

**ESA DATALABS: Towards a collaborative e-science platform for ESA.** C. Arviset<sup>1</sup>, V. Navarro<sup>1</sup>, R. Alvarez<sup>1</sup>, D. Basso<sup>2</sup>, S. del Rio<sup>3</sup>, M.A. Diego<sup>4</sup>, M. Lopez Caniego<sup>4</sup>, A. Lousa Marques<sup>5</sup>, F. Marinic<sup>1</sup>, A. Pereira<sup>2</sup>, N. Ramos<sup>5</sup>, V. Zlobin<sup>6</sup>, <sup>1</sup>ESA (Madrid, Spain, [Christophe.Arviset@esa.int](mailto:Christophe.Arviset@esa.int) and [Vicente.Navarro@esa.int](mailto:Vicente.Navarro@esa.int)), <sup>2</sup>Uninova, Lisbon, Portugal, <sup>3</sup>RHEA for ESA, Madrid, Spain, <sup>4</sup>AURORA for ESA, Madrid, Spain, <sup>5</sup>Edisoft, Lisbon, Portugal, <sup>6</sup>CGI, Tartu, Estonia.

**Introduction:** ESA space science missions are continuously producing data which is being hosted at the ESAC Science Data Centre and made available via archive services to the astronomical, planetary and heliophysics science community.

The increased volume and complexity of such science data calls for a paradigm shift in data analysis approach, enabling science users to bring their code to the data, rather than bringing the data to the users.

Leveraging on big data, cloud, container and artificial intelligence technologies, ESA Datalabs opens a new world for e-science collaboration platform that will provide scientists and engineers within the ESA science operations teams and users around the world with a seamless Open Science environment to process, analyse and visualise the sheer volume of science data.

Furthermore, ESA Datalabs has been built in a discipline agnostic way, support initially astronomy, planetary and heliophysics from ESA Space science missions and the Science data generated by the GNSS Science Support Centre, and potentially any other ones in the future.

**Data and Storage Services:** Data services allow users to add new data volumes to their ESA Datalabs environment, materializing the code to the data paradigm. These services provide a smooth integration of the Science Archives at ESAC. Thus, scientists only need to select among the list of available archive volumes and their algorithms will have read access to the datasets without the need to copy over anything.

For example, users will be able to mount data volumes from the Planetary Science Archive, hence having immediate access to all data from ESA planetary missions (ie Rosetta, Mars Express, BepiColombo to name a few).

In addition to the catalogue of hosted volumes, these services offer multiple options to connect to other data volumes the user may have access. Beyond standard protocols like WebDav, NFS or FTP, it is possible to connect to popular services like DropBox or Google Drive.

**Analysis and Desktop Services:** ESA Datalabs Analysis and Desktop Services provide on-demand web-access to exploitation tools, the Datalabs. Datalabs range from general domain systems that are becoming

de-facto standards for data processing, to domain-specific ones designed to tackle a particular problem.

Datalabs can be provided as Jupyter notebooks, enabling interactive data analysis (such as the JUICE icy moons' coverage and opportunity analysis tool), with data mounted from predefined well trusted and updated data such as the latest SPICE kernels provided by the ESA SPICE Service.

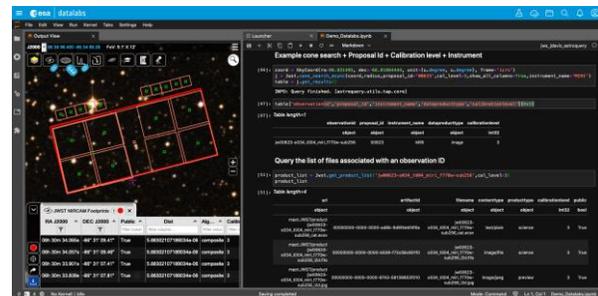


Figure 1: ESASky Jupyter notebook in ESA Datalabs

Datalabs can embed specific desktop applications (ie Topcat, well known astronomical tool to manipulate and visualize tables) without the need to actually install the applications on the end user machine.

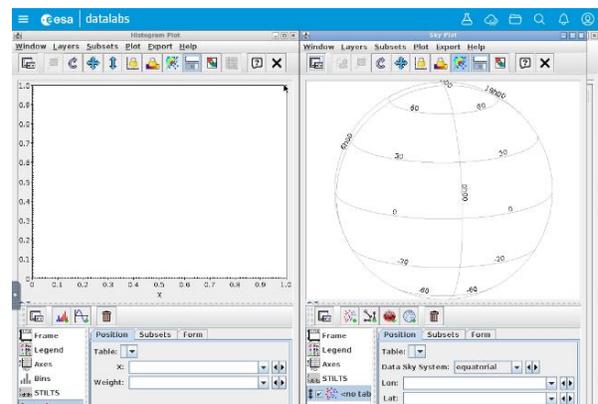


Figure 2: Desktop Application Topcat as a Datalab

Beyond support for native web-based systems, ESA Datalabs ensures long-term preservation of legacy software, enabling re-execution of desktop-based data processing systems that encapsulate key product generation algorithms. ESA Datalabs solution repackages these systems into runnable web-based applications.

At present, ESA Datalabs catalogue includes contributions from ESA Early Adopters, delivering immediate access to ESA Archives exploitation tools in domains like Astronomy, Planetary, Earth Observation, or Satellite Navigation.

**Pipeline Services:** In a data-intensive domain, a finding is the outcome of a pipeline of computations, applied to some large existing dataset, or to a dataset created by researchers from multiple other large datasets, combined in unexpected and complex way. ESA Datalabs Pipeline Services allow data integration, transformation and analytics based on a set of processing assets. Two major parts can be identified in this area: Pipeline Editor and Pipeline Executor.

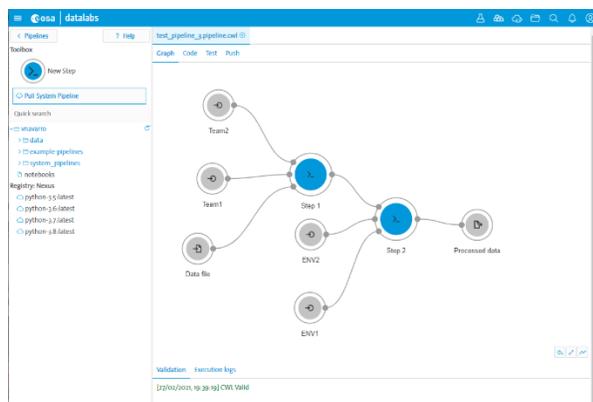


Figure 3: ESA Datalabs Pipeline Editor

The Pipeline Editor provides an integrated visual development environment for users to create a workflow of computing elements. The editor drives the user through the development cycle simplifying the creation process and transforming the graphical representation of the pipeline into its underlying code in CWL.

The Pipeline Executor provides overall orchestration through the Pipeline Launcher and Runs viewer. The first allows users to search a catalogue of Pipelines while the second allows control and visualization of running pipelines

**Status and Future Work:** ESA Datalabs, already available at [datalabs.esa.int](http://datalabs.esa.int), has successfully completed its core development phase implementing key features and a preliminary catalogue of Datalabs.

These contributions have demonstrated ESA Datalabs suitability to act as a catalyzer for ESA users, students and external organisations. These groups can build Datalabs that, sitting on top of ESA archives, address scenarios like:

- Fast development of prototype ideas requiring initial provisioning of complex systems;
- Software-as-a-Service delivery of desktop applications and systems;
- Training framework for coding competitions, trainees, small external developments
- Outreach and demonstration of systems requiring complex initial configuration.



Figure 4: Initial sets of datalabs

ESA Datalabs next steps gear towards the evolution from currently restricted, pre-operational system available at [datalabs.esa.int](http://datalabs.esa.int), into a public environment for open-science and innovation across multi-disciplinary research communities.

Moreover, throughout 2021, ESA missions and projects are to contribute to build up ESA Datalabs catalogue with new applications, notebooks and pipelines capable to exploit the full potential of this platform for Artificial Intelligence.

**Acknowledgments:** The authors would like to thank all the people making ESA Datalabs a reality, in particular designers and developers from Edisoft and Uninova in Portugal, Thales in France, CGI in Estonia, as well as the Science and Operations technical IT Unit at ESAC, and not to forget the members of ESA Space Science missions and the ESAC Science Data Centre who are building datalabs to be deployed on the platform.

#### References:

- [1] V. Navarro et al (2019), "ESA Datalabs Multi Mission Science Exploitation and Preservation Platform" ADASS 2019.
- [2] C. Arviset et al (2019), "Use cases for the ESAC Science Exploitation and Preservation Platform" BiDS 2019.