Euclid Processing Pipeline Operations (from raw telemetry up to mass mapping of the universe)

The complete Euclid Processing Pipeline covers a wide scope of data products: from raw telemetry received from the spacecraft to the 3D mass mapping of the universe or the catalog of clusters of galaxies right through calibration data, catalog of galaxies, calibrated spectro images or stacked images.

This wide zoology of data products is processed through a complex data and control flow that have to deal with some various constraints or opportunities such as:

- Taking advantage of a distributed processing and storage infrastructure in heterogeneous environment
- Taking into account dependencies between the different elements of the Euclid pipeline
- Considering different processing stages triggering strategies based on business rules
- Automating as much a possible the execution of the different pipelines instances
- Assessing and monitoring the quality of the products
- Monitoring the data production rate to ensure the delivery of data Releases on due dates
- Considering the indispensable reprocessing of some stages to fix some bugs or to reach the specified quality requirements of the data products
- Monitoring and Controlling the log reports from the different applications including processing stages and infrastructure services
- Analyzing these logs manually or automatically and react accordingly
- Supervising the operations through different dashboards: infrastructure, logs, data rate production

We address in this presentation the different perspectives of the Processing Operations and the different technical and organizational solutions put in place.

We'll first introduce the different definitions:
Processing Plan: selection of a pipeline definition and a set of input data products
Processing Orders: for a given Processing Plan the list of the elementary pipeline run
Pipeline Script: python script that defines the job control flow of the pipeline definition with the different patterns implemented: parallelization, synchronization, split, conditional triggering …
Pipeline Runner: the infrastructure component that executes the Pipeline Script while submitting the different jobs to the queue of the computing center.
Infrastructure dashboard services: the component that centralizes the status of the different infrastructure services on the different computing centers
Production Plan dashboard monitoring: the user interface allowing to trigger, to monitor the Porcessing Plans and its Processing Orders.

We'll depict some operational scenarii and the tools and dashboards made available to the profiles involved in these operations.
We'll come back on the different deliverables from the scientific development teams and how we orchestrate the distribution of the data and the executables in a distributed environment.
We'll illustrate the different levels of monitoring and control: scientific components, infrastructure components and how we oversee the correct execution of the pipelines.