



Euclid Science Ground Segment (SGS) Processing Operations Concept



Maurice Poncet,¹ Christophe Dabin.²
Centre National d'Etudes Spatiales (CNES), Toulouse, France

John Hoar³
European Space Agency (ESA), Madrid, Spain

Andrea Zacchei⁴
National Institute of astrophysics (INAF), Trieste, Italy

and
Marc Sauvage⁵
Atomic Energy Commission (CEA), Saclay, France

On behalf of the Euclid Science Ground Segment Team

Euclid is a high precision survey mission developed in the framework of the Cosmic Vision Program of ESA in order to study the Dark Energy and the Dark Matter of the Universe. This project is endorsed both by ESA and by a European + US Scientific Euclid Consortium (EC). Euclid Science Ground Segment (SGS) will have to deal with around 175 Petabytes (PB) of data coming from Euclid satellite raw data, complex processing pipeline, external ground based observations or even simulations. It will produce namely an output catalog containing the description of more than 10 billion of sources with hundreds of attributes each. Thus the design, implementation, operation and maintenance of the SGS are a real challenge at least in terms of organization and architecture. This paper addresses the following topics: project & mission, processing pipeline, operational constraints, architecture solutions, development process, main benefits & concerns and next steps.

I. Nomenclature

N/A

II. Project & Mission

Euclid [1] is a high-precision survey space mission designed to answer fundamental questions on Dark Energy and Dark Matter. Euclid satellite will map the large-scale structure of the Universe over the entire extragalactic sky out to a redshift of 2 (about 10 billion years ago). It will then cover the period over which dark energy accelerated the universe expansion. This will be accomplished thanks to a payload which consists of a 1.2 m telescope, a visible imaging instrument (VIS) and a near-infrared instrument (NISP) with photometric and spectrometric capabilities. The satellite is planned to be launched end of 2020 on a Soyuz rocket from Kourou spaceport for a six years nominal mission. The spacecraft will be placed in a large second Sun-Earth Lagrange point (SEL2) halo orbit.

¹ Euclid SDC-FR Lead & SGS Common Tools Lead, Science Ground Segment department.

² Euclid Consortium SGS System Team Lead, Science Ground Segment department.

³ Euclid SOC Development Lead, European Space Astronomy Centre (ESAC).

⁴ Euclid Consortium SGS Project Office Lead, Osservatorio Astronomico di Trieste (OAT).

⁵ Euclid Consortium SGS Scientist, Institute of Research into the Fundamental Laws of the Universe (Irfu).

During the Euclid mission, the extragalactic sky survey operated by the Euclid satellite will cover 15 000 deg² and accurate photometric information for about $1.5 \cdot 10^{10}$ galaxies and spectra for about $3.0 \cdot 10^7$ of them, while the deep survey will cover 40 deg² around ecliptic poles and accurate photometric information for several tens of million galaxies and spectra for several tens of thousands of them. The wide survey will imply to take 45,200 acquisitions. The Euclid processing pipeline will then have to process visible images (~2 GB each), near-infrared images (~130 MB each) and near-infrared spectrum (~130 MB each) coming from the satellite instruments telemetry and reference data coming from ground based observations and will, in particular, generate a catalog of galaxies. See Fig. 1 – Euclid Mission at a glance.

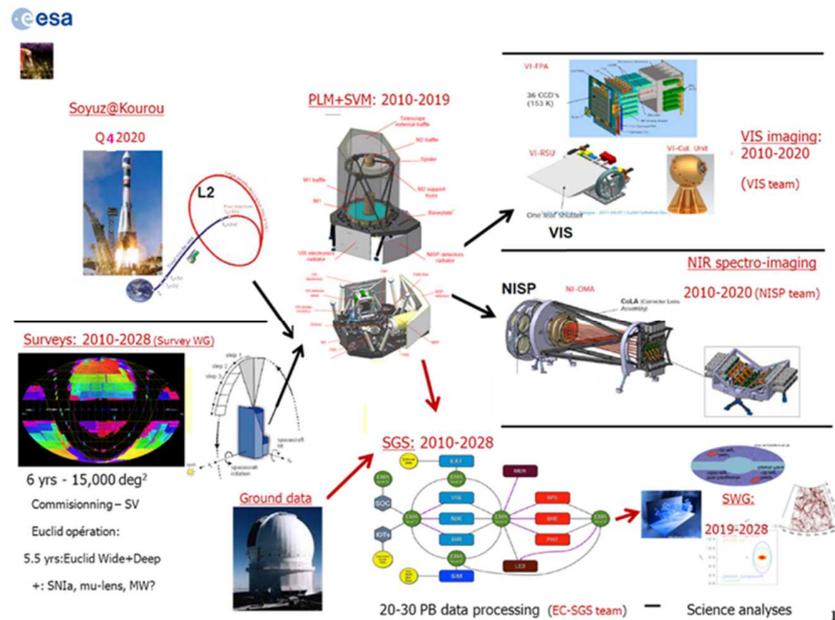


Fig. 1 Euclid Mission at a glance.

This process will be achieved through a workflow of 11 Processing Functions (PF) – see Fig. 2 - from Level1 (raw telemetry processing) up to Level3 (enhanced catalogs):

- VIS: Production of fully calibrated images from the VIS instrument, and astrometric reference,
- NIR: Production of fully calibrated photometric images from the NISP instrument, and pre-processed spectroscopic exposures,
- EXT: Production of fully calibrated images in bands covering the VIS band as well as ancillary data for spectroscopic and shear measurement calibration,
- MER: Construction of the object catalog and photometric measurements for all detected sources,
- SIR: Production of 1D spectra for all sources in the catalog,
- SPE: Measurement of spectroscopic redshifts, with quality information, identification of lines and flux measurements,
- PHZ: Derivation of photometric redshifts for all sources, with quality information, production of physical parameters,
- SHE: Measurement of shape parameters for weak-lensing sources. Definition and maintenance of the VIS PSF model,
- LE3: Production of all products related to the weak-lensing and galaxy clustering probes, production of other high-level science products,
- SIM: Production of all the simulated data needed either to develop and test the SGS pipeline, to calibrate observational or method biases, or to validate science results.

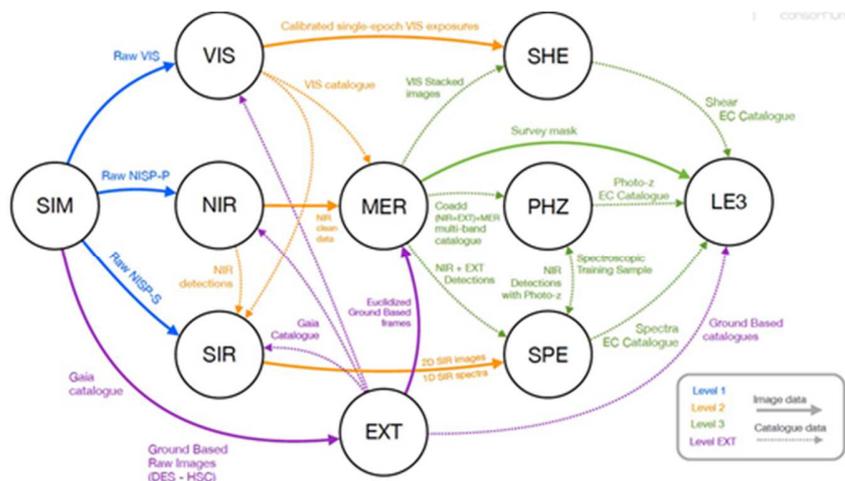


Fig. 2 Euclid Processing Pipeline.

The plan is then to deliver 3 Data Releases (DR) to the scientific community every 2 years after the nominal start of the mission, covering respectively: DR1 $\sim 2,500 \text{ deg}^2$, DR2 $\sim 7,500 \text{ deg}^2$ and DR3 $15,000 \text{ deg}^2$. In addition to that, we have also to consider the heavy simulations and reprocessing that will be needed in order to improve the processing algorithm and to tune the processing parameters and data calibration.

Having begun with an overall description of the Euclid mission, the paper will then focus more on the ground segment development challenges and the associated design approach.

III. Euclid Ground Segment

The Euclid ground segment, as shown in Fig. 3, is divided into the Operation Ground Segment (OGS) and the Science Ground Segment (SGS). The SGS [2] consists of a Science Operation Centre (SOC) and of several national Science Data Centers (SDCs). Both OGC and SOC are under ESA responsibility while the SDCs are endorsed by the Euclid Consortium (EC) – constituted by national space agencies and scientific institutes and labs. The SDCs are in charge of the science data processing (Processing Functions) and simulation software development, the science data processing operations and the generation of the data products. Eight SDCs over Europe and one in the US have already been planned, for which these responsibilities will be allocated or distributed.

At science side, there are 10 transnational Organization Units (OUs) covering the different processing steps and thematic. These OUs are mainly devoted to algorithm specifications and selection, prototyping and corresponding pipelines validation.

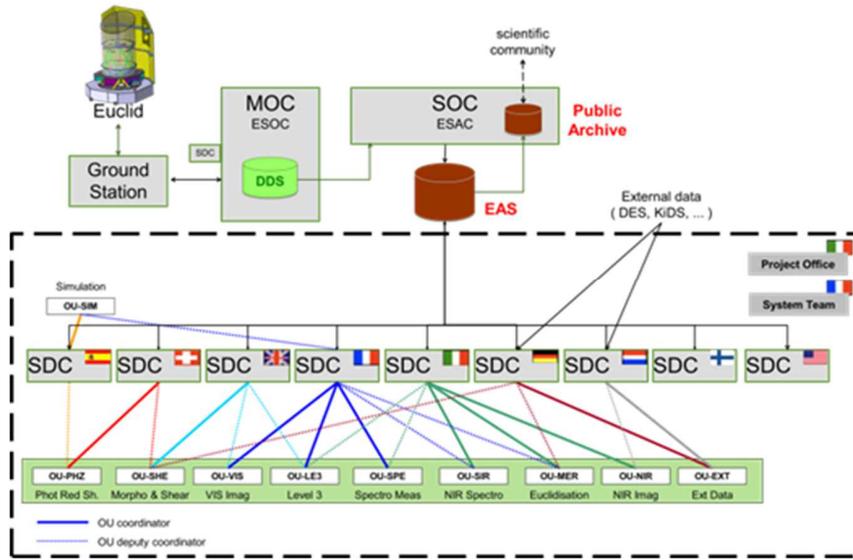


Fig. 3 Euclid Ground Segment.

During its lifetime, the SGS will have to handle an unprecedented volume of data for such a space mission: about 850 Gbit of compressed data will be downlinked every day. To properly perform the multi-wavelength photometric measurements, the Euclid data need to be complemented by ground-based data: this means that a total amount of at least 13 PBytes will be processed, transferred and stored for a given release of scientific data.

We expect to produce and store at least 175 PB of data in total - see Fig. 4 - and to provide a catalog containing the description of more than 10 billion sources with hundreds of columns.

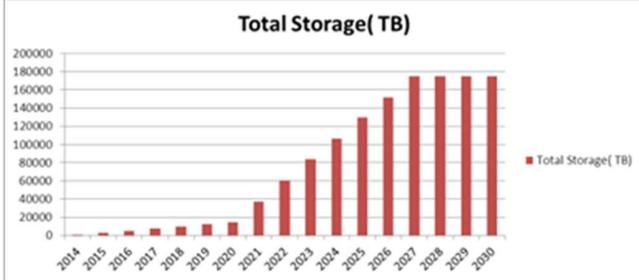


Fig. 4 Estimation of total amount of storage.

According to the current processing needs estimation, the overall Euclid data processing will require up to 22,500 cores, as shown in Fig. 5.

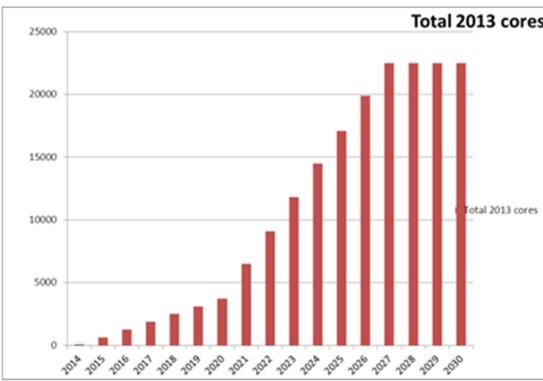


Fig. 5 Estimation of processing cores needs.

IV. Euclid SGS Architecture

The Euclid SGS development is therefore a real challenge [3] both in terms of architecture design (storage, network, processing infrastructure) and of organization. To achieve this challenge, 9 Euclid SDCs will have to be federated, ensuring an optimized data storage and processing distribution and providing sufficient networking interconnection and bandwidth. In terms of organization, 16 countries and 220 laboratories/institutes are already involved in the project and ~1500 non-necessarily co-located people are working together on scientific, engineering and IT aspects.

A. Euclid SGS Reference Architecture

In particular, the reference architecture [4], currently proposed for the SGS, is based on a set of services which allows a low coupling between SGS components: e.g. metadata query and access, data localization and transfer, data storage and data processing orchestration and M&C:

- A Euclid Archive System (EAS), composed of a single metadata repository which inventories, indexes and localizes the huge amount of distributed data,
- A Distributed Storage System (DSS) providing a unified view of the SDCs' distributed storage and managing the data transfers between SDCs,
- A Monitoring & Control Service allowing to monitor the status of the SGS as a whole or at SDC level,
- A COmmon ORchestration System (COORS) managing the distribution of the storage and the processing among the different SDCs (ensuring the best compromise between data availability and data transfers) following the paradigm: "move the process not the data",
- An Infrastructure Abstraction Layer (IAL) allowing the data processing software to run on any SDC independently of the underlying IT infrastructure, and simplifying the development of the processing software itself. It shows generic interfaces to the processing software and isolates it from the "the nuts and bolts" of the system (e.g. it gathers input data, publishes output data on behalf of the processing S/W),
- A common Data Model (CDM) shared by all software. This CDM relies on a common dictionary of data types from the simplest to the most complex. This CDM is the remit for the definition and implementation of any Euclid data product. The implementation of the EAS is based on the CDM.

B. Current Euclid SGS infrastructure

The current SGS infrastructure is depicted in the Fig. 6 below.

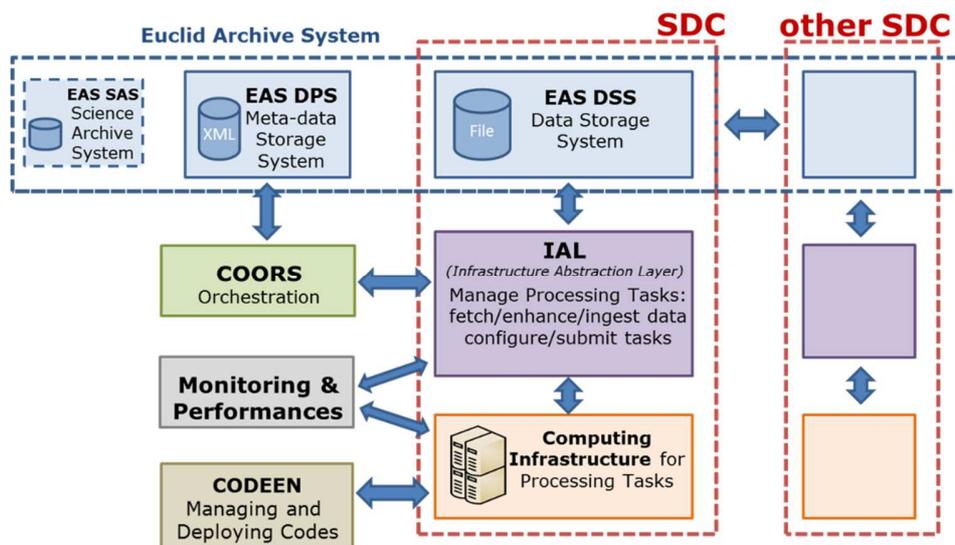


Fig. 6 Current SGS architecture infrastructure.

The current EAS instance is being developed and hosted by the Dutch SDC. It is based on a relational database that implements the Euclid common data model. This database stores the queryable metadata and the corresponding data location among the SDCs. It offers namely the following services:

- Metadata ingestion,
- Metadata browsing,
- Metadata and data location query,
- Cutout services.

During operations, the main EAS instance will reside at ESAC with a mirror at SDC-NL.

The current DSS instance is being developed by the Dutch SDC. This DSS prototype relies on DSS servers installed at each SDC. These servers drive the corresponding SDC local storage and ensure requested data transfers between the SDCs. A DSS server implements a set of basic file based operations: store, retrieve, copy and delete. At the moment, it supports sftp and Grid Storage Element protocols. Other protocols may be added by simply providing the corresponding driver that complies with the corresponding DSS interfaces and plugs into the DSS server.

The current Monitoring instance covers both inter-SDCs network bandwidth and SDC local resources monitoring. The network monitoring relies on iperf while the resources monitoring relies on Icinga. For both a dynamic kibana dashboard is available. This Monitoring Prototype is handled by the German SDC.

A COORS prototype is under development at the English SDC. It will allow to formerly define Processing Plans and to dispatch them on the SDCs through the IAL prototype as Data Distribution Orders and Pipeline Processing Orders.

The current IAL instance is being develop at the Fachhochschule Nordwestschweiz (FHNW) in Switzerland. According to the formal description of a given Processing Function, and a corresponding Processing Order, it takes care of:

- Creating a working directory,
- Fetching locally all the input data described in the Processing Order, through the EAS and the DSS,
- Launching the processing steps through the SDC infrastructure batch manager, as per the Processing Function workflow description (including sequences, parallel split, join),
- Monitoring the processing logs and exit status,
- Storing the output data at the SDC local storage from the working directory,
- Submitting to the EAS the corresponding metadata for ingestion,
- Removing the Working Directory.

C. Euclid SGS Challenges

This architecture concept has already been partially validated through "SGS Challenges". The SGS challenges are a kind of real scale proof of concept through prototypes. Thus, the first challenges namely allowed to distribute and to execute preliminary simulation prototypes on any of the SDCs thanks to a first version IAL and EAS prototypes. This challenge approach allows deploying working prototypes at early stages and is a great factor of motivation for the teams disseminated among different laboratories and Computing Centers around Europe and in the US.

Moreover, the challenges are not standalone and ephemeral but allow an iterative and incremental approach: each challenge stays in place and running after its achievement; for the benefits of the following ones that can rely on it. These challenges will also be the mechanism through which the whole SGS will be incrementally integrated, verified and validated.

Consequently, the Euclid SGS integration, verification and validation plan [13] is driven by the sequence of alternatively infrastructure (ITn) and scientific (SCn) challenges [5]. Infrastructure challenges are more devoted to the integration or enhancement of SGS technical components, while Scientific challenges are focused on Processing Functions integration and/or improvement.

The first SGS Infrastructure challenge (IT1), driven by the German SDC, has been held in 2012 and its objective was to put in place a network bandwidth monitoring between the SDCs. This monitoring is based on iperf and a centralized dashboard is now available. It helps to detect bottlenecks and to investigate in case of failure.

The second SGS Infrastructure challenge (IT2), driven by the French SDC, has been held during the first half of 2013. Its main objective was to check the feasibility of a first level of federation of the SDCs. It has been achieved through the deployment and the launch of the NISP simulator prototype on SDCs local processing nodes (centOS, SL and Suse) from a central point. This has been achieved thanks to the Euclid continuous integration and deployment platform (CODEEN) based on the Jenkins engine and on the deployment of Jenkins slaves at the SDCs.

The third SGS Infrastructure challenge (IT3), driven by the French SDC, has been held in 2013/2014. It aimed to go a step further by deploying a VM embedding the IAL agent prototype itself on any SDC and to allow it to fetch input data from the EAS prototype (hosted in SDC-NL), to launch Euclid VIS and NISP instruments simulation prototypes on the local infrastructure, to store the output product at the SDC local storage area and to send to corresponding metadata to the EAS prototype in order to ingest it.

The next IT challenges (IT4 to IT7) – from 2014 to 2017 - have been devoted to SGS infrastructure components (IAL, EAS, DSS, M&C, CODEEN) improvement, as well as scalability and performance tests.

The first Scientific Challenge (SC1) has been held in 2014/2015, led by the Spanish team. SC1 aimed at a simulation of the Euclid satellite VIS and NISP instruments outputs on a restricted area of the sky (~ 20 deg²).

The Scientific Challenge #2 (SC2) has been held in 2016 and was driven by the Italian SDC. The main objectives of SC2 was to design and develop the first prototypes of the VIS, NIR and SIR Processing Functions, to integrate them with the SGS infrastructure components (Data Model, IAL), to deploy each PF on any SDC and to process the VIS and NISP raw frames simulated by the SIM PF and archived in the EAS. A more detailed list of SC2 goals is the following:

- a preliminary validation of the scientific output of each PF prototype.
- SIM to generate a validation test set of simulated Level 1 raw frames (science and calibration frames) for VIS, NIR and SIR and ingest them in the Euclid Archive System. Few square degrees are considered sufficient for the challenge purposes.
- A verification/validation of the SIM simulated instrument systematic effects and background noise necessary to test the processing elements covered by each PF.
- A performance evaluation of each PF prototype, to identify the minimum processing and storage requirements on the SDCs infrastructure
- Test of the interfaces between PFs and the SGS components involved in the challenge. They include:
 - Data Model definitions provided as output of the SIM PF and expected in input to the VIS, NIR and SIR PFs. Data Model definitions of the VIS, NIR and SIR outputs
 - a consolidation of the Mission Data Base (MDB) information content
 - a common source catalogue format definition
- Check requirements coverage of the SGS infrastructure used in the challenge

The Scientific Challenge #3 (SC3) has been held in 2017 and was led by French and Italian teams. Its main objectives were to improve already integrated PFs (SIM, VIS, NIR and SIR) and to integrate and validate a first version of the EXT and MER pipelines. This implied namely the following activities:

- To design and implement the first prototypes of the MER and EXT (EXT-DES and EXT-KIDS) Processing Functions (PFs)
- To integrate them with the SGS Infrastructure components (Euclid Archive System, IAL) and with the upstream PF's: VIS/NIR/SIR/EXT KIDS/EXT DES sourced with SIM data products
- To deploy each PF on every SDC Prod through the CODEEN system with the CernVM-FS framework
- To run these PFs with input simulated data products: VIS/NIR/SIR/EXT on the various SDCs: NL/DE/FR/SP/CH/US/UK/IT/FI
- To analyze / check the output data products of MER
- To validate the scientific requirements
- To store the output data in the EAS

The current ongoing challenge in 2018 is the Scientific Challenge #456 (SC456). It is led by the French team. The main goals of SC456 are to improve the PFs already involved during the previous challenges and to integrate the downstream PFs: SHE (SC4), PHZ (SC5) and SPE (SC6). It will also introduce a new EXT observatory with LSST data simulations.

The next Scientific Challenges will be:

- SC7 in 2019: Update pipeline releases to meet more consistent requirements coverage w.r.t. the previous challenges and assembly the whole Euclid pipeline (except LE3).
- SC8 in 2020: Introduction of LE3 PF, quality of LE3 data products shall be challenged according to the corresponding scientific requirements.

All of these challenges were almost successful and rich in lessons. They allowed to raised many issues during early stages and to fix most of them. They are also – but not least - a great factor of federation not only at a technical level but also at the teams' level.

V. Processing Functions Development

Another challenge of the Euclid project is how to organize the Processing Functions development at SGS level knowing that people and teams are distributed across Europe and US and have different “customs and practices” ?

The first answer coming for the Common Tools group of the SGS System Team has been to provide Euclid developers commonalities in terms of rules, tools and best practices.

A. Software Stack

The pipeline software stack is basically organized into a set of independent software layers that include applications and supporting libraries that are logically connected. The following figure 7 depicts the different components of the architecture.

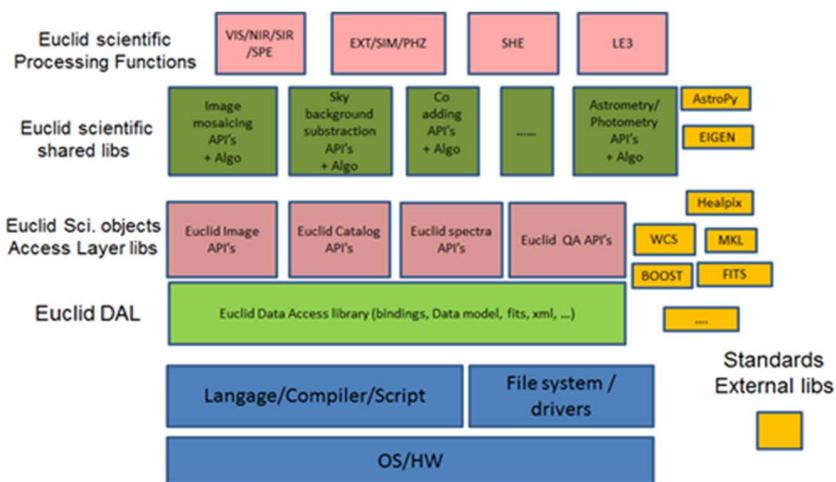


Fig. 7 SGS Software Stack.

B. Common Referential

First, a common referential, named EDEN (Euclid Development Environment), has been established such as all development would be held using the same restricted and supported set of O/S (CentOS7), languages (C++/Python), libraries and tools.

As the developers are using different kind of laptops (e.g. Windows, Linux, Mac) and thus it was difficult to support different development flavors, it has been then decided to provide an EDEN compliant VM for local development. This turnkey VM called LODEEN (Local Development Environment) has been adopted in the community and is maintained by the Common Tools team. Moreover, a Dockerized version of LODEEN will be soon available and will be a lightweight solution with no virtualization overhead.

C. Continuous Integration & Delivery

Once the local development issue covered, the need is then to gather the source code centrally and to facilitate its configuration management, the build of binaries, the generation of documentation, the tests, the quality check, the packaging and the delivery.

This has been solved by providing a unique configuration management tool (the Euclid Consortium GitLab) and a central continuous integration platform. This EDEN compliant platform named CODEEN (Collaborative Development Environment) is hosted by the French SDC – with a failover mirror at the UK SDC - and is based on the following main components:

- o Jenkins as main engine
- o Jenkins slaves for scalability
- o SonarQube and C++/Python plugins for quality check
- o Doxygen/Sphinx for documentation generation
- o Pytest and Boost:test for unit and smoke tests
- o RPM as packaging format and tool
- o Nexus as RPMs repository

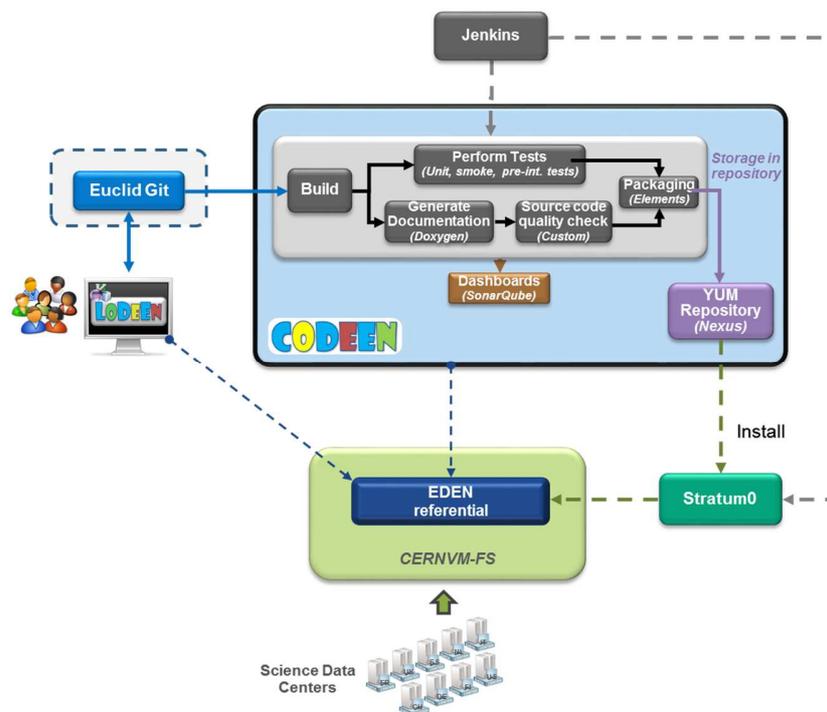


Fig. 8 Euclid SGS Development Environment.

Thus, this platform – see figure 8 - allows relying on continuous integration and delivery principles and best practices for the Euclid PF software development.

D. Continuous Deployment

The Nexus repository allows deploying any Euclid software with a single yum command including the automatic installation of the requested version of any dependency. It is convenient for local and single deployment, but it is not well suited to massively and frequently deploy new versions of Euclid software in the 9 SDCs since it requires the manual intervention of local system administrators and privileged access. Moreover, most of the Euclid SDCs rely on pre-existing shared cluster infrastructures and cannot easily install custom libraries in system spaces.

As a consequence, we had a look at different solutions in order to automate the deployment of Euclid pipeline inside the SDCs clusters. Most of them (e.g. Puppet, Chef, Ansible...) are too intrusive and raise some security issues.

We then studied the solution that has been developed and put in place at CERN for the LHC experiments called CernVM-FS.

The CernVM-File System (CernVM-FS) provides a scalable, reliable and low-maintenance software distribution service. It was developed to assist High Energy Physics (HEP) collaborations to deploy software on the worldwide-distributed computing infrastructure used to run data processing applications. CernVM-FS is implemented as a POSIX read-only file system in user space (a FUSE module). Files and directories are hosted on standard web servers and mounted in the universal namespace /cvmfs. Files and file metadata are downloaded on demand and aggressively cached. The Euclid cvmfs deployment architecture – refer to figure 9 – relies on the following components:

- A unique central repository, called stratum 0, on which the software and their dependencies are deployed (with yum command)
- A ring of servers – called stratum 1 – mirroring the stratum 0, for scalability. At the time being there is one stratum 1 at SDC-FR and one at SDC-UK (one is foreseen at the SDC-US).
- At each SDC, at least one Squid Proxy is installed, attached to the stratum 1 ring, acts as a local cache and serves the local cluster.
- At each of the local cluster processing nodes, a cvmfs client is installed, configured and connected to the local Squid Proxy

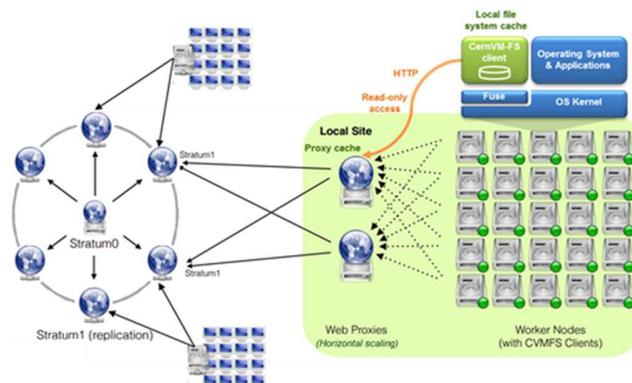


Fig. 9 Euclid SGS cvmfs at a glance.

This cvmfs deployment infrastructure is now in place on any of the 9 SDCs and has been successfully tested during the last SGS Architecture Challenge – see [9][10].

The next step has been to connect the continuous integration & delivery platform to the deployment one in order to be able to continuously deploy Euclid pipeline into the SGS.

As shown in figure 10, two branches of each software are deployed in such a way:

- The development branch in order to deploy current under development version of the code and to launch integration and validation test jobs on the target infrastructures,
- The master branch in order to deploy official tagged version of the code and to launch production jobs.

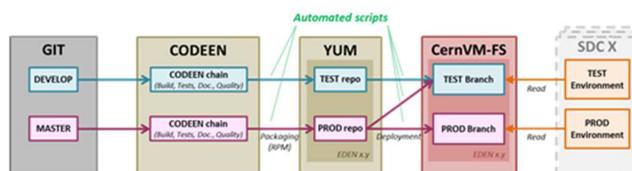


Fig. 10 SGS Software deployment on SDCs with cvmfs.

E. Run Time Environment

Another factor of potential complexity is the fact that most of the SDCs rely on existing Computing Centers that usually do not share the same infrastructure and operating system. Rather than having to setup, test and maintain

different targets for the Euclid software, the choice has been made to rely on virtualization in order to be able to deploy the same guest operating system and the same Euclid software distribution on any of the host operating systems and infrastructures of the SOC and the (currently) 9 Euclid SDCs. This virtual processing node image also called “EuclidVM” will simplify a lot both the development of the Euclid processing software and its deployment. At the time being, two kinds of VMs are available:

- A KVM instance,
- A μ CernVM instance [14].

SDCs have also the alternative to provide EDEN compliant physical processing nodes as they wish.

Another candidate technology that is also under investigation is the “Docker” platform [16]. Docker is an open-source project that automates the deployment of applications inside independent “software containers” on a single Linux instance. It avoids thus the overhead of running a hypervisor and of instantiating VMs. It is an extension of the Linux standard container (LXC).

A second alternative is to use Singularity containers that would be even simpler than Docker container and would not require privileges at all. Singularity containers are now more and more adopted in HPC/HTC centers. This approach is currently under study.

VI. Data & Processing Distribution Principles

In terms of Euclid Data & Processing distribution, the following approaches are adopted:

- Co-locating the data and processing in the same place avoiding large data transfers
- Use of the infrastructure abstraction layer (IAL) that isolates and decouples the SDC infrastructure from the rest of the SGS.

This is practically implemented through the concept of dividing the work to be done by each SDC according to sky patches, not Processing Functions; that is, it should be possible to run any Processing Function on any SDC computing facility and the data for each sky area should remain at the SDC responsible for that sky area.

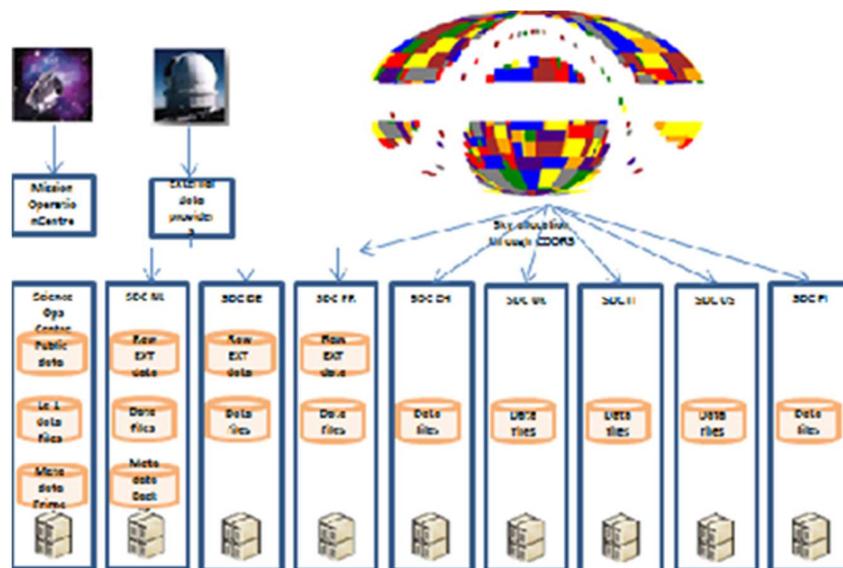


Fig. 11 SGS Observations Data Distribution.

Moreover, a major part of the EUCLID data processing pipeline [16] (up to Level 3 processing) is parallelizable at the granularity of a given observation. Hence, it has been possible to implement the “Big Data paradigm” as a multi-centers scale Map Reduce [8].

A. The Map Stage

First, each SDC is allocated areas of the survey (sky patches) for which it is responsible. The total area allocated to a SDC is defined as a function of the amount of resources available at that SDC since the size of the area of sky directly determines both the data volume output and the CPU resources necessary. A similar allocation can also be applied for simulation data generation of large sky areas. There is some overlap between patches for processing reasons. In order to minimize data transfers, the system should maximize the amount of processing that can be done on that sky patch with the least additional data. Once an area of the sky is allocated to an SDC, the selected SDC will be responsible for running the entire processing tasks over that area of sky (except Level 3) and for maintaining the output data files. A second copy will be maintained at another SDC for robustness. To maintain some degree of flexibility a second SDC can be assigned a sky patch.

B. The Reduce Stage

At Level-3 stage, the processing pipeline changes its shape. All the catalogued galaxies with the shear measurements, the spectra, the redshift and related attributes processed from single observations at level 2 are stored in the various SDCs. The relevant inputs needed for Level 3 are filtered out and retrieved to be processed by Level 3 processing functions in order to obtain high level scientific data products. This processing can be seen as a Reduce step (as per Map/Reduce paradigm) at a multi-center level. The typology of LE3 processing algorithm is different from previous stages since they require to compute 'all the sky available' in a single run being able to produce some: 2 points correlation functions, 3 points correlation functions, E mode auto correlation, E and B mode cross correlation, covariance matrix... for the various science domains that are: galaxy clustering, weak lensing, clusters of galaxies, time domain, milky way and nearby galaxies... The challenge is now to cope with High Performance Computing codes on a reduced set of input data required by LE3 algorithms.

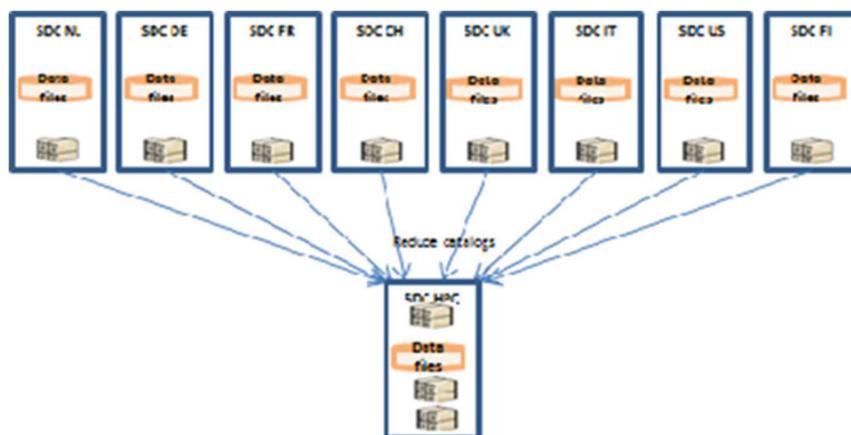


Fig. 12 SGS LE3 Reduce Stage.

VII. Conclusion

Euclid is a very challenging and innovative project that will involve a large European + US scientific consortium beyond the next decade. The draft architecture and general organization are now outlined.

The main concerns we encountered so far are among others the following:

- Huge effort required to develop and run simulation codes at early stages,
- Difficulties to define and stabilize the Common Data Model,
- Difficulties to integrate and stabilize both SGS components and PFs,
- Difficulties to provide helpful documentation,
- Some performances issues at IAL/DSS level.

While the main benefits of our approach are:

- Boost of the interaction of different stakeholders,
- Smoothly move from a prototype to an operational SGS,
- Early assessment of PF algorithms and SGS components functionalities,
- Early assessment of flexibility, maintainability and scalability of the different codes and SGS components.

The ongoing and forthcoming SGS challenges will help us to consolidate the pillars services of this architecture. They will allow progressively federating the different SDCs up to the operational SGS, for the benefit of the scientific community in Astrophysics.

Acknowledgments

The authors are indebted to all the individuals participating in the Euclid SGS development within ESA and the EC, too many to be listed here.

References

- [1] Laureijs, R., et al., “Euclid: ESAs mission to map the geometry of the dark universe”, Proc. SPIE 8442.
- [2] F. Pasian et al, “The Science Ground Segment for the ESA Euclid mission”, *SPIE*, 2012.
- [3] M. Poncet et al, “The challenge of Euclid Scientific Ground Segment Architecture”, GSAW, Los Angeles, 2012.
- [4] M. Poncet et al, “Euclid Scientific Ground Segment Organisation & Architecture”, ESAW, Darmstadt, 2013.
- [5] F. Pasian & al, “Development Plans for the Euclid Science Ground Segment”, ADASS XXIII, Hawaii, 2013.
- [6] M. Poncet & al, “Euclid: Big Data from Dark Space”, BiDS 2014, ESRIN, Frascati, IT
- [7] G Buenadicha & al, “Euclid: Big Data from Dark Space”, Big Data Spain 2015, Madrid, ES
- [8] C. Dabin & al, “Euclid: Orchastrating the Software Development and the Scientific Data Production in a Map Reduce Paradigm”, BiDS 2016, Tenerife, ES
- [9] Q. Le Boulc’h & al, “Using CernVM-FS to deploy Euclid processing software on Computing Centres”, CernVM Workshop, Rutherford Appleton Laboratory, 2016
- [10] M. Poncet & al, “Using CernVM-FS to deploy Euclid processing S/W on Science Data Centres”, ADASS 2016, Trieste, IT
- [11] M. Poncet & al, “Euclid: Big Data from Dark Space”, ESAW 2017, Darmstadt, DE
- [12] M. Poncet & al, “Euclid Science Ground Segment Operations Concept”, SciOps 2017, ESAC, Madrid, ES
- [13] C. Dabin & al, “Euclid Science Ground Segment Integration, Verification & Validation”, ADASS 2017, Santiago, CL
- [14] J Blomer et al, “Micro-CernVM: slashing the cost of building and deploying virtual machines”, http://iopscience.iop.org/1742-6596/513/3/032009/pdf/1742-6596_513_3_032009.pdf.
- [15] “Docker: Build, Ship and Run Any App, Anywhere”, <https://docker.com/>
- [16] M. Sauvage, “Data and Processing Flow document”, EUCL-OTS-ADD-8-002