

Pipeline computing in ESA Datalabs

To turn data from the archives of European Space Agency into valuable source of knowledge about Earth and the universe a robust, yet accessible data processing toolset is needed. The PIPEMAN component of ESA Datalabs is a new central tool within this powerful toolset. PIPEMAN functions allow the researcher to create, edit and launch flexible, re-usable, multi-input, multi-stage, multi-output data processing workflows also called *pipelines*. As a result of using pipelines the processing of raw or already processed data so that it can reveal answers to research questions becomes accessible to the less technical researchers. It also simplifies and speeds up the data processing work of the more technically trained researchers. It allows the researcher to focus more on research and less to the data processing technical tool creation and maintenance. Additionally, it allows the researcher community to share their data processing pipelines and algorithms which allows for yet another reduction of time needed for getting the answers from the data.

During the presentation an overview of pipeline computing functions and features soon to be available in ESA Datalabs is presented. We will look at how to create, edit and launch a pipeline. Versioning, scheduling, monitoring and stopping pipelines is presented as well. Also, the publication (and limiting access) of pipelines and steps into a common catalogue is shown and the searching for a suitable pipeline or step from the catalogue is looked at. Usage of a pipeline in the context of the notebook development is touched upon.

Plans for the release of the features are going to be laid out.

Presenter: Alo Joosepson, Senior Analyst-Consultant, CGI Estonia

Presentation duration: 20 minutes