planck

euclid

Planck experience
and
Euclid expectations

A. Zacchei

Planck- DPCM, Euclid SGSM

INAF-OATs

on behalf of the Planck and Euclid Collaborations

Overview

- This is the LAST year for Planck project, we are packaging and delivering the last data.

- Planck was MY FIRST Space based mission and, the first mission will always be the *first love.* I hope, for this reason, that my experience is NOT too much biased.

- Most of the *Lessons Learned* points reported are extracted from the *Planck Lesson Learned* Document by Jan Tauber (PST Chair) and approved by the PST.

- I summarized the main points related to the Management, Operations and SGS structure reporting what we learn from Planck and what has been implemented/designed for Euclid

The Planck mission

- The Planck Satellite was launched on 14 May 2009 and concluded its operations on October 2013.
- It carried on board two Instruments: LFI (Low Frequency Instrument) and HFI (High Frequency Instrument) that observed the entire sky for about 4 years (we had two extension) in 9 different frequencies (from 30 to 857 Ghz).
- Its scope, in one sentence, was to map the microwave sky with an unprecedented combination of sky coverage, frequency range, calibration accuracy, independence from systematic errors, stability and sensitivity, to provide the most precise CMB and foregrounds maps.
- Planck was a "P.I. Mission" coordinated by ESA. Two consortia lead by two P.I. had the responsibility to operate the two instruments, process the data and deliver to ESA the agreed *products.* ESA, through its archive system at ESAC, was then in charge of making those products available.
- More than 140 papers has been published till now, covering all scientific aspects of the mission, with an impressive *impact factor.*

The Euclid mission

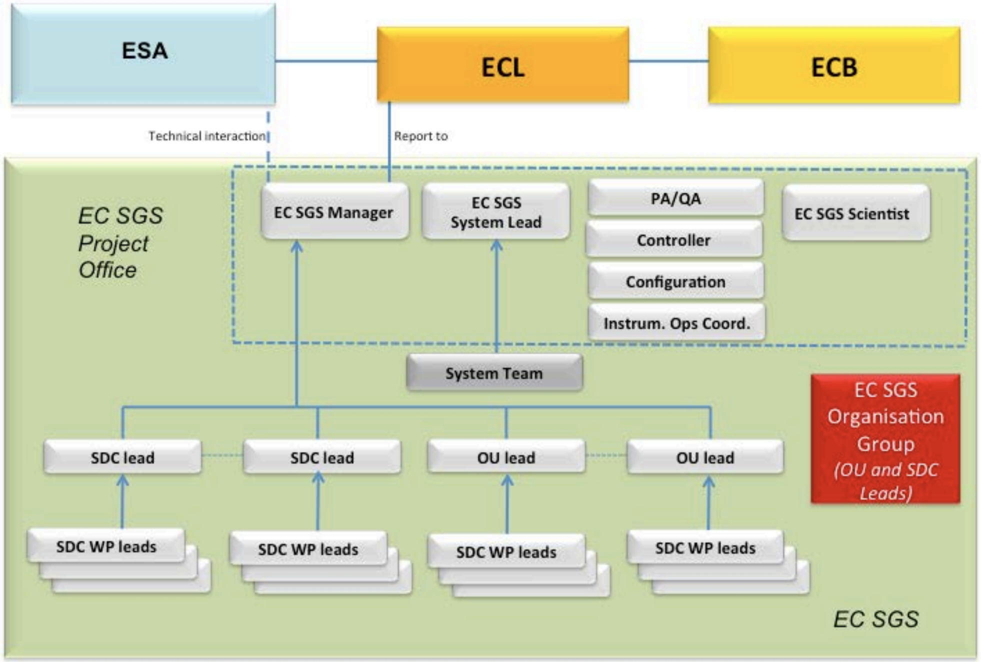
- Euclid is a space-borne survey mission dedicated to investigate the origin of the Universe's accelerating expansion and the nature of dark energy, dark matter and gravity. Euclid will characterize the signatures of dark energy on the 3D distribution of cosmic structures. In 2012, Euclid was approved as the second Medium Class mission (M2) in the Cosmic Vision Programme.
- Will be launched on 2021 and operate (baseline) for about 7 years.
- It will carry on board two instrument VIS (Visible Imager) and NISP (Near Infrared Spectrum Photometer).
- Euclid is a ESA Mission where the Data analysis and Instrument realization are in charge to a Consortium. The Consortium has the responsibility to reach the scientific goals and to provide to ESA the agreed *products.* ESA, through the SOC, will operate the instruments and through its archive system at ESAC, will be then in charge of making the products available.

The Planck management LL

- Planck was characterized by the large size of the two Consortia put together to develop and deliver to ESA the two Planck Instruments and generate the data products (dozens institutes with about 600 scientist at the peak).
- This heterogeneity caused many difficult problems that needed resolution in a day-to-day level. Before the operations the Herschel/Planck Project Manager instituted REGULAR meeting of the Consortium PIs and PMs together with representative of the major funding agency. This worked VERY well as a problem solving mechanisms during the development phase.
- The Planck Science Team, which included representatives of the two consortia and from the reflector provider consortium chaired by ESA, acted as a similar problem and conflict resolution entity for all matters related to Science and SGS.
- A very large amount of document has been generated over the life time of Planck , it seems that a centralized repository it is essential and a “centralized” librarian should be dedicated to make it work.
- Communicate with a so large project is very hard, each consortia held for many years a 2-days meeting with a typical interval of two months.
- Wiki has been found to be a very good way to record, discuss any item (avoiding mass e-mail).

The Euclid Consortia Management

- Euclid Consortium is bigger than Planck and is NOT PI centered.
- The main entity is formed by the ECB (Euclid Council Board) that is a forum where all the countries are represented.
- The Science Ground segment is leaded by the EC SGS Project Office that has the responsibility to develop and operate the SGS.



On the Instruments

- It is very difficult for mission like Planck, which depend on exquisite control of systematics, to know in advance which will be the most significant. An instrument systematics error budget should be develop already during the design phase and updated throughout development, operation and data analysis.
- Develop very early an instrument model which include all possible systematics, including those considered negligible.
- In Euclid we already starts creating a simulation machine able to incorporate instrument related systematics.
- Keep a margin for systematics in the performance budgets as long the pipelines are not fully developed.
- Ensure that will be possible to access in flight scientific data in a very raw state and be sure that all the HK will be available.
- Ensure that the on-ground development/test teams continue to support the flight operations and the data analysis. Those are the once that will maintain the Instrument History and knowledge.

Optical and pointing Information

- In Planck the optic knowledge was essential, is then suggested to establish early the complete chain leading from optical modelling to the use of this info in to the science analysis.
- The ability to extract optical information from the sky data must be evaluated at a very early stage.
- For Survey mission the stability is a key requirement, in Planck Thermoelastic deformation of the Star Tracker structure lead to apparent pointing variation of 20''. Understanding and correction of this effects takes time, it is then suggested to take into account in the mission planning that long timescale are needed to measure and analyze subtle pointing impacts on thermoelastic effects.
- The quality of the pointing reconstruction benefit significantly from close interactions of FD, MOC and Scientific users. It is crucial to keep them in sync and plan adequate resources.

The Planck Data Processing

- Two Planck DPC (Data Processing centers) have been the responsible of the operations and data analysis. Both follow the same overall approach to the data reduction with specific tasks aimed to correct instrument dependent systematics;

- In the initial design phase for efficiency/redundancy/cross-checking purpose was proposed that each DPCs should have been able to analyze the data of the other instruments. This deal to the construction of a Data Management Component (DMC), software language and pipeline builder (Process Coordinator) and a Federation Layer;

- Process has been then logically divided in four main levels:
  - Level 1 that was responsible to get directly the data form MOC, produce the Daily quality Report, operate the Instrument and transform HK and Science telemetry in raw timeline and store in a dedicated database with time information associated;

The Planck Data Processing

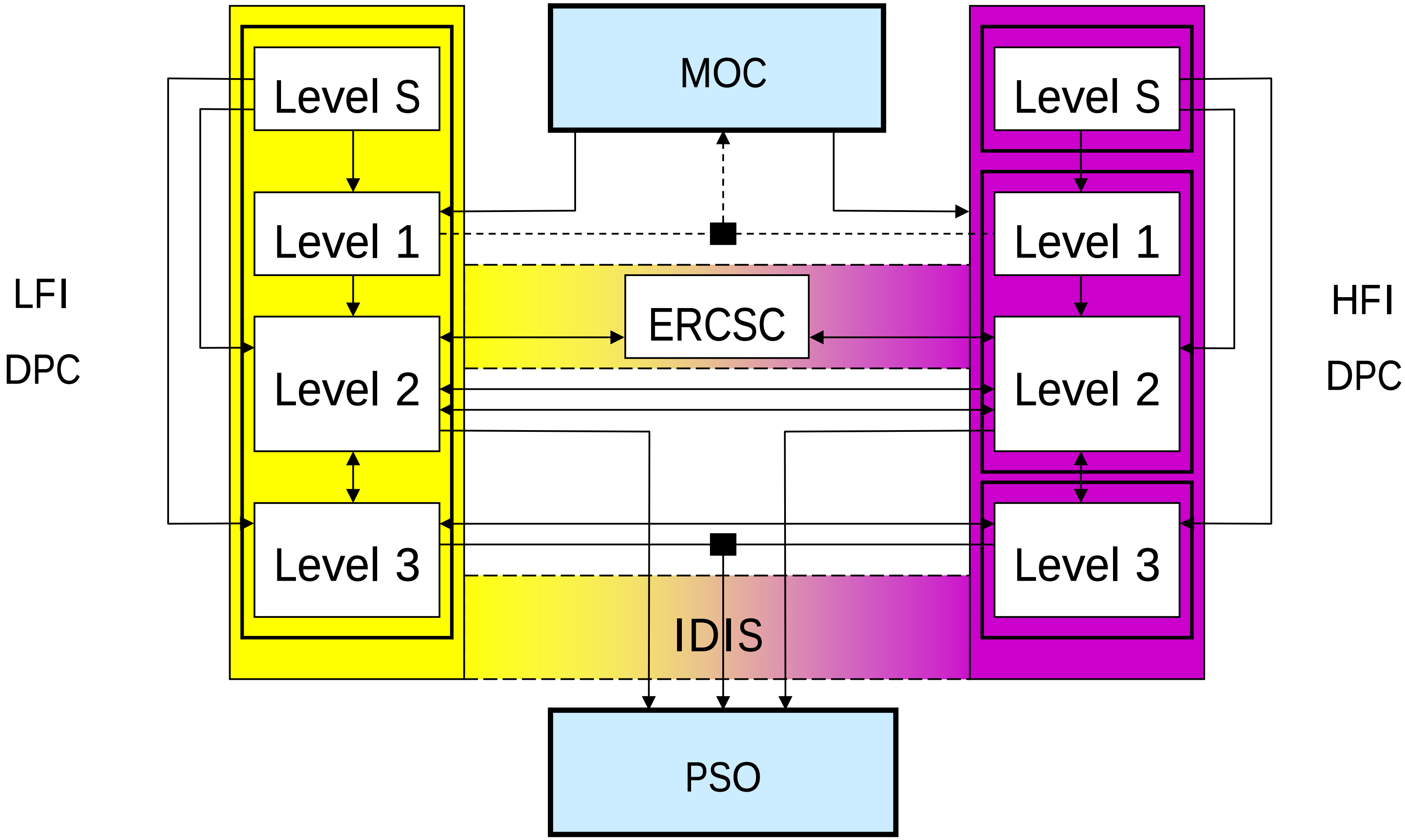
- Level 2 was dedicated to synthesize the instrument information in the Instrument Model, remove the systematics, flag the data that are considered not usable, calibrate the data and finally create the maps and all associated products;

- Level 3 was dedicated to the more scientific analysis with the responsibility to separate components into catalogues and specific astrophysical emissions, and evaluate CMB spectra and likelihood.

- Level S responsible to produce the required simulation needed to validate the pipelines.

- For every essential step of a pipeline for which a proven method does not exist, we develop at least two independent methods. Is important to make clear from the start how the final method will be selected and promote cooperation over competition.
- Each step was internally validated and most of the DPCs time was spent to cross check all the results first internally and then between instruments.

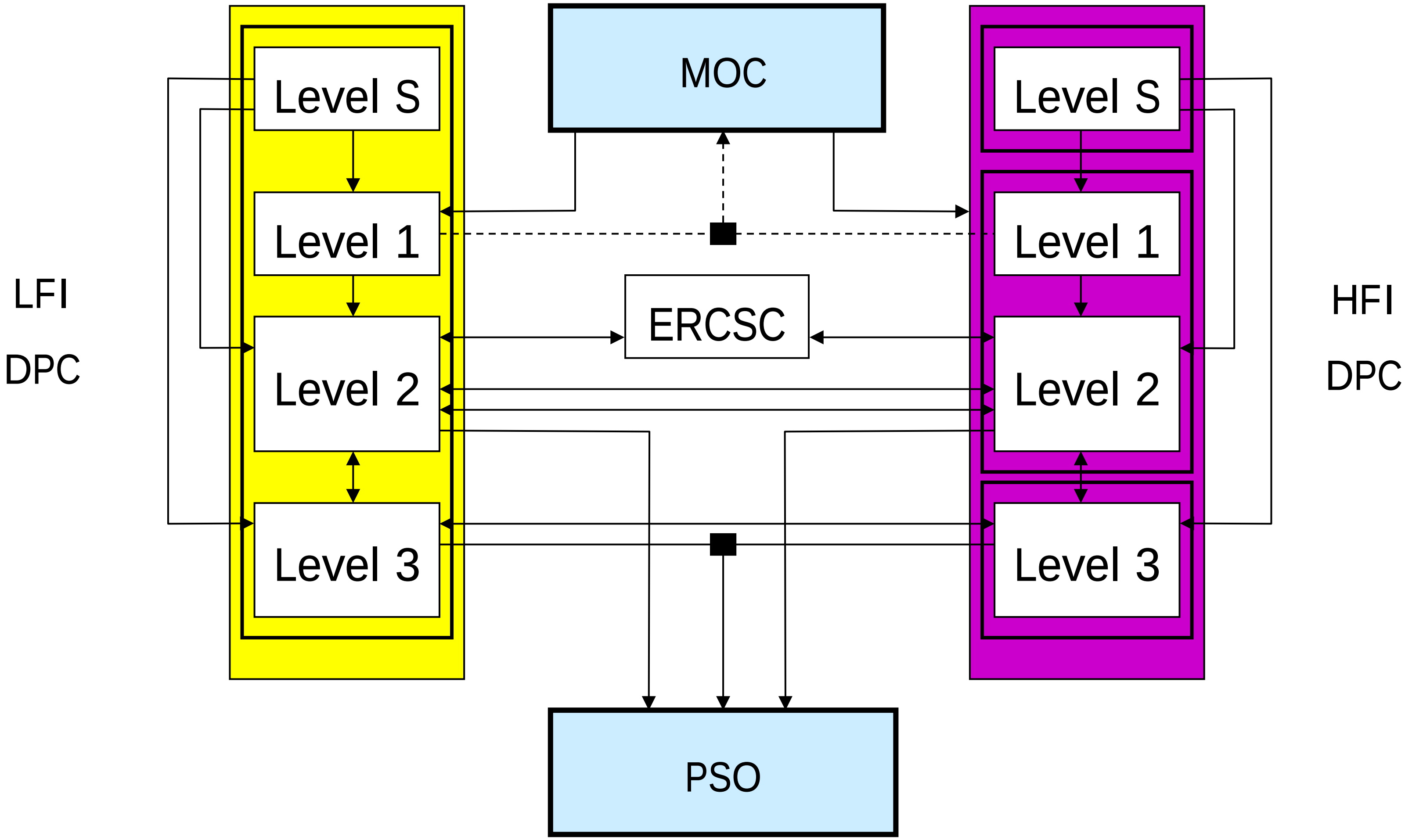
The Planck Data Processing how was designed



The Planck Data Processing

- After a long development period it became clear that some of the building blocks of the joint infrastructure were NOT meeting the needs of some of the parts.
- The principally cause was an ineffective management, the fact that Planck was a TWO PIs mission, the complexity of the prototype (killing a fly with a cannon) and the instrument knowledge geographically located.
- Only few features of the common development remained as the common simulation Tool (Basic-Level S).
- The common Infrastructure was then substituted by fixed interfaces that define the data to be exchanged at each level.
- The cost of this change on the design was paid in the Massive Montecarlo simulation needed to asses the scientific results, it was necessary to setup two different infrastructure.
- The advantage was that in such way pipeline and entire process was developed independently at each DPC adding **strong** value to the cross-instrument validation.

The Planck Data Processing how was Implemented



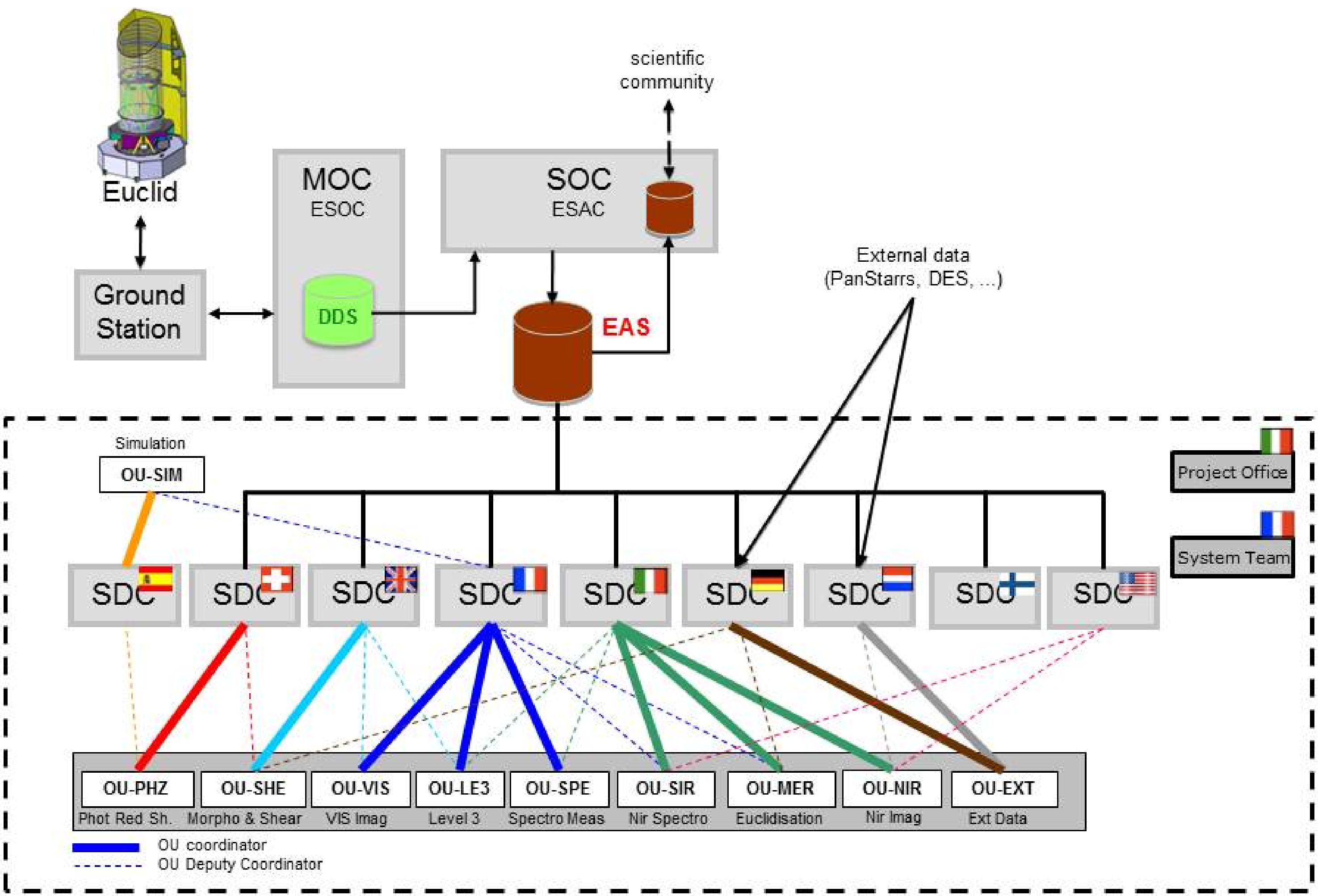
The Euclid Data Processing

- The SGS is then essentially a distributed structure.
- Euclid will produce and use a big amount of data (estimated to be at the end of the mission of the order of hundreds PB). It will be then essential to avoid excessive data transfer, to develop a structure where the code will be moved instead of the data.
- The various Science Data Center is providing different hardware
- Two languages (C++ and python) has been selected a Common Data Model and Common Infrastructure (hardware independent) has been built → EACH Science Data Center should be able to run the same code to process the data in parallel and be redundant.

The Euclid Data Processing

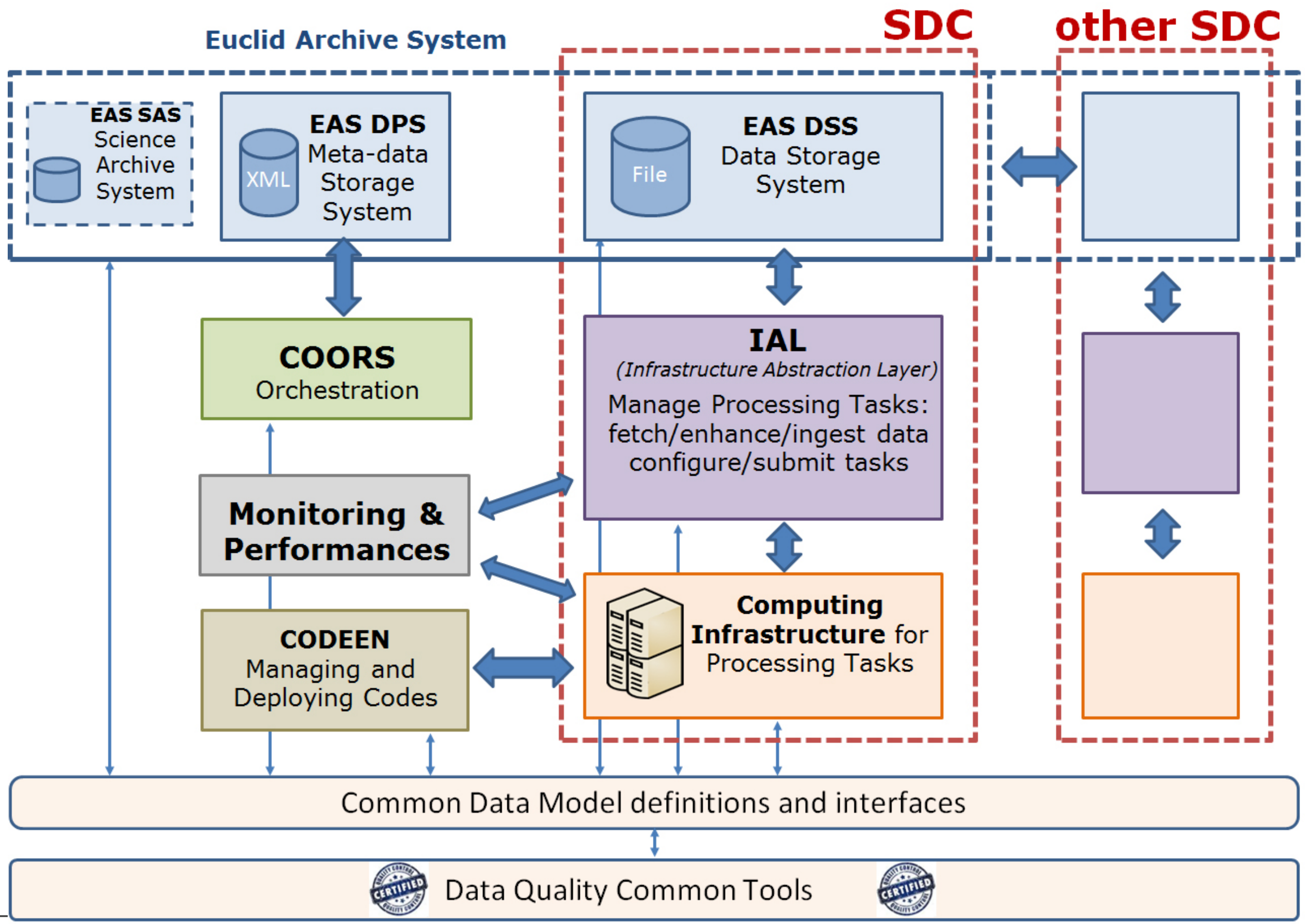
- The creation of a common infrastructure is very manpower demanding and require lot of test to make it SIMPLE. For this reason in Euclid we set different Instrument technology test that verify the entire infrastructure in each SDC.
- At the same time to facilitate the scientific code integration we institute the year based Developers Workshop with the aim to be a tutorial for Euclidian developers.
- The Data processing pipeline in Euclid will be a series of Processing Functions: designed by the OUs (Organization Units, scientist), developed in collaboration between the OUs and SDC developers , integrated by the SDCs, and running on the SDCs infrastructure.
- Processing Function are tested in to the SGS in growing complexity Scientific Challenges.

The Euclid Data Processing



The Euclid System Team

System Team actors in the SGS



The Planck PSO and Euclid SOC

- The Planck SGS was based on the concept that the entire data processing and instrument operations would be carried out by the PIs teams. This means that Planck didn't need a SOC. However, during the development, it was clear that a certain amount of coordination was required between the two instrument and MOC. This situation led to the creation of gradual strengthening of the Planck Science Office at ESAC. PSO had then the responsibility to distribute to the science community the Products.

- In case of Euclid SOC at ESAC play, from the beginning, a central role as unique interface with the MOC, responsible of the Level 1 and the centralized Euclid Archive System.

Simulations

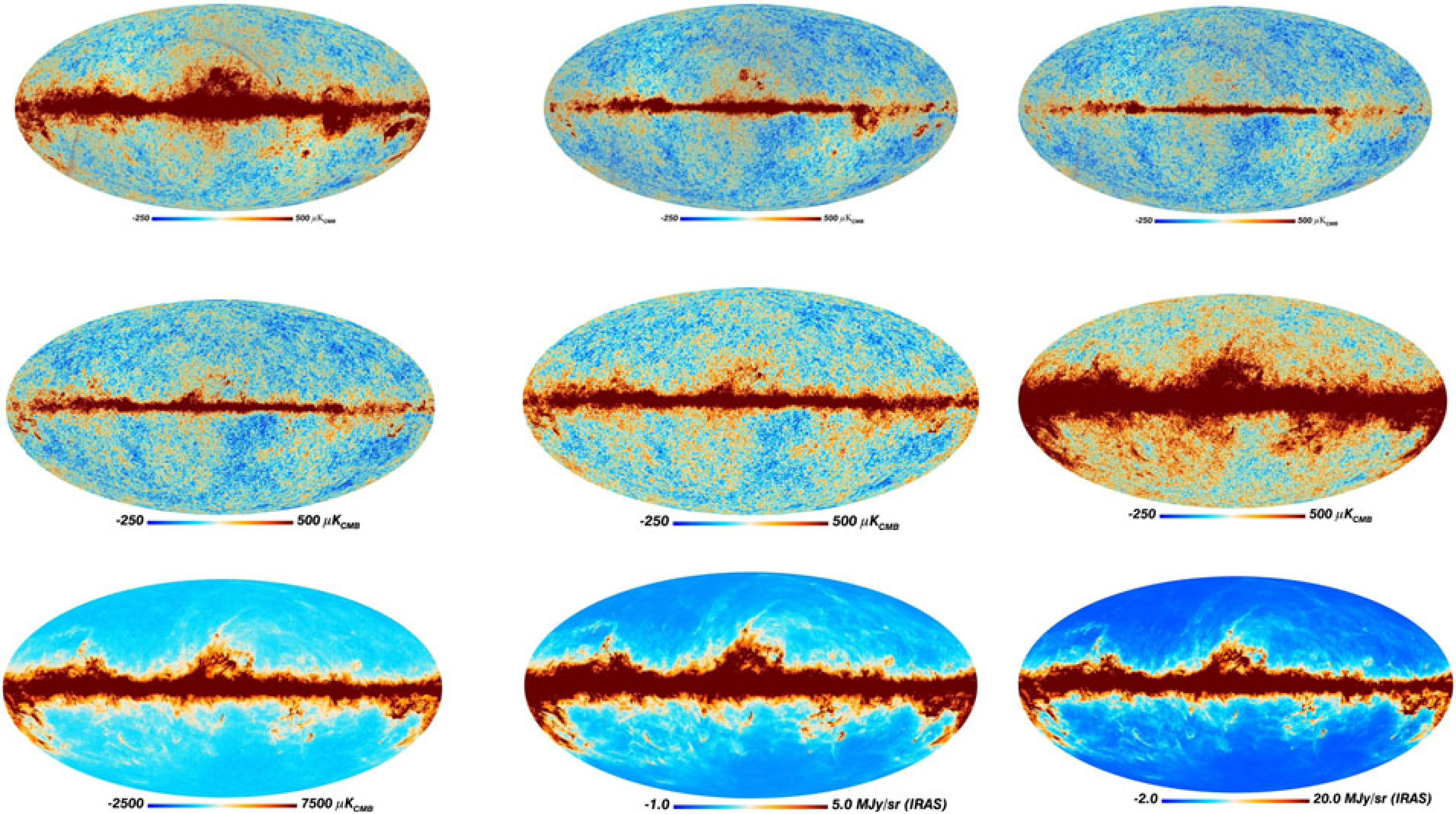
- Simulations are required in all stages of the mission development. Planck, by today standard, is quite a modest instrument, and yet the amount of supercomputing resources that it has used is huge and they constitute a limitation on the quality of some of the scientific results.

- Defining a long term plan for simulation at an early stage, and dedicate adequate resources and management to this effort will have an effect on the science.

Releasing the data and Publishing Paper

- Planck Data was distribute to the Public via the PLA developed by ESA's Archive Team (SAT) under the direction of PSO. Together with the products PLA made available an Explanatory Supplement with the scope to describe in details the products. The ES suffer the overlap with the paper production.
- The ultimate objective of the Planck Collaboration was to write scientific Paper. Managing the production of so big number of publication was not easy. The PST put in place an internal review system (Editorial Board) supported by the PSO. An SVN repository was used for all papers, a *Planck* Stile was established in a Planck Style Guide and Managed via latex tools, author list were centrally managed.
- This process had as a consequence the production of a very high quality paper and the organization of them together with the data release.

Planck- Temperature maps



Euclid SCh 3 Simulations

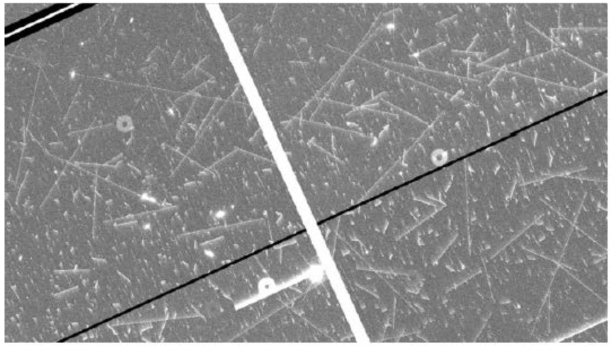
- Simulation from the last Euclid SCh3.

On the top

- VIS Frame

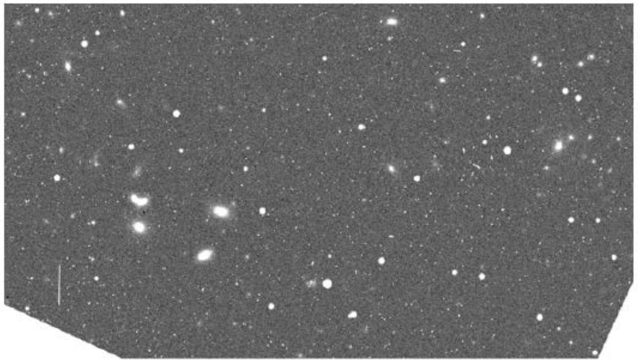
On the bottom from left to right

- NISP-Photometric
- NISP-Spectroscopic



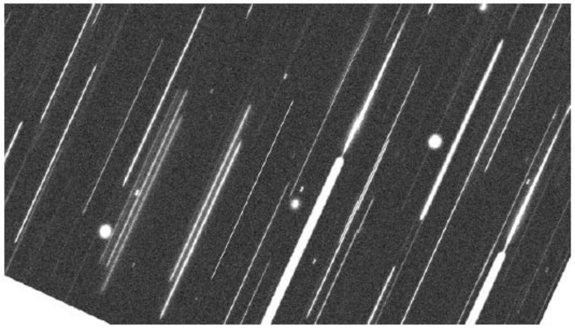
VIS
(Science + Flats + Bias)

Stars (r<24.0 - 24.5)
Galaxies (r<24.0 - 24.5)
Cosmic Rays
Photon Noise
Optical Ghosts
Zodiacal Light
AOCS PSF
Optical PSF
Geometric Distortions
Calibration Lamp
ADC Conversion
ADC Offset
Discretization
ROE non linearity
PRNU
CTI
Saturation Bleeding
Pre/Overscans
Readout Noise



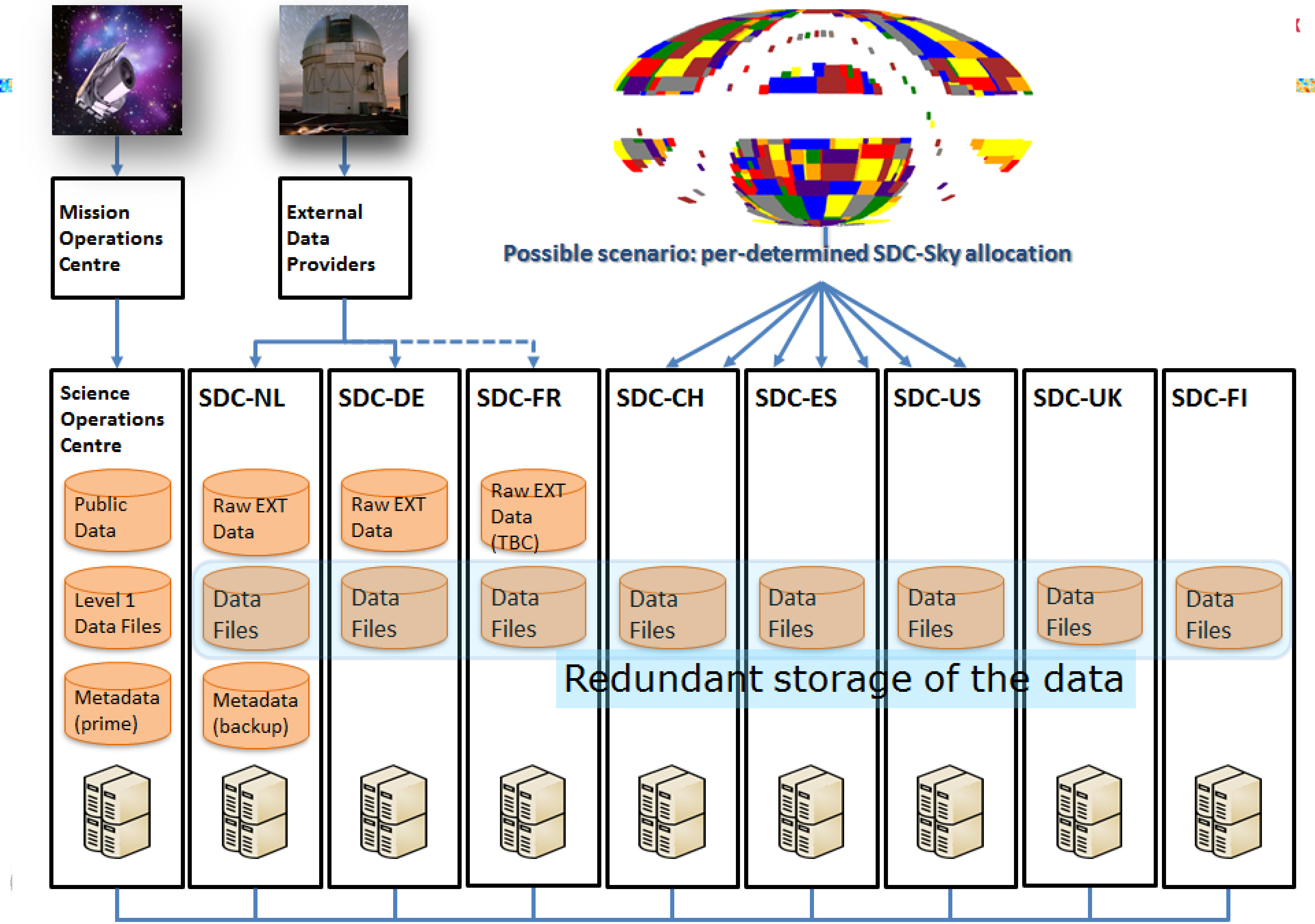
NISP-P
(Science + Dark + Flats)

Stars (r<24.0 - 24.5)
Galaxies (r<24.0 - 24.5)
Cosmic Rays (integration and readout)
Diffuse Scattered Light
Thermal BG
Zodiacal Light
AOCS PSF
Optical PSF
Geometric Distortions
FPS Illumination
Crosstalk + IPC
QE non linearity
Discretization
IntraPixel QE variation
Cosmetic Defects
Dark Current
Saturation Bleeding
Readout Noise
Poisson Noise



NISP-S
(Science+chi2)

Stars (H<22.5 - 23.0)
Galaxies (H<22.5 - 23.0)
Diffuse Scattered Light
Thermal BG
Zodiacal Light
Optical PSF
0th, 1st and 2nd Grism orders
FPA Metrology
Shutter
QE non linearity
Cosmetic Defects
Dark Current
Saturation
Readout Noise
Poisson Noise


Mission Operations Centre
External Data Providers
Possible scenario: per-determined SDC-Sky allocation
Science Operations Centre
Public Data
Level 1 Data Files
Metadata (prime)
SDC-NL
Raw EXT Data
Data Files
Metadata (backup)
SDC-DE
Raw EXT Data
Data Files
SDC-FR
Raw EXT Data (TBC)
Data Files
SDC-CH
Data Files
SDC-ES
Data Files
SDC-US
Data Files
SDC-UK
Data Files
SDC-FI
Data Files
Redundant storage of the data