18th International Conference on Space Operations, Montreal, Canada, 26 - 30 May 2025.

SpaceOps-2025, ID # 329

Euclid Science Ground Segment concepts and operations

Guillaume Libet[a]*

[a] Euclid *Science Ground Segment Production Manager, Centre National des Etudes Spatiales, France*, guillaume.libet@cnes.fr
* Corresponding Author

Abstract

The Euclid mission, an ESA project designed to explore the nature of dark energy and dark matter, relies on a complex Science Ground Segment (SGS) for data processing. This manuscript describes the structure and operation of the Euclid Pipeline, which transforms raw data from the Euclid spacecraft into scientific products. The study highlights the multi-stage processing, encompassing Levels 1 to 3, and the challenges of automating Processing Function execution across varying spatial scales. The manuscript also details the tools used by the Operations Team to manage data production and monitor pipeline operations. The Euclid Pipeline efficiently converts raw data into valuable scientific products for cosmological research.

Keywords: Euclid mission, Science Ground Segment, data processing, Euclid Pipeline, automation, data management

Acronyms/Abbreviations

Baryon Acoustic Oscillations (BAO)

Centre National d'Études Spatiales (CNES)

COmmon DEvelopment ENvironment (CODEEN)

COmmon ORchestration System (COORS)

Data Processing System (DPS)

Distributed Storage System (DSS)

Euclid Consortium (EC)

European Space Agency (ESA)

External Ground Based Surveys (EXT)

Infrastructure Abstraction Layer (IAL)

Level 1 (LE1)

Level 2 (LE2)

Level 3 (LE3)

Merged (MER)

Near-Infrared Spectrometer and Photometer (NISP)

Near-Infrared Photometry (NIR)

Processing Functions (PF)

Photometric Redshift (PHZ)

Performance Verification (PV)

Science Data Centers (SDC)

Science Ground Segment (SGS)

Shape Measurement (SHE)

Near-Infrared Spectroscopy (SIR)

Science Operations Centre (SOC)

Spectroscopic Redshift (SPE)

Visible Imager (VIS)

1. Introduction

The Euclid mission, launched by the European Space Agency (ESA), aims to investigate the nature of dark energy and dark matter by simultaneously employing two primary cosmological probes: weak gravitational lensing and baryon acoustic oscillations (BAO). To achieve these scientific goals, Euclid will observe galaxies and clusters of galaxies out to a redshift of z~2, utilizing a wide-angle extragalactic survey covering 14,000 deg² and a deep survey covering 40 deg² with multiple visits. The Euclid spacecraft carries two scientific instruments: the Visible imager (VIS) and the Near-Infrared Spectrometer and Photometer (NISP). Following its launch on July 1, 2023, and the subsequent commissioning and Performance Verification (PV) phase, the nominal survey acquisition commenced on February 14, 2024 (see [1] Mellier et al., Euclid Collaboration, 2024 for a comprehensive mission overview).

2. Euclid Science Ground Segment

The Euclid Science Ground Segment (SGS) represented in Fig. 1) is responsible for processing the vast amount of data generated by the Euclid mission. The SGS comprises the Science Operations Centre (SOC), operated by ESA, and a network of nine Science Data Centers (SDCs). These SDCs, provided by the Euclid Consortium (EC), a global collaboration of over 110 institutes across 15 countries, perform the critical task of data processing. The Operations Team, based at CNES, oversees the configuration, triggering, monitoring, and reporting of the data processing within the Euclid Pipeline.

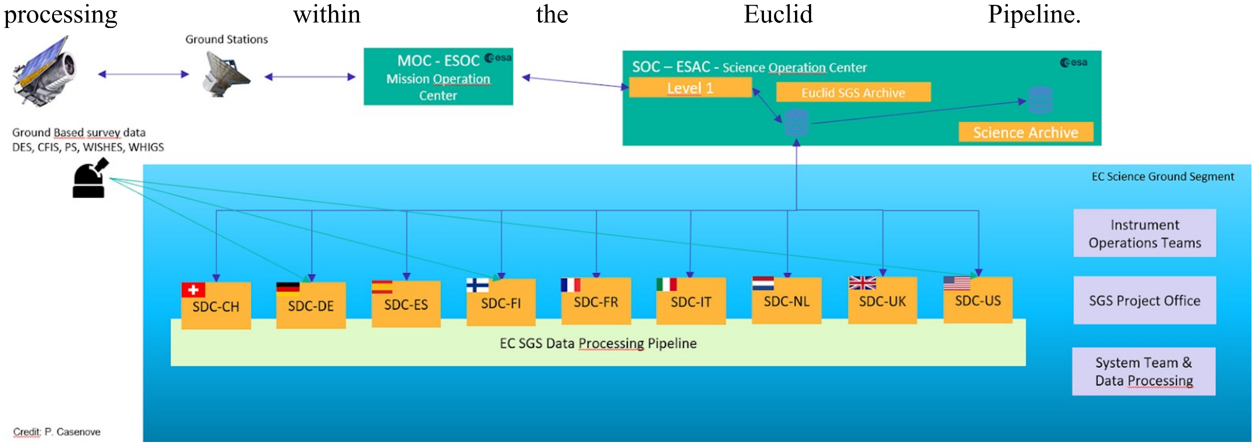

Fig. 1. SGS Overview

To minimize the need for extensive data transfers between geographically dispersed centers, each SDC autonomously executes the complete suite of Processing Functions (PFs) that constitute the Euclid Pipeline, employing a strategic sky sharing mechanism. This mechanism is detailed in "The Automation Imperative" chapter.

2. The Euclid Pipeline

2.2 Overview:

The Euclid Pipeline is a multi-stage process that transforms raw data from the Euclid instruments into calibrated scientific data products.

It is represented in Fig. 2 :

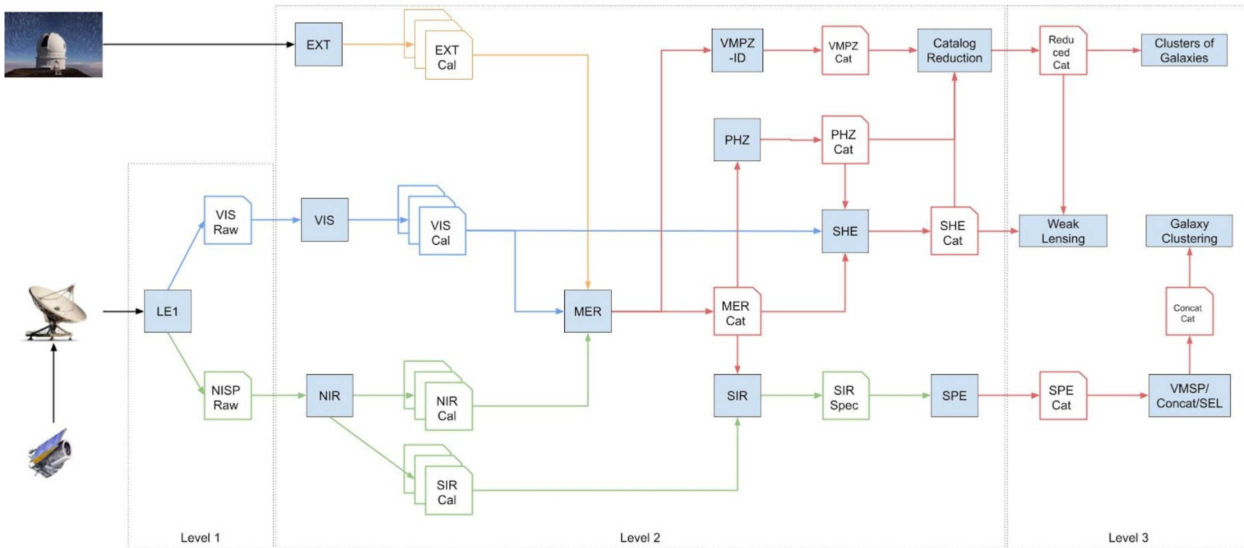

Fig. 2 : The Euclid Pipeline

The pipeline spans three levels of processing:

- **Level 1:** The primary goal of Level 1 processing is to convert the raw, unprocessed instrument data into calibrated engineering units. This involves decommutating the raw data stream from the instruments.
- **Level 2:** At Level 2, raw frames from the VIS and NISP instruments are processed to generate calibrated images. This includes visible-band photometry (VIS), near-infrared photometry (NIR), and near-infrared spectroscopy (SIR). The MER processing function then creates a comprehensive catalog by integrating these calibrated data with images from external, ground-based surveys (EXT). From this step, two main pipelines are executed:
    - The photometric pipeline, composed of PHZ, which measures the photometric redshift for observed galaxies, and SHE, which measures the cosmic shear.
    - The spectroscopic pipeline, composed of SIR and SPE, which measures the spectroscopic redshift.
- **Level 3:** The final Level 3 processing functions synthesize the outputs of the photometric and spectroscopic pipelines. This synthesis enables the mapping of the three-dimensional distribution of matter in the Universe, providing crucial insights into the nature of dark energy and dark matter.

2.2 The Automation Imperative: Navigating Multi-Scale Data Processing:

A significant challenge in the Euclid Pipeline execution is the automation of Processing Function execution as data becomes available. The pipeline operates across varying spatial scales. Some PFs are triggered at the level of individual Euclid Observations (whose footprints differ between VIS and NISP), while others are executed at the scale of MER Tiles.

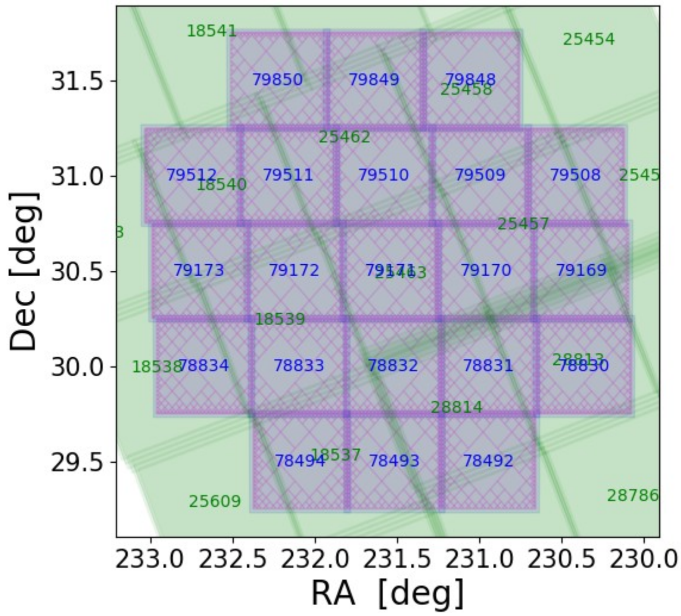

Fig. 3 : Mapping between tiles and observations

As illustrated in Fig. 3, a single MER Tile (e.g., index 78832) can be covered by multiple Euclid Observations (e.g., 18539, 25463, 28814, and 18537). Critically, these observations are not acquired – and thus processed at the VIS/NIR level – simultaneously.

A naive automation approach, triggering processing each time a new observation intersects a tile, would lead to a massive waste of computational resources through redundant processing. Conversely, waiting until a tile is fully covered could introduce significant delays in data production, potentially spanning months if the final covering observation is scheduled far in the future.

Euclid Pipeline Automation is specifically designed to address the question of when the optimal time to trigger processing. This decision point arises each time the pipeline transitions between different spatial scales (MER, SIR, SHE). The automation relies on the use of Intersection Maps.

2.3 Intersection Maps and Pipeline Triggering:

The Euclid Pipeline automation requires the use of Intersection Maps. In brief, the Intersection Maps are a table that stores the relations between the Euclid VIS observations and the tiles, with coverage ratio, planned observation acquisition dates, and processed observation and tiles.

The Intersection Maps are generated and updated as follows:

- From any Euclid survey and the MER tiling, the *ST_IntersectionMaps* program computes the theoretical overlap between the Euclid Observations and the MER Tiles.
- This process generates a *DpdIntersectionMaps* product, giving the complete list of planned intersections for the upcoming survey.
- When Level 2 processing occurs and products are ingested, the *DpdIntersectionMaps* is updated by the DPS during Level 2 products ingestion. This update retrieves the processed spatial footprint of the product and computes the exact observation/tile intersections.

To define when a pipeline has to be triggered, in addition to the Intersection Map, operators ingest business rules (or advanced triggering criteria) based on the content of the processed Intersection Map. These rules are stored in the archive as a *DpdTriggeringRules* product. Initially, these rules are written solely by the Operations team, with updates that take into account scientific feedback.

The last required component of the automation is the Distribution Map, that assigns Euclid Observation and MER Tiles to a SDC. In Euclid, the same pipeline is executed in all SDCs, the sky is shared among the SDCs, according to the high level share between countries in the SGS.

Based on the survey, *ST_Intersectionmap* is able to create a *DpdDistributionMap*, a data product that will cover two needs:

- Define the SDC that will used to process a given Euclid Observation or a MER Tile
- Define how the LE1 output will be stored in a Primary SDC (the one that will do the processing) and in a secondary SDC (used for storage backup)

For a given product the secondary SDC is not used for processing but only to have a backup copy of LE1 outputs.

3. Tools for data production

To effectively manage the complex data production flow, the Operations Team utilizes a suite of dedicated tools:

- COORS (COmmon ORchestration System): A web-based platform for defining and managing the execution of the entire SGS Production Pipeline through user-defined execution plans.
- Euclid Archive System: A hierarchical system comprising:
  - DPS (Data Processing System): A central database referencing all data products.
  - DSS (Distributed Storage System): Local file storage deployed at each data center for locally produced data.
  - Euclid Science Archive: A public-facing archive for the dissemination of validated data products to Science Working Groups and the broader scientific community.
- IAL (Infrastructure Abstraction Layer): The middleware responsible for executing Pipeline Processing Orders across the diverse computing infrastructure within each Data Center.
- CODEEN (COmmon DEvelopment ENvironment): The continuous integration platform ensures the robustness and reliability of the pipeline software.

These tools are represented on Fig. 4

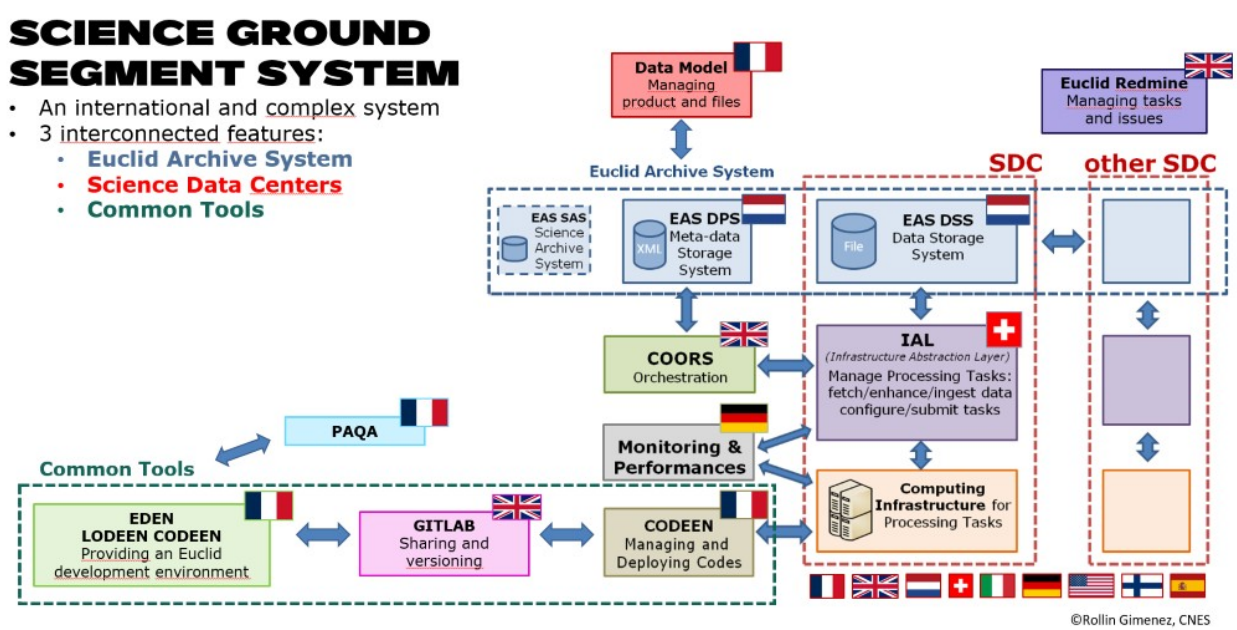

Fig. 4 : Euclid Science Ground Segment System overview

4. Operations Monitoring

The Euclid Science Ground Segment incorporates a robust suite of monitoring tools to ensure efficient and reliable operations:

- ELK (Elastic Stack) & Grafana: A powerful combination for centralized log management and the creation of dynamic dashboards to track processing execution and critical performance metrics.
- Stellarium: A visualization tool that allows operators to track the creation and spatial distribution of data products by mapping them onto the celestial sphere.

Grafana empowers operators to monitor the execution of individual Pipeline Processing Orders or entire execution Plans in real-time. Furthermore, it provides valuable insights into resource utilization and detailed profiling information for individual Processing Functions and the overall performance of each Science Data Center.

Fig. 4 : Screen Capture of the SGS Monitoring dashboard in Grafana

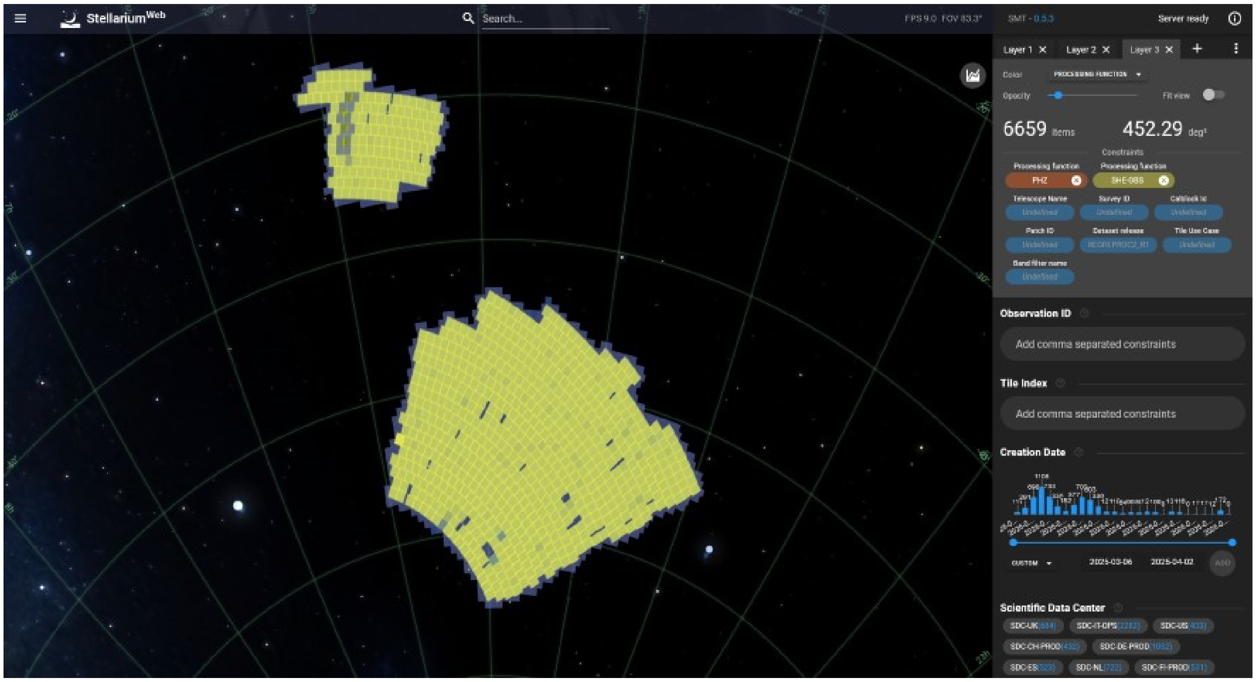

Fig. 5 : Screen Capture of the Stellarium view of Non-Regression Reprocessing 2 production

Stellarium implements several layers to improve data visualization. For example in Fig.5, VIS/NIR is superimposed on layer 1, MER on Layer 2, and PHZ and SHE (at the observation scale) on layer 3. Processing functions are discriminated using different colors and transparency between layers.

5. Conclusion

The Euclid Science Ground Segment and its associated data processing pipeline are critical for the success of the Euclid mission. The multi-stage pipeline, the automation of processing, and the suite of dedicated tools ensure the efficient and reliable transformation of raw data into valuable scientific products for cosmological research.

References

[1] Mellier et al., Euclid Collaboration, Overview of the Euclid mission, September 25 2024, https://arxiv.org/abs/2405.13491