

## Mission Exploitation Platform PROBA-V

Jeroen Dries, Erwin Goor, Dirk Daems

TAP - unit Earth Observation

VITO NV – Flemish Institute for Technological Research

Mol, Belgium

e-mail - jeroen.dries@vito.be, erwin.goor@vito.be, dirk.daems@vito.be

**Abstract**— VITO and partners developed recently an end-to-end solution to drastically improve the exploitation of the PROBA-V EO (Earth Observation) data archive and derived vegetation parameters from the Copernicus Global Land Service by researchers, service providers and thematic users. The analysis of time series of data (+1PByte) is addressed, as well as the large scale on-demand processing of the complete archive, including near real-time data. Several applications will be released to the users, e.g., a time series viewer, a full resolution viewing service, pre-defined on-demand processing chains and virtual machines with powerful tools and access to the data. 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, e.g., Landsat-7/8 and Sentinel-2/3 data, will be addressed as well.

**Keywords** - MEP Mission Exploitation Platform; PROBA-V; vegetation; data analytics; on-demand processing; Web Services.

### I. OBJECTIVES AND BENEFITS

The PROBA-V MEP (Mission Exploitation Platform) builds further on the R&D (Research & Development) results, from the ESA ‘ESE’ project. These results consist of prototypes which were further refined in several other projects thanks to the active involvement of these projects in the ESE pilots activities. The paper is organized as follows: in this section I we discuss the objectives of the platform. In section II the technical solution is described and section III contains the conclusions and future work.

The PROBA-V MEP has the ambition to complement the PROBA-V [1] user segment by building an operational Exploitation Platform (EP) on the data, complementary data and derived products, addressing hereby the wider vegetation user community with the final aim to ease, and increase, the use of PROBA-V data. The data offering will consist of the complete archive from SPOT-VEGETATION, PROBA-V and bio-geophysical parameters from the Copernicus Global Land Service [2].

The reasons for deploying a MEP dedicated to the PROBA-V mission are numerous:

- The data and specifically the time series of daily / ten-daily data from 1998 till present is too big to be downloaded to and processed on the users’ premises, at least for the majority of the users.
- On top of the Earth Observation (EO)-data mentioned above, the platform can co-locate as well

complementary data in a way that it is easily accessible. Furthermore tools, libraries and applications, which can be used by the large community will be provided. This includes as well the data needed for calibration and validation activities.

- The platform can stimulate collaboration between the users, as we bring together services from various users on the same platform with a number of tools to support the publishing of and to provide feedback on these services. A further focus on documentation, knowledge sharing and user support complements this.
- The platform goes beyond offering standard products by offering in a first place applications to visualize and analyze large time series of data and pre-defined on-demand processing services, which deliver user-tailored products. In a next step we will gradually deploy a Virtual Research Environment, being a platform, which allows users to develop – debug – test an application on an infrastructure at VITO with access to the complete data archive. Successful applications from third-parties can then be offered as an operational on-demand processing service to the user community on the same platform.
- As an Exploitation Platform (EP) with a focus on open interfaces, we position the PROBA-V mission in an ecosystem of TEPs (Thematic EPs), REPs (Regional EPs) and other MEPs. In the future, the PROBA-V MEP can be integrated gradually in a federation of different platforms, including as well Sentinel Collaborative Ground Segments, in line with the current ESA strategy on the ‘EO Ground Segment Evolution’.

During the PROBA-V MEP project, which will at least last till the end of the PROBA-V mission in May 2018, several third-party service projects will develop and operate applications on the operation MEP platform. We will address their user requirements to implement the shift of paradigm from “data to user“ to “user to data”, bridging the gap between the traditional EO ground segment and the scientist or value added industry by providing a one stop shop for access to the full PROBA-V Mission data (including derived parameters) and to external repositories of similar missions/sensors (including Landsat and Sentinel).

## II. TECHNICAL SOLUTION

The PROBA-V MEP will provide scalable processing facilities with access to the complete data archive and a rich set of processing algorithms, models, open source processing libraries/toolboxes and public/collaborative software. The platform becomes the processing infrastructure hub of the mission by functioning as a powerhouse system and open access development environment.

To realize this, the platform consists of the following components:

- The existing Product Distribution Facilities [3] and [4], are serving the access to the data archive, both via a Web portal as well as standardized discovery, viewing and data access interfaces. More evolutions on these standardized machine-to-machine interfaces are planned in the near future.
- Hadoop [7], as a platform for data-intensive distributed applications, is designed to process large amounts of data by separating the data into smaller chunks and performing large numbers of small parallel operations on the data. It is applied often for processing big data and is applied in this context for the on-demand processing of EO data, as prototyped successfully in the ESE project. Oozie [5] is used as a workflow processing engine to design an EO-application as a workflow of multiple processes. Spark [6] is used intensively to allow analytics on large time series of data. The Hadoop ecosystem provides furthermore a rich and still growing set of tools, which are used to provide fast access to the data in a format needed by the specific application.
- The EO raster data is accessible via NFS (Network File System) and possibly uploaded to the Hadoop Distributed Filesystem (HDFS) using a Data Manager. This Data Manager also integrates with several catalogues implementing different protocols, so that third party-data can be ingested into the platform when needed by a specific user.
- Cloud computing technology enables dynamic resource provisioning and is therefore providing a flexible and scalable solution. OpenStack [8] is chosen as cloud middleware. Pre-configured virtual machines will be offered and can run on the OpenStack cluster at VITO, providing the environment needed for users to work with the data and develop/deploy applications on the platform, i.e., containing IDE's, a rich set of tools and access to the complete data archive.
- Interactive Web-based dashboards are designed to provide user-tailored information from the EO data archives of VITO and other providers, by combining existing components such as AngularJS, Javascript libraries and GIS components into one single

solution. The combination of these different components, applied on data available in disparate data stores, offers powerful Web portals to the users in order to make vast amounts of data understandable. We can easily design user-tailored Web-based dashboards, which offer at any time near real-time information for the regional extent of interest to the user and in the format chosen by the user.

- A Web portal provides access to all applications and tools offered by the PROBA-V MEP and to the cloud consoles. Furthermore the portal provides all information on the data and components available on the platform and offers tools for e-collaboration and knowledge sharing amongst the users.
- A main concern is security since we allow users to develop and execute their applications on the platform. Their IPR shall be properly protected and the activities of individual users cannot influence the stability of the system and the work of other users. Single sign-on and proper monitoring of used resources are further requirements.

## III. CONCLUSIONS AND FUTURE WORK

The platform was launched in January 2016 at the PROBA-V conference in Ghent, Belgium. Three iterations are planned to gradually expand the capabilities of the system and provide new features, in close collaboration with the first third-party projects working on the platform.

The impact of this PROBA-V 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 operational platform is based on recent R&D activities and is in line with the new ESA strategy on the 'EO Ground Segment Evolution'. Hence, as future work, the integration of the platform within a federation needs to be addressed. More applications and users will be integrated in the platform to enrich the content and enlarge the user community. Furthermore the evolutions in Big Data analytics and processing will be followed closely and integrated in the platform where relevant.

## REFERENCES

- [1] <http://proba-v.vgt.vito.be/> [accessed: 2016-03-19].
- [2] <http://land.copernicus.eu/global/> [accessed: 2016-03-19].
- [3] <http://www.vito-eodata.be> [accessed: 2016-03-19].
- [4] <http://land.copernicus.vgt.vito.be/PDF/> [accessed: 2016-03-19].
- [5] <http://oozie.apache.org/> [accessed: 2016-03-19].
- [6] <http://spark.apache.org/> [accessed: 2016-03-19].
- [7] <http://hadoop.apache.org/> [accessed: 2016-03-19].
- [8] <http://www.openstack.org> [accessed: 2016-03-19].