Publication Year	2016
Acceptance in OA@INAF	2020-05-04T09:50:08Z
Title	Management of the science ground segment for the Euclid mission
Authors	ZACCHEI, Andrea; Hoar, John; PASIAN, Fabio; Buenadicha, Guillermo; Dabin, Christophe; et al.
DOI	10.1117/12.2232934
Handle	http://hdl.handle.net/20.500.12386/24416
Series	PROCEEDINGS OF SPIE
Number	9913

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Management of the science ground segment for the Euclid mission

Zacchei, Andrea, Hoar, John, Pasian, Fabio, Buenadicha, Guillermo, Dabin, Christophe, et al.

Andrea Zacchei, John Hoar, Fabio Pasian, Guillermo Buenadicha, Christophe Dabin, Anna Gregorio, Oriana Mansutti, Marc Sauvage, Claudio Vuerli, "Management of the science ground segment for the Euclid mission," Proc. SPIE 9913, Software and Cyberinfrastructure for Astronomy IV, 99131B (26 July 2016); doi: 10.1117/12.2232934

SPIE.

Event: SPIE Astronomical Telescopes + Instrumentation, 2016, Edinburgh, United Kingdom

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Management of the Science Ground Segment for the Euclid mission

Andrea Zacchei[a], John Hoar[b], Fabio Pasian[a], Guillermo Buenadicha[b], Christophe Dabin[c], Anna Gregorio[d,a], Oriana Mansutti[a], Marc Sauvage[e], Claudio Vuerli[a]

[a] INAF-OATs, Via G.B.Tiepolo 11, I-34143 Trieste, Italy
[b] European Space Agency / ESAC, Villanueva de la Cañada, E-28692 Madrid, Spain
[c] CNES, 18 Av. Edouard Belin, F-31401 Toulouse Cedex 9, France
[d] University of Trieste, P.le Europa 1, I-34128 Trieste, Italy
[e] CEA, Orme des Merisiers, F-91191 Gif-sur-Yvette, France

ABSTRACT

Euclid is an ESA mission aimed at understanding the nature of dark energy and dark matter by using simultaneously two probes (weak lensing and baryon acoustic oscillations). The mission will observe galaxies and clusters of galaxies out to z~2, in a wide extra-galactic survey covering 15000 deg², plus a deep survey covering an area of 40 deg². The payload is composed of two instruments, an imager in the visible domain (VIS) and an imager-spectrometer (NISP) covering the near-infrared. The launch is planned in Q4 of 2020. The elements of the Euclid Science Ground Segment (SGS) are the Science Operations Centre (SOC) operated by ESA and nine Science Data Centres (SDCs) in charge of data processing, provided by the Euclid Consortium (EC), formed by over 110 institutes spread in 15 countries. SOC and the EC started several years ago a tight collaboration in order to design and develop a single, cost-efficient and truly integrated SGS.

The distributed nature, the size of the data set, and the needed accuracy of the results are the main challenges expected in the design and implementation of the 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 distributed storage to avoid data migration across SDCs.

This paper describes the management challenges that the Euclid SGS is facing while dealing with such complexity. The main aspect is related to the organisation of a geographically distributed software development team. In principle algorithms and code is developed in a large number of institutes, while data is actually processed at fewer centers (the national SDCs) where the operational computational infrastructures are maintained. The software produced for data handling, processing and analysis is built within a common development environment defined by the SGS System Team, common to SOC and ECSGS, which has already been active for several years. The code is built incrementally through different levels of maturity, going from prototypes (developed mainly by scientists) to production code (engineered and tested at the SDCs). A number of incremental challenges (infrastructure, data processing and integrated) have been included in the Euclid SGS test plan to verify the correctness and accuracy of the developed systems.

Keywords: Euclid mission, ground segment, distributed development, distributed processing, management.

1. INTRODUCTION: THE EUCLID MISSION

Euclid is the second medium-sized mission (M2) of the ESA Cosmic Vision 2015-2025 Plan, planned to be launched in Q4 of 2020, and 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 spacecraft operates in L2, the second Sun-Earth Lagrange point. The Euclid survey will nominally last 6 years, and will observe galaxies and clusters of galaxies out to z~2 in a wide extra-galactic survey covering 15000 deg², more than 35% of the entire sky. A deep survey will also be performed, covering an area of 40 deg². The payload is composed of a 1.2 m three-mirror Korsch telescope with a stable and large flat field of view, and two instruments: an imager in the

Software and Cyberinfrastructure for Astronomy IV, edited by Gianluca Chiozzi, Juan C. Guzman, Proc. of SPIE Vol. 9913, 99131B · © 2016 SPIE
CCC code: 0277-786X/16/$18 · doi: 10.1117/12.2232934

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

visible domain (VIS) and an imager-spectrometer (NISP) covering the near infrared. The instruments and the related data processing are provided by a Consortium (the Euclid Consortium – EC) composed of more than 110 laboratories from 14 European countries (Austria, Belgium, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Portugal, Romania, Spain, Switzerland, UK) plus the United States.

With its X and K band transponders to support the tele-commanding and the science data transfer to ground, respectively, Euclid will deliver an unprecedented large volume of data for astronomical space missions: more than 1 Pbit of data per year, about 4 times more data than Gaia. It is to be noted however that the broadband Euclid imaging data alone are not sufficient to achieve the required photometric redshift accuracy and precision, which means that additional ground-based data are required. A large volume of ground-based data from optical surveys like DES, KIDS/VIKING and others is used for calibrations, quality control tasks and scientific data reduction, specifically for obtaining photometric redshifts. The ground based data have to undergo Euclid specific processing (such as the conversion to common astrometric and magnitude reference systems) in order to be consistently handled with the Euclid data.

Details on the Euclid and its science case are available in Laureijs et al. [1], on the mission design and its status in Racca et al. [2]; information on the Euclid instruments can be found in Cropper et al. [3] (imaging) and Maciaszek et al. [4] (spectral).

2. THE EUCLID GROUND SEGMENT

As described in Pasian et al. [5], the main elements of the Euclid Ground Segment are

- the Ground Stations, a network of Deep Space antennas that connect to the spacecraft during periods of 4 hours each for the uploading of tele-commands (TC) and downlink of telemetry (TM);
- the Mission Operations Centre (MOC), in charge of all mission operations planning, execution, monitoring and control of the spacecraft;
- the Science Operations Centre (SOC), in charge of scientific operations planning, performance monitoring of the payload using spacecraft and instrument files delivered by the MOC, and science data archiving and distribution;
- the Euclid Consortium Science Ground Segment (ECSGS), responsible for the production of the science ready calibrated images, source catalogues with astrometry, photometry, morphometry, spectra and redshifts, and all relevant quality control and meta-data that are necessary for the scientific exploitation of the Euclid mission; two Instrument Operation Teams (IOTs), one for each instrument, are also part of the ECSGS and are in charge of maintaining the payload procedures, databases, software and operational modes, and of monitoring the health of the payload throughout the mission, in coordination with the SOC.

The Ground Stations and the MOC compose the Operations Ground Segment (OGS); the SOC and the ECSGS compose the Science Ground Segment (SGS). The Ground Stations, the MOC and the SOC are provided by ESA, the ECSGS by the Euclid Consortium.

The MOC (via the Data Distribution System - DDS) interfaces only to the SOC and provides raw telemetry and all auxiliary information necessary to manage the mission from the scientific point of view. The SOC provides the MOC with information related to observation planning and instruments commanding. The overall planning of the survey is organised by the Euclid Science Team, who provides the relevant input for the SOC. The SOC is also in charge of mission planning, of the first consistency and quality checks and of the production of quick-look-quality data for public distribution.

The Science Operations Centre (SOC) operated by ESA and nine Science Data Centres (SDCs) are the elements of the Euclid SGS. The SDCs are provided by the member of the Euclid Consortium and form the ECSGS. They are in charge of instrument-related processing, production of science data products, simulations, ingestion of external data and in general all science-driven data processing. They will furthermore support the computational needs of the IOTs. SDCs are located in Finland, France, Germany, Italy, the Netherlands, Spain, Switzerland, UK, and United States. Other SDCs may be added if needed and/or relevant.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

3. THE EUCLID SGS AS A DISTRIBUTED ENVIRONMENT

3.1 Data flow and processing

From the above description, it is clear that the SGS activity is geographically spread as part of its very design. Conversely, it is quite clear that, to achieve the ambitious results of the Euclid mission, the SOC and the ECSGS need to design and develop a single, cost-efficient and truly integrated SGS.

The SOC develops and operates survey-planning tools to enable routine execution of the sky survey and incorporate additional input to the mission planning process, such as instrument commanding requests or additional observing requests. The monitoring process starts with Level 1 production activities followed by the Quick Look Analysis, which include telemetry checking and handling, assessment of the science telemetry and housekeeping and the production of daily reports, thus closing the short-term survey-planning loop at the SOC. The software for Level 1 processing is developed by the ECSGS for operation at the SOC. The elements of the instrument-monitoring environment such as the quick-look analysis system (QLA) are deeply linked to the instrument knowledge available at the relevant ECSGS SDCs, therefore it is expected a high level of interactions SOC-SDC. The SOC designs, implements and operates the Science Archive System component; the data and meta-data repository of the Euclid Archive System (EAS), which forms the basis of the final archive of Euclid after the completion of the mission.

The ECSGS develops and operates the data processing that convert the telemetry delivered by the SOC into the final science products of the mission. This activity is spread over the various teams contributing to the ECSGS. The processing steps required within the SGS is shown in Figure 1

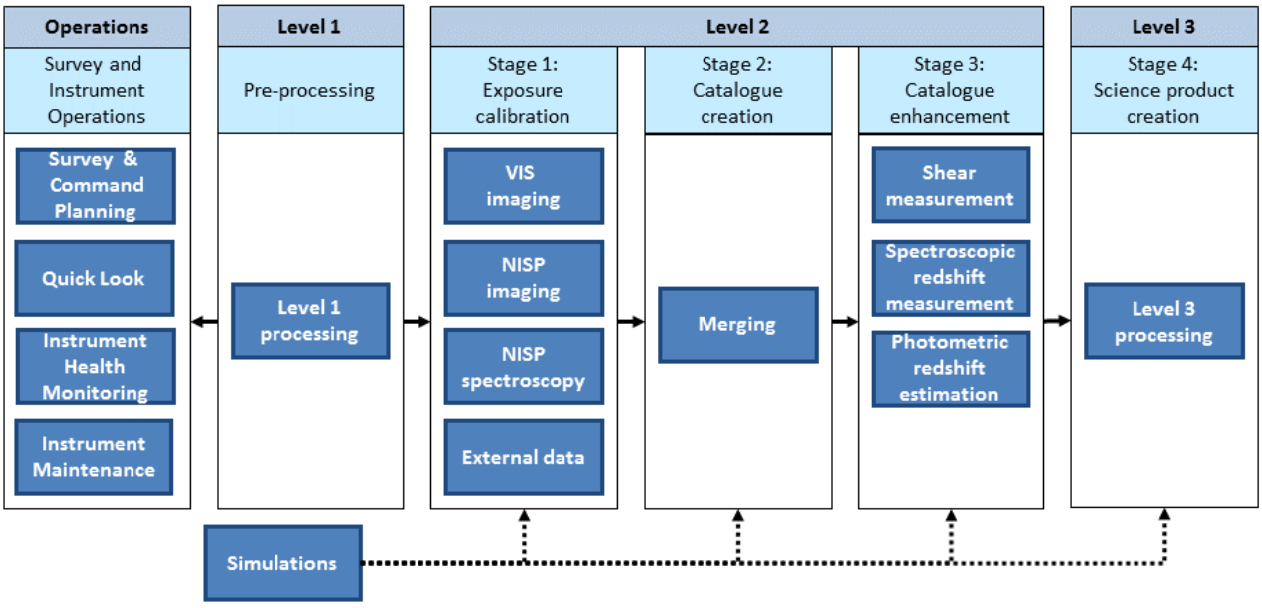

Figure 1 – Processing steps needed in the Euclid SGS.

The processing of science data can be decomposed in eleven logical data Processing Functions, defined by considering that they represent self-contained processing units, i.e. they represent the highest-level breakdown of the complete pipeline that can be achieved with units that communicate only with the help of the archive. The identified Processing Functions are listed in the following, and the interactions between them are shown in Figure 2.

- Telemetry processing to be operationally performed at SOC [LE1];
- Processing the Visible imaging data from edited telemetry to fully calibrated images, as well as source lists for quality check purposes only [VIS].
- Processing the Near-Infrared imaging data from edited telemetry to fully calibrated images as well as source lists, used to check quality and to allow spectra extraction [NIR].

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

- Processing the Near-Infrared imaging data from edited telemetry fully calibrated spectral images and extraction of the spectra in the slitless spectroscopic frames taken by the NISP [SIR].
- Entering in the EA all of the external data that are 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 [EXT].
- Implementing the simulations needed to test, validate and qualify the whole pipeline [SIM].
- Merging of all the information produced by the Processing Functions above; it is in charge of providing stacked images and source catalogues where all the multi-wavelength data (photometric and spectroscopic) are aggregated [MER].
- Computing spectroscopic redshifts from the spectra produced by SIR [SPE].
- Computing photometric redshifts from the multi-wavelength imaging data [PHZ].
- Computing shape measurements on the visible imaging data [SHE].
- Computing all the high-level science data products (Level 3), from the fully processed shape and redshift measurements and any other possibly needed Euclid data [LE3].

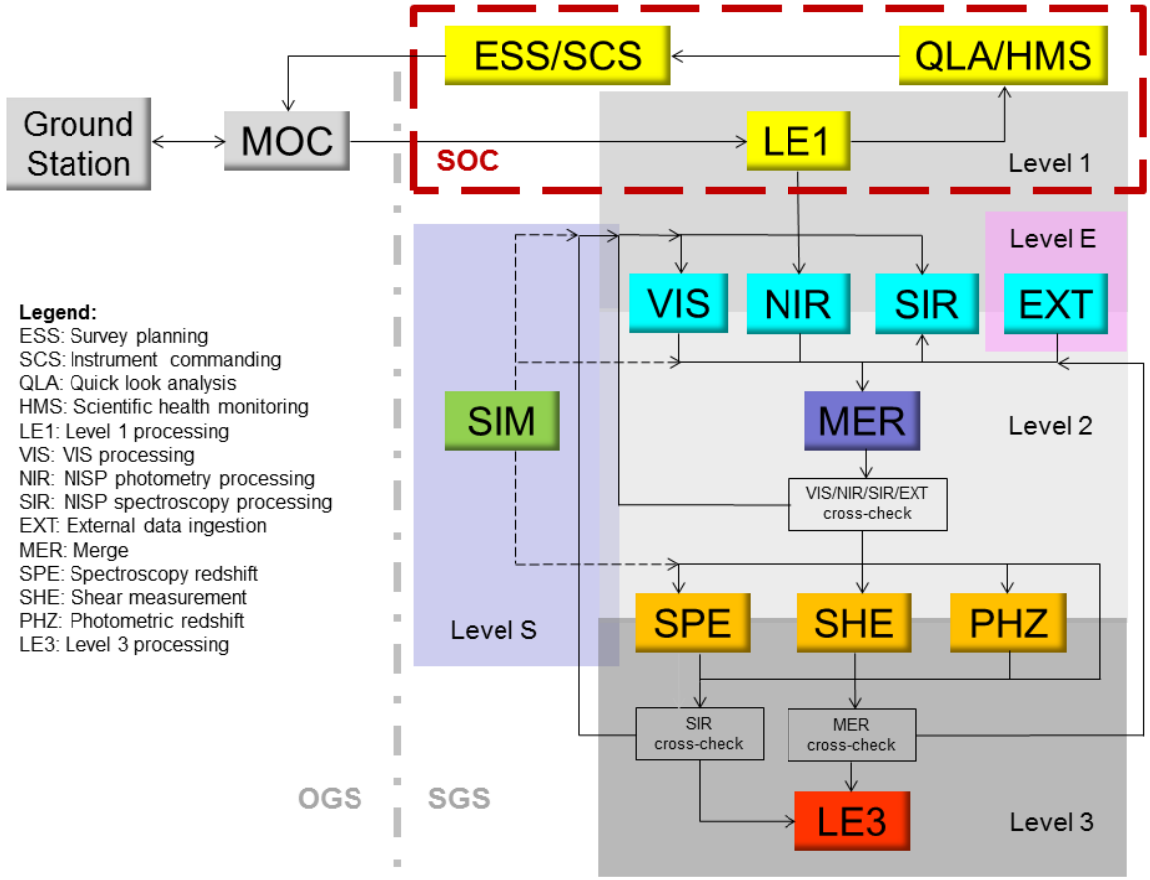

Figure 2 – The interactions between Euclid SGS Processing Functions, including data flow (also shown in [2]).

3.2 A distributed software development process

Achieving the precisions needed by the final products of the mission call for a high level of refinement of the processing code. Therefore, the data processing algorithms at the core of the Processing Functions need to be written by the best scientists available within the Consortium. Gathering the best professionals in a single institute or location is quite impractical: this means that the teams developing algorithms even for a single Processing Function are distributed throughout the various countries participating in the Euclid Consortium.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

The first management challenge that the Euclid is facing in dealing with the complexity of its SGS is therefore related to the organisation of a geographically distributed software development team. In principle, algorithms and code are developed in a large number of institutes, while data is actually processed at fewer centers (the national SDCs) where the operational computational infrastructures are maintained.

Any new pipeline element is developed around the following scheme:

- The high-level scientific requirements of the mission are defined in the Science Requirements Document, which flow down to a set of Ground-based Data Processing requirements, to which the SGS pipeline code shall comply. Responsibility for checking compliance with such requirements mainly rests with the Euclid Science Working Groups (SWGs), which are specialized teams of scientists in the EC that have been built around the Euclid science objectives.
- As a flow-down and a refinement of the above high-level approach, requirements are placed on the various Processing Functions under consideration. These can come from studies published in the refereed literature, research on signal processing, and so on. Additional requirements related to the implementation part of the algorithm (e.g. interfaces, parallelization capabilities, computing time, etc.) are specified as well. These requirements are accompanied with the descriptions of key tests that can be performed to check the validity of the pipeline elements. Once requirements are agreed upon, they can be turned into prototypes.
- Algorithmic research proceeds by designing prototypes, performing numerical tests, and comparing the results with the original requirements. There are no formal requirements placed on the infrastructure and languages choices that can be used at that stage of the research, in order to maximize creativity. Responsibility for this algorithmic research rests in the Organization Units (OUs), which are again teams of EC scientists, grouped along their interest/competences for one for each of the individual data Processing Functions defined above and depicted in Figure 2.
- Once validated by the proper OU, the prototype is passed to an SDC, along with a test harness, and the development team at the SDC (SDC-DEV) turns this prototype into a full-fledged Euclid pipeline element. This includes complying strictly with common coding standards, using pre-defined input and output mechanisms, i.e. homogenization and configuration control of the Euclid pipelines is a responsibility of the SDCs. It is to be noted that it is a task of SDC-DEV to optimize the algorithms with reference to the hardware/software environment.
- Once available as a pipeline element, further tests will have to occur to fully validate the new element and in particular decide to include it in the production chain or not. The responsibility for the validation of the pipeline element rests with the OU, but these decisions will involve the SGS management as well, which will resolve possible conflicts.

It must be emphasized at this stage that the above description is a formal representation of the logical steps to be followed to produce new pipeline elements. Rather than defining teams in the sense of groups of people, it defines functions inside the SGS:

- the functional role of the SWGs is to provide scientific requirements for the pipeline development;
- the functional role of the OU teams is to turn these requirements into code prototypes and assess whether the requirements are met;
- the functional role of the SDC-DEVs is to turn these prototypes into pipeline modules respecting the architecture and interface definitions, to include unit and integration tests and to put the modules under configuration control;
- the software delivered by the ECSGS to the SOC is validated by the SOC itself.

This distinction however is not to be interpreted strictly. As a matter of fact, it is to be noted that individual Euclid scientists may belong to more than one of the above groups. This has in important consequence in the arrangements made to avoid over-formalisation of interfaces between SWGs, OUs and SDCs, as shown and explained in Pasian et al. [6]. In most cases, no interfaces will exist between OU and SDC-DEV, but rather a joint development will take place. On the other hand, interactions between OU and SWG will occur only for validation of the resulting Processing Functions against high-level requirements and not on a day-by-day interaction basis.

3.3 Distributed processing

As mentioned earlier, centralization of data processing in one single center is not feasible for political and economic reasons. On the other side, a data flow analysis has demonstrated that a strict coupling of Processing Functions with

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

dedicated SDCs leads to a non-sustainable data transfer across the SGS. These two constraints need to be taken into account in the design of the Euclid data processing.

Since it is undesirable to dedicate SDCs to specific steps of the data processing, it was decided that there is in principle no SDCs dedicated to specific tasks. SDCs are considered as generic resource providers capable of providing the services requested by the SGS as needed. As a consequence, any pipeline will be in principle able to run on any SDC. Since the hardware/software infrastructure across the various SDCs is inhomogeneous, each SDC is expected to be capable of running the same code through virtualization techniques. The basic concept is quite logical: move the code, not the data. In other words, the pipeline will be run wherever the main input data is stored, and will store as well the related output data. The data processing mechanism could boil down to a kind of Map/Reduce mechanism.

In the first place, the concept of "tile", i.e. the minimal processable set of data covering a given sky area, has been introduced. In principle, lower level pipelines (i.e. starting from raw data and arriving to the point where catalogues of objects are built) could operate on a single tile. In practice, to reduce processing overheads, a number (still TBD) of tiles will be merged in an appropriate chunk of the sky to be processed. A higher level of processing is then performed, based on data cross-matching, and this run across different SDCs with a certain amount of data exchange, estimated to be about two orders of magnitude lower than for raw or calibrated data.

3.4 Distributed location of Instrument Operation Teams

After the completion and delivery of VIS and NISP, the two Euclid instruments, the Euclid Consortium will be in charge of defining and maintaining the instruments modes of operation in order to maximize the scientific return of the mission, and of supporting the SOC in monitoring the health of the payload and assessing the quality of data. For this purpose, two Instrument Operation Teams (IOTs) will operate, that build on the knowledge of the Instrument Development Teams. IOTs play a crucial role in the successful mission execution. They assist ESA in planning the mission while instruments may need various interventions: update parameters and flight operations procedures, new on-board software updates, others. They will interact with the MOC through the SOC in routine phase of the mission, and directly in case of instrument contingencies.

An additional complexity is given by the fact that the IOTs are geographically distributed in different institutes belonging to the Euclid Consortium. The activity of the IOTs, and of the instrument Calibration Teams, needs to be coordinated in order to avoid possible clashes or inconsistencies in operating the two instruments, while providing a single operational interface from the ECSGS to SOC.

4. MANAGING THE EUCLID SGS COMPLEXITY

4.1 Management and coordination teams

The SOC Development Manager and the ECSGS Manager lead the SOC management team and the ECSGS Project Office (PO), respectively. They liaise to coordinate the management of the overall SGS.

The first example of the tight collaboration established between SOC and the ECSGS to develop a single, cost-efficient and truly integrated SGS is a common documentation tree and work breakdown structure. In practice, most of the high-level documents, which define the organization and management of the SGS, are in common. The only exceptions are the SOC Management Plan and the ECSGS Management Plan, which are bound to be different since they deal with manpower, recruitment, procurement, funding, internal reporting, that are organization-specific. This situation is depicted in Figure 3. Deriving from this common documentation is a coordinated activity in the fields of Quality Assurance (QA), Configuration Management, Risk Management, etc., which can be considered to be performed the usual way for a space mission.

The SOC team is primarily composed of two groups, the SOC Operations Team consisting of scientific and engineering staff that executes and monitors the Euclid survey, and rapidly analyze science data from the instruments in order to minimize loss to the survey. The Operations Team liaises with the MOC and Instrument Operations Teams for routine operational matters and are supported by the Survey and Calibration Working Groups for longer term planning and coordination. The SOC Development Team is responsible for the design and implementation of the SOC systems that support the Operations Team in the course of their work. The SOC Development Manager is responsible for the overall

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

management and planning of the SOC before launch with the SOC System Engineer coordinating the technical implementation.

Members of the ECSGS PO, under the coordination of the ECSGS Manager, are the System Team Lead, the ECSGS Scientist, the Quality Assurance (QA) Lead, the Configuration Management Lead, the Project Controller, and the IOTs Coordinator. The latter is the point of merging between the activities of the Instrument Operations Teams and of the Calibration Teams, and the single interface of the ECSGS to SOC for operations. As for QA coordination, the QA Lead chairs a group formed by the reference persons in charge of QA at the various SDCs: this group is in charge of gathering and monitoring risks, and performing all related QA-related actions.

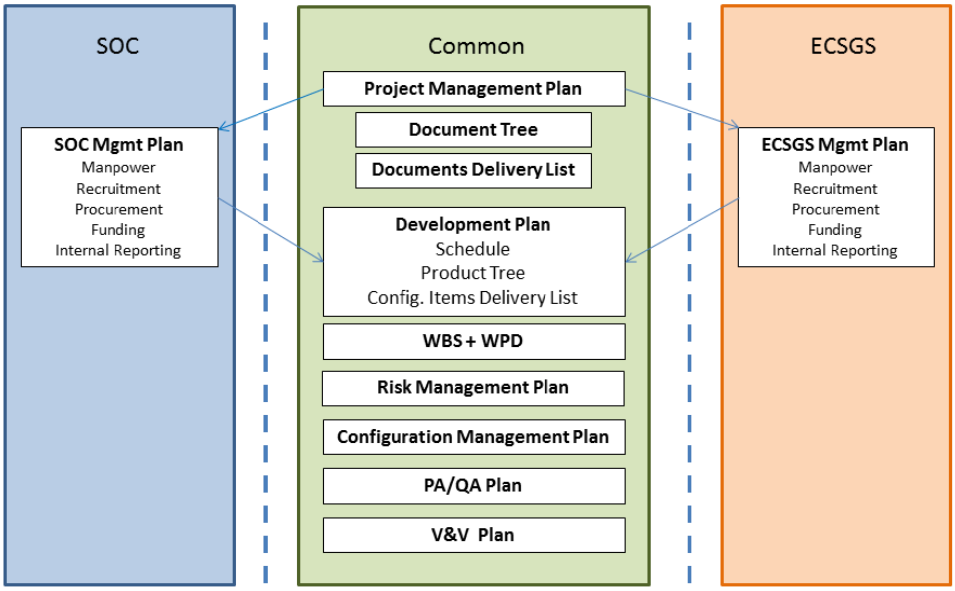

Figure 3 – Most of the high-level Euclid SGS documentation is in common between SOC and ECSGS.

A SGS System Team (SGS ST) common to both SOC and ECSGS has been active for several years, taking the lead in helping the SGS to define the overall data processing philosophy, architecture and strategy. The SGS ST provides tools, standards and support to the code development. Among the tasks of the SGS ST there are preparing coding guidelines; defining and implementing tools to support software tests and integration; defining, designing, implementing and testing common software (e.g. interfaces, transfer systems and common toolboxes); defining, designing, implementing and testing the archive; designing and implementing tools to define and maintain an Euclid common Data Model; estimating and allocating the processing budget continuously throughout the project. Commonality will be enforced wherever needed. It is important to note that SDC Leads are full-fledged members of the ST, and this simplifies the flow of information on the SGS architecture to the code developers, besides helping in the creation of consensus within the SGS.

SDC Leads are also members, together with the OU Leads and the ECSGS PO, of the Organization Group (OG). The OG is a forum where the problems related to data processing are discussed: currently, focus is especially on algorithms preparation, code integration and testing. In the near future, popular topics will be data flow and processing operations. OG face-to-face meetings are scheduled three or four times per year.

Another important forum for the ECSGS are the so-called Garage Days, where OUs discuss with the SWGs on aspects such as flow-down of requirements from scientific to technical, algorithms, validation. They are usually coupled with OG meetings.

There is SGS participation in collaboration-wide working groups. OU and IOT members participate in the Calibration WG that discusses instrument calibration issues during ground testing and operations.

The SOC Development Manager and the ECSGS Manager participate in the Extended Ground Segment Progress Meetings (EGSPMs), another important forum of discussion where the managers from the Euclid Project, the EC, MOC,

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

SOC and ECSGS discuss interfaces and commonalities between the various elements of the whole Ground Segment, both OGS and SGS.

Within the EC an important role is played the Euclid Consortium Coordinating Group (ECCG), a forum where the leaders of instrument development, SWGs, ECSGS, survey planning etc. discuss items relevant to the interaction of the various systems of the Consortium. A subset of the ECCG with executive responsibilities is the EC Management Group composed of the EC Lead and the managers of the two instruments and the ECSGS.

4.2 A data-centric approach: the Euclid Archive

Key features of Euclid are the amount of data that the mission will generate, the heavy processing needed to go 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 Archive System (EAS), 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 EAS is performed by a set transversal components that allow insertion of processing orders and smooth interaction with the processing elements at SDC, as well as monitoring and control functions and data quality verification.

As mentioned in the previous section, the huge volume of data to be processed in the SDCs requires a distributed storage to avoid data migration across SDCs. In the Euclid SGS there will be distributed data and processing: each SDC is both a processing and a storage node contributing to the overall SGS tasks. This means that the EAS archive will have bulk data distributed over several locations.

There is separation of metadata from data: a centralized metadata repository will be available (at SOC), containing "pointers" to the actual pixel data distributed geographically across the SDCs. For integrity purposes, both the metadata repository and the bulk data are mirrored. Furthermore, the EAS is logically composed of a set of functionalities supporting the data processing (DPS) and others supporting the scientific exploitation (SAS), plus storage and data transfer support (DSS).

4.3 Standards and guidelines

To manage the distributed software development process, standards and guidelines have been decided and implemented. They help software developers in taking the right decisions: e.g. by showing how/where to improve the code to meet the demanding requirements of the Euclid data processing, by encouraging the use of best practices and by providing tools to help developers improving their code.

The SGS uses a single development platform specifying operating system, programming language and support libraries. CODEEN is the Euclid collaborative development and continuous integration platform. It was important to define this environment early, since the cost of fixing bugs increases as the system integration approaches completion. Its usage is therefore mandatory for the main processing software.

Python and C++ have been adopted as the allowed languages for pipeline development, the drivers being an increased flexibility about who can contribute to development, and the long-term trends of programming in the field in astronomy.

A preliminary explicit Data Model (DM) has been built by the OUs to describe the output of their processing functions (therefore input to other Processing Functions in most cases). DM Workshops have been held with wide participation from OUs and System Team. The first iterations of the DM seem very promising, since real data products are being defined. The work being performed now is to increase the coverage to all products and maintain a flexible process to allow the DM to evolve in a controlled way along with the Processing Functions.

It is to be noted that the OUs are formed by EC scientists with knowhow in code development, and that the prototypes could well be developed using directly the common coding standards defined by the SGS System Team. This includes complying strictly since the very beginning with the common coding standards defined by the SGS System Team, using pre-defined input and output mechanisms.

4.4 Quality control and testing

In the SGS framework, quality control plays a crucial role. Strict Quality Assurance (QA) methods, coordinated between SOC and ECSGS, are applied on both code development and data quality checking. The QA procedures jointly agreed

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

include Configuration Management, and a Risk Management mechanism involving a team including one responsible for every Processing Function and coordinated by the ECSGS QA Lead, member of the Project Office.

The software produced for data handling, processing and analysis is built within a common development environment defined by an SGS System Team, which has been already active for several years. The code is built incrementally through different levels of maturity going from prototypes, developed mainly by scientists, to production code, engineered and tested at the SDCs. This approach is shown in Table 1.

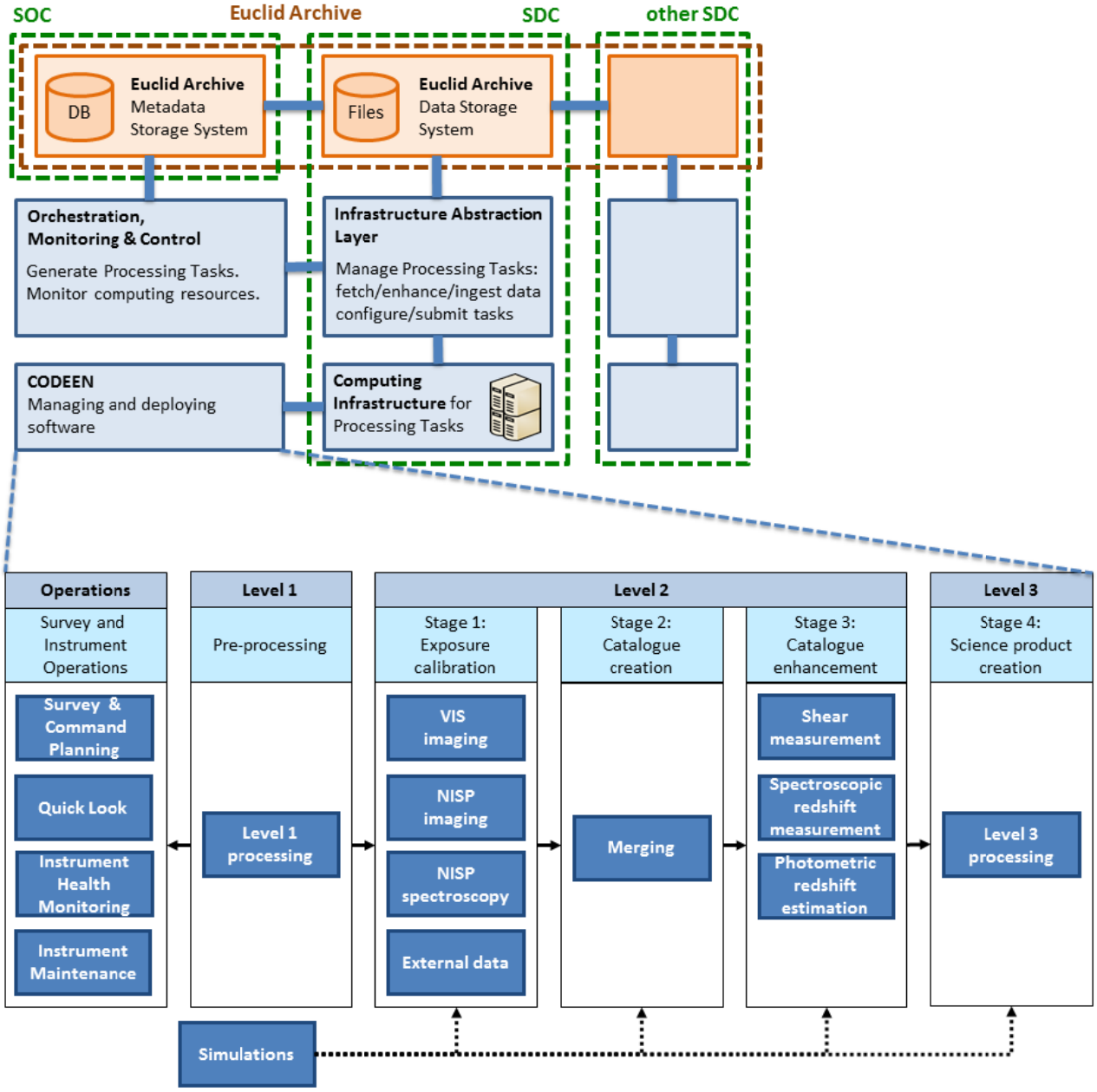

Figure 4 – On the top of the figure, the common elements in the Euclid SGS: the Archive, the Orchestration and Monitoring & Control elements, and the Infrastructure Abstraction Layer (IAL) that allows interfacing with the different non-homogeneous computing infrastructures of the SDCs. Finally, CODEEN is the software use to manage and deploy the software across the various sites (SOC and SDCs) and the various Processing Functions composing the Euclid SGS.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Maturity level (ML)	Status of the PF code vs the targeted production code	V&V level of the code	Compliance level of the code vs coding rules
0	Algorithm defined and/or identified	Algorithm assessed by paper analysis and/or simulation on preliminary test data	None
1	Algorithm code implemented in a high level language in an OU specific environment	(1A) Algorithm code validated in an OU specific environment with a set of OU internal test data	None
		(1B) Algorithm code validated in an OU specific environment with a set of representative test data	
2	Algorithm code implemented in C++ and/or Python, hosted by CODEEN, SGS DM compliant and IAL interface compliant. Local use of standalone libraries punctually admitted.	(2A) Algorithm code validated in the CODEEN environment with a local copy of representative test data	"Prototype" profile
		(2B) Algorithm code validated in a partial SDC-PROD environment with representative test data (managed by EAS) and dedicated V&V tools	
3	Algorithm code based upon the Euclid reference libraries, and potentially official 3rd party libraries	Algorithm code hosted by CODEEN and validated in the complete SDC-PROD environment with an extended set of representative test data and dedicated V&V tools	(3A) "Development" profile
			(3B) "Production" profile

Table 1 – Code maturity levels

The soundness of this development scheme is periodically checked. A set of formal reviews and technical checkpoints have been defined: Reviews follow the normal path (requirements-design-implementation-readiness), while the Technical Checkpoints are less formal and aimed at assessing the status of development of the SGS on the basis of presentations.

A number of incremental Challenges (Infrastructure, Data Processing and Integrated) have been furthermore included in the Euclid SGS test plan to verify the correctness and accuracy of the developed systems. This mechanism is aimed at verifying practically that the system being designed and incrementally implemented is sound, technically feasible, scalable and capable of being run in a distributed environment providing consistent results. Challenges are implemented at all SDCs, and are not considered as passed until all SDCs fulfil the requirements. Five IT Challenges and the first Science Challenge have been successfully completed, while the sixth IT Challenge and second Science Challenge are underway at the time of writing. Table 2 contains a list of the challenges completed and underway. In Figure 5 the Reviews, Technical Checkpoints and Challenges are shown on a time axis, from mission adoption until launch.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

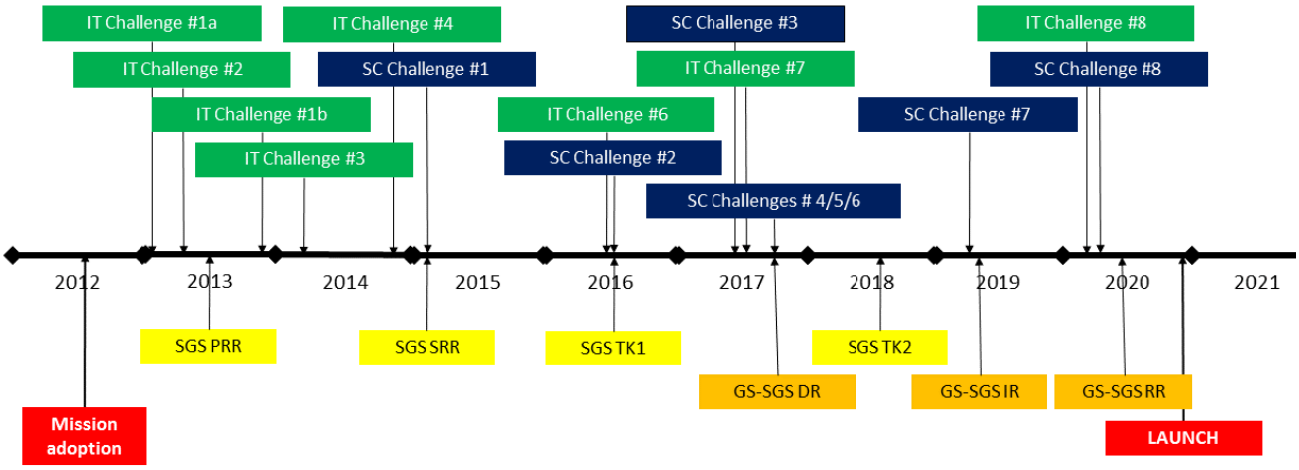

Figure 5 – Reviews and Technical Checkpoints (in yellow for SGS only, in orange for both OGS and SGS), and Challenges (in green the IT ones, in dark blue the Science ones), from mission adoption until launch.

Challenge	Purpose	End Date
IT Challenge #1 (SDC-DE)	Test network bandwidth between SDCs and with SOC	10 Jan 2013 (1a) 31 Jan 2014 (1b)
IT Challenge #2 (SDC-FR)	Check the feasibility to deploy and run a prototype (TIPS) from CODEEN through a Jenkins slave or manually	24 Apr 2013
IT Challenge #3 (SDC-FR)	Deploy IAL (Infrastructure Abstraction Layer) Virtual Machines into SDCs, allow IAL to fetch input data, launch SIM pipelines, publish results	7 Feb 2014
IT Challenge #4 (SDC-UK)	Euclid Archive System (EAS), Monitoring & Control (M&C), COORS prototypes	15 Dec 2014
SC Challenge #1 (OU-SIM)	Release of VIS, NISP-S and NISP-P field simulations from "true universe" and mission parameters (set of data releases)	1a – Sep 2014 1b – Oct 2015 1c – Dec 2015 2 – Apr 2016
IT Challenge #6 (SDC-UK)	IAL, EAS DPS & DSS, M&C improvement, COORS Processing Plan execution, performance & scalability test	29 Apr 2016
SC Challenge #2 (SDC-IT)	VIS, NIR & SIR Processing Functions prototypes integrated in SGS infrastructure	31 May 2016
SC Challenge #3 (ST + OU-MER)	Integration and validation of the Euclid Process Pipeline including SIM/VIS/NIR/SIR/EXT/MER	May 2017 (estim.)

Table 2 – IT and Science Challenges completed and underway, the teams responsible for their execution (in parentheses), their purpose, and the dates of completion.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

5. CONCLUSIONS

The main challenges in the design and implementation of the Euclid SGS are expected to be the distribution of the development teams in many institutes throughout Europe and the United States, plus the distributed nature of the data processing during operations. The size of the data set, the complexity of the processing, and the needed accuracy of the results add to the complexity of the system being built. 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. Data management will be performed in a distributed environment, implying quite a complexity in the design and implementation of the data handling and processing facilities.

To cope with this complexity, the Euclid SGS management has established an organization based on collaborative software development, virtualization, distributed processing and a data-centric approach to the system architecture. This organization is supported by a management structure where SOC and ECSGS has started a tight collaboration since the proposal phase of the mission, and jointly participate in a SGS System Team that defines standards and provides a single development platform specifying operating system, programming language and support libraries: a collaborative development and continuous integration platform.

The code is built incrementally through different levels of maturity going from prototypes, developed mainly by scientists, to production code, engineered and tested at the SDCs. A number of incremental challenges (infrastructure, data processing and integrated) have been included in the Euclid SGS test plan to verify the correctness and accuracy of the developed systems.

ACKNOWLEDGEMENTS

The authors are indebted to Giuseppe Racca, René Laureijs, and Yannick Mellier, who lead different aspects of the Euclid project, to Cyril Colombo and Jerome Amiaux for many useful suggestions, and to all the individuals participating in the SGS development within ESA and the Euclid Consortium, too many to be listed here. The participation in the Euclid SGS activities is supported by the National Space Agencies; in particular, the Italian team acknowledges the financial contribution from the agreements ASI/INAF I/023/12/0 and I/023/12/1.

REFERENCES

[1] Laureijs, R. et al., “Euclid Definition Study report”, ESA report ESA/SRE(2011)12, arXiv:1110.3193 (2011)
[2] Racca, G. et al., “The Euclid mission design”, Proc. SPIE 9904 (2016), in press
[3] Cropper, M. et al., “VIS: the visible imager for Euclid”, Proc. SPIE 9143 (2014)
[4] Maciaszek, T., et al., “Euclid Near Infrared Spectro Photometer instrument concept and first test results at the end of phase B”, Proc. SPIE 9143 (2014)
[5] Pasian, F. et al., “Science Ground Segment for the ESA Euclid mission”, Proc. SPIE 8451 (2012)
[6] Pasian, F. et al., “Organization of the Euclid Data Processing: dealing with complexity”, Proc. ADASS XXIV, ASP Conference Series 495 (2015)

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 21 Apr 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use