Science Ground Segment

M. Sauvage - SGS Scientist
Euclid France 2015

Foreword - What is the SGS?

- The SGS does not produce measurements of the Dark Energy EoS, or any statement regarding alternate theory of Gravitation.
- The SGS does not produce statements regarding Galaxy Evolution or the Primordial Universe.
- However, the SGS is tasked with turning the measurements made by Euclid (wide-field photometric exposures and slit-less grism exposures) into data products from which the above results can directly be extracted.
  - Correlation functions, power spectra (and associated “errors”) for shear and positions.
  - Source catalogs containing, photometry, spectroscopy (lines and fluxes), redshifts (photometric and spectroscopic), shapes (ellipticities, morphologies), physical parameters (for legacy studies).

This is the “science” part of the SGS task

- To be blunt, the SGS performance is not measured by its ability to reveal anything about the underlying laws of physics, but rather by its capacity to extract a pre-defined set of observables from a given sky and with given instruments (including margins on these assumed elements), with a pre-defined level of precision and accuracy.

Operational structure within the SGS

- For the implementation phase, the SGS consists mainly of two series of entities:
  - The Organization Units (OUs)
  - The Science Data Centers (SDCs)

Operational structure within the SGS

- For the implementation phase, the SGS consists mainly of two series of entities:
  - The Organization Units (OUs)
  - The Science Data Centers (SDCs)

The SDCs are built around national HPC facilities, they gather IT support as well as developper expertise.
SDCs are in charge of the pipeline development (software & support architecture).
Production software will run in the SDCs.
For France the SDC is the CCIN2P3.

The OUs group EC members according to their data processing expertise.
OUs are in charge of analyzing the science data processing requirements, and of producing the pre-integration version of the pipeline modules.
Demonstrating that the DP requirements are met is an OU responsibility.

Operational structure within the SGS

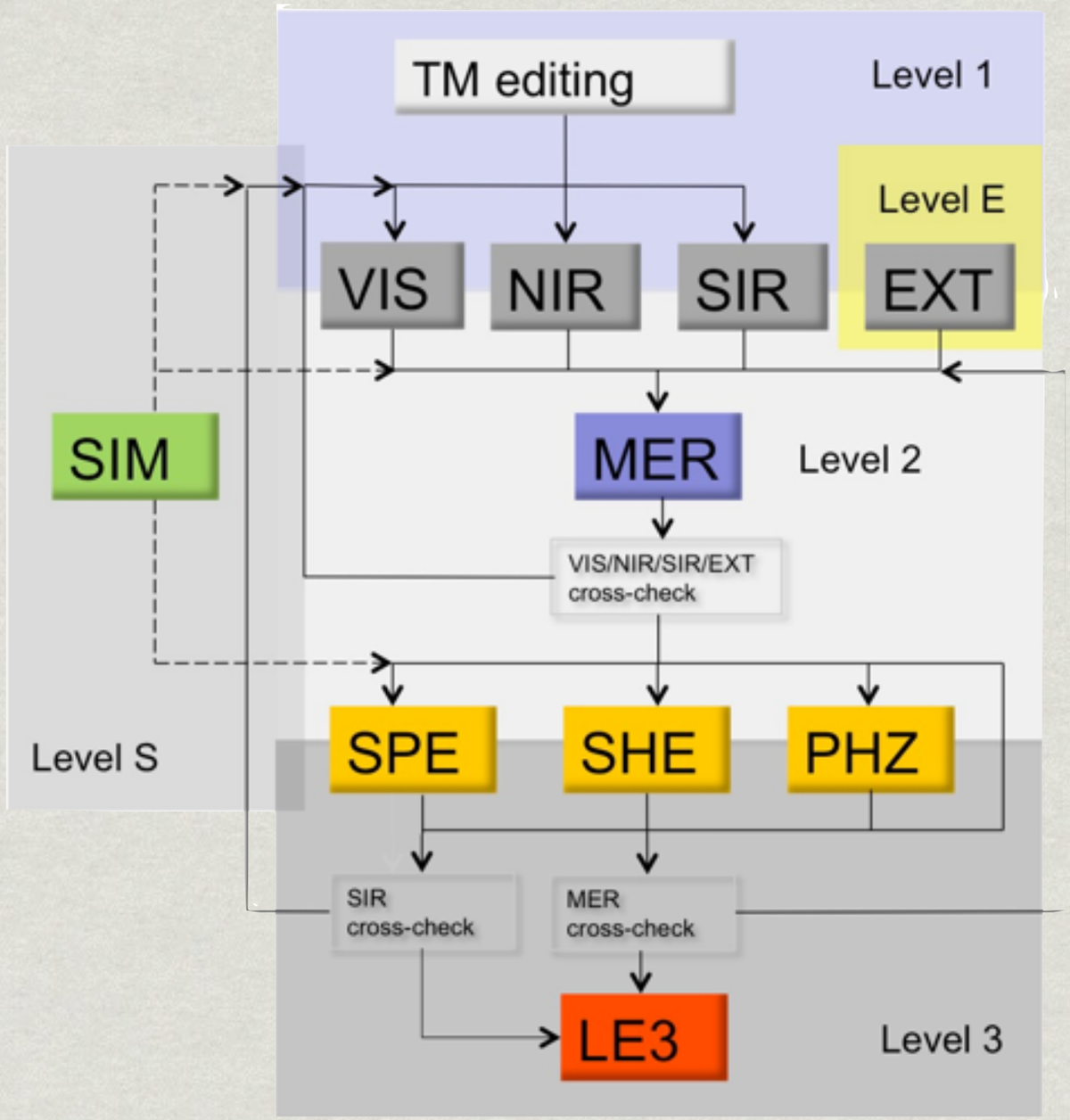
- For the implementation phase, the SGS consists mainly of two series of entities:
  - The Organization Units (OUs)
  - The Science Data Centers (SDCs)

The SDCs are built around national HPC facilities, they gather IT support as well as developper expertise.
SDCs are in charge of the pipeline development (software & support architecture).
Production software will run in the SDCs.
For France the SDC is the CCIN2P3.

The OUs group EC members according to their data processing expertise.
OUs are in charge of analyzing the science data processing requirements, and of producing the pre-integration version of the pipeline modules.
Demonstrating that the DP requirements are met is an OU responsibility.

OUs have been created following a plausible but a priori division of the pipeline in logical blocks.

The mapping of OUs on the pipeline



- VIS, NIR, EXT: production of fully calibrated photometric exposures from Euclid and ground-based surveys
- SIR: production of fully calibrated 1D spectra extracted from the NISP spectroscopic exposures.
- MER: production of a source catalog containing consistent photometric and spectroscopic measurements.
- PHZ: production of the photometric redshift for all catalogued sources.
- SPE: production of spectroscopic redshifts for all sources with spectra.
- SHE: measurements of galaxy shapes.
- LE3: production of all high-level science products.
- SIM: production of all the simulated data necessary to validate the data processing stages, and to calibrate observational or method biases.

From OUs to processing functions

- Yet another acronym… but at attempt at clarification:
  - OUs and SDCs are groups and structures (sometimes physical structures). They collaborate to design and build a system, which is going to be the Euclid data processing pipeline and its infrastructure.
  - The pipeline elements, which are the result of the joint OU and SDC efforts are called the Processing Functions.
    - As we cannot fully redeem ourselves, the OU and the processing function it creates have the same name…

The Euclid data processing pipeline will be:
a series of Processing Functions,
designed by the OUs,
developed in collaboration between the OUs and SDC developers,
integrated by the SDCs,
and running on the SDCs infrastructure.

A personal aside

- The initial construction of the SGS with OUs in charge of prototyping the data processing algorithms and SDCs in charge of running the pipelines was unfortunately a way to hide how the pipelines themselves would be produced.

- The SGS is hitting that phase now, the development, and is feeling strongly the scarcity of software developers in the consortium.
  - This is probably an item that is going to appear often in discussions with funding agencies.

Developments since the last EF meeting

- System Requirements Review passed
- Participation of the SGS in Science Performance Reviews
- Future reviews of the SGS

System Requirements Review is passed!

- SRR objectives: verify that at the SGS level (OUs, SDCs, System team), the requirements coming from the flow-down analysis have been correctly analysis and transferred to the relevant units.
- This has been an extremely heavy process:
  - 66 documents in the data pack, 430 RIDs, 300 Actions.
  - Closed for almost all of them in July 2015 (more than 6 months between the kick-off and the closure, not counting preparation).
- Not always efficient:
  - It is hard for us to meet the panel's expectations at colocations meetings (for lack of enough time for preparation or discussion).
  - Document review is inhomogeneous (depending on the reviewer).
  - The system we use to track our exchanges with the panel is not really helping (not responsive).

SRR - Main recommendation

- Two main concerns were identified:
- Organize the LE3 OU so that its management principles are clear and compatibility with likely resources can be demonstrated.
  - At the time of the SRR, the LE3 processing function appeared as a very large collection of functions in different states.
  - Sizing for the computing resources requirements were showing LE3 as an outlier in any dimensions.
  - This has been addressed both in term of restructuring the LE3 tasks, and in terms of sizing studies.
- Interfaces between the SGS and its “partners” within and without the Euclid Consortium.
  - Identify the actual partners of the interfaces (IDT, Telescope, Pointing systems…)
  - Identify the nature of the interfaces (what is exchanged between these partners)
  - Identify the mechanisms of these interfaces (e.g. the Mission Data Base) and how the SGS can use them.

Interfaces within and without the SGS are going to be the top-level action of PO

Future reviews of the SGS

- We have proposed a modification of the review process to make it more efficient with respect to our development needs.
  - Merge the Preliminary Design Review and Consolidated Design Review into in single Design review. This DR will follow the classical review process.
    - Given that a classical review is typically a 6-month halt for the “executive” structure of the SGS, we want to keep them to a minimum number.
    - Since the number of documents generally increases with time, the less formal the reviews, the more manageable they are (alternative is to go for subsystem reviews which we would like to avoid).
  - Create new rendez-vous with the review panel in the form of Technical Key-points.
    - Light-weight data pack, concentrating on the top level documents of the SGS.
    - Extended co-location meetings with the panel focusing in details on the OU and SDC work.
    - Preparation of the co-location meeting in advance to make sure the content of the co-location meeting is in line with the panel expectation.
  - Schedule:
    - TK1 - July 2016
    - DR - March 2017
    - TK2 - March 2018

Science Performance Reviews

- At the Mission PDR level (see J. Amiaux tomorrow), the question arose of how well the mission is doing with respect to its science objectives.
  - This is considered at the mission level and thus combines instrument performances to on-board and ground-based data processing.
- We are asked to contribute to the future mission reviews with quantitative elements on the performance of the data processing systems (e.g. the processing functions).
  - Not exactly the End-to-End philosophy as it will use the actual PF (or their prototypes).
  - We are not required to produce a proto-pipeline (or are we?), disconnected runs of individual PFs are acceptable.
  - We need to define/agree on which input/simulated real-universe data are meaningful in this context.
  - This is committing ourselves at a level which is larger than before, and a schedule over which we have little to no control.

Practically - I

- At mission level, Science Performance Review Cycles are defined
  - SPR01 - Post-Adoption performance review - January 2012.
  - SPR02 - Feb-Sep 2017.
  - SPR03 - Sep 2018 - May 2019.
  - At the beginning of the cycle the objectives would be defined, and the cycle length is meant to accommodate the production of data (simulation) and their processing.

- General method foreseen:
  - Agree on the status of elements affecting the performance (Payload Module, VIS, NISP, AOCS).
  - Generate simulated data taking into account this state of the hardware (OU-SIM+sim WGs). These simulations will likely include short-cut simulations to directly feed higher-level PFs.
  - Demonstrate performance of PF with respect to requirements in whatever environment is suitable at the time (SGS).
    - *This implies that requirements are quantitatively defined (SWG).*

Practically - II

- How does it fit with our current schedule?
  - New SGS review schedule has now identified:
    - Technical key point 1 - July 1st 2016
    - SGS Design Review - October 15th 2017 **<- Conclusion of SPR02 September '17**
    - Technical key point 2 - July 1st 2018
    - SGS Implementation Review - April 25th 2019 **<- Conclusion of SPR03 May '19**
    - SGS readiness Review - May 10th 2020
  - We have a series of Integration and Science Challenges
    - In 2017 we have one integration challenge and three science challenges that would overlap with the SPR02 cycle.
    - Situation is less well matched for SPR03 as in 2019 we plan only one scientific challenge, all the integration challenges are in 2020.
  - Are there elements of activity planning on the SWG side?
    - If the plan is adopted, we have to make sure we have a consistent schedule regarding:
      - The existence of relevant quantitative requirements.
      - the availability of "realistic-enough" cosmological simulations.

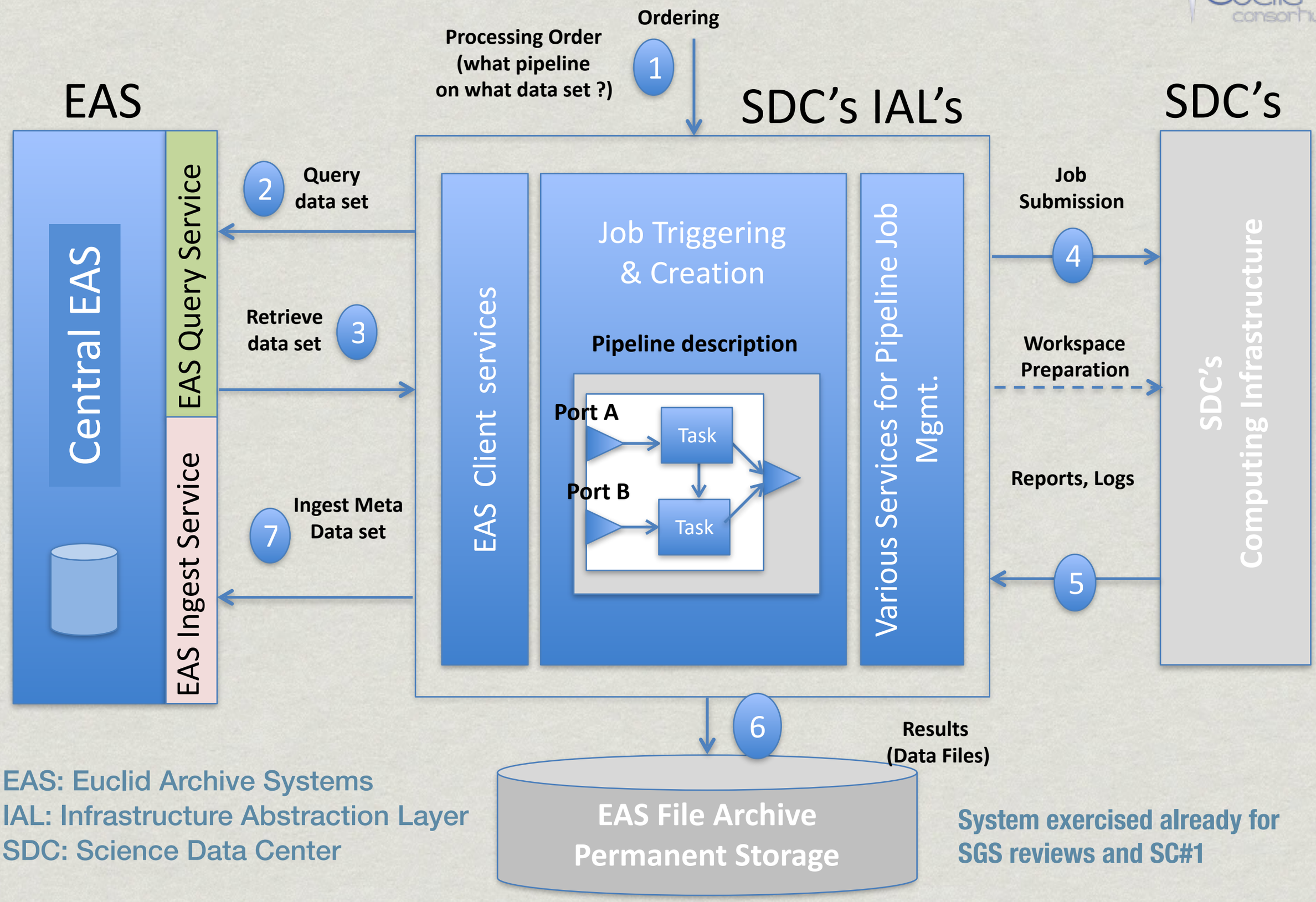
Practically - II

- How does it fit with our current schedule?
  - New SGS review schedule has now identified:
    - Technical key point 1 - July 1st 2016
    - SGS Design Review - October 15th 2017 **<- Conclusion of SPR02 September '17**
    - Technical key point 2 - July 1st 2018
    - SGS Implementation Review - April 25th 2019 **<- Conclusion of SPR03 May '19**
    - SGS readiness Review - May 10th 2020
  - We have a series of Integration and Science Challenges
    - In 2017 we have one integration challenge and three science challenges that would overlap with the SPR02 cycle.
    - Situation is less well matched for SPR03 as in 2019 we plan only one scientific challenge, all the integration challenges are in 2020.
  - Are there elements of activity planning on the SWG side?
    - If the plan is adopted, we have to make sure we have a consistent schedule regarding:
      - The existence of relevant quantitative requirements.
      - the availability of "realistic-enough" cosmological simulations.

In the opinion of the PO, these Science Performance Review are replacing what was formerly described as the E2E activity, as far as the SGS is concerned.

News from the Data Processing Systems

Data processing mechanism



EAS: Euclid Archive Systems
IAL: Infrastructure Abstraction Layer
SDC: Science Data Center

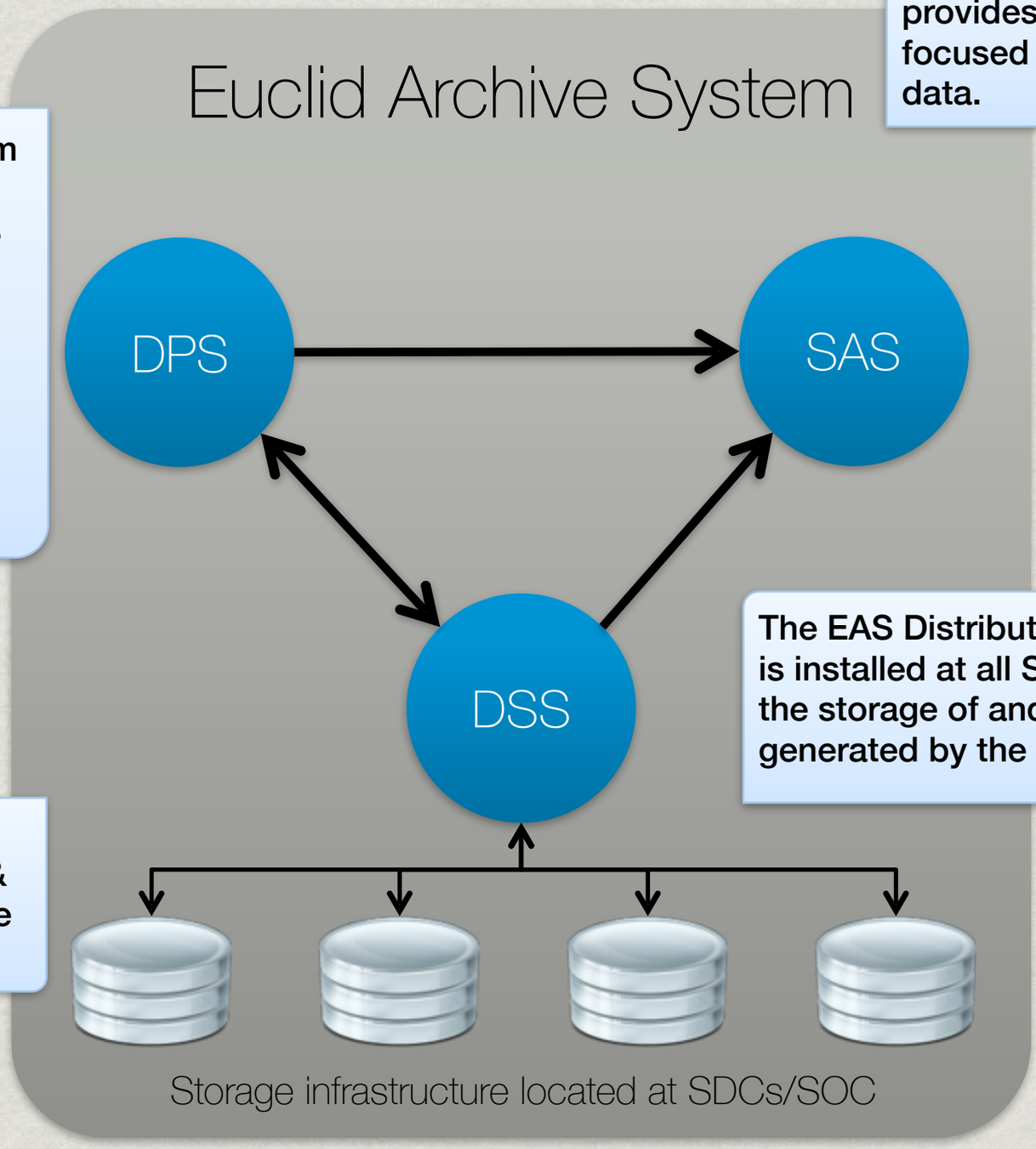
System exercised already for SGS reviews and SC#1

Archive overview

The EAS Science Archive System (SAS) provides access to the Euclid data focused on the scientific use of the data.

The EAS Data Processing System (DPS) stores the data product metadata including the locations of the data files.

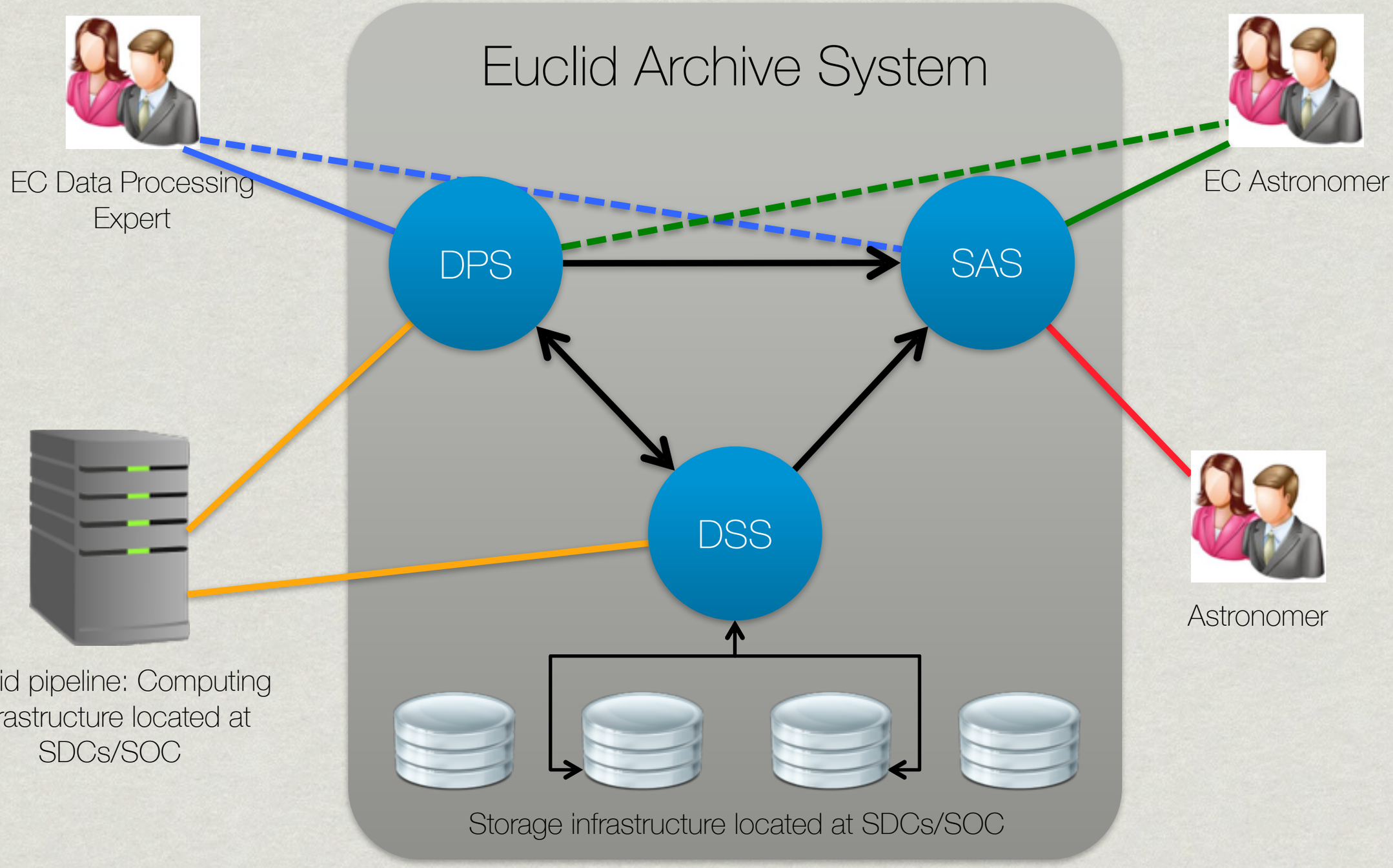
It provides access to the data products to the EC members, including processing coordination, quality control and processing history tools.



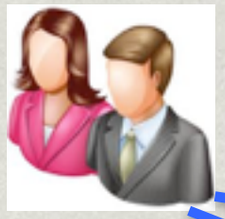
The EAS Distributed Storage System (DSS) is installed at all SDCs & SOC and manages the storage of and access to the data files generated by the pipelines.

Storage systems are provided by all the SDCs & SOC. All public data will be located at ESAC.

Major Archive users



Major Archive users



EC Data Processing Expert

Euclid Archive System

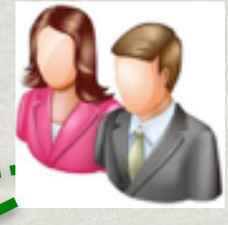
DPS

SAS

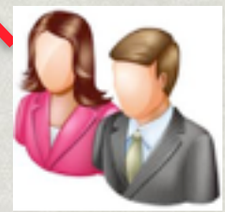
DSS

Euclid pipeline: Computing infrastructure located at SDCs/SOC

Storage infrastructure located at SDCs/SOC



EC Astronomer



Astronomer

We capture their needs (functionalities) through the Euclid Archive User's group

The Euclid Archive exists!

The Euclid Archive exists!

[euclid-ec-sgs-sys] [EAS-DSS] SDC-NL DSS server is down — Corbeille

A.N. Belikov

aujourd'hui 10:02

À : euclid-ec-sgs-sys@rssd.esa.int

[euclid-ec-sgs-sys] [EAS-DSS] SDC-NL DSS server is down

AB

Dear all,

sorry to inform you that due to GPFS crash DSS server in SDC-NL is down. The crash affects only production environment (http://application14.target.rug.nl:8008)

DSS server in production environment was restarted with the use of local storage which is limited to 100GB but files previously stored are temporally unavailable.

Regards,
Andrey

_____
euclid-ec-sgs-sys mailing list - euclid-ec-sgs-sys@rssd.esa.int

The Euclid Archive exists!

[euclid-ec-sgs-sys] [EAS-DSS] SDC-NL DSS server is down — Corbeille

A.N. Belikov aujourd'hui 10:02

À : euclid-ec-sgs-sys@rssd.esa.int

[euclid-ec-sgs-sys] [EAS-DSS] SDC-NL DSS server is down

AB

Dear all,

sorry to inform you that due to GPFS crash DSS server in SDC-NL is down. The crash affects only production environment (http://application14.target.rug.nl:8008)

DSS server in production environment was restarted with the use of local storage which is limited to 100GB but files previously stored are temporally unavailable.

Regards,
Andrey

_____
euclid-ec-sgs-sys mailing list - euclid-ec-sgs-sys@rssd.esa.int

More seriously the KIDS data (part of the ground-based data) are already released to the consortium through the Euclid Archive.