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

SpaceOps-2025, ID # 472

The EUCLID Italian Operational Science Data Centre

Lorenzo Bramante[a]*, Federico Farinetto[a], Lucio Vincenzo Costa[a], Alberto Alessio[a], Stefano Lanza[a], Matteo Del Giudice[a], Stefano Giannuzzi[a], Marco Richichi[a], Rosario Messineo[a], Filomena Solitro[a], Marco Frailis[b], Daniele Tavagnacco[b], Erik Romelli[b], Thomas Gasparetto[b] , Andrea Zacchei[b] , Elisabetta Tommasi[c]

[a] *ALTEC S.p.A, Turin, Italy, 10146,* lorenzo.bramante@altecspace.it, federico.farinetto@altecspace.it, luciovincenzo.costa@altecspace.it, alberto.alessio@altecspace.it, stefano.lanza@altecspace.it, matteo.delgiudice@altecspace.it, stefano.giannuzzi@altecspace.it, marco.richichi@external.altecspace.it, rosario.messineo@altecspace.it, filomena.solitro@altecspace.it
[b] *INAF, Trieste, Italy, 34149,* marco.frailis@inaf.it, daniele.tavagnacco@inaf.it, erik.romelli@inaf.it, thomas.gasparetto@inaf.it, andrea.zacchei@inaf.it
[c] *ASI, Rome, Italy, 00133,* elisabetta.tommasi@asi.it
* Corresponding Author

Abstract

Euclid is an ESA mission designed to map the geometry of the Universe and better understand dark matter and dark energy. The satellite hosts the Visible Instrument (VIS) and the Near Infrared Spectrometer and Photometer (NISP). The raw data acquired by the satellite is processed within the Euclid Science Ground Segment (SGS) through scientific pipelines running on HPC infrastructures available across nine Science Data Centres (SDCs).

SDC-IT-PROD is the Italian Euclid Science Data Centre, funded by the Italian Space Agency, where the mission data is processed and the scientific outputs are produced. It was designed and implemented to validate the SGS systems before the satellite launch and to carry out the mission operations, contributing 25% of the entire SGS's processing and storage resources. SDC-IT-PROD also hosts tools used by the NISP Instrument Operations Team.

SDC-IT-PROD uses advanced High Performance Computing (HPC) infrastructures and is divided into an integration and test platform and a production platform. The centre's architecture combines modern HPC and HTC paradigms to perform efficient data processing and manage the massive data volumes generated during the mission. The large data volumes led to the creation of a two-tier hierarchical storage system. Tier 1 is high-performance storage supporting real-time processing, while Tier 2 serves as long-term archive. Data in both tiers is automatically managed based on mission requirements. Tools for data exploitation are available to the scientific team to perform troubleshooting.

An operational team was established months before the satellite launch to define and prepare the tools needed to manage the Italian data processing centre and familiarize with the instruments provided by the Euclid SGS. The team focuses on:

- controlling data processing platform components,
- monitoring the execution of scientific pipelines,
- verifying the status of software for scientific algorithms,
- exchanging data with other SDCs and archiving data,
- checking the status of instruments used for data exploitation,
- quickly identifying issues through automatic alerts,
- executing Data Quality Check pipelines for algorithms managed by the Italian data processing centre.

The advanced technologies and architectural design of SDC-IT-PROD have been crucial to meet the mission requirements, allowing the centre to evolve according to mission needs, optimize resource utilization, and enhance operational efficiency.

Keywords: data processing operations, science mission, big data, data management, HPC & HTC infrastructure.

Acronyms/Abbreviations

ALTEC:	Aerospace Logistic Technology Engineering Company
ASI:	Agenzia Spaziale Italiana (Italian Space Agency)
DSS:	Distributed Storage System
EAS:	Euclid Archive System
ECSGS:	Euclid Consortium Science Ground Segment

EDEN:	Euclid Development ENvironment
ESA:	European Space Agency
HPC:	High Performance Computing
HTC:	High Throughput Computing
HTTP:	HyperText Transfer Protocol
IAL:	Infrastructure Abstraction Layer
ICR:	Instrument Command Request
INAF:	Istituto Nazionale di Astrofisica
IODA:	Instrument Operations Data Analysis
IOT:	Instrument Operation Team
MOC:	Mission Operation Centre
NISP:	Near Infrared Spectrometer and Photometer
OGS:	Operational Ground Segment
PF:	Processing Function
SDC:	Science Data Centre
SGS:	Science Ground Segment
SOC:	Science Operation Centre
VIS:	Visible Imaging System

1. Introduction

Euclid is an ESA mission defined within the ESA Cosmic Vision 2015-2025 program, aimed at mapping the geometry of the Universe and gaining a better understanding of the mysterious dark matter and dark energy.

The satellite was launched in July 2023 and will acquire data for a six-year nominal phase, followed by two years of post-processing, resulting in the production of an unprecedented volume of data for a space mission.

Euclid is equipped with two instruments: the Visible Imaging System (VIS) and the Near Infrared Spectrometer and Photometer (NISP). The data acquired by these two instruments will be processed within the Euclid Science Ground Segment (SGS) through the activation of scientific pipelines. These pipelines are interconnected through a complex mechanism of dependencies, running on the HPC infrastructures available in the nine data processing centres visible in Fig. 1, known as Science Data Centres [1, 2].

The vast data volume characterizing the Euclid mission is the main challenge that led the SGS to select specific data processing strategies and prompted the Science Data Centres to design infrastructures capable of handling such large amounts of data, both in terms of processing power and storage capacity.

1.1 Euclid Ground Segment

The Euclid Ground Segment is composed by:

- **Operational Ground Segment** (OGS)
- **Science Ground Segment** (SGS)

The OGS consists of the Mission Operations Centre (MOC) and the Ground Station. The MOC is responsible for spacecraft operations and for delivering telemetry and attitude data to the Euclid Science Ground Segment.

The SGS includes:

- **Euclid Consortium Science Ground Segment** (ECSGS).
- **Science Operations Centre** (SOC).

SOC acts as the interface between the MOC and ECSGS and is responsible for Euclid survey implementation, the production and delivery of Level 1 data, and for monitoring the health of the instruments.

Nine **Science Data Centres** (SDC) compose the ECSGS (Fig. 1), one for each country of the Euclid Consortium.

Each SDC is responsible for executing the Euclid scientific pipelines, called Processing Functions (PF), as well as for the storage and distribution of the produced output.

Two SDCs host the **Instrument Operations Team** (IOT), one for each Euclid instrument. The IOT is responsible for the maintenance and operation of the Euclid instruments. The NISP IOT is located within the Italian data processing centre.

The **Italian data processing centre** (SDC-IT) is composed by three processing infrastructures:

- SDC-IT-DEV: development platform located at INAF OATs premises.
- SDC-IT-INT: integration and test platform located at ALTEC S.p.A premises.
- SDC-IT-OPS: operational platform located at ALTEC S.p.A premises.

The development platform is used to implement algorithm, support the evolution and maintain the three Euclid scientific PFs in charge of Italy (NIR, SIR and MER). The team at INAF OATs has also the responsibility for the high level coordination of the entire SDC-IT and develops part of the LE3 PFs and the Data Quality pipelines necessary to validate the NIR and SIR results.

SDC-IT-INT and SDC-IT-OPS are the reference platforms for the execution of the mission operations and validation of the SGS systems. This paper outlines the processes undertaken to design, realize and operate the integration & test and operational platforms within the data processing centre named **SDC-IT-PROD**.

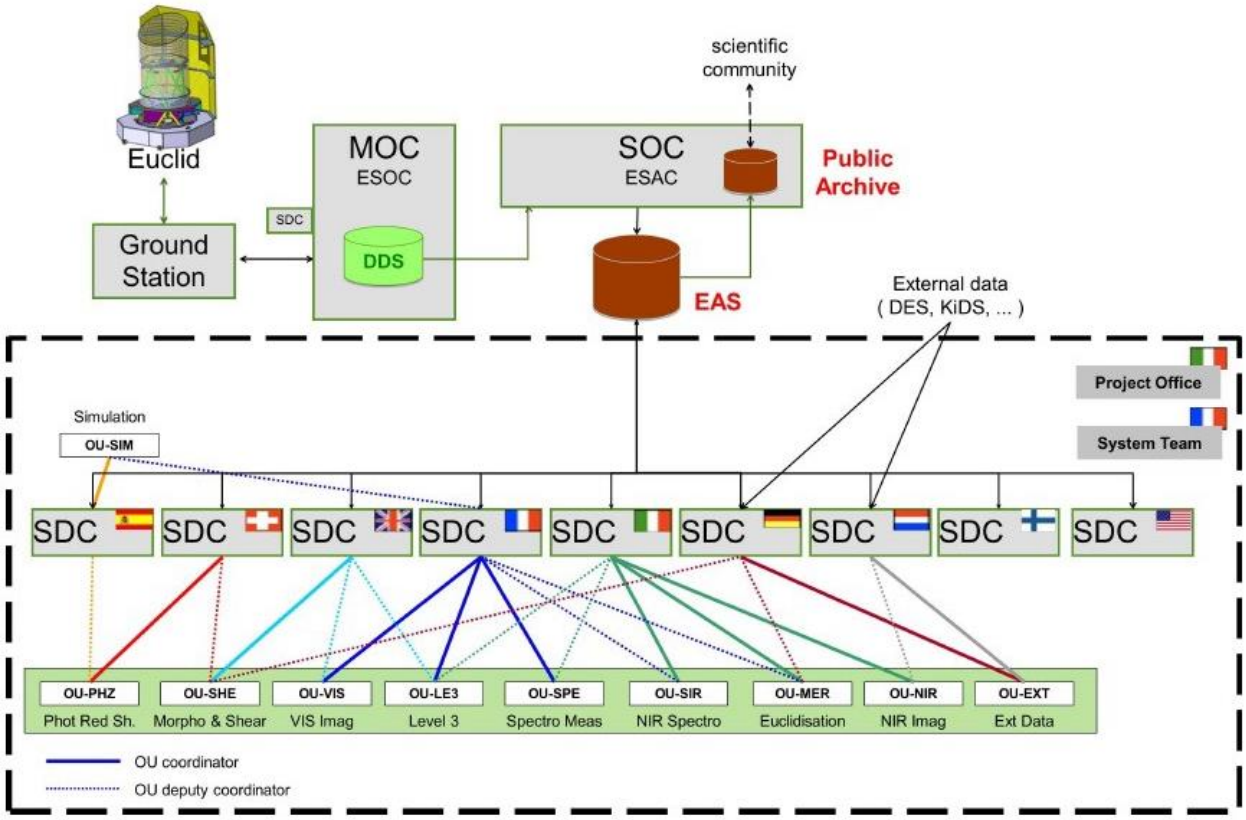

Fig. 1. Euclid Ground Segment

1.2 Mission Timeline

The Euclid Mission timeline consists of four main phases: the Launch and Early Operations Phase, the Commissioning Phase, the Routine Operations Phase and the Post-Operations Phase.

The **Launch and Early Operations Phase** covers the initial period of mission operations from launch to a few days afterward.

The **Commissioning Phase** is composed by two sub-phases: Spacecraft Commissioning and Performance Verification Phases. During the Spacecraft Commissioning Phase the instruments are switched-on and the checkout and commissioning are performed in order to verify that the satellite is ready to take science exposures.

The scope of the Performance Verification Phase is to establish the initial in-flight calibration and confirm that the instrument performance meets the science goals.

Euclid's nominal mission is known as the **Routine Operations Phase**. During this period, the science data that will form the Data Set Releases will be produced and made available to the scientific community after an internal validation process. At the time of writing, the mission is in this phase, and the SGS is preparing for the production of the first Internal Data Set Release scheduled for fall 2025 with the data expected to be made publicly available in fall 2026.

The **Post-Operation Phase** starts after the reception of the last scientific data on ground. The last Euclid Data Release will be produced in this period and the mission will enter the **Active Archive Phase**.

2. SDC-IT-PROD

SDC-IT-PROD is the Italian data processing centre located at ALTEC S.p.A. where mission operational data will be processed and scientific outputs produced. It was designed and built to assist in the validation of the SGS systems prior to the satellite launch and to support the execution of scientific mission operations throughout the entire satellite's lifecycle.

SDC-IT-PROD consists of two main infrastructures: the **integration and test platform**, called SDC-IT-INT, and the **production platform**, SDC-IT-OPS.

SDC-IT-INT serves as the reference platform for verifying and validating the scientific and infrastructure software products of the SGS, as well as supporting their maintenance and evolution. It is also used to validate the SDC-IT-PROD systems before their release into the production environment.

SDC-IT-OPS is the platform dedicated to the execution of Euclid mission operations, including data exchange, processing, storage, IOT activities, and data analysis performed by the scientific teams.

2.1 SDC-IT-PROD goals

The main goal of SDC-IT-PROD is to provide 25% of the total SGS processing and storage mission resources. This objective is achieved through the establishment of the SDC-IT-OPS HPC & HTC infrastructure.

In the months leading up to the nominal mission, SDC-IT-PROD, in collaboration with the other SDCs, has been responsible for validating the SGS systems that are currently used in the execution of routine operations. It also contributes to the validation of the NIR, SIR, and MER scientific pipelines, as well as the validation of their data products.

An objective of SDC-IT-PROD is to assist scientific teams in the data analysis process when issues are identified during the execution of the mission operations. This includes ensuring fast data access and providing the necessary tools for performing data analysis.

Additionally, it is equipped with all necessary tools to operate the NISP instrument, perform health checks through plots and trend analysis, and generate Instrument Command Requests (ICR) for instrument planning updates.

The LE1 NISP Processor, running at the SOC, is also developed within the Italian data processing centre and is currently under maintenance.

2.2 SDC-IT-PROD architecture

The high-level architecture of SDC-IT-PROD is shown in Fig. 2. It consists of the subsystems and functionalities described in this chapter, which operate seamlessly through the infrastructures presented in paragraph 2.3.

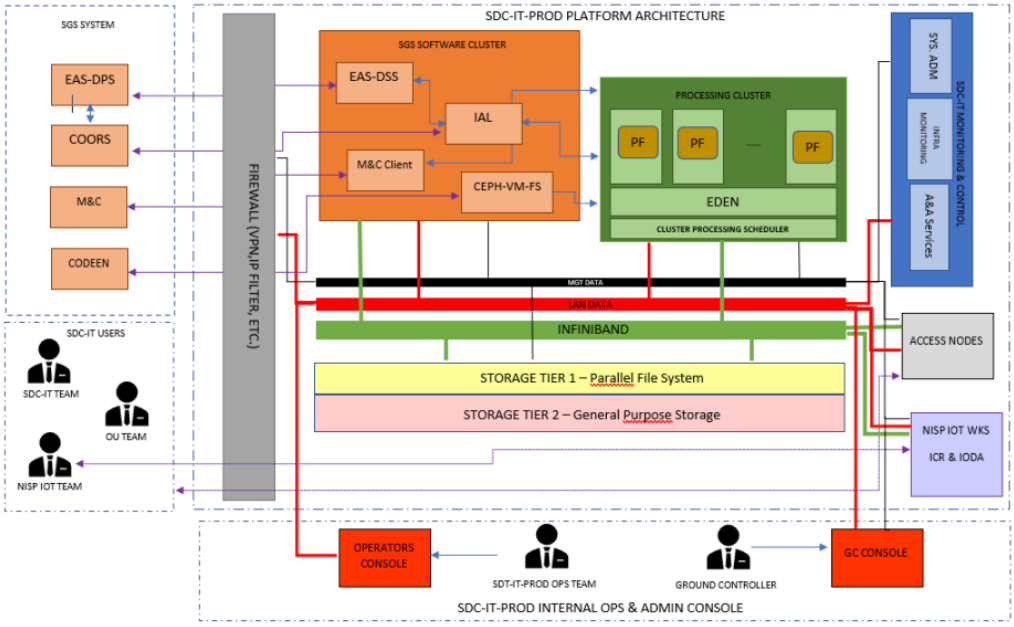

Fig. 2. SDC-IT-PROD Architecture

The **Data Exchange** subsystem enables the exchange of input/output data products between the SDCs and their release into the Euclid Science Archive System (SAS). Each SDC hosts its own instance of the Data Storage System (DSS), contributing a specific percentage to the overall distributed storage architecture which is a key component of the mission's data management. DSS is the software developed at SGS level that allows the retrieval of input data files for the pipelines execution and the storage of output data in the local archive [3]. All DSS instances are interconnected and continuously exchange data over the HTTP protocol.

The **Data Persistence** subsystem consists of data stores configured within the SDC-IT-PROD hierarchical storage. The first level, known as Tier 1, is a high-performance storage system that supports real-time processing. The second level, referred to as Tier 2, serves as a long-term data archiving solution characterized by high capacity.

The **Data Processing** functionality aims to provide the ability to execute the scientific pipelines. The processing subsystem, deployed on top of the HPC cluster, is called IAL (Infrastructure Abstraction Layer) and is a tool provided by the SGS [4] that works in combination with a workload manager. The IAL is responsible for activating processing within the HPC infrastructure, collecting the necessary input data through interaction with the DSS, submitting processing jobs to the HPC workload manager and handling output data storage by interacting once again with the DSS.

The scientific pipelines are made available within the data processing centre thanks to the Cern Virtual Machine File System (CVMFS) that is mounted along the entire infrastructures. Updated versions of the pipeline are released within CVMFS by a mechanism of continuous integration and continuous deployment.

The processing is carried out within EDEN, the Euclid development and processing environment, which is accessible through CVMFS, ensuring that the pipelines are executed under consistent conditions across all data processing centres.

All SDCs are capable of executing all scientific pipelines (available through CVMFS) within the same data processing environment (EDEN) using the same data processing software (IAL). This data processing strategy addresses the large volume of data associated with the mission by minimizing data sharing moving software between the SDCs

The **Monitoring and Control** tools are foreseen at SGS level to monitor the overall status of the processing infrastructures, processing distribution and data exchange. Alongside with the SGS monitoring and control software a local set of tools, such as Grafana, Elasticsearch and Zabbix have been deployed and configured to provide a more precise overview of the SDC-IT-PROD status. These tools are used to monitor the hardware and software performances and to detect in advance possible anomalies that could affect the execution of the mission operations.

The activities that the **IOT** carries out through SDC-IT-PROD include the planning, maintenance, and monitoring of the NISP instrument. Two dedicated workstations, equipped with the IOT SGS tools, are installed and configured within SDC-IT-PROD for these purposes.

The **Science Support** subsystem includes hardware, software, and configurations designed to provide the scientific team with rapid access to data generated within the SDC-IT-PROD. Its primary goal is to enable scientists to efficiently review scientific data and quickly address any issues that may arise.

The **Mission Support** Tools is a subsystem composed by software used to support the execution of the data processing centre operations. It includes Time Display Application, Mission Console Log and an issue tracking system.

The **Internal Services** are used by the other subsystems to perform basic operations such as authentication, software access, configuration versioning and automation tools.

2.3 IT Infrastructures

SDC-IT-PROD is a data processing centre built on two High Performance Computing (HPC) architectures, SDC-IT-OPS and SDC-IT-INT, designed to scale both computationally and in storage capacity throughout the mission. It ensures also a fast data access for quick mission data inspection in case of contingencies.

The two platforms mirror each other, as the integration and test platform is designed to validate the software products and configurations of the operational platform.

The elements that characterize the infrastructures and how they are integrated with each other are presented in Fig. 3. There are components specific to one of the two platforms and parts that are shared between them.

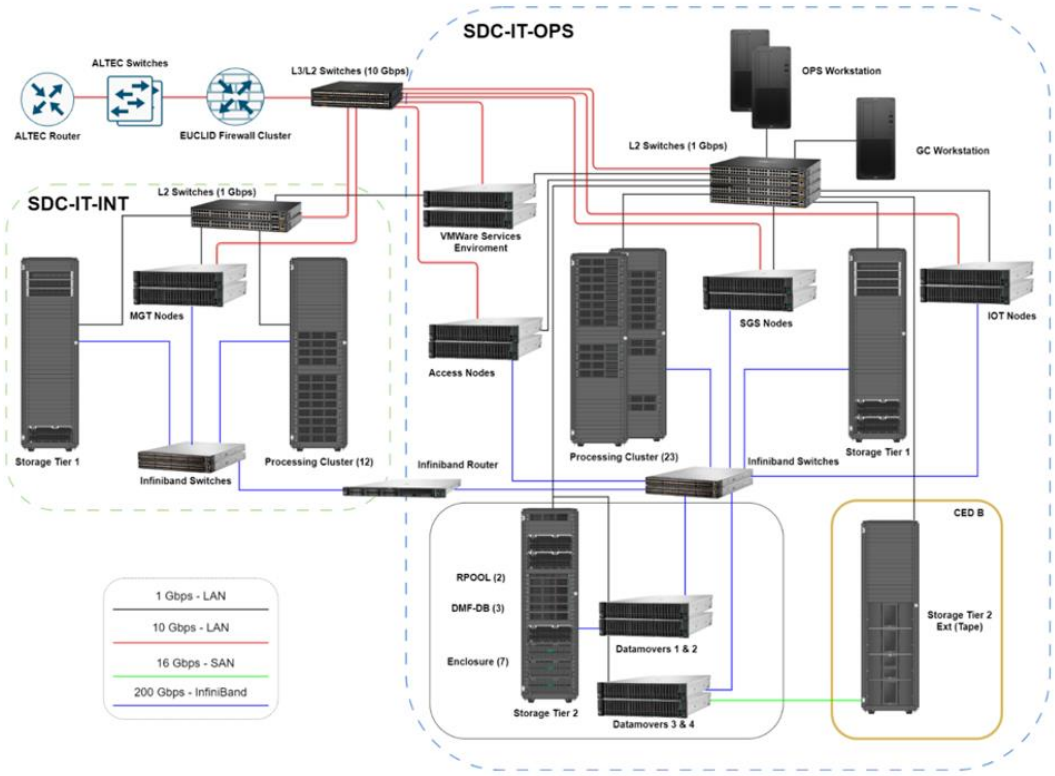

Fig. 3 – SDC-IT-PROD platforms

The current configuration of the SDC-IT-OPS platform is equipped with:

- **10 Gbps link**, provided by GARR, that ensures a fast connection with the others data processing centre. This specific link connects all components that exchange data with the outside and all elements where the instruments configured for the SDC-IT-PROD end-users are available.
- **HPC processing clusters** consisting of 23 nodes, utilizing general-purpose hardware, with Slurm as the workload manager and xCAT for the cluster management [7]. The total amount of physical cores is 1824 which support 2 thread per core, for a total amount of 3648 available CPUs. This subsystem will scale along the mission reaching the number of cores visible in Fig. 4.
- **SGS cluster** hosts the SGS software responsible for data exchange (DSS), storage (DSS), and processing (IAL). Each software is assigned to a dedicated server.
- **Hierarchical Storage with two tiers**:
    - Tier 1 is implemented with the storage appliance machine ClusterStor E1000 with lustre [6] and it is used to support real time processing with 1.5 PB of high performance usable space.
    - Tier 2 is implemented with the Zero Watt Storage technology and it is used for long term archive with a total capacity of 4.4 PB [5]. It also includes a Tape Library, installed in a different server room, to increase the level of protection of the data. Tier 2 is shared between the operational and integration and test infrastructures.
    - HPE Data Management Framework (DMF) is designed to enable transparent data management between the two storage levels, following policies defined by the data processing operational team that are based on file name convention, age, size and location [5].
- **Internal Network up 200 Gbps** with InfiniBand, a high-performance, low-latency interconnect network providing fast and efficient communication between servers and storage systems.
- **Management Cluster**: a virtual environment implemented with VMWare technology used to install and configure monitoring and administration services for both infrastructures.

The SDC-IT-PROD procurement plan foresees regular storage and computation resources upgrade every 2-3 years in order to follow the mission needs. Upon completion of its expansion, SDC-IT-PROD will have approximately 6000 cores and 12000 TB of storage capacity. The increase in resources is shown in Fig. 4.

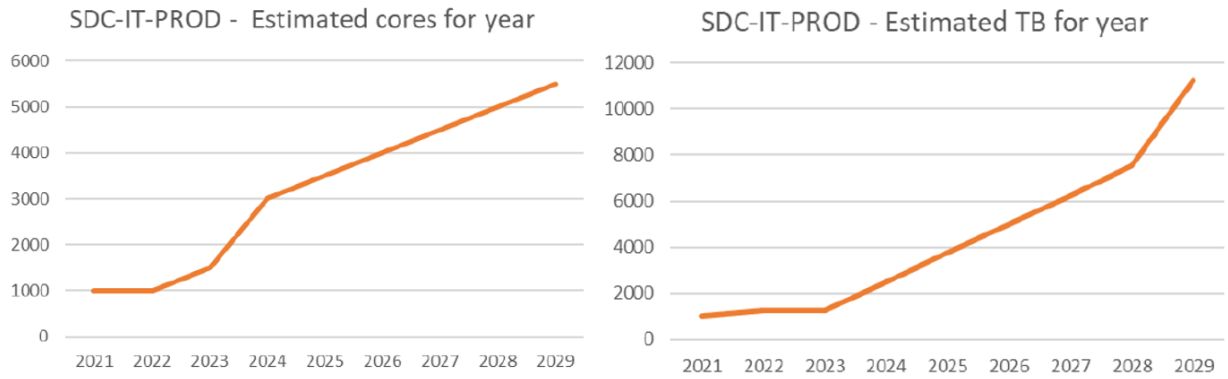

Fig. 4. SDC-IT-PROD resources evolution

The SDC-IT-PROD end users can access the data stored inside the data processing centre and the software necessary to perform analysis and check through a dedicated Virtual Private Network, using:

- **Access Nodes**: two HPE ProLiant DL385 Gen10 Plus with Dual AMD EPYC 7313 (3.0GHz 16-core 155W) and 128 GB of RAM configured with data access and analysis tool.
- **IOT Workstations**: two HPE ProLiant DL385 Gen10 Plus servers with Dual AMD EPYC 7413 (2.65 GHz 24-core 155W) and 256 GB of RAM configured to host IOT tools.

The SDC-IT-PROD operational team can manage the data processing centre using two workstations installed within the ALTEC MultiMission Science Data Centre [8] visible in Fig. 5.

Fig. 5. ALTEC S.p.A. MultiMission Science Data Centre

2.4 SDC-IT-PROD Validation

The SDC-IT-PROD validation process is performed at two different levels. The validation of the two HPC processing infrastructures is carried out initially, when each component is verified as a standalone element, while the integration of all components is checked only at the end.

The SDC-IT-PROD System Validation Campaign begins once the two infrastructures have been declared ready to operate. At this stage, the integration with all SGS entities is verified. This includes checking the connection with other SDCs, the ability to execute all Euclid scientific pipelines and storing the data produced within both the SDC local storages and the Euclid Archive System. The accessibility of the SDC-IT-PROD elements, designed and configured for end-users to access and inspect data, is also monitored during this phase.

The goal of the System Validation Campaign is to declare the data processing centre ready to support the Euclid mission. System testing is conducted whenever there is a major upgrade to any of the systems involved in the data processing centre, or upon the completion of each infrastructure expansion phase. This ensures that all updates and expansions are thoroughly validated to maintain the integrity, performance, and reliability of the entire infrastructure.

The first major System Validation Campaign was conducted prior to the satellite launch and another one was performed at the end of the first infrastructure expansion completed at the end of 2024 while the mission was running.

2.5 SDC-IT-PROD Operations

An operational team was established prior to the satellite launch to define the primary objectives and clearly identify all the interfaces required to ensure the operability of the Science Data Centre. Once the main goals were identified, the team was responsible for selecting all the tools needed to meet these objectives. These tools includes the mission log, an issue tracking system, dashboard for monitoring infrastructure performances and system trends, alert systems designed to quickly highlight any issues and a procedure handbook. During this period, all procedures for operating the data processing centre were documented and training sessions were conducted to ensure familiarity with the operational tools.

Following the satellite launch, the SDC-IT-PROD entered its operational phase. The operational team took responsibility for monitoring the status of the data processing centre on a daily basis and publishing this status in the mission log. This process allows the team to promptly identify potential issues and take the necessary corrective actions. The data processing infrastructure is also monitored throughout the workday to assess system performance, enhance understanding of the components in operation and promptly detect any emerging issues. The elements monitored by the operational team, highlighted in Fig. 6, include the components that make up the infrastructure, the software used for data exchange and processing, the tools that are specific to the IOT team and the scientific processing conducted within the Science Data Centre. On a daily basis, the team is also responsible for running the Data Quality Check software specific to certain Euclid pipelines (NIR, SIR and MER). The execution of this task generates reports that are inspected by the scientific team to provide feedback at SGS level on the quality of the data produced along the mission.

The operational team uses the SGS planning tools to stay updated on the mission schedule. This approach enables them to plan the data processing centre maintenance and evolution activities minimizing the impact on the mission schedule by avoiding downtime during critical data processing periods, such as when Data Releases are produced.

After the first year of the mission, the data processing centre operations entered into a new phase aimed at refining the execution of daily tasks through automation and reducing the time required to complete all activities. The tools used and the alerts configured within the infrastructures have been further perfected based on the experience gained throughout the mission and will be improved along the entire satellite lifecycle.

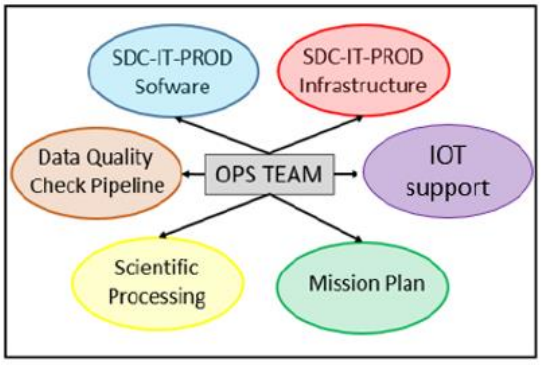

Fig. 6. SDC-IT-PROD OPS team Goal

3. Conclusion

The success of the Euclid mission is strongly dependent on the capabilities of the SDC-IT-PROD, which plays a pivotal role in processing, storing and analysing the mission's data. Equipped with advanced High-Performance Computing (HPC) infrastructure, SDC-IT-PROD contributes 25% of the total data processing and storage resources, ensuring the efficient handling of the substantial data volumes generated by the mission.

The operations of the Science Data Centre (SDC) are essential to ensuring the continuous functionality and efficiency of the data processing centre. The operational team is responsible for daily monitoring of the infrastructure, overseeing the execution of scientific pipelines and ensuring the quality of the data produced. Regular updates and improvements to the tools used, based on real-time mission experience, enhance the overall efficiency of the centre.

SDC-IT-PROD's scalability and flexibility in adapting to the mission's evolving needs are crucial to sustaining ongoing data processing activities. Its integration with other SDCs, along with its ability to execute complex data processing workflows while ensuring data integrity, has been crucial in maintaining mission continuity. The operational team's proactive approach to system maintenance and evolution, aligned with the mission schedule, along with their continuous efforts to improve data processing operations, are key to the success of the Science Data Centre.

In conclusion, the operational efficiency and continuous enhancement of SDC-IT-PROD are fundamental to the success of the Euclid mission, ensuring that its objectives are met.

Acknowledgements

This SDC-IT-PROD has been developed in the frame of the Italian Space Agency contract 2024-90-I.0 “Attività industriali per la fase operativa del Science Data Centre italiano della missione Euclid”.

SDC-IT-PROD team gratefully acknowledges ASI and INAF for the constructive cooperation and the support in setting up this project.

References

[1] Racca, G., et al., “The Euclid mission design” Proc. SPIE 9904 (2016), in press
[2] Pasian, F., et al., “Science Ground Segment for the ESA Euclid mission”, Proc. SPIE 8451 (2012).
[3] Belikov, A.N, et al., “The Role of the EUCLID Archive System in Distributed Data Management & Processing”, (2017).
[4] Melchior, M., et al., “IAL Overview & Pilots”, (2019).
[5] “HPE Data Management Framework technical white paper”, hpe.com[…]a50001461enw, accessed June 10th, 2023.
[6] “Introduction to Lustre Architecture”, wiki.lustre.org[…]/LustreArchitecture-v4, accessed June 9th, 2023.
[7] “xCAT Overview” https://xcat-docs.readthedocs.io/en/stable/overview/index.html, accessed October 4th, 2023
[8] R. Messineo, A. G. Villa, M. Martino, M. Del Giudice, F. Solitro, “MULTI-MISSION MSC & SDC: SHARED INFRASTRUCTURES, FRAMEWORKS AND FACILITIES FOR GROUND SEGMENT”. IAC 2024 Conference in Milan. 2024.