



Contribution ID: 75

Type: **Presentation**

From Jupyter notebooks to web dashboards for big geospatial data analysis

Wednesday, January 29, 2020 3:20 PM (20 minutes)

From Jupyter notebooks to web dashboards for big geospatial data analysis

The Joint Research Centre (JRC) of the European Commission has set up the JRC Big Data Platform (JEODPP) as a petabyte scale infrastructure to enable EC researchers to process and analyse big geospatial data in support to EU policy needs[1]. One of the service layer of the platform is the JEO-lab environment[2] that is based on Jupyter notebooks and the Python programming language to enable exploratory visualization and interactive analysis of big geospatial datasets. JEO-lab is set-up with deferred processing, using multiple service nodes to execute the Jupyter client processing workflow starting from data stored in the CERN EOS distributed file system deployed on the JEODPP. In this context, many new applications and services were recently added in order to expand the platform attractiveness towards data scientists and researchers. The presentation will make a tour of the many new features added to the JEO-lab, providing use cases and demos that will include topics like:

- **Sentinel2explorer: an advanced remote sensing application that fully exploits the Jupyter widgets**

It allows users to browse, search and display the full set of Copernicus Sentinel-2 images stored in the JEODPP platform. Selecting any band combination, calculating vegetation or water indexes, creating videos, animations or other types of exports, drawing vector features on top of the displayed images, extracting the full story of the images covering a polygonal feature are among the many functions available, which were created by using at their maximum extent the ipywidgets[3] collection of standard GUI elements as well as some other Jupyter widgets[4]. The outcome is an application that helps end-users to easily navigate inside the many petabytes of Sentinel-2 images available in the JEODPP platform.

- **From interactive to distributed computing of land parcel signatures using HTcondor**

A demonstration of an integrated solution, which comprises interactive and heavily parallel batch processing, to support the new CAP (Common Agricultural Policy) in the monitoring of agricultural parcels at regional or national scale. Using HTcondor orchestrator, a batch extraction of yearly vegetation profiles over millions of polygons is launched and the results are visualized and assessed in the JEO-lab interactive environment. The users can easily view the full story of any by accessing a single, indexed, multi GBytes binary file containing all the results of the batch extraction.

- **ML classification inside a Jupyter notebook using server-side injection of custom Python code**

An example of interactive training for a Symbolic Machine Learning (SML)[5] algorithm inside a Jupyter notebook, exploiting the capability of JEO-lab to execute any custom python code inside the server-side processing chain, via the on-the-fly creation of a Python interpreter inside the server C++ tile engine. Users can profit from the pyjeo[6] EO Python library to execute complex tasks, like image classification or segmentation, thus greatly expanding the analytic capabilities of JEO-lab.

- **Dynamic API to browse and display the full catalogue of Sentinel-2 data in geo-spatial web portals**

The Biodiversity and Protected Areas Management (BIOPAMA) Programme[7] assists the African, Caribbean and Pacific countries to address their priorities for improved management and governance of biodiversity and natural resources. BIOPAMA provides a variety of tools, services and funding to conservation actors in the African, Caribbean and Pacific (ACP) countries. Inside its web portal, a new service provided by JEODPP is implemented: web maps can now show the full JEODPP Sentinel-2 catalogue by using a dedicated REST API that provides discovery, query and fast display capabilities. Inside a Mapbox client, the JEO-lab tile engine dynamically serves TMS layers.

• **Porting notebooks and applications to Voilà to grant access without authentication**

Voilà[8] turns Jupyter notebooks into standalone web-dashboard applications; it supports Jupyter interactive widgets, while not permitting arbitrary code execution, thus posing less security threats. Many applications developed inside the JEO-lab environment are going to be brought into the Voilà world, where they will be accessible without the need for user authentication, and thus greatly expanding the impact of the JEODPP platform and providing an easy way to publish complex interactive visualization environments.

JEODPP platform is a living demonstration of a complex ecosystem of cloud applications and services that allows data scientists' navigation inside a petabyte scale world. In particular, the exploratory visualization and interactive analysis tools in the JEO-lab component can run custom code to prototype the generation of scientific evidence as well as create GUI applications that can be used by end-users ranging from policy makers to citizens.

[1] P. Soille, A. Burger, D. De Marchi, P. Kempeneers, D. Rodriguez, V.Syrris, and V. Vasilev. "A Versatile Data-Intensive Computing Platform for Information Retrieval from Big Geospatial Data". *Future Generation Computer Systems* 81.4 (Apr. 2018), pp. 30-40. <https://doi.org/10.1016/j.future.2017.11.007>.

[2] D. De Marchi, A. Burger, P. Kempeneers, and P. Soille. "Interactive visualisation and analysis of geospatial data with Jupyter". In: *Proc. of the BiDS'17*. 2017, pp. 71-74. <https://zenodo.org/record/3248741#.XeDvSuhKg2w>.

[3] <https://ipywidgets.readthedocs.io/en/latest/>

[4] <https://github.com/quantopian/qgrid>

[5] M. Pesaresi, V. Syrris and A. Julea. "A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning". *Remote Sens.* 2016, 8(5), 399; <https://doi.org/10.3390/rs8050399>

[6] P. Kempeneers, O. Pesek, D. De Marchi, P. Soille. "pyjeo: A Python Package for the Analysis of Geospatial Data" *ISPRS International Journal of Geo-Information*, Volume 8, Issue 10, October 2019. <https://doi.org/10.3390/ijgi8100461>

[7] <https://www.biopama.org/>

[8] <https://blog.jupyter.org/and-voil%C3%A0-f6a2c08a4a93>, <https://github.com/voila-dashboards/voila>

Primary authors: Mr BURGER, Armin (European Commission - Joint Research Centre); Mr DE MARCHI, Davide (European Commission - Joint Research Centre); Mr SOILLE, Pierre (European Commission - Joint Research Centre)

Presenter: Mr DE MARCHI, Davide (European Commission - Joint Research Centre)

Session Classification: Fabric and platforms for Global Science

Track Classification: User Voice: Novel Applications, Data Science Environments & Open Data