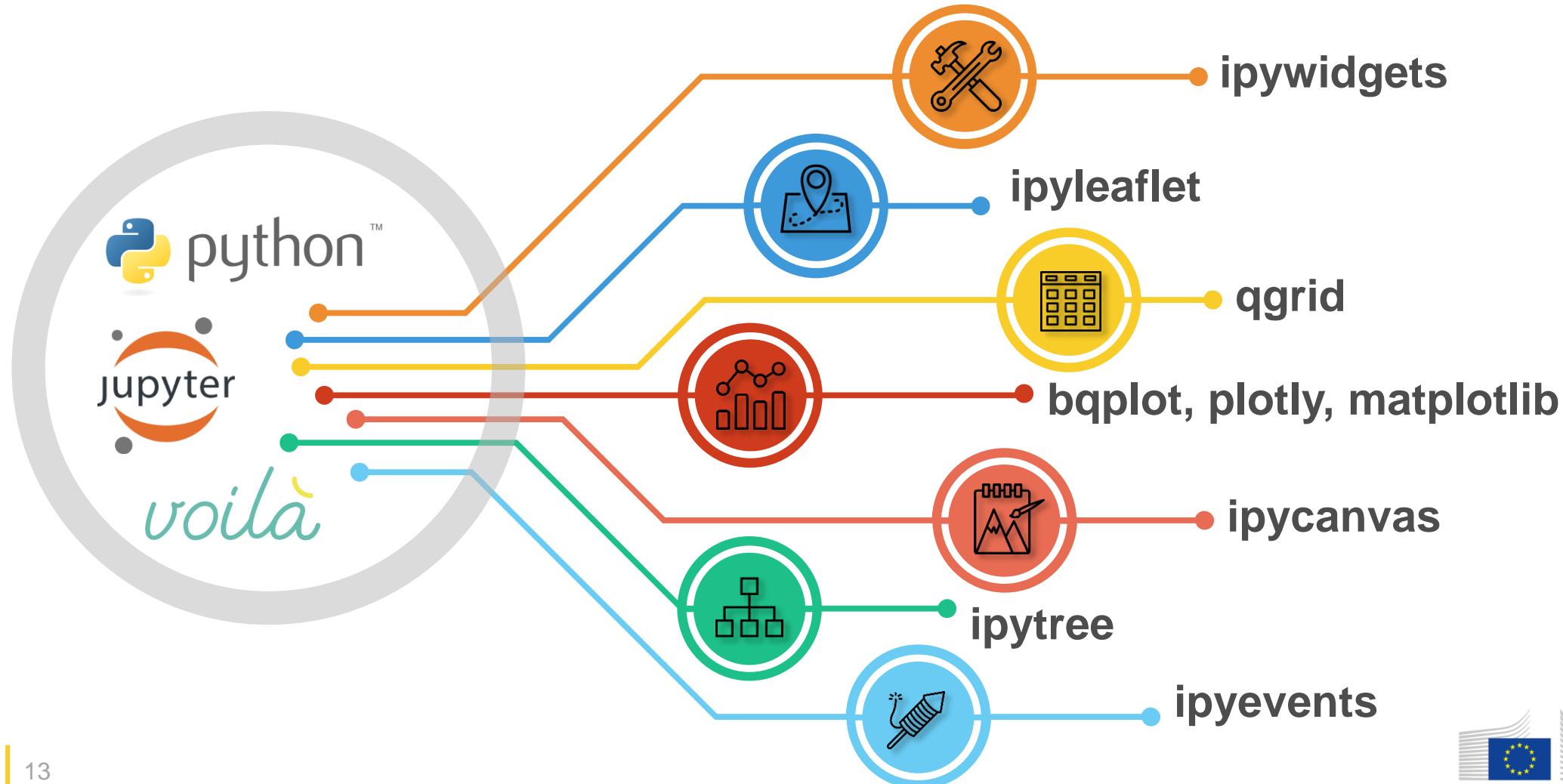


Voilà dashboarding



- Voilà is a Jupyter notebook extension to automatically create standalone applications and dashboards.
- Notebooks are rendered by showing only the output of the cells, while the code is hidden.
- **Suitable for non-technical experts** for communicating insights and foresight to a **wider audience**.
- Single environment for full data analytics workflows from research and innovation to outreach engaging policy makers and citizens.

Widgets and libraries used to create Voilà dashboards



Voilà dashboard: Collections Explorer, disaster monitoring

Browse geospatial datasets, make comparison using the “split window”, assess damages of natural events

→

»

Search Display Stretch Zoom Measure Vector Extract Clouds

Stretch images on current displayed area

Red: 0 - 2500

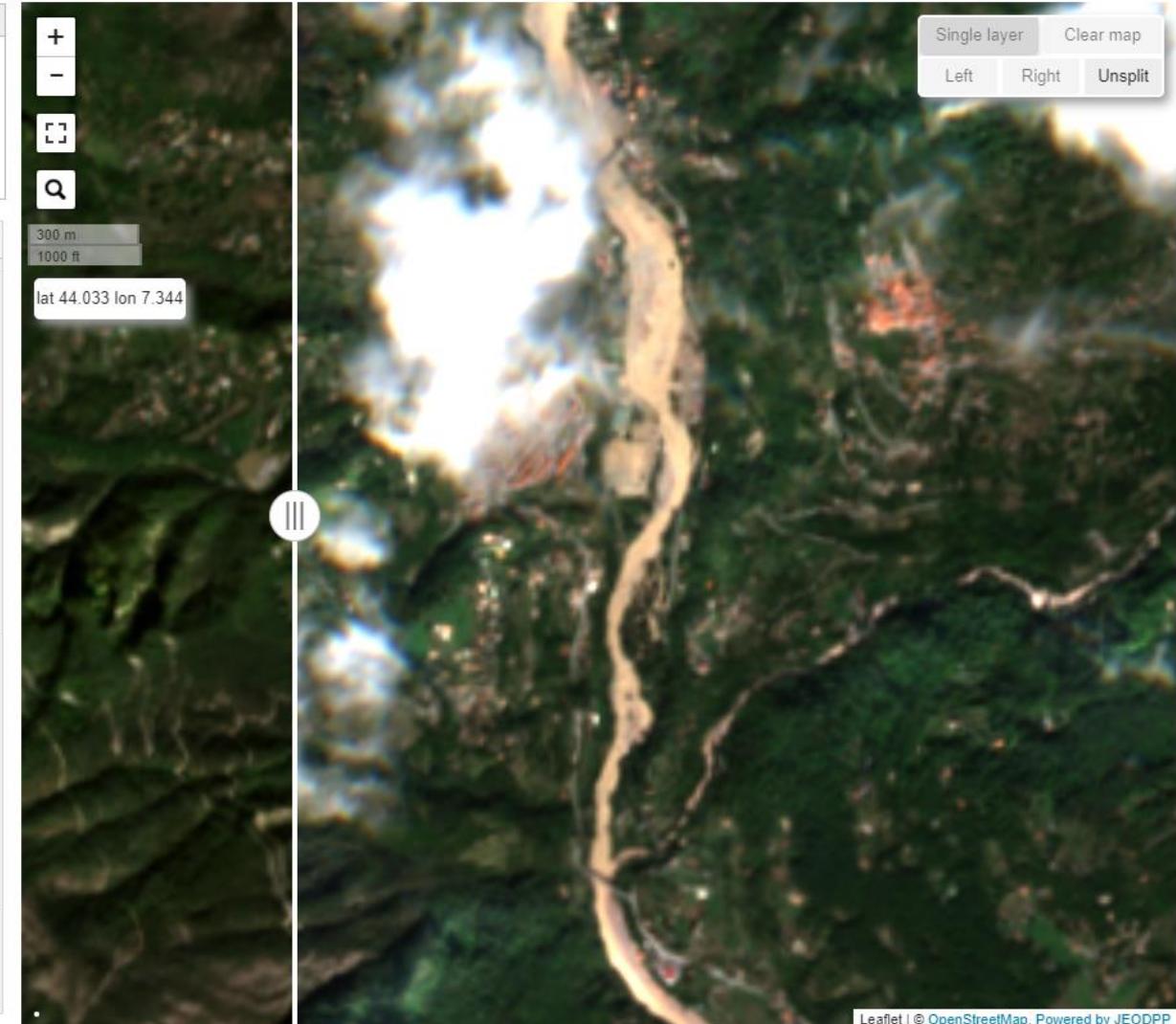
Std. devs. 2.00

Green: 0 - 2500

Blue: 0 - 2400

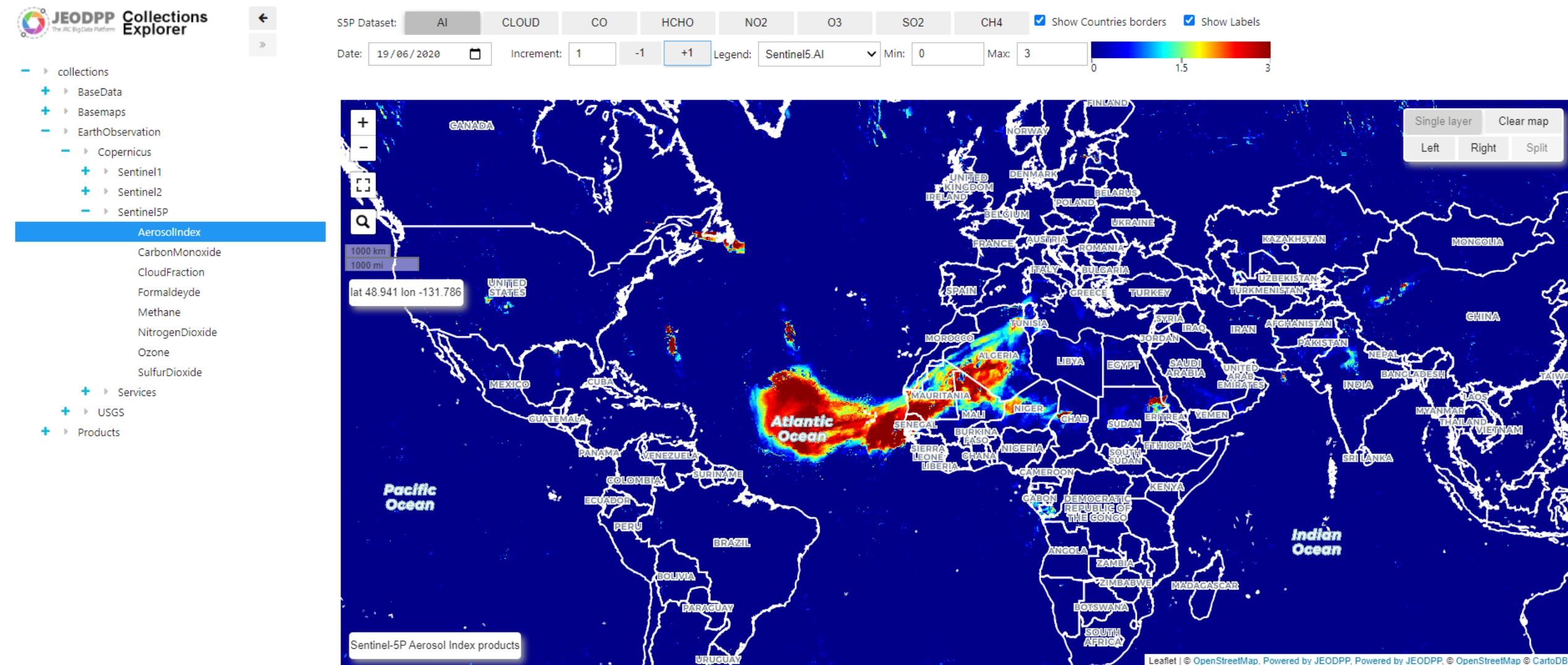
Remove fixed display parameters

	Date	Orbit	Type	Sat	Full	SizeMB	Cloud	UTM	MGRS
26	2020-10-16	108	L2A	B	✓	983	45	32N	32TLP
27	2020-10-13	65	L2A	B	✓	954	5	32N	32TLP
28	2020-10-08	65	L2A	A	✓	944	15	32N	32TLP
29	2020-10-03	65	L2A	B	✓	981	10	32N	32TLP
30	2020-10-01	108	L2A	A	✓	1068	32	32N	32TLP
31	2020-09-28	65	L2A	A	✓	921	1	32N	32TLP
32	2020-09-26	108	L2A	B	✓	1076	1	32N	32TLP
33	2020-09-23	65	L2A	B	✓	948	51	32N	32TLP
34	2020-09-21	108	L2A	A	✓	960	49	32N	32TLP
35	2020-09-18	65	L2A	A	✓	965	15	32N	32TLP
36	2020-09-16	108	L2A	B	✓	1060	2	32N	32TLP
37	2020-09-13	65	L2A	B	✓	957	0	32N	32TLP
38	2020-09-08	65	L2A	A	✓	974	4	32N	32TLP
39	2020-09-06	108	L2A	B	✓	1088	11	32N	32TLP
40	2020-09-03	65	L2A	B	✓	942	11	32N	32TLP
41	2020-09-01	108	L2A	A	✓	1054	24	32N	32TLP
42	2020-08-27	108	L2A	B	✓	1079	5	32N	32TLP
43	2020-08-24	65	L2A	B	✓	947	34	32N	32TLP
44	2020-08-22	108	L2A	A	✓	1098	1	32N	32TLP
45	2020-08-19	65	L2A	A	✓	977	28	32N	32TLP
46	2020-08-17	108	L2A	B	✓	1088	23	32N	32TLP



Voilà dashboard: monitor air quality with Sentinel5-P

Godzilla dust cloud from Sahara covers the Caribbean in once-in-50-years event



Voilà dashboard: custom SVG drawing for Gender Equality

Gender Equality Index: measure the progress of gender equality in the EU, developed by EIGE

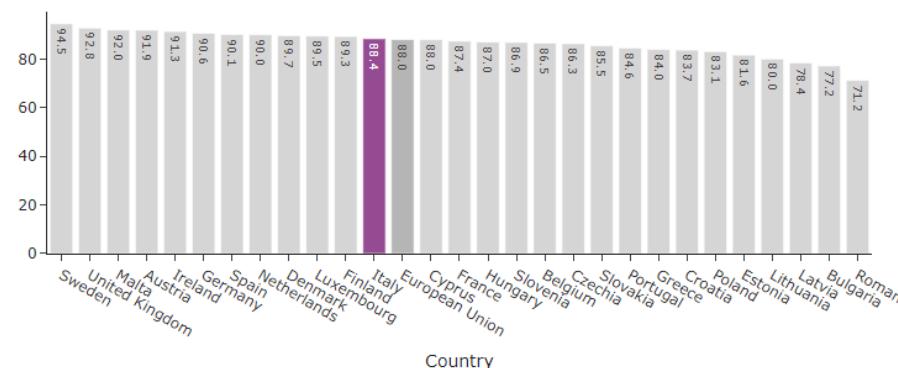
Source data: <https://eige.europa.eu/gender-equality-index/2020> - European Institute for Gender Equality



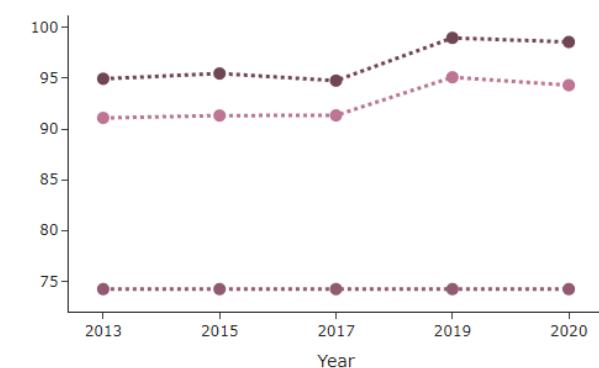
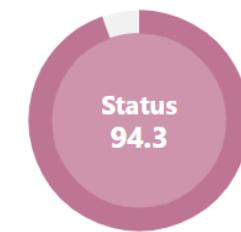
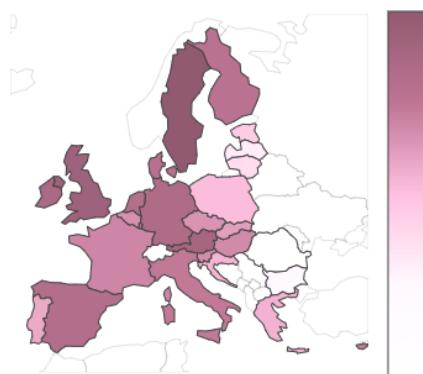
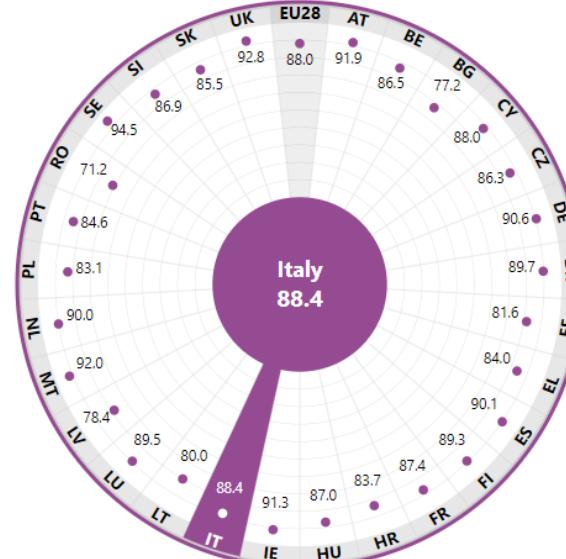
Select an index/domain/subdomain from the tree:

- + 2013
- + 2015
- + 2017
- + 2019
- 2020
- Gender Equality Index
 - + Work
 - + Money
 - + Knowledge
 - + Time
 - + Power
- Health
 - + Status
 - + Behaviour
 - + Access

Health for Italy in 2020



Select a country by clicking on the wheel:



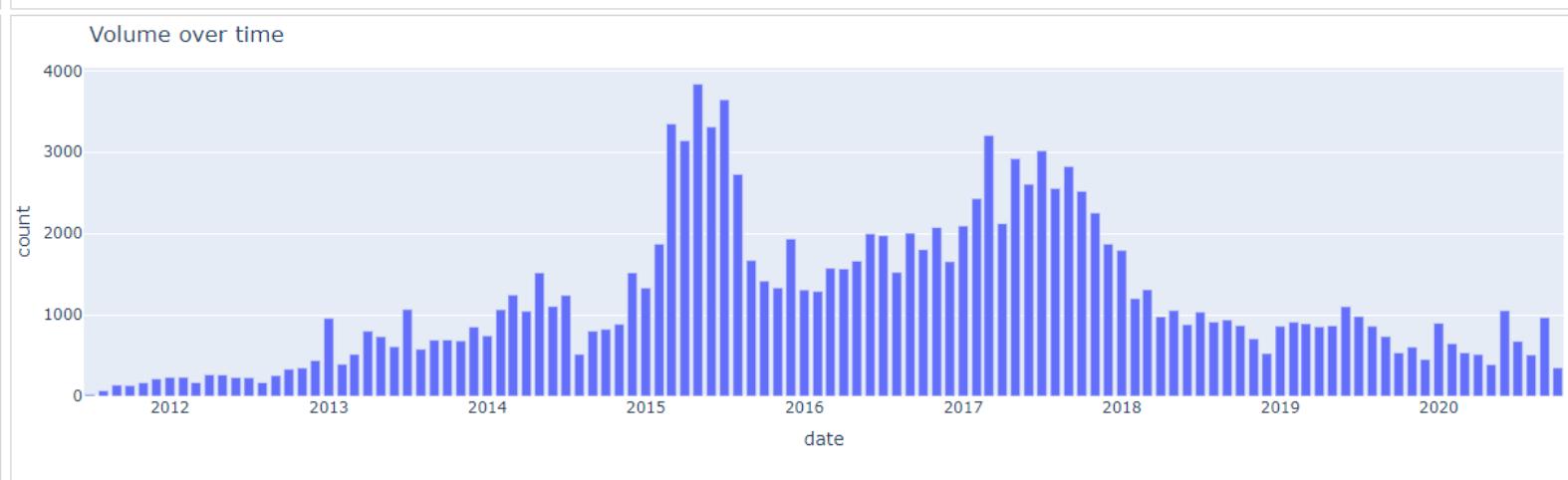
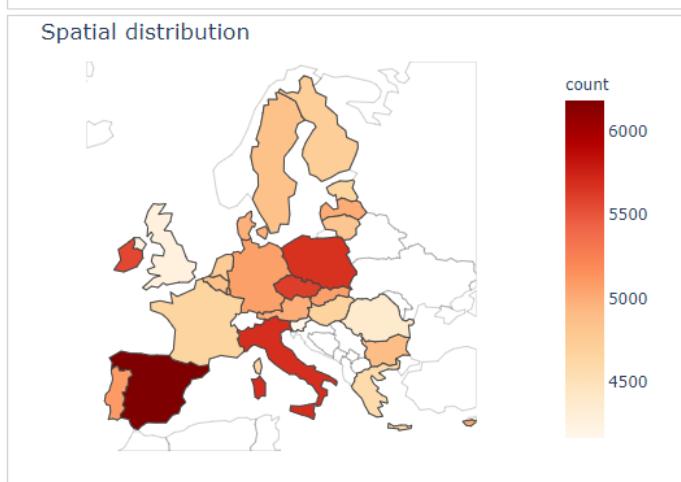
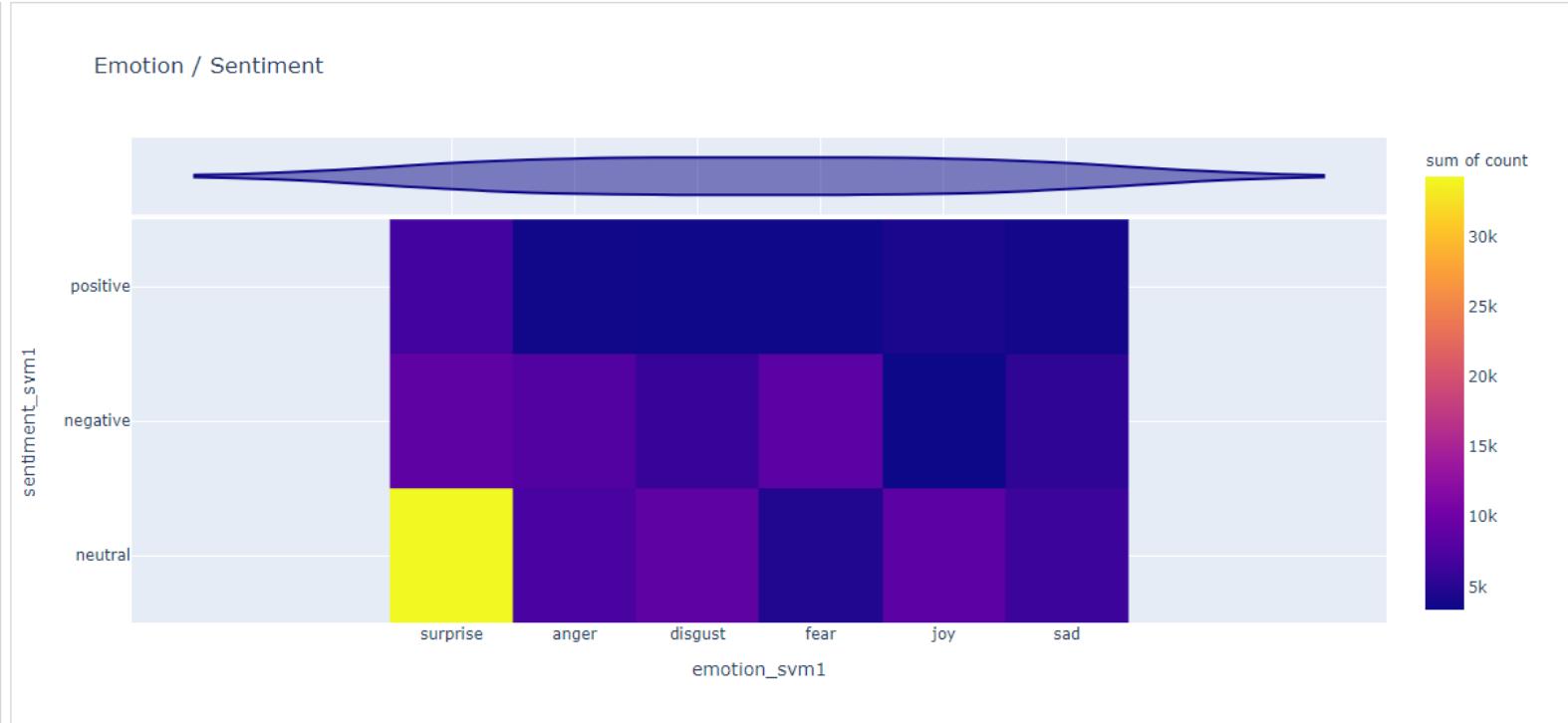
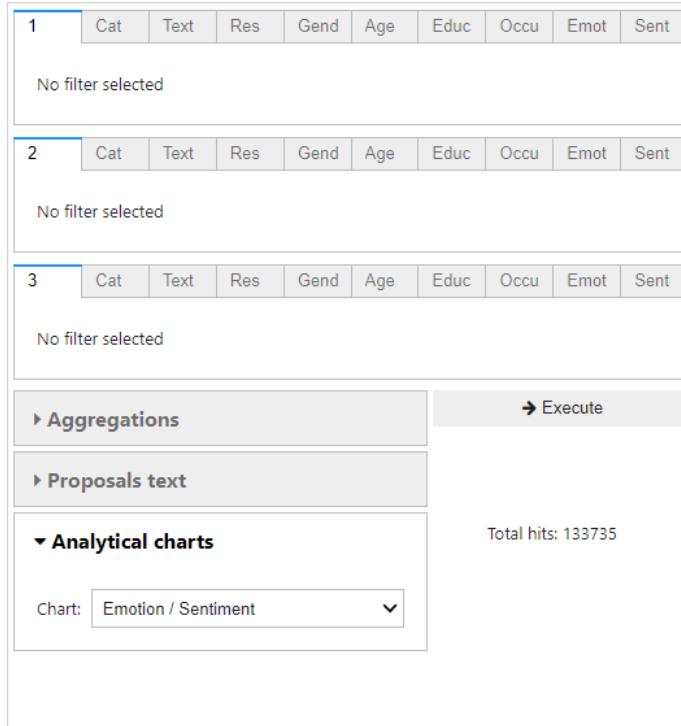
Voilà dashboard: custom SVG drawing for Covid19 monitoring

COVID-19 in geographic Europe: evolution in time of cases, deaths and recovered

Source data: <https://github.com/CSSEGISandData/COVID-19 - Johns Hopkins University - Eurostat>



Voilà dashboard: plotly charts for citizen science



Takeaway

- JRC is providing CS3MESH4EOSC project with a solid use case in EO where the federation of services can be tested on real scientific applications
- Linking interactive and batch processing (JupyterLab $\leftarrow\rightarrow$ HTCondor) will provide a prototyping environment coupled with a processing playground
- Voilà is rapidly gaining traction for dashboarding thanks to:
 - Flexibility with fast growing charting and visualisation libraries
 - Option for custom visualisations/interaction in SVG/ipyevents if needed
 - Scaling with full dockerisation
 - Suitable for big data analytics and lazy computations

Keep in touch



JRC: ec.europa.eu/jrc

Big Data Platform: jeodpp.jrc.ec.europa.eu



JRC: @EU_ScienceHub

Personal: @demarchidavide



EU Science Hub – Joint Research Centre



EU Science, Research and Innovation



Eu Science Hub

Thank you



© European Union 2021

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](#) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.