Mission Exploitation Platform PROBA-V

Erwin Goor

Belgium (erwin.goor@vito.be)

VITO and partners developed an end-to-end solution to drastically improve the exploitation of the PROBA-V EO-data archive (http://proba-v.vgt.vito.be/), the past mission SPOT-VEGETATION and derived vegetation parameters by researchers, service providers and end-users. The analysis of time series of data (+1PB) is addressed, as well as the large scale on-demand processing of near real-time data.

From November 2015 an operational Mission Exploitation Platform (MEP) PROBA-V, as an ESA pathfinder project, will be gradually deployed at the VITO data center with direct access to the complete data archive. Several applications will be released to the users, e.g.

- A time series viewer, showing the evolution of PROBA-V bands and derived vegetation parameters for any area of interest.

- Full-resolution viewing services for the complete data archive.

- On-demand processing chains e.g. for the calculation of N-daily composites.

- A Virtual Machine will be provided with access to the data archive and tools to work with this data, e.g. various toolboxes and support for R and Python.

After an initial release in January 2016, a research platform will gradually be deployed allowing users to design, debug and test applications on the platform. From the MEP PROBA-V, access to Sentinel-2 and landsat data will be addressed as well, e.g. to support the Cal/Val activities of the users.

Users can make use of powerful Web based tools and can self-manage virtual machines to perform their work on the infrastructure at VITO with access to the complete data archive. To realise this, private cloud technology (openStack) is used and a distributed processing environment is built based on Hadoop. The Hadoop ecosystem offers a lot of technologies (Spark, Yarn, Accumulo, etc.) which we integrate with several open-source components.

The impact of this MEP on the user community will be high and will completely change the way of working with the data and hence open the large time series to a larger community of users. The presentation will address these benefits for the users and discuss on the technical challenges in implementing this MEP.