Science ground segment for the ESA Euclid Mission

Fabio Pasian, John Hoar, Marc Sauvage, Christophe Dabin, Maurice Poncet, et al.

Fabio Pasian, John Hoar, Marc Sauvage, Christophe Dabin, Maurice Poncet, Oriana Mansutti, "Science ground segment for the ESA Euclid Mission," Proc. SPIE 8451, Software and Cyberinfrastructure for Astronomy II, 845104 (24 September 2012); doi: 10.1117/12.926026

Event: SPIE Astronomical Telescopes + Instrumentation, 2012, Amsterdam, Netherlands

Science Ground Segment for the ESA Euclid mission

Fabio Pasian*[a], John Hoar[b], Marc Sauvage[c], Christophe Dabin[d], Maurice Poncet[d], Oriana Mansutti[a]
[a]INAF - Osservatorio Astronomico di Trieste, Via G.B.Tiepolo 11, 34143 Trieste, Italy; [b]European Space Astronomy Center, Villanueva de la Cañada, E-28692 Madrid, Spain; [c]CEA, Laboratoire AIM, Irfu/SAp, Orme des Merisiers, F-91191 Gif-sur-Yvette, France; [d]Centre National d'Etudes Spatiales, 18 Avenue Edouard Belin, F-31401 Toulouse Cedex 9, France

ABSTRACT

The Scientific Ground Segment (SGS) of the ESA M2 Euclid mission, foreseen to be launched in the fourth quarter of 2019, is composed of the Science Operations Center (SOC) operated by ESA and a number of Science Data Centers (SDCs) in charge of data processing, provided by a Consortium of 14 European countries. Many individuals, scientists and engineers, are and will be involved in the SGS development and operations. The distributed nature of the data processing and of the collaborative software development, the data volume of the overall data set, and the needed accuracy of the results are the main challenges expected in the design and implementation of the Euclid SGS. In particular, the huge volume of data (not only Euclid data but also ground based data) to be processed in the SDCs will require a distributed storage to avoid data migration across SDCs. The leading principles driving the development of the SGS are expected to be the simplicity of system design, a component-based software engineering, virtualization, and a data-centric approach to the system architecture where quality control, a common data model and the persistence of the data model objects play a crucial role. ESA/SOC and the Euclid Consortium have developed, and are committed to maintain, a tight collaboration in order to design and develop a single, cost-efficient and truly integrated SGS.

Keywords: Euclid mission, dark energy, ground segment, instruments monitoring and control, data processing, software architecture

1. INTRODUCTION – THE EUCLID MISSION

Euclid is the second medium-sized (M2) mission of the ESA Cosmic Vision 2015-2025 Plan, aimed at understanding the nature of dark energy and dark matter by accurately measuring the accelerated expansion of the Universe. By measuring two probes (weak lensing and baryon acoustic oscillations) simultaneously, Euclid will constrain dark energy, general relativity, dark matter and the initial conditions of the Universe with unprecedented accuracy.

The mission will observe galaxies and clusters of galaxies out to z~2, in a wide extra-galactic survey covering 15000 deg^2, plus a deep survey covering an area of 40 deg^2. Besides the primary objectives of the mission, Euclid will also produce a massive legacy of deep images and spectra over at least half of the entire sky. This will be a unique resource for the astronomical community and will impact upon all areas of astronomy.

The launch is planned in the fourth quarter of 2019. The payload is composed of a 1.2 m Korsch telescope and two instruments, an imager in the visible domain (VIS) and an imager-spectrometer (NISP) covering the near-infrared. The launch vehicle will be a Soyuz, and the orbit will be located in L2, the second Sun-Earth Lagrange point. The spacecraft will roll around its vertical axis to keep the Sun direction as the normal of the solar array, within 1 degree to maintain thermal stability. The rotation along the second axis allows the spacecraft to yaw up to 360 degrees to observe a full strip of the sky. The rotation along the third axis allows to pitch backwards an angle of maximum 30 degrees to allow for corrections and a more complete coverage of the sky. The observing will follow a step-and-stare mode: the sky is covered by a set of strips that are re-composed as a mosaic. The scanning strategy allows to cover the foreseen 15000 deg^2 extragalactic survey in the 6-years duration of the mission

---

*fabio.pasian@inaf.it; phone 39 040 3199180; fax 39 040 309418; http://www.ts.astro.it/en

Software and Cyberinfrastructure for Astronomy II, edited by Nicole M. Radziwill, Gianluca Chiozzi,
Proc. of SPIE Vol. 8451, 845104 · © 2012 SPIE · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.926026

Weak lensing and baryon acoustic oscillations measurements require a very high level of accuracy. Weak gravitational lensing requires extremely high image quality because possible image distortions by the optical system must be suppressed or calibrated out to be able to measure the true distortions induced by gravity. On the other hand, the Euclid baryonic acoustic oscillations experiment involves the determination of the redshifts of galaxies to better than 0.1%, and this can only be accomplished through spectroscopy.

It is also to be noted that the broad-band Euclid data alone are not sufficient to achieve the required photometric redshift accuracy and precision, which means that additional ground-based data are required. The Euclid survey area (covering 15000 deg^2) needs to be imaged from the ground using at least 4 filters, covering at least the full wavelength range 420−930 nm, with an overlap between the filters less than ~10%. Collaborations are underway with ground-based surveys to obtain external data that comply with the required depth and wavelength coverage.

The Euclid science requirements were flowed down first to define the instrument characteristics and performances. But also data processing is a critical aspect of the mission. A specific set of detailed requirements was therefore imposed on the data processing facilities, to ensure that the accuracy and precision of the processing is appropriate to measure the faint features that will be observed.

The mission was selected in October 2011. Details on Euclid, its instruments and the survey are available in the Euclid Definition Study Report[1] and in several presentations within the SPIE Astronomical Telescopes + Instrumentation 2012 Conferences (e.g.[2,3,4,5]).

2. EUCLID DATA PRODUCTS

A whole set of data products are expected from the Euclid mission. They can be roughly divided in "levels" as follows.

Level 1 data: raw VIS and NISP images; processed housekeeping telemetry and associated ancillary information such as pointing history files.

Level 2 data: calibrated and co-added images from VIS and NISP, validated for cosmology analysis; PSF model and optical distortion maps; co-added spectra.

Level 3 data: catalogues (including redshift, ellipticity, shear, etc.); dark matter mass distribution; shear and galaxy correlation functions and covariance errors; additional science catalogues; ground based information which was used in the derivation of the data products.

Transients: transient events data products, including the derived transient category (e.g. supernova, solar system object, etc.) and brightness, target position and possible finding chart.

Level Q data: products defined so that they are suitable for most purposes in Astronomy, except for the main cosmological goals of the mission.

The release of the first year of Level 1, Level 2 and Level 3 data will occur 26 months after the start of the survey, and subsequently every year. Level Q is expected to contain "quick-release" data: the first release will occur 14 months after the start of the survey, and subsequently every year.

The Euclid data processing system is organized in sequential processing steps of increasing sophistication. With each step is associated a data processing level or "data level". Data levels consist of all data produced by the corresponding data processing step including intermediate data. Each data level has corresponding quality controls. To the products listed above (currently a tentative list) one should add also **Level E**, i.e. quality-controlled external data from existing missions and ground-based surveys which are used for calibrations and photometric redshift derivations, and **Level S** data, i.e. pre-launch simulations and modeling impacting on calibrations and observing strategies.

Data include not only processed data but also the quality control information associated with them. The quality control information ensures traceability of input data sets as well as the processing steps applied.

All intermediate and final data-set and associated quality control and processing information are stored into the Euclid Mission Archives (EMA). The EMA constitutes the "working" repository of the mission and is used for disseminating data within the Euclid collaboration. The Euclid Legacy Archive (ELA) will provide access to the final validated products to the general scientific community.

3. THE EUCLID GROUND SEGMENT

3.1 Structure

The Euclid Ground Segment is structured as in Figure 1.

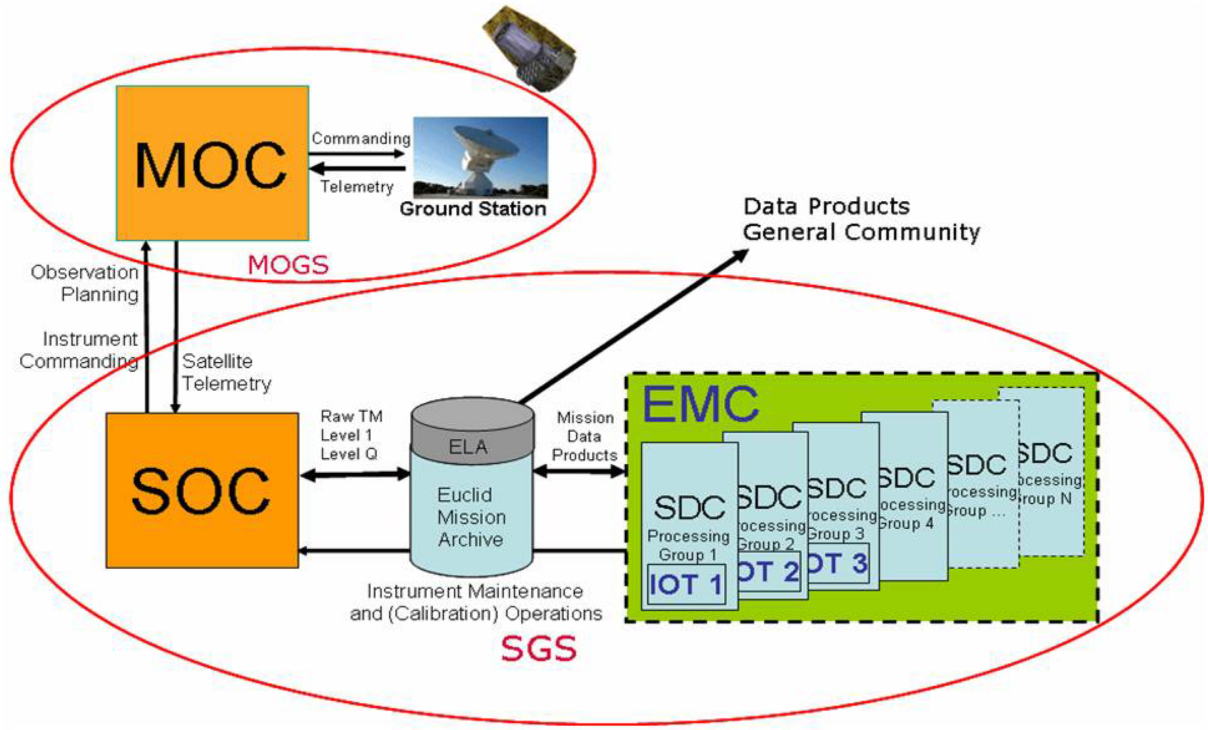

Figure 1. The Euclid Ground Segment. Explanations in the text.

The spacecraft, operating in L2, will be connected to one or two Ground Stations (operated by ESA) during Daily Tele-Communication Periods (DTCPs) of 4 hours each in which the telecommands will be uploaded and the telemetry downloaded from the spacecraft.

The Mission Operations Center (MOC) located at ESA's Space Operations Center (ESOC) monitors the spacecraft health and safety and the instrument safety, controls the spacecraft attitude, and handles telemetry and telecommands for spacecraft and instruments. MOC and Ground Station form the Mission Operations Ground Segment (MOGS) which is completely under the control of ESA.

The Science Operations Center (SOC) is located at ESA's Space Astronomy Center (ESAC) and acts as the single interface to MOC. It is the central node for the mission planning, executes the planned surveys, performs an initial quality check and prepares the daily quality reports. From the processing point of view, SOC implements Level 1 by preparing edited telemetry; it is furthermore in charge of running Level Q processing and of distributing to the scientific community the relevant quick-release data. Finally, SOC manages EMA and operates ELA.

The first of the Euclid Mission Consortium (EC) duties is to maintain the instruments, monitor their health, perform trend analysis, and produce weekly instrument reports: these tasks are going to be performed by dedicated Instrument Operations Teams (IOTs), in principle composed (after instruments delivery to ESA) of most of the scientists and engineers who had been involved in the development of the instruments themselves. The EC provides as well a number of Science Data Centers (SDCs), which provide different functions: instrument-oriented SDCs host the IOTs and are in charge of instrument calibration activities (Level 2 data processing); data processing SDCs perform science processing and create science-ready data products (Levels 2 and 3); finally, science support SDCs provide simulated data (Level S) or reprocessed external data (Level E). Quite naturally, an individual SDC can provide more than one of these functions.

SOC and the SDCs form the infrastructure of the Science Ground Segment (SGS).

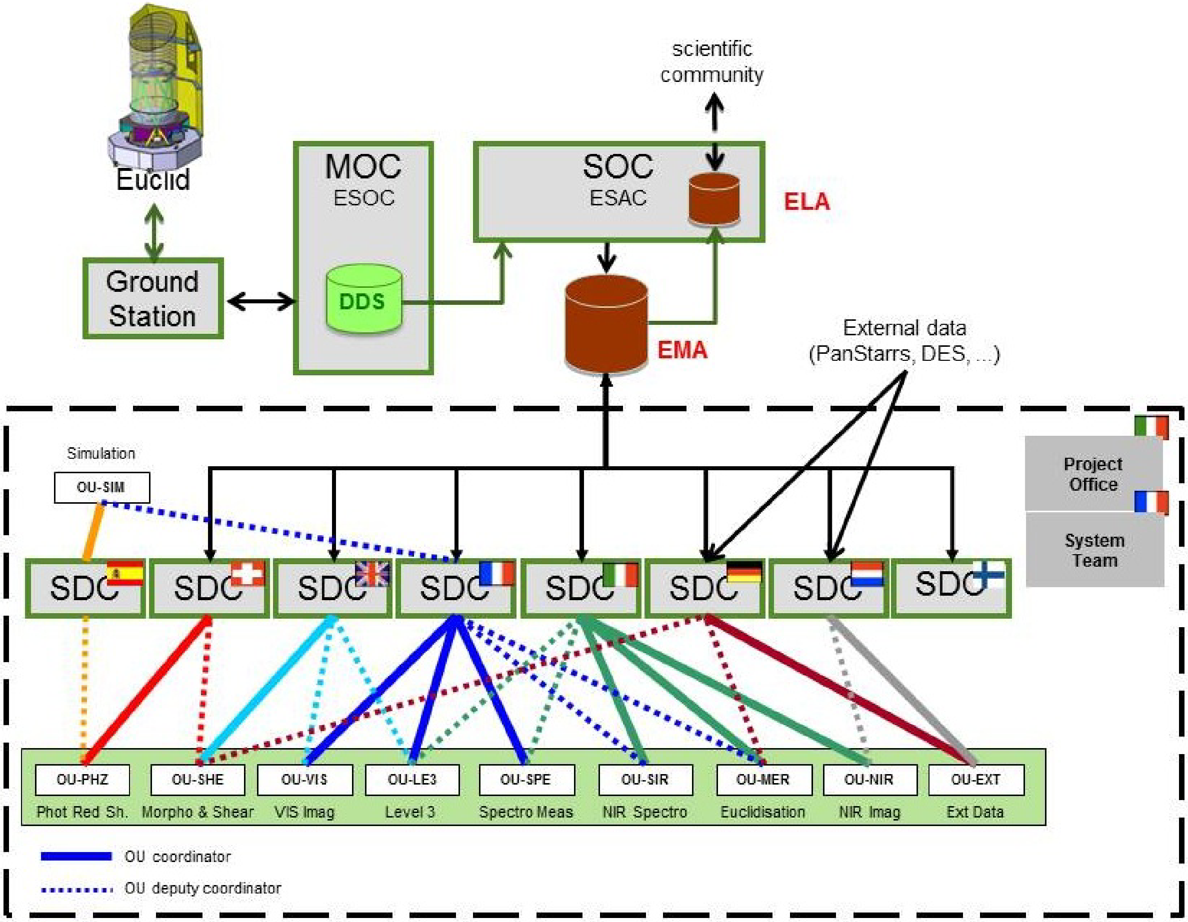

Figure 2. Schema describing the Euclid Ground Segment organization. The Organization Units (OU) are transnational and provide the national SDCs with the algorithmic definition of the processing to be implemented. The Science Data Centers (SDCs) implement and run the data processing pipelines, then the OUs validate the implementation. Additional SDCs may be added to the eight shown and currently foreseen. The EMA is built jointly by the EC and ESA, and is managed by SOC.

3.2 Mission Planning Concept

The main task of the Ground Segment is to correctly operate the mission. In Euclid the organization of mission planning, also derived from previous experiences such as the operations of the Planck/LFI mission[6], is organized as follows.

The Euclid Science Team (EST) through the Euclid Project Scientist (PS) provides the SOC with the survey strategy. On their side, the IOTs maintain a routine calibration plan which is delivered to the SOC for execution and submit unplanned calibration requests to SOC as observation requests.

The SOC gathers these inputs and implements the survey strategy: generates the long term plan, and derives a series of daily/weekly/monthly observation sequences. The MOC provides planning information, including a predicted orbit, the planned events at spacecraft and Ground Segment level, and tools/data to correctly plan the spacecraft pointing.

The IOTs receive back from SOC the Long-Term Plan for the survey and the executed History File containing the list of the actions performed by spacecraft and instruments.

3.3 Organization

A more detailed schema describing the Euclid Ground Segment organization is shown in Figure 2. The upper part of the diagram depicts the fact that the EMA is central to the Euclid SGS: it receives the original spacecraft and instrument telemetry and auxiliary data, and stores any intermediate data set, from edited telemetry to calibrated imaged and spectra, to catalogues and final products. The EMA is built jointly by the EC and ESA, and is managed by SOC.

The data distributed through the ELA is logically an EMA subset; this subset is a formal EC delivery to ESA and is distributed by SOC to the scientific community through mechanisms compliant with the international Virtual Observatory standards.

The lower part of the diagram shows how the development of the SGS occurs within the EC. The organization is based on the decomposition in transnational Organization Units (OU), each corresponding to a subset of overall Euclid Data Processing. Each OU produces algorithms which are integrated and executed in the SDCs, which are essentially tied to national locations and funding. Besides the eight national SDCs currently foreseen, there are other national contributions to the SGS in the form of participation to OU activities.

In other words, the Organization Units provide the algorithmic definition of the processing to be implemented by the SDCs and validate the implementation. The Science Data Centers implement and run the data processing pipelines as specified by the OUs, procuring the needed local hardware and software resources. SDCs carry out different activities: SDC-DEV (development – i.e. transforming algorithms into robust pipeline code) and SDC-PROD (production – i.e. integration on the local infrastructures, production runs of the pipelines).

The EC Science Working Groups (SWGs) do not belong to the SGS. However, their influence is quite strong, since they are in charge of turning science objectives into requirements placed on the pipeline products and performances, and of verifying that the requirements are met (basically, they define the Validation & Verification procedures).

It is understood that individual Euclid scientists may belong to more than one of the above groups (OUs, SDCs, SWGs).

3.4 EMA and ELA

Key features of Euclid are the amount of data that the mission will generate, the heavy processing needed from raw data to science products, and the accuracy and quality control required at every step.

Data are central for the SGS. The design of the SGS is therefore based on a *data-centric* approach: all SGS operations logically revolve around the Euclid Mission Archive (EMA), which is a logical, rather than physical, entity giving access to all mission-related analyses and a storage and inventory of the data products and their metadata including quality control. The orchestration of data exchange and metadata update involving SOC and SDCs through the EMA is performed by a monitoring and control function.

EMA is physically composed of distributed data sets and centralised metadata, which contain information on the location of the actual data files to allow easy retrieval. Proper integrity and security measures will be put into effect to prevent damage, loss of data or unauthorised access.

The Euclid Legacy Archive (ELA) is a public archive and is the unique distribution channel of Euclid data products to the scientific community.

The criteria for data availability in the ELA are defined by the EST and the EC, and are implemented in the EMA. After approval of the data products for public release, the relevant data shall be delivered to ESA for public distribution through the ELA.

3.5 Data processing functions

The flow of data processing functions is shown in Figure 3.

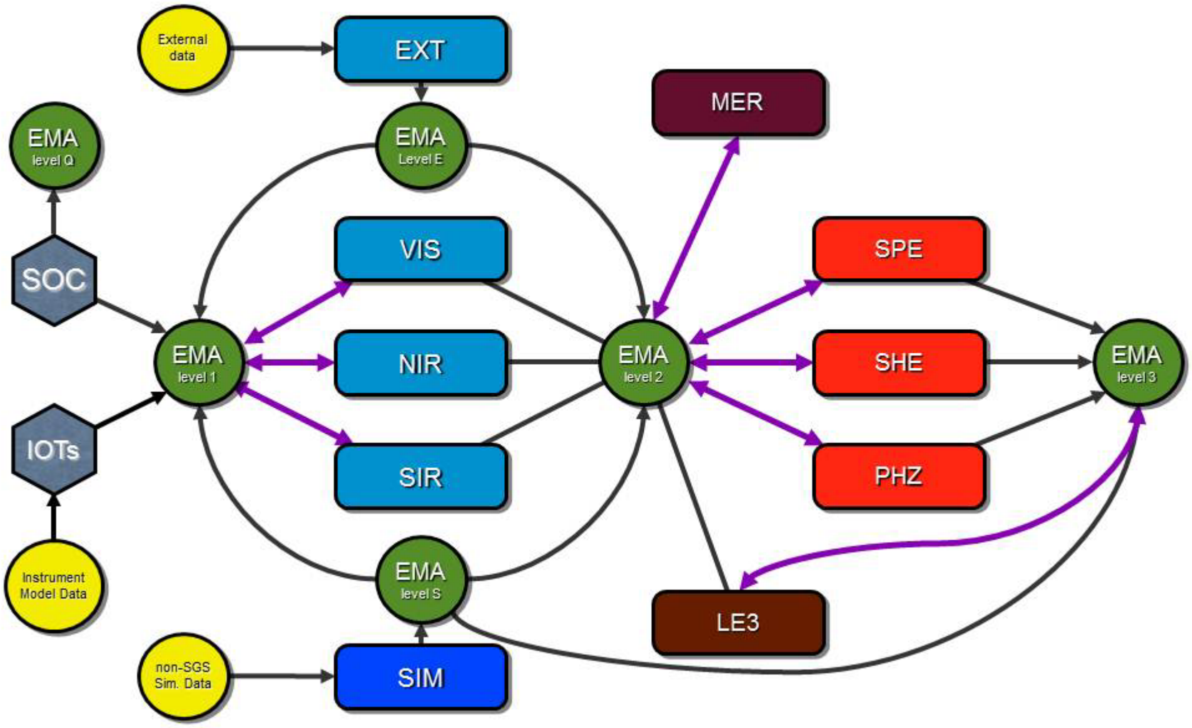

Figure 3. The flow of data processing functions in the Euclid SGS. The sections of the EMA in which data are ingested, or from where data are retrieved, is evidenced.

In the figure the different data processing levels (as defined above) are connected with logical data processing functions. These logical functions, or modules, are defined by considering that they represent self-contained units, i.e. they represent the highest-level break-down of the complete processing that can be achieved with units that communicate only with the help of the EMA (and in that respect they constitute indeed a first step into the realization of a distributed pipeline development). They are listed briefly hereafter.

VIS: is in charge or processing the Visible imaging data from edited telemetry to level 2, i.e. it produces fully calibrated images, as well as source lists (for quality check purposes only).

NIR: is in charge of processing the Near-Infrared imaging data from edited telemetry to level 2, i.e. it produces fully calibrated images as well as source lists (for quality check purposes and to allow spectra extraction).

SIR: is in charge of processing the Near-Infrared imaging data from edited telemetry to level 2, i.e. it produces fully calibrated spectral images and extracts the spectra in the slitless spectroscopic frames taken by the NISP.

EXT: is in charge of entering in the EMA all of the external data that will be needed to proceed with the Euclid science. This is essentially multi-wavelength data for photo-z estimation, but also spectroscopic data to validate the spectrometric redshift measurement tools.

SIM: implements the simulations needed to test, validate and qualify the whole set of pipelines.

MER: implements the merging of all the level 2 information. It is in charge of providing stacked images and source catalogues where all the multi-wavelength data (photometric and spectroscopic) are aggregated.

SPE: extracts spectroscopic redshifts from the level 2 spectra.

PHZ: computes photometric redshifts from the multi-wavelength imaging data.

SHE: computes shape measurements on the visible imaging data.

LE3: is in charge of computing all the high-level science data products, from the fully processed shape and redshift measurements (and any other possibly needed Euclid data).

During the Operations Phase, the Euclid SGS will produce the results of the processing functions defined above. Such data products, both intermediate and final, will be validated by the SWGs, stored in the EMA and documented, according to the schedule and the content defined by the Project Scientist (PS) and the Euclid Science Team.

4. SGS ARCHITECTURE

4.1 Drivers for the SGS

The drivers for the design and development of the SGS is led by a number of drivers, namely: instrumental, data, simulations**,** interfaces and data model, optimisation.

Instrumental drivers: in order to achieve an optimal data processing by the SGS, a close collaboration between the SGS and the instrument development teams (which will eventually become the core of the IOTs) needs to be established as soon as possible. An important subset of the IOTs will be made up with scientists experienced both in the instruments and in the systems used to process the data. Although it is unlikely that the SGS will be operational enough for the first instrument-level test campaigns, a goal is to gradually use the SGS systems to support these campaigns. This will bring the added benefit that the instrument test data are readily available in the EMA. Following the experience of GAIA, instrument parameters have been included into a common facility (the "Instrument Parameters Database") since the earliest phases of the instruments development. The parameters will evolve in time, from expected values useful for simulation purposes, to more stable ones when the performance of the instrument will be estimated, to real values when measures of the actual instruments are made on ground and in flight. This approach allows a smooth transition between the planning, development and operation phases of the instrument, and at the same time guarantees a tight connection with the SGS for what concerns the production of simulated data and the acquisition of real data, and the capability of the SGS pipelines to process both types of data in a satisfactory manner.

Data drivers: the core Euclid science cannot be achieved without ground-based survey projects. Agreements are being pursued with the KIDS, DES[8] and Pan-Starrs[9] surveys, and some discussion is occurring with LSST[10] as an additional possibility, to be able to integrate these data into the Euclid system (and the EMA). This activity is performed by the EXT processing function. The schedules of the aforementioned projects show that their data will start to be available before the launch of the Euclid mission, and therefore the activities of EXT will start right away. This will have the double benefit of spreading the computing needs over time - rather than having them compete for resources along with the processing of the Euclid data themselves - and provide input about which part of the complete SGS to train, e.g. the merging activity (MER), the derivation of photometric redshifts (PHZ), and other SGS functions.

Simulation drivers: Simulations will play a key role in Euclid science, in order to discriminate between an actual signal of interest and instrumental or data reduction artefacts. Simulations will also be at the heart of the SGS development, and so the activities of the OU-SIM team will be among the first to be started in the SGS. Indeed SIM integrates or interfaces with the simulation activities that are already taking place in the Instrument Development Teams, so as to provide the VIS, NIR and SIR OUs with input data. Higher level simulations are needed early on as well, so that the high-level OUs, such as PHZ, SPE, SHE or LE3, can soon start defining their activities and researching their methods. As for the possibility of producing also telemetry level simulations at an early stage from an instrument simulator comprising observational data, housekeeping data and auxiliary data, it is to be noted that telemetry will be completely defined only rather late and will be needed only for the purpose of testing the pipelines.

Interfaces and data model: a key item to implement an efficient SGS will be the correct and complete identification and efficient management of data interfaces: EC-SOC, but also between OUs, between OUs and SDCs, between OUs and Instrument Development/Operations Teams. It is clear, from the very beginning, that the SGS will need to manage the description of these interfaces through a single and consistent Data Model (DM) in an automated and fully electronic form.

Optimisation: a number of processing steps or functions (e.g. astrometry, PSF-homogenization, stacking, photometry, data quality checking, etc.) need to be performed within different instances and there is the obvious need to avoid duplication. On the other hand, processing functions may apparently need the same tools, but there might be subtle effects that might be detected by specific tools. There is a delicate balance to be kept between these two aspects. In particular, transversal (global) data quality tools (e.g. Data Quality Mining - DQM) do not overlap with other elementary quality checking steps. Global techniques, such as those based on machine learning and data mining methods, do not affect data themselves but complement quality masks and other quality information provided by each data processing step.

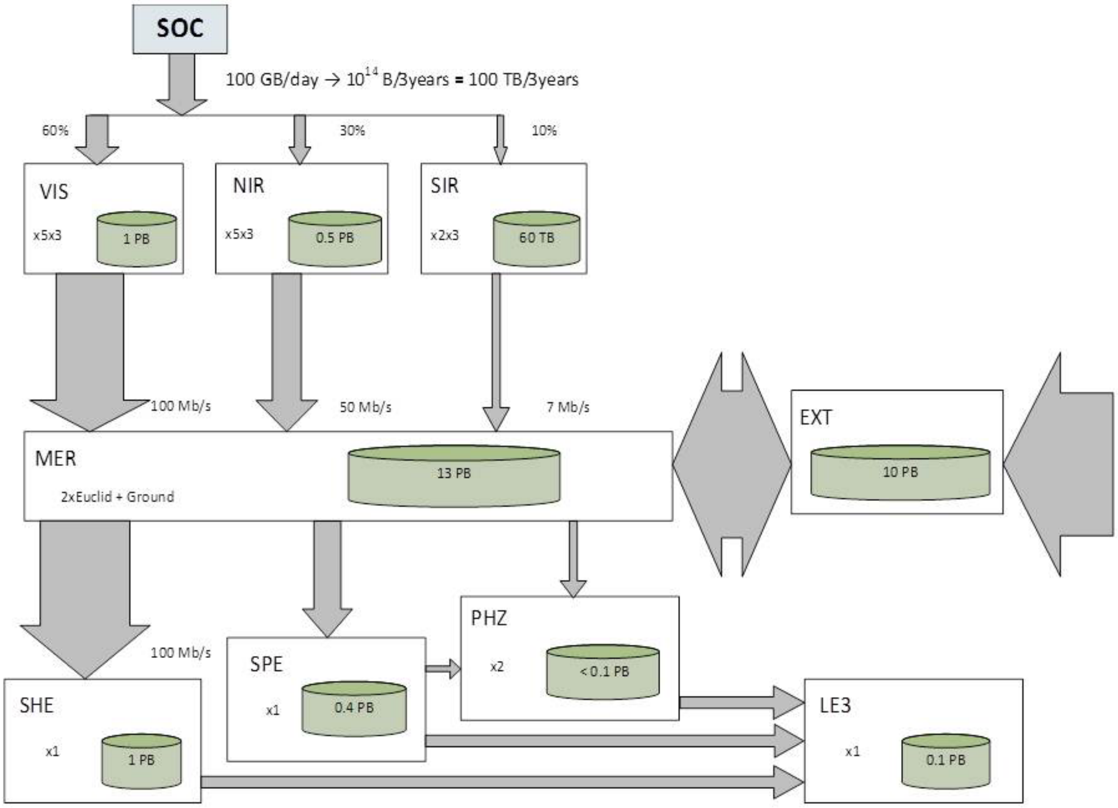

Figure 4. The data flow between the various processing functions of the SGS, and the related foreseen size. The size of the arrows representing the data flow are not to scale.

4.2 Data flow

The data flow between the various processing functions of the SGS, and the related size, is shown in Figure 4, which provides a visual estimation of the Euclid SGS data flow, given by the amount of data that, provided by a data processing function, are ingested by another data processing function for further analysis. The arrows representing the data flow are not to scale. EXT data can be considered “off-line” and do not contribute directly to the day-by-day data flow.

4.3 Design concepts

The main concepts at the basis of the SGS design are summarized hereafter.

- Minimisation of data transfers.
- A concept of distributed data products storage (bulk data products are stored at least twice among the SDCs and metadata are indexed inside the EMA) avoiding the unnecessary movement of huge amounts of data between the SDCs and the EMA/SDC.
- A single EMA metadata repository which inventories and indexes all metadata (and corresponding data locations).
- A concept of software layers inside the SGS: metadata access layer (query/retrieve), data product access layer (open, read/write, get info,..), data processing layer.
- A design allowing the flexibility to implement new software pipeline releases without redesigning the SGS architecture.
- In the distribution of work on the SDCs, we are already considering the data transfer aspects. For the data processing functions that will use or generate very large amounts of data, we have a minimum number (2 for redundancy) of SDCs in order to minimize the constraints on transfer.

4.4 Development principles

The leading principles driving the development of a cost-efficient and coordinated Euclid SGS are listed in the following.

- Simplicity of system design.
- Component-based software engineering. This is a modular approach to software development: each module can be developed independently and wrapped in the language adopted as the standard for the system (C/C++ and PYTHON as scripting language have been chosen) to form a pipeline or workflow. The concept is already in use in working systems for astronomy, for both ground-based and space-borne observations.
- Virtualisation: executing pipeline software on virtual machines and separating pipeline software from the underlying hardware resources. These technologies should make easier the deployment and run of any pipeline software on any SDC infrastructure. Since one of the main principles of the SGS is to move the data as little as possible, we plan to use a scheme where a code is developed in an SDC, and virtual machine images are created and transferred into the SDCs holding the data this part of the code needs. This also has advantages on the development side since, provided we have strict guidelines on the data model and interface tools, it allows each pipeline module to be developed independently and integrated as a suite of virtual machines at any SDC.
- A common data model for each module, application and pipeline. This means that each module, application and pipeline will deal with the unified data model for the whole cycle of the data processing from the raw data to the final data product.
- Persistence of the data model objects: each frame in the data processing chain is described by the common data model and saved in the EMA along with all the parameters used for the data processing. Finding a compromise between the number of persistent objects and the required storage will be part of the architecture design during the Implementation Phase.

These principles for the development of the data processing software combined with the EMA allow parallel and independent data processing on different levels of data, in the cases where redundancy and cross-check have been identified as desirable. They also enable access to quality controls to all participants. The distribution of all data-items facilitates the analysis and cross-checking of results by several independent groups, which is crucial for the redundancy of data quality controls and to secure the validation of critical scientific results, like the complex shear measurements or the determination of cosmological parameter values.

4.5 SGS logical architecture

The SGS is based on the logical architecture[11] summarized in the following and shown in Figure 5.

- A single metadata repository which inventories, indexes and localizes the huge amount of distributed data.
- A distributed storage of the data over the SDCs (ensuring the best compromise between data availability and data transfers).

- A set of services (Service-Oriented Architecture – SOA) which allows a low coupling between SGS components: e.g. metadata query and access, data localization and transfer, data processing monitoring and control (M&C), …
- 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.
- A common Decentralized Processing Control, data and event driven, deployed on each SDC.
- An automatic approach to Data Quality Control, to be performed at every processing step.

In the Science Implementation Plans (SIP)11,12 the plan to migrate from a logical architecture to a physical architecture has been described.

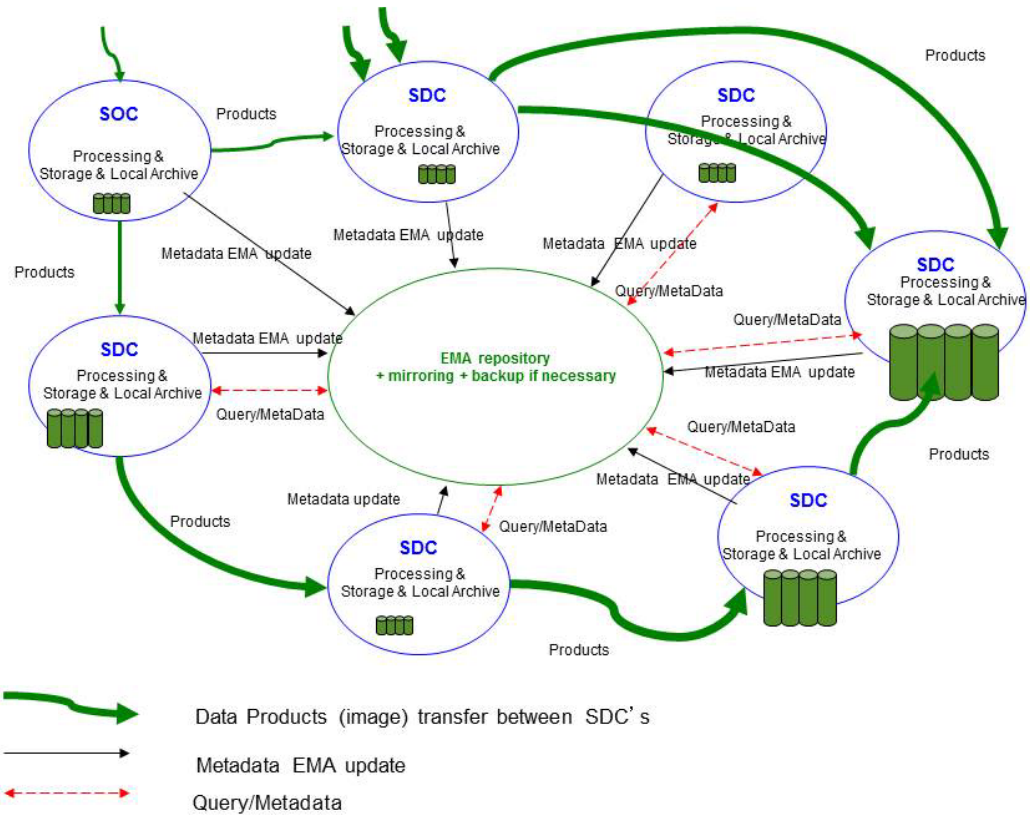

Figure 5. The logical architecture of the Euclid SGS. The different sizes of the storage symbols represent the different capabilities of the SDC resources. Bold arrows represent bulk data transfers, thin arrows metadata exchanges (queries are dashed, EMA updates are solid).

4.6 Technology watch

To avoid getting tied too early to specific technical solutions, the SGS team needs to pay attention during implementation phase to the technologies trends (evolving very fast) and, uppermost, to refine requirements in terms of querying models, ingestion/retrieval metadata throughput and performances requirements[14].

A technology watch (open to commercial COTS, thus not restricted to open source solutions) has been set up by partners to perform benchmarks from candidate technologies. This technology watch must integrate all elements of decision support: technical, operating costs, administration costs, etc.

The SGS System Team is developing within an experimental team (composed of OUs and SDC-DEV staff) a mock-up of the Infrastructure Abstraction Layer.

This could help as a proof of concept (among other things):

- to define the interfaces of pipeline;
- to anticipate problems of integration on existing infrastructures;
- to start pipeline software development taking into account external interfaces;
- to procure a stand-alone development frameworks for pipeline developers.

5. CONCLUSIONS

The success of the Euclid mission heavily relies on careful design and implementation of its ground segment facilities. The Ground Station(s) and the Mission Operations Center (MOC), both operated by ESA, are the elements of the Mission Operations Ground Segment (MOGS). The Science Operations Center (SOC) operated by ESA and a number of Science Data Centers (SDCs) in charge of data processing, provided by a Consortium of 14 European countries, are the elements of the Euclid Scientific Ground Segment (SGS).

The SOC acts as the central node for the mission planning, performs an initial quality check and processing of the data and makes the telemetry available to the remainder of the SGS; the SOC is also responsible for developing and maintaining the Euclid Legacy Archive (ELA) and for delivering the data products to the general scientific community. The Euclid Consortium provides: support to instrument maintenance and operations, the SDCs responsible for instrument specific data processing and the production of quality-controlled processed data and higher level results which are delivered to ESA for ingestion into the ELA, plus simulations aimed at verifying the end-to-end performances of the mission and validating the data processing, and any external ancillary data set that is required to achieve the mission's scientific objectives.

The distributed nature, the huge data volume of the overall data set (Euclid plus ancillary data), and the needed accuracy of the results are the main challenges expected in the design and implementation of the SGS. The leading principles driving the development of the Euclid SGS are expected to be the simplicity of system design, a component-based software engineering, virtualization, and a data-centric approach to the system architecture where quality control, a common data model and the persistence of the data model objects play a crucial role.

ESA/SOC and the Euclid Consortium have developed, and are committed to maintain, a tight collaboration in order to design and develop a single, cost-efficient and truly integrated SGS.

ACKNOWLEDGEMENTS

The authors act on behalf of the many people working for the Euclid SGS: in particular the Euclid Consortium Lead Yannick Mellier (IAP), the PO Support Team and the SGS System Team. Among others, thanks are due to: Claudio Vuerli (INAF), Anna Gregorio (Univ. Trieste), Marco Frailis (INAF), Andrea Zacchei (INAF), Pasquale Panuzzo (CEA), Keith Noddle (UoE), Jean-Marc Delouis (IAP), Laurent Vibert (IAS), Rees Williams (RuG), Christian Neissner (PIC), Johannes Koppenhoefer (MPG), Joseph Mohr (LMU), Christian Surace (LAM), Pierre Dubath (Unige), Stéphane Paltani (Unige), Martin Melchior (ETH), Stefan Muller (ETH), Elina Keihanen (UHelsinki), Massimo Brescia (INAF), Luigi Paioro (INAF), and all of the OU and SDC leaders and teams.

The participation in the Euclid phases A/B1 has been supported by the National Space Agencies. In particular in Italy by ASI contracts I/031/10/0 and I/039/10/0.

REFERENCES

[1] Laureijs, R.; Amiaux, J.; Arduini, S.; Auguères, J. -L.; Brinchmann, J.; Cole, R.; Cropper, M.; Dabin, C.; Duvet, L.; Ealet, A.; Garilli, B.; Gondoin, P.; Guzzo, L.; Hoar, J.; Hoekstra, H.; Holmes, R.; Kitching, T.; Maciaszek, T.; Mellier, Y.; Pasian, F.; Percival, W.; Rhodes, J.; Saavedra Criado, G.; Sauvage, M.; Scaramella, R.; Valenziano, L.; Warren, S.; Bender, R.; Castander, F.; Cimatti, A.; Le Fèvre, O.; Kurki-Suonio, H.; Levi, M.; Lilje, P.; Meylan, G.; Nichol, R.; Pedersen, K.; Popa, V.; Rebolo Lopez, R.; Rix, H. -W.; Rottgering, H.; Zeilinger, W.; Grupp, F.; Hudelot, P.; Massey, R.; Meneghetti, M.; Miller, L.; Paltani, S.; Paulin-Henriksson, S.; Pires, S.; Saxton, C.; Schrabback, T.; Seidel, G.; Walsh, J.; Aghanim, N.; Amendola, L.; Bartlett, J.; Baccigalupi, C.; Beaulieu, J. -P.; Benabed, K.; Cuby, J. -G.; Elbaz, D.; Fosalba, P.; Gavazzi, G.; Helmi, A.; Hook, I.; Irwin, M.; Kneib, J. -P.; Kunz, M.; Mannucci, F.; Moscardini, L.; Tao, C.; Teyssier, R.; Weller, J.; Zamorani, G.; Zapatero Osorio, M. R.; Boulade, O.; Foumond, J. J.; Di Giorgio, A.; Guttridge, P.; James, A.; Kemp, M.; Martignac, J.; Spencer, A.; Walton, D.; Blümchen, T.; Bonoli, C.; Bortoletto, F.; Cerna, C.; Corcione, L.; Fabron, C.; Jahnke, K.; Ligori, S.; Madrid, F.; Martin, L.; Morgante, G.; Pamplona, T.; Prieto, E.; Riva, M.; Toledo, R.; Trifoglio, M.; Zerbi, F.; Abdalla, F.; Douspis, M.; Grenet, C.; Borgani, S.; Bouwens, R.; Courbin, F.; Delouis, J. -M.; Dubath, P.; Fontana, A.; Frailis, M.; Grazian, A.; Koppenhöfer, J.; Mansutti, O.; Melchior, M.; Mignoli, M.; Mohr, J.; Neissner, C.; Noddle, K.; Poncet, M.; Scodeggio, M.; Serrano, S.; Shane, N.; Starck, J. -L.; Surace, C.; Taylor, A.; Verdoes-Kleijn, G.; Vuerli, C.; Williams, O. R.; Zacchei, A.; Altieri, B.; Escudero Sanz, I.; Kohley, R.; Oosterbroek, T.; Astier, P.; Bacon, D.; Bardelli, S.; Baugh, C.; Bellagamba, F.; Benoist, C.; Bianchi, D.; Biviano, A.; Branchini, E.; Carbone, C.; Cardone, V.; Clements, D.; Colombi, S.; Conselice, C.; Cresci, G.; Deacon, N.; Dunlop, J.; Fedeli, C.; Fontanot, F.; Franzetti, P.; Giocoli, C.; Garcia-Bellido, J.; Gow, J.; Heavens, A.; Hewett, P.; Heymans, C.; Holland, A.; Huang, Z.; Ilbert, O.; Joachimi, B.; Jennins, E.; Kerins, E.; Kiessling, A.; Kirk, D.; Kotak, R.; Krause, O.; Lahav, O.; van Leeuwen, F.; Lesgourgues, J.; Lombardi, M.; Magliocchetti, M.; Maguire, K.; Majerotto, E.; Maoli, R.; Marulli, F.; Maurogordato, S.; McCracken, H.; McLure, R.; Melchiorri, A.; Merson, A.; Moresco, M.; Nonino, M.; Norberg, P.; Peacock, J.; Pello, R.; Penny, M.; Pettorino, V.; Di Porto, C.; Pozzetti, L.; Quercellini, C.; Radovich, M.; Rassat, A.; Roche, N.; Ronayette, S.; Rossetti, E.; Sartoris, B.; Schneider, P.; Semboloni, E.; Serjeant, S.; Simpson, F.; Skordis, C.; Smadja, G.; Smartt, S.; Spano, P.; Spiro, S.; Sullivan, M.; Tilquin, A.; Trotta, R.; Verde, L.; Wang, Y.; Williger, G.; Zhao, G.; Zoubian, J. and Zucca, E., [Euclid Definition Study Report], ESA/SRE(2011)12, eprint arXiv:1110.3193 (2011).

[2] Laureijs, R., et al., "Euclid: ESAs mission to map the geometry of the dark universe", Proc. SPIE 8442, in press

[3] Cropper, M., et al., "VIS: the visible imager for Euclid", Proc. SPIE 8442, in press

[4] Prieto, E., et al., "Euclid near-infrared spectrophotometer instrument concept at the end of the phase A study", Proc. SPIE 8442, in press

[5] Amiaux, J., et al. "Euclid Mission: building of a reference survey", Proc. SPIE 8442, in press

[6] Valenziano, L., et al., "Spaceborne survey instrument operations from Planck/LFI to Euclid NISP: lessons learned and new concepts", Proc. SPIE 8448, in press

[7] Hanisch, R.J., "The Virtual Observatory: Retrospective and Prospectus", Astronomical Data Analysis Software and Systems XIX, ASP Conference Series 434, 65 (2010)

[8] Lin, H., Flaugher, B. and the Dark Energy Survey Collaboration, "The Dark Energy Survey", Bulletin of the American Astronomical Society, 41, 669 (2009)

[9] Kaiser, N., Burgett, W., Chambers, K., Denneau, L., Heasley, J., Jedicke, R., Magnier, E., Morgan, J., Onaka, P. and Tonry, J., "The Pan-STARRS wide-field optical/NIR imaging survey", Proc. SPIE 7733, 1-14 (2010).

[10] Axelrod, Tim S.; Becla, J.; Connolly, A.; Dossa, D.; Jagatheesan, A.; Kantor, J.; Levine, D.; Lupton, R.; Plante, R.; Smith, C.; Thakar, A.; Tyson, J. A. and LSST Data Management Team, "The LSST Data Challenges", Bulletin of the American Astronomical Society, 39, 983 (2007)

[11] "Euclid SOC Science Implementation Plan", Euclid_SO_Dc_00007, v. 0.5 (2011)

[12] "Euclid Consortium SGS Science Implementation Plan", EUCL-OTS-SGS-PL-00003, v.2.0 (2011)

[13] "Architecture Definition Study Report", EUCL-CNE-SYS-TN-00007, v. 0.2 (2011)

[14] Poncet, M., GSAW 2012 Conference, in press