The Euclid Science Ground Segment

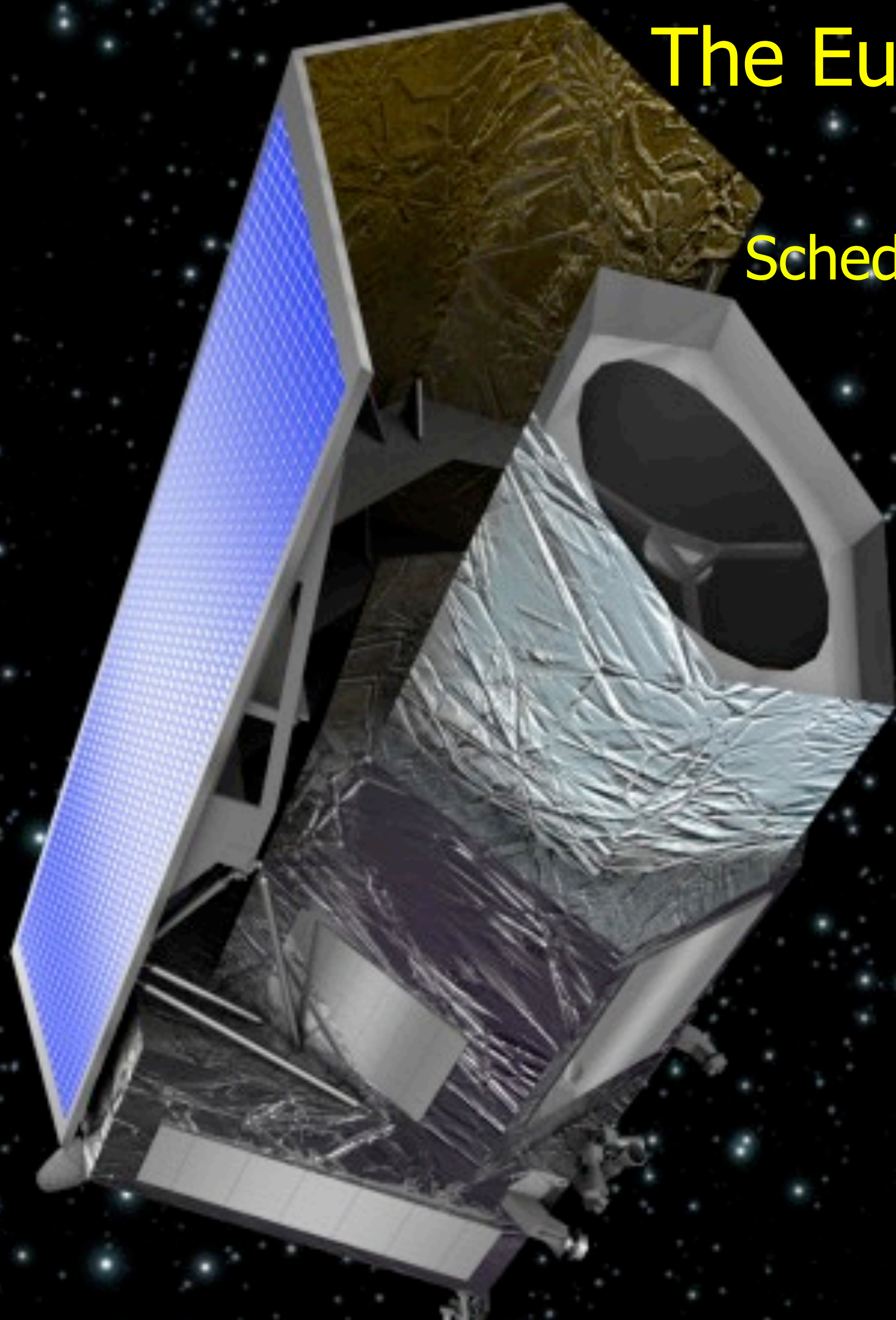
Marc Sauvage (EC SGS Scientist)
with contributions from
Fabio Pasian (EC SGS Manager)
John Hoar (SOC Development Manager)
Christophe Dabin (SGS System Team Lead)

(on behalf of the many scientists and engineers working on the Euclid SGS at SOC and in the EC)

The Euclid space mission

M-class mission

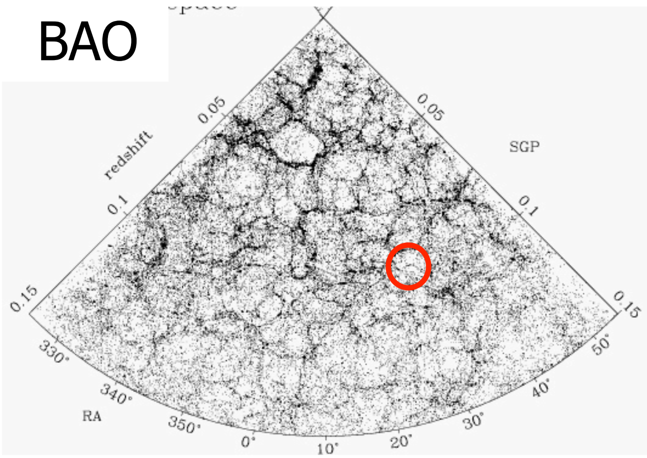
Scheduled for launch in 2020

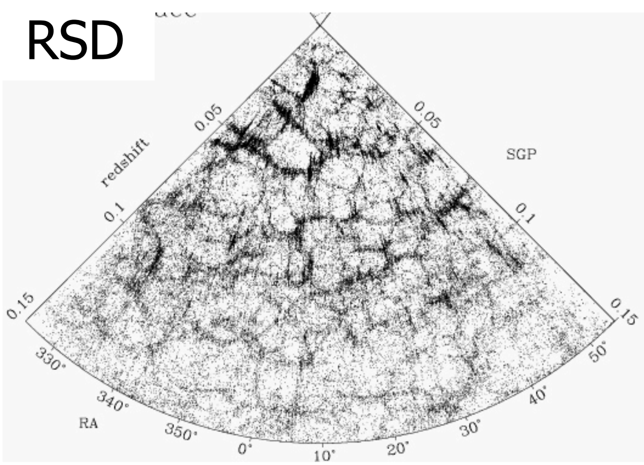


Euclid primary probes:

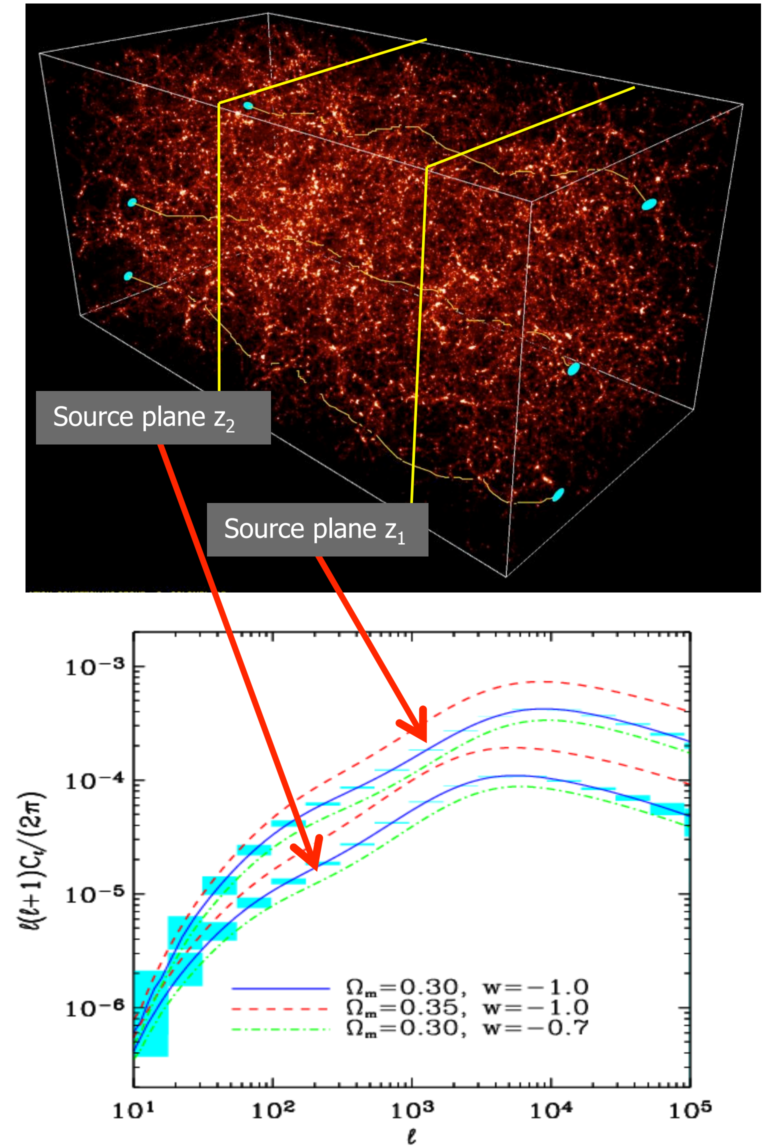
BAO, RSD and WL over 15,000 deg^2

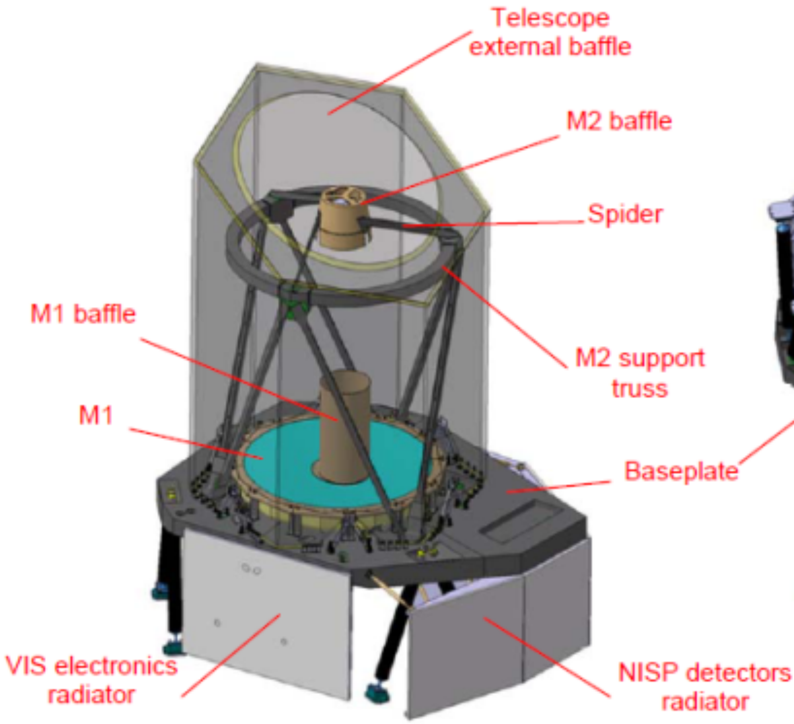
75 million galaxies with redshifts

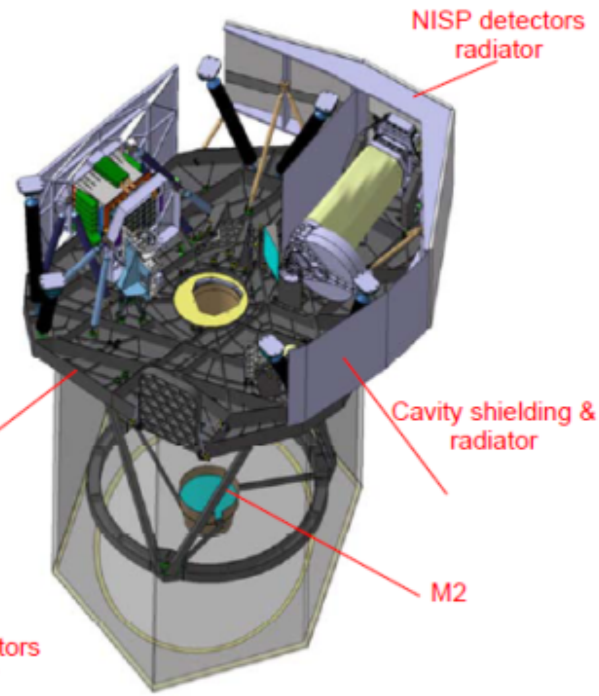




1.5 billion sources with shapes, 10 slices







Euclid:

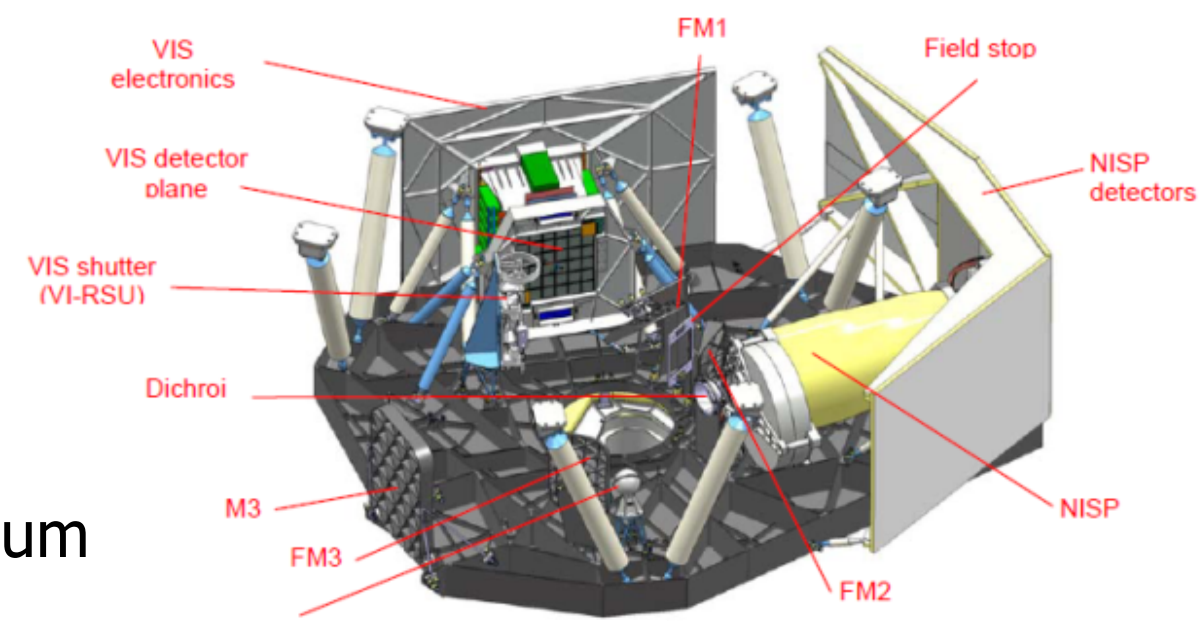
telescope and instruments

- Stabilisation:

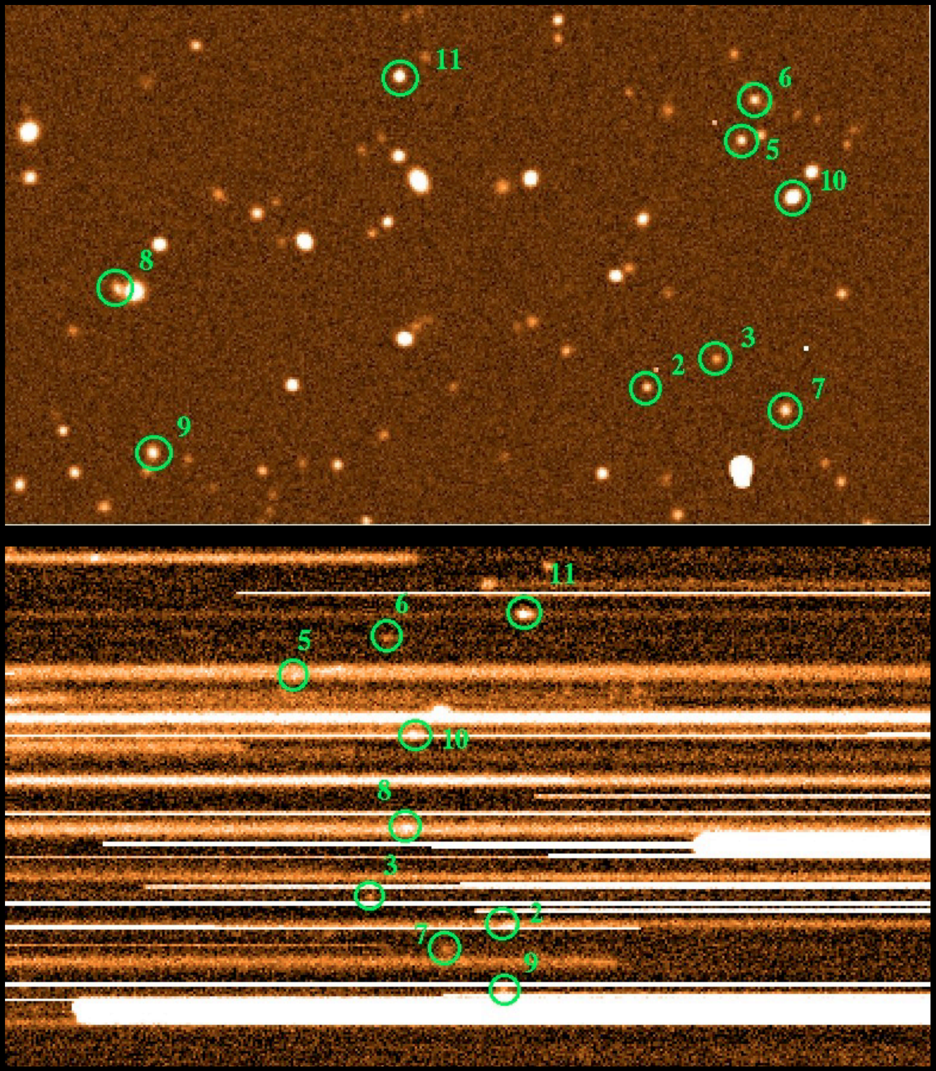
Pointing error along the x,y axes= 25mas over a period 600 s.

- FoV:

Common visible and NIR Fov = 0.54 deg^2



Astrium



Euclid NISP-P NISP-S

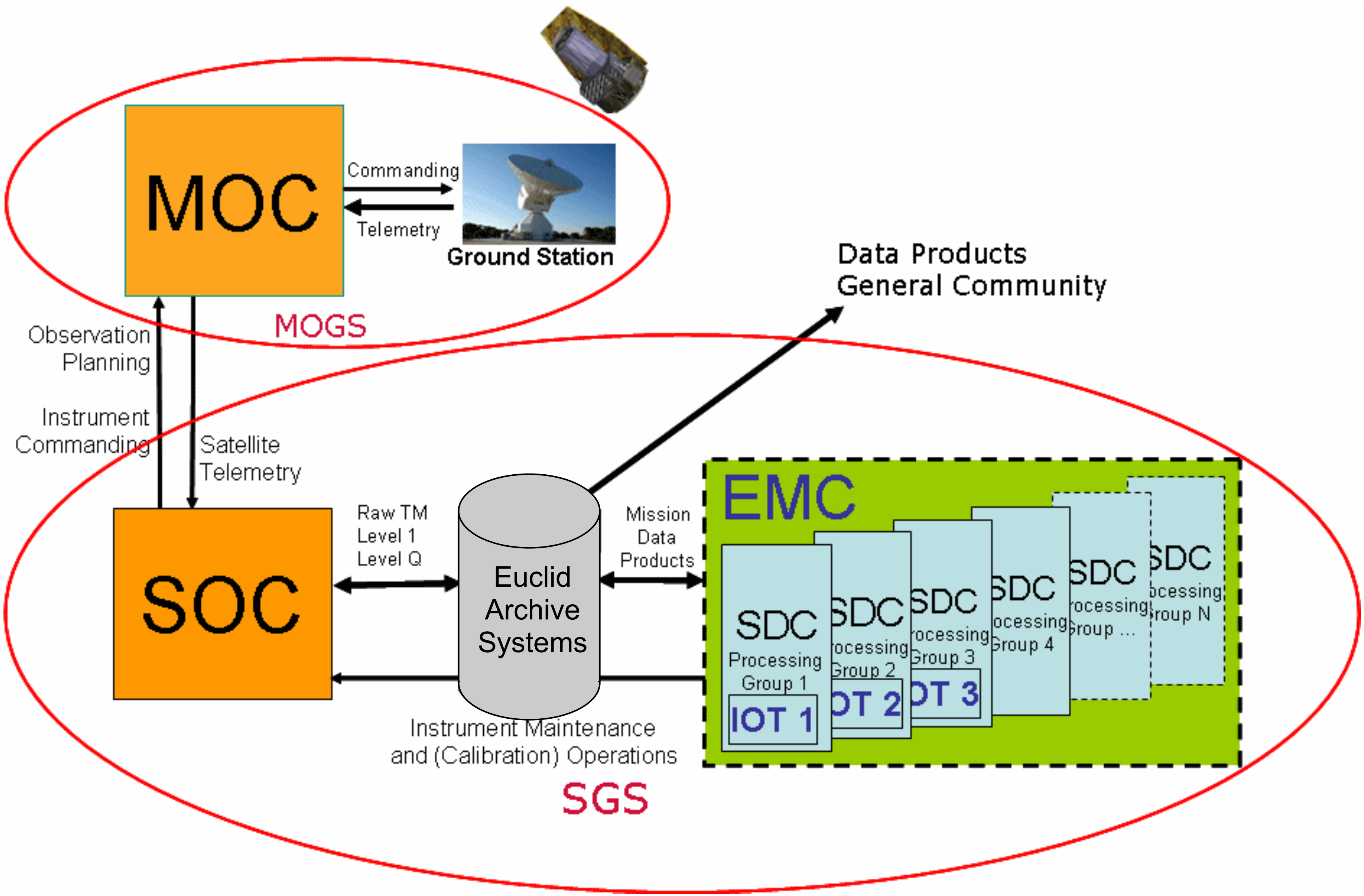
Simulation of a Euclid NIR image with a simulation of one grism exposure

Euclid VIS+NISP

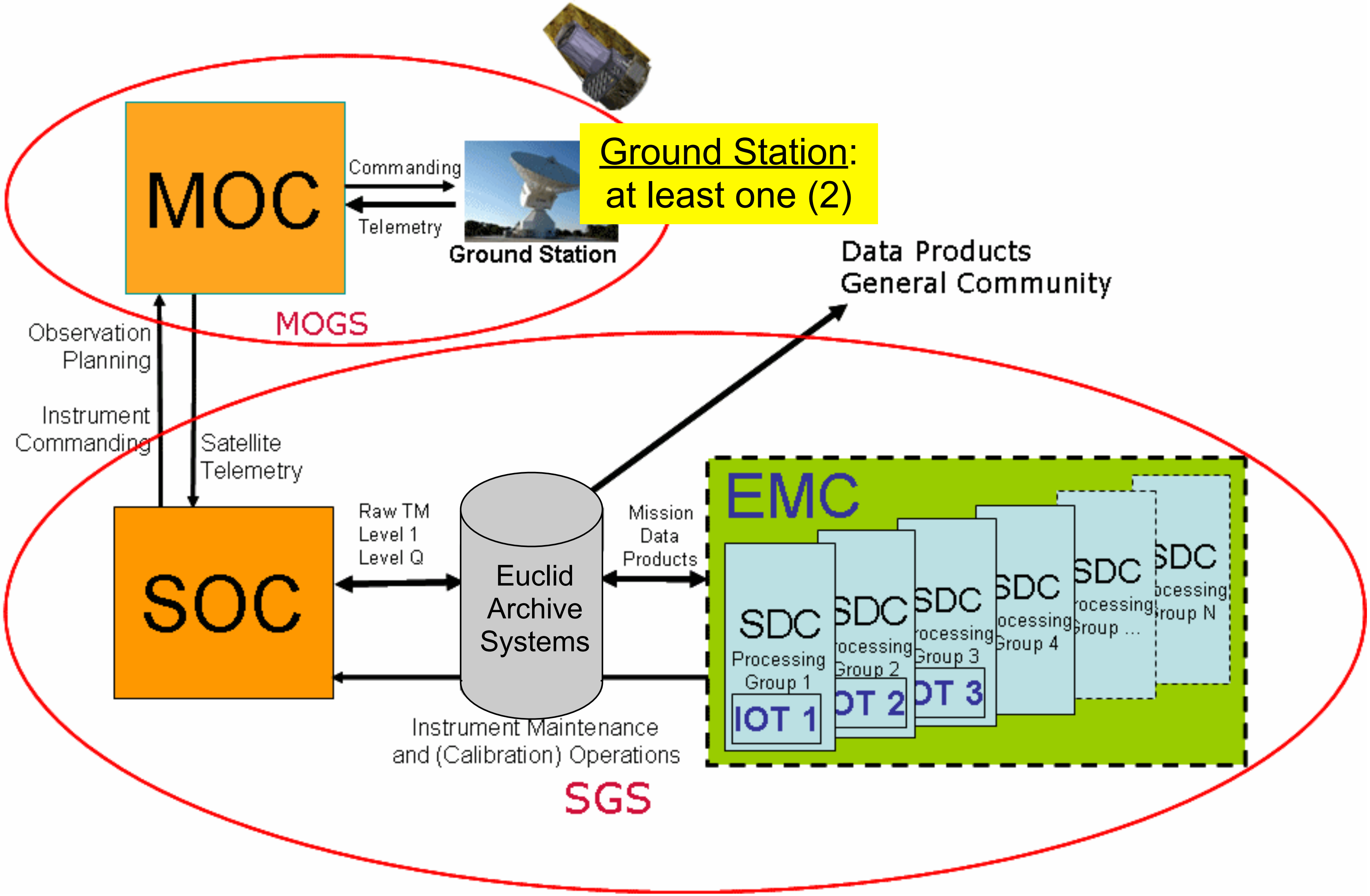
Simulation of a Euclid visible+Y+J+H combined image of a lensing cluster of galaxies

The Euclid Science Ground Segment

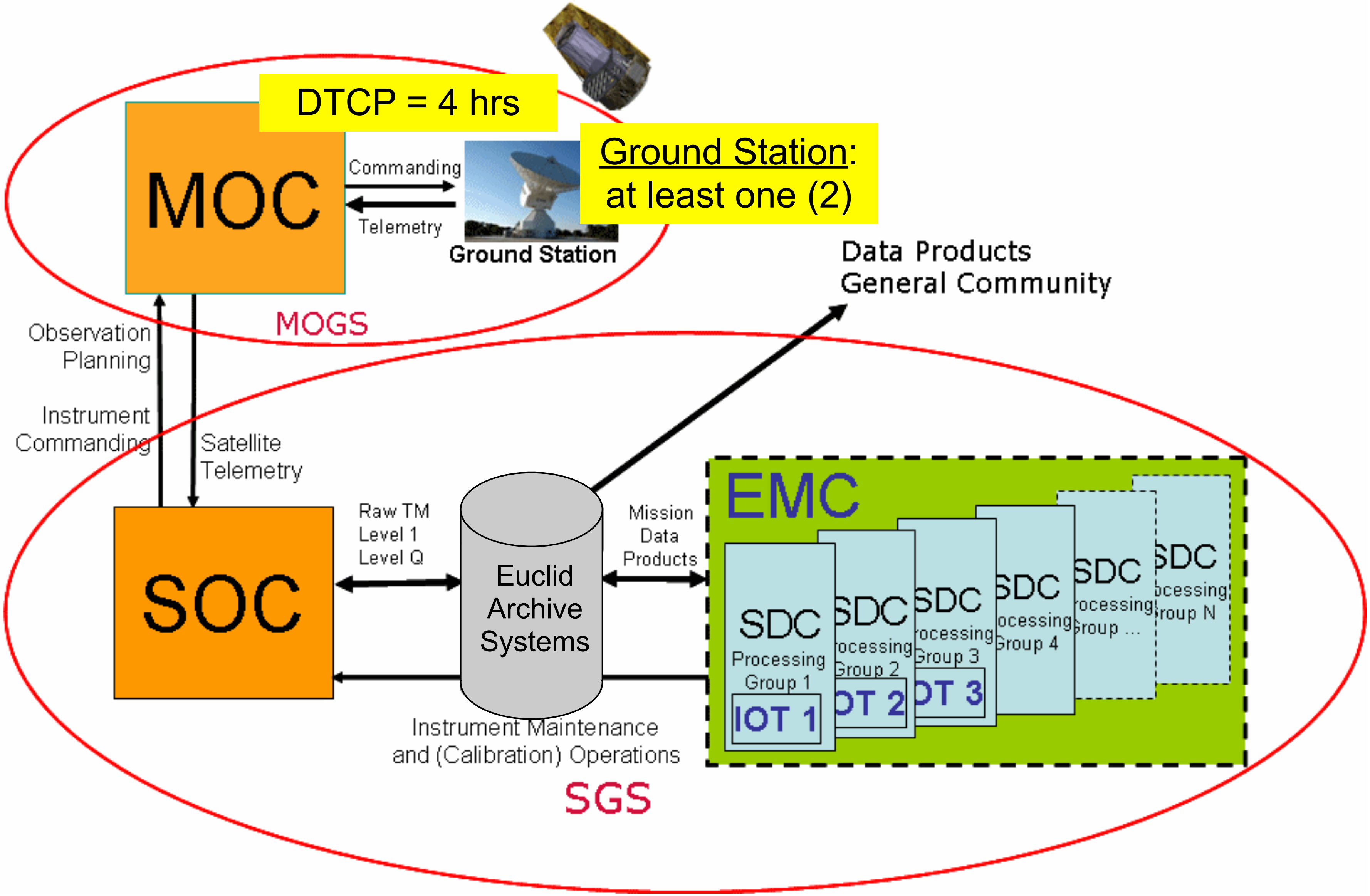
The Euclid Ground Segment



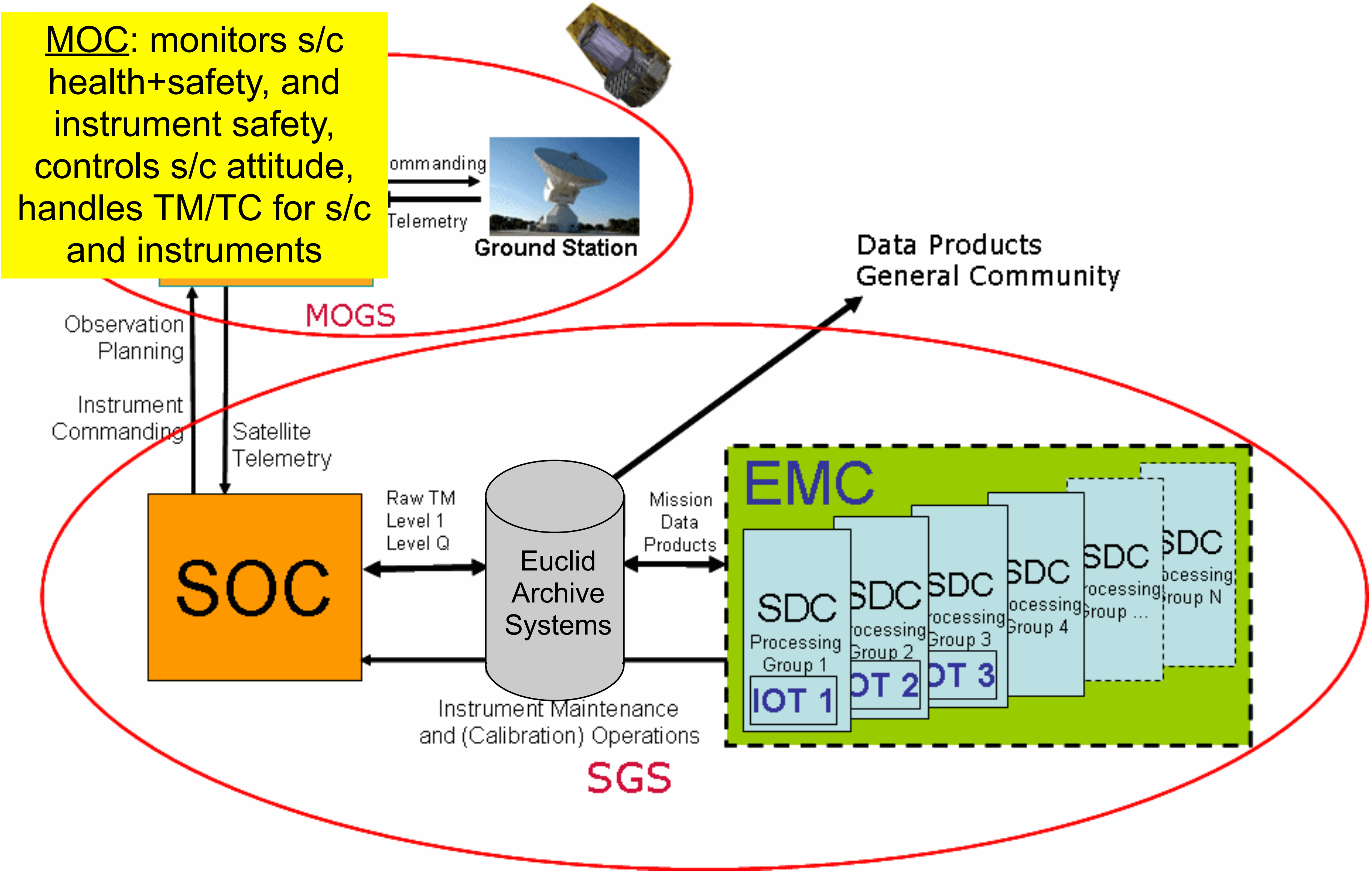
The Euclid Ground Segment



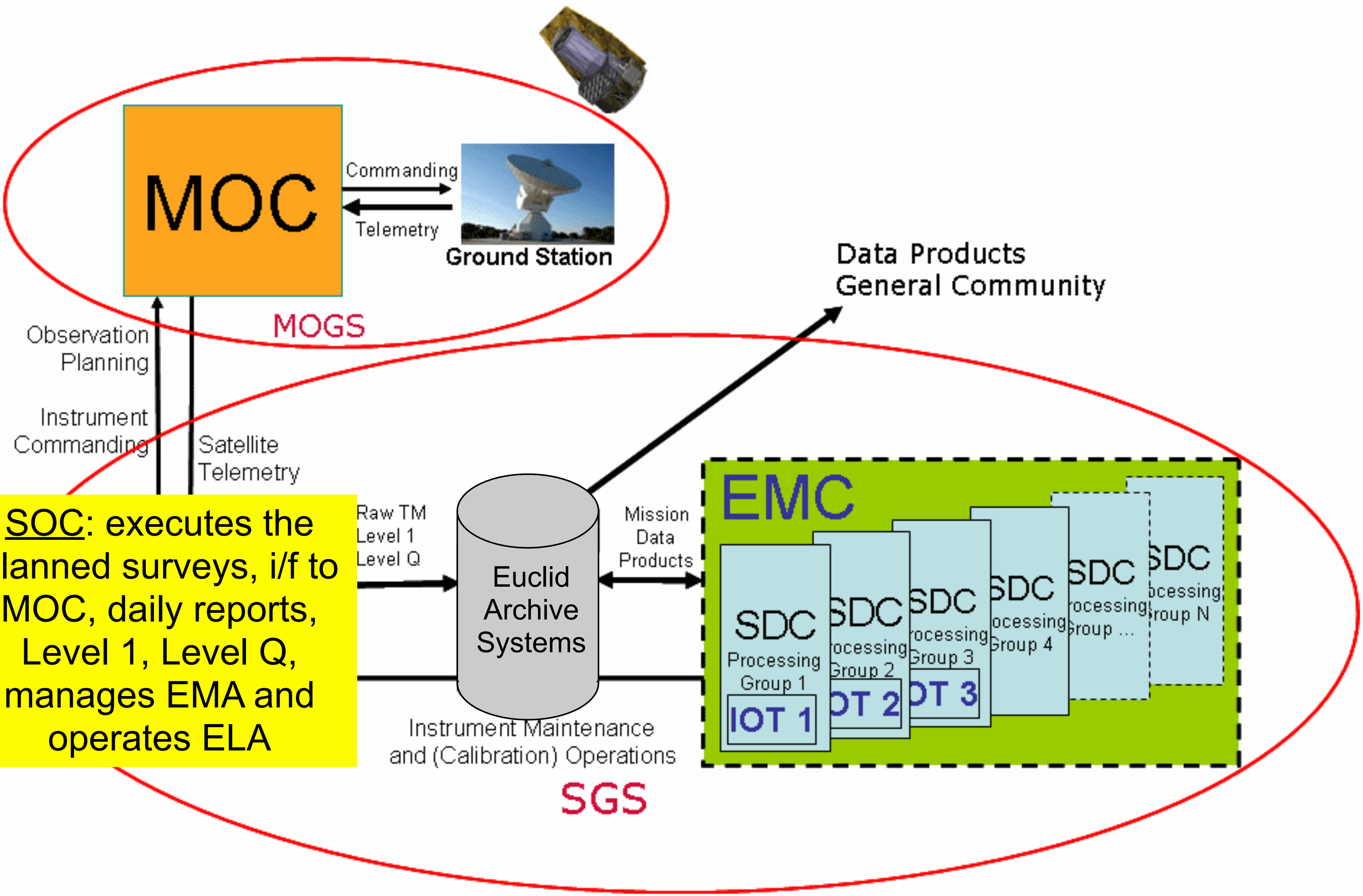
The Euclid Ground Segment



The Euclid Ground Segment

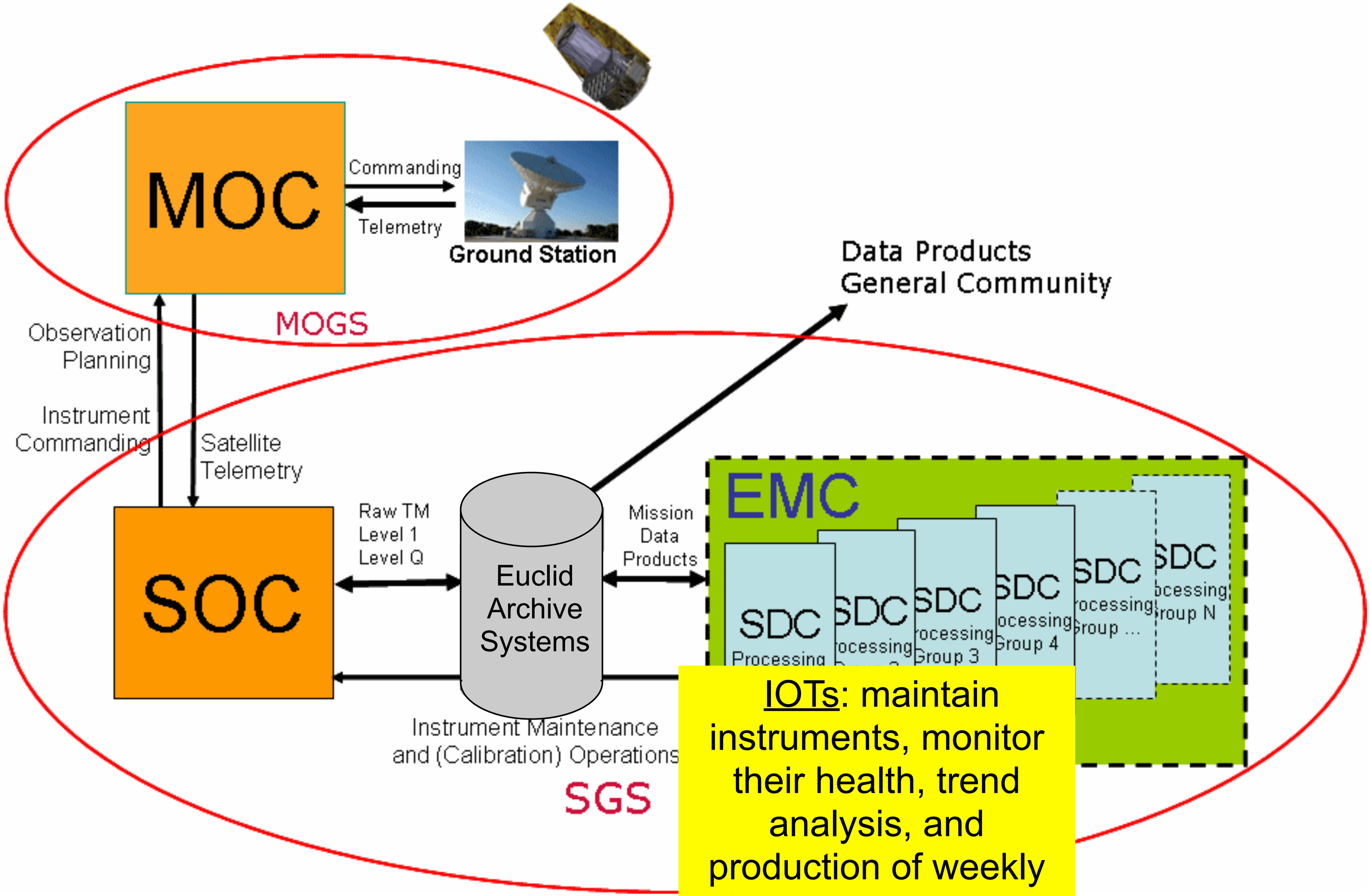


The Euclid Ground Segment



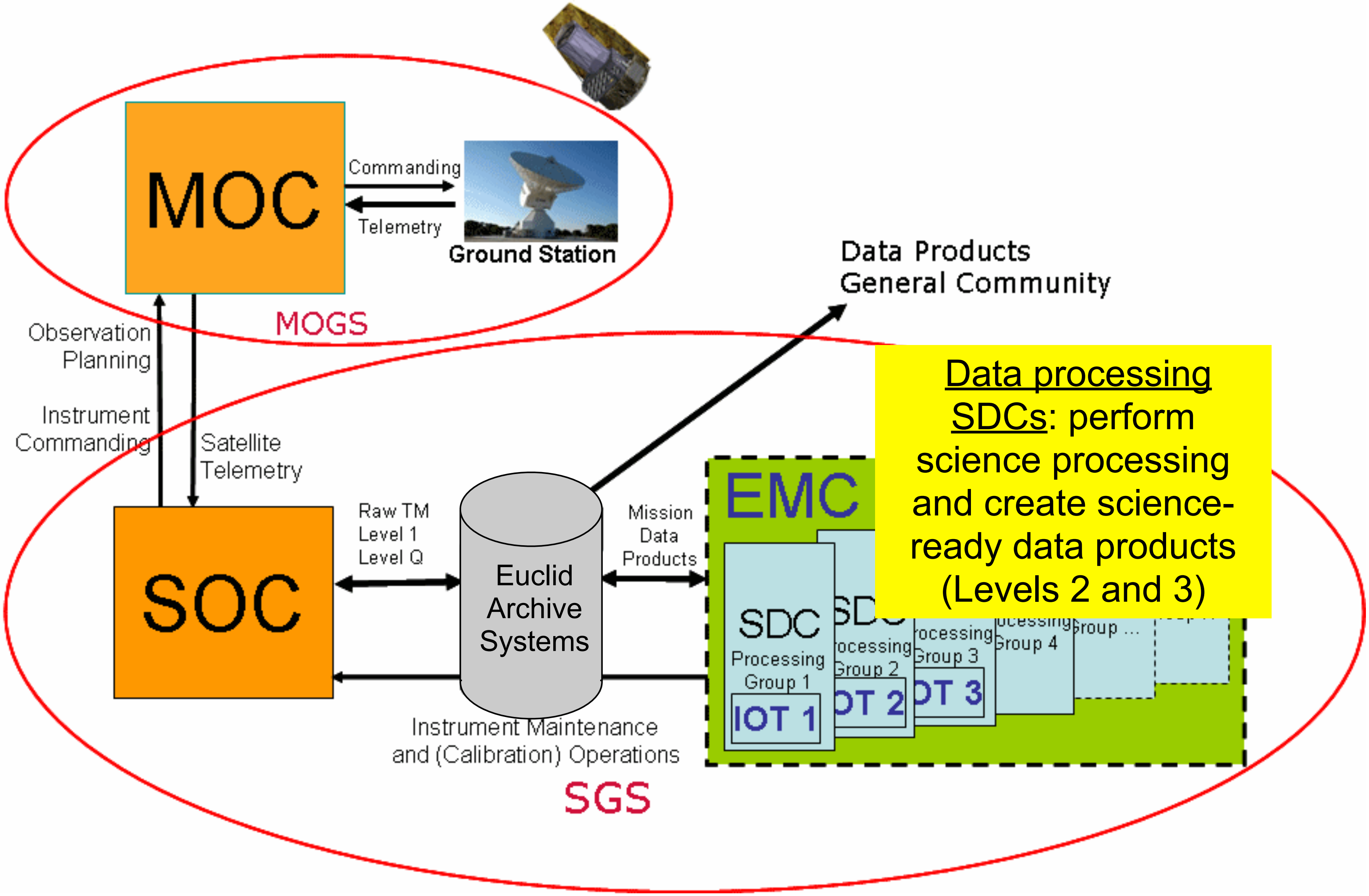
<u>SOC</u>: executes the planned surveys, i/f to MOC, daily reports, Level 1, Level Q, manages EMA and operates ELA

The Euclid Ground Segment



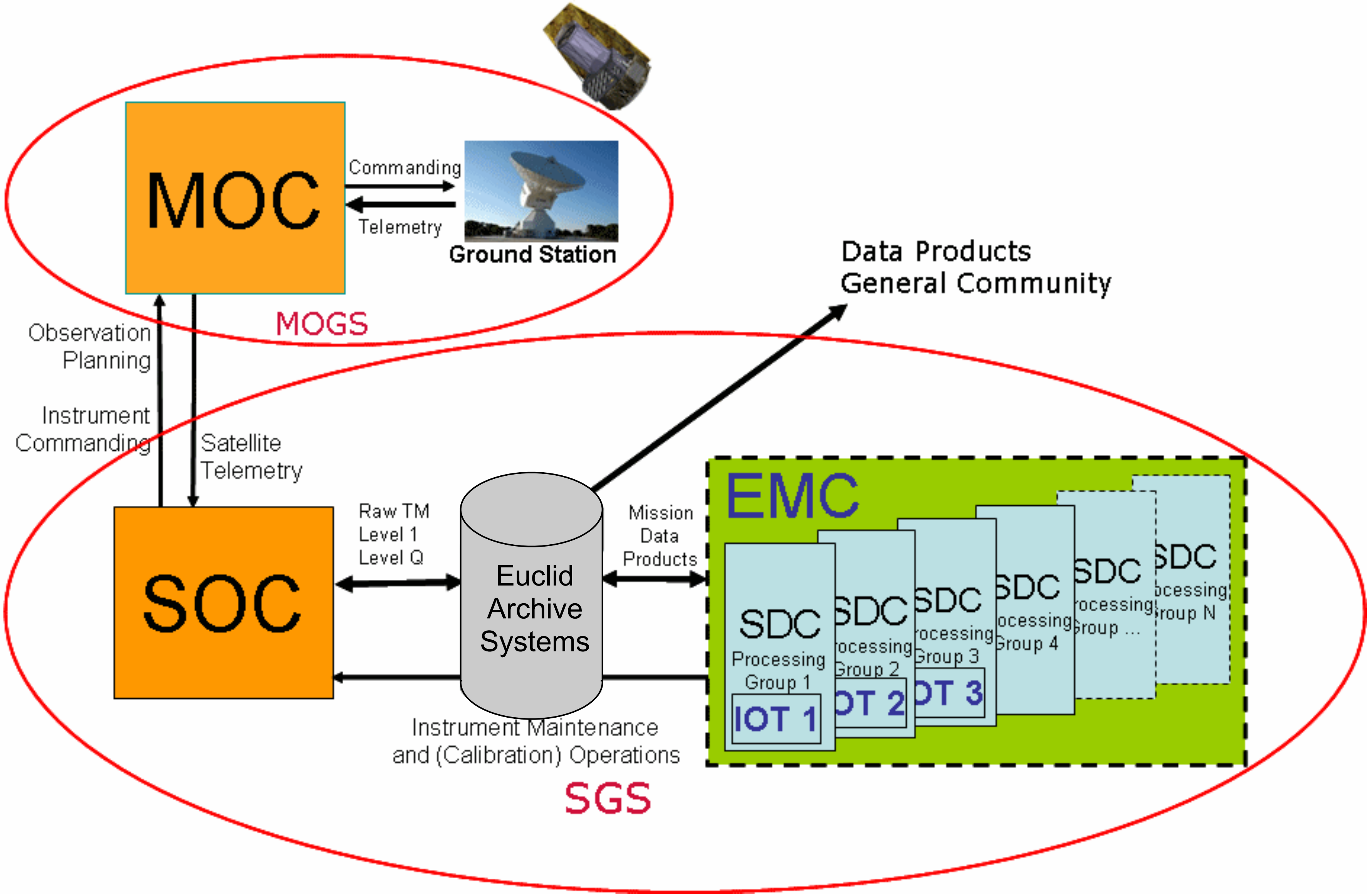
<u>IOTs</u>: maintain instruments, monitor their health, trend analysis, and production of weekly instrument reports

The Euclid Ground Segment

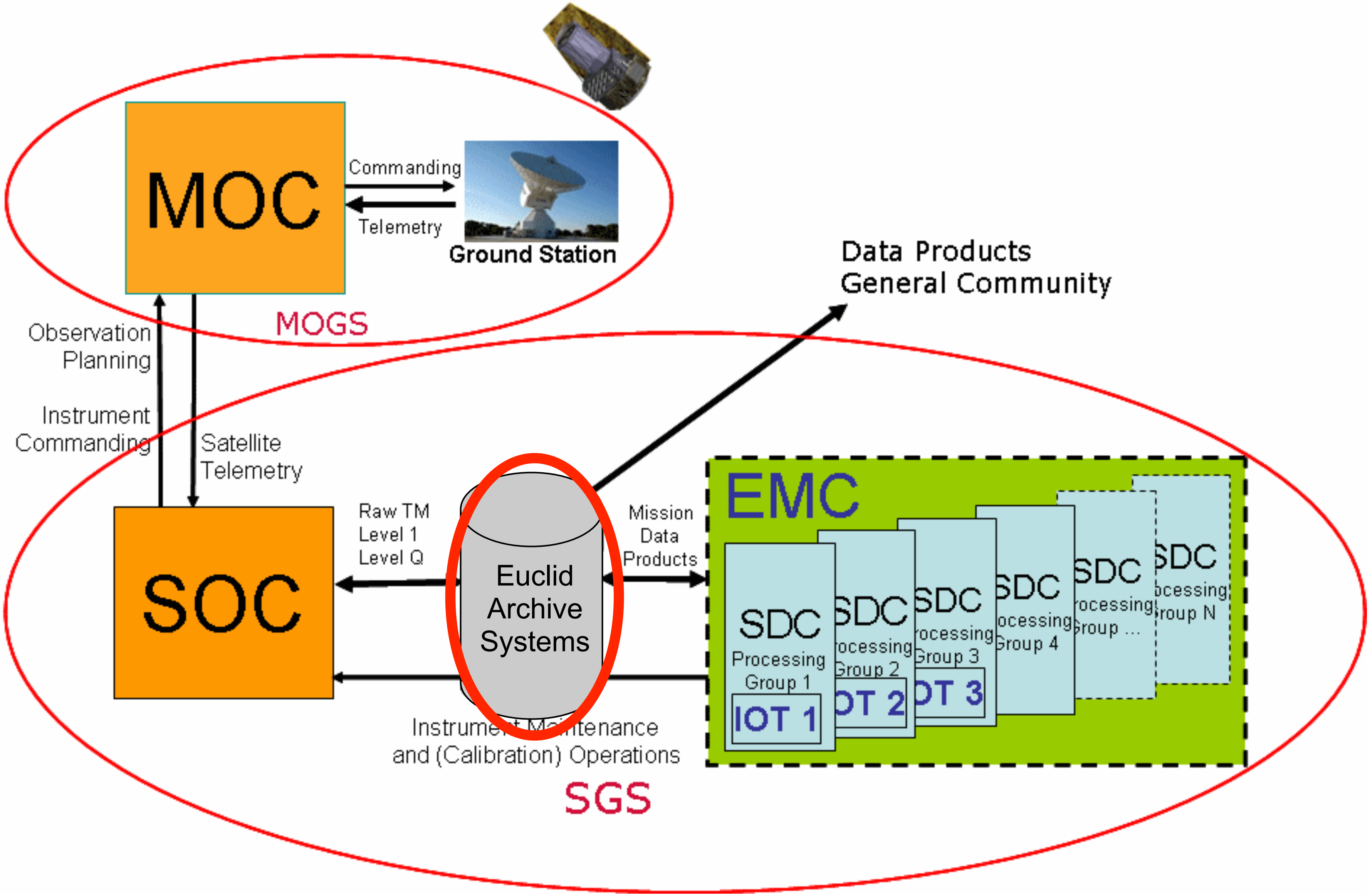


Data processing SDCs: perform science processing and create science-ready data products (Levels 2 and 3)

The Euclid Ground Segment



The Euclid Ground Segment



Ground Segment drivers

- Operational:
  - There is the need to operate the instruments and to verify their full functionality.
  - The satellite will generate ~ 22h of science observations per 24h period. An organisation is needed to face and control this flow.
  - We have an imposed data release schedule.

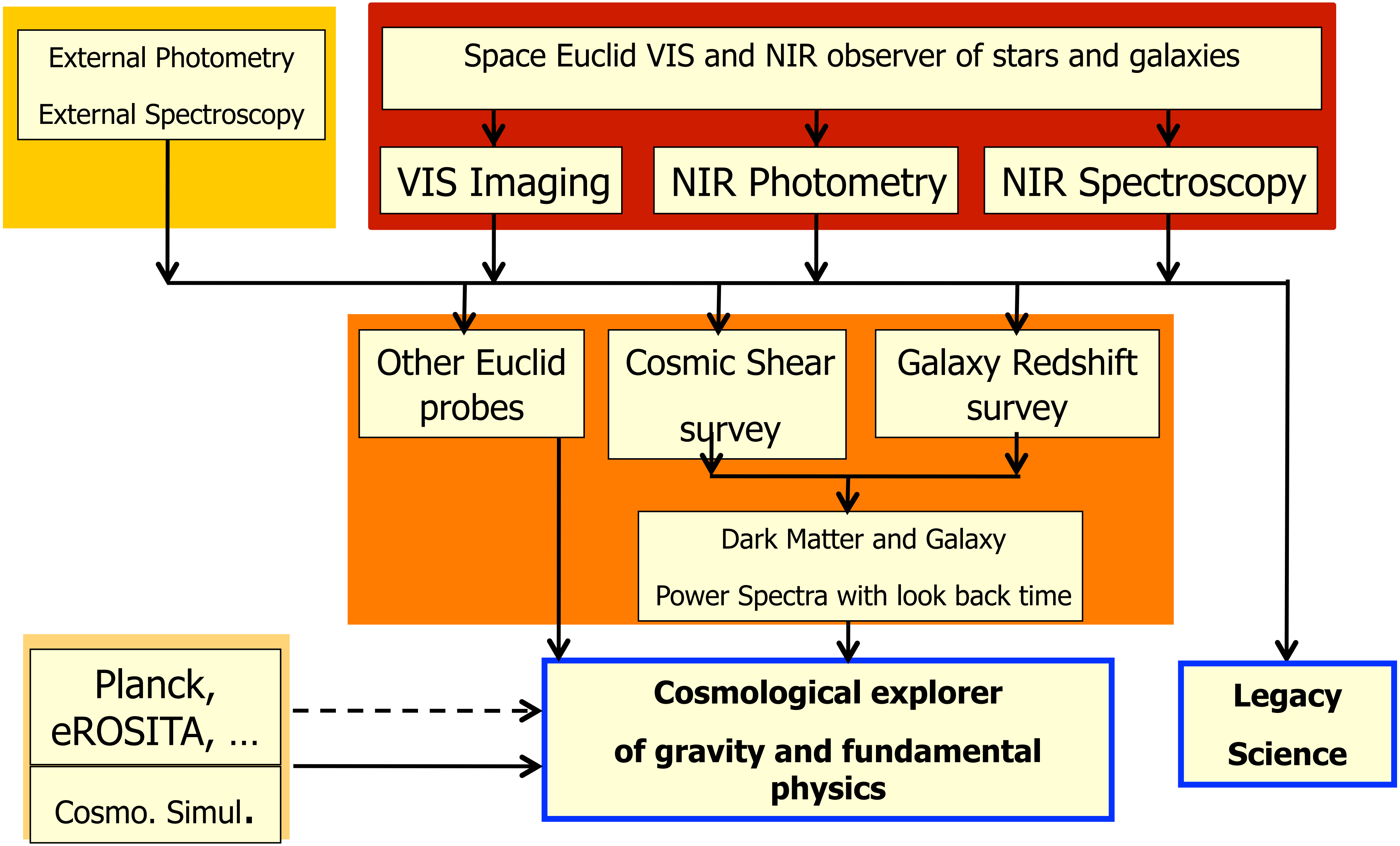
- Scientific:
  - The amount of processing required to go from the raw data to the information that can be interpreted in a cosmological context is large and complex.
  - The volume of data is very large, and all processing steps have to be planned well in advance.

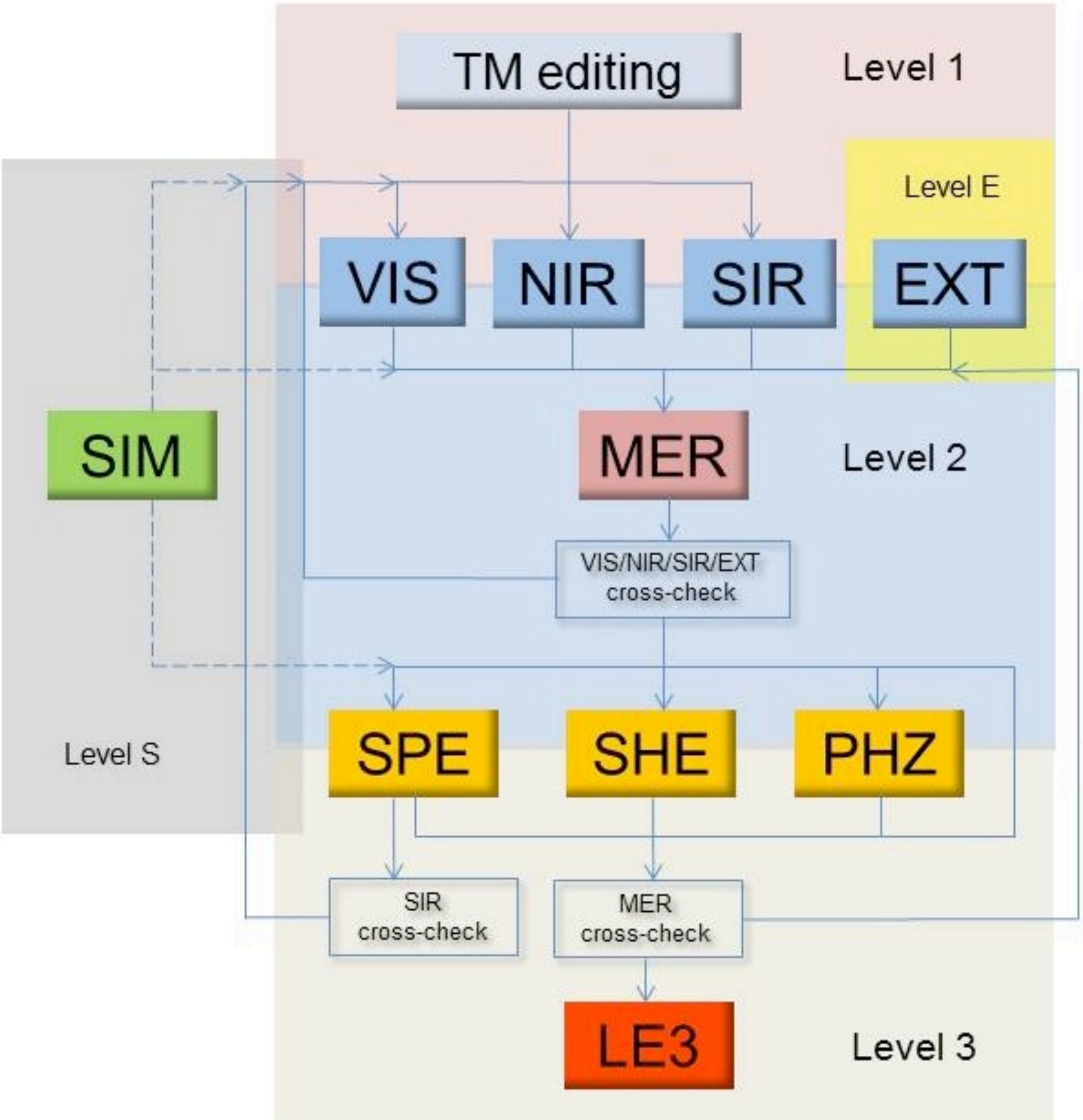
Ground Segment drivers

- Operational:
  - There is the need to operate the instruments and to verify their full functionality.
  - The satellite will generate ~ 22h of science observations per 24h period. An organisation is needed to face and control this flow.
  - We have an imposed data release schedule.
- Scientific:
  - The amount of processing required to go from the raw data to the information that can be interpreted in a cosmological context is large and complex.
  - The volume of data is very large, and all processing steps have to be planned well in advance.

- A «smooth transition» through testing-commissioning-operations
- Development of data processing software has to be started early and tap into the scientific expertise of the consortium
- Data processing involves geographically distributed centres

The *Euclid Machine*





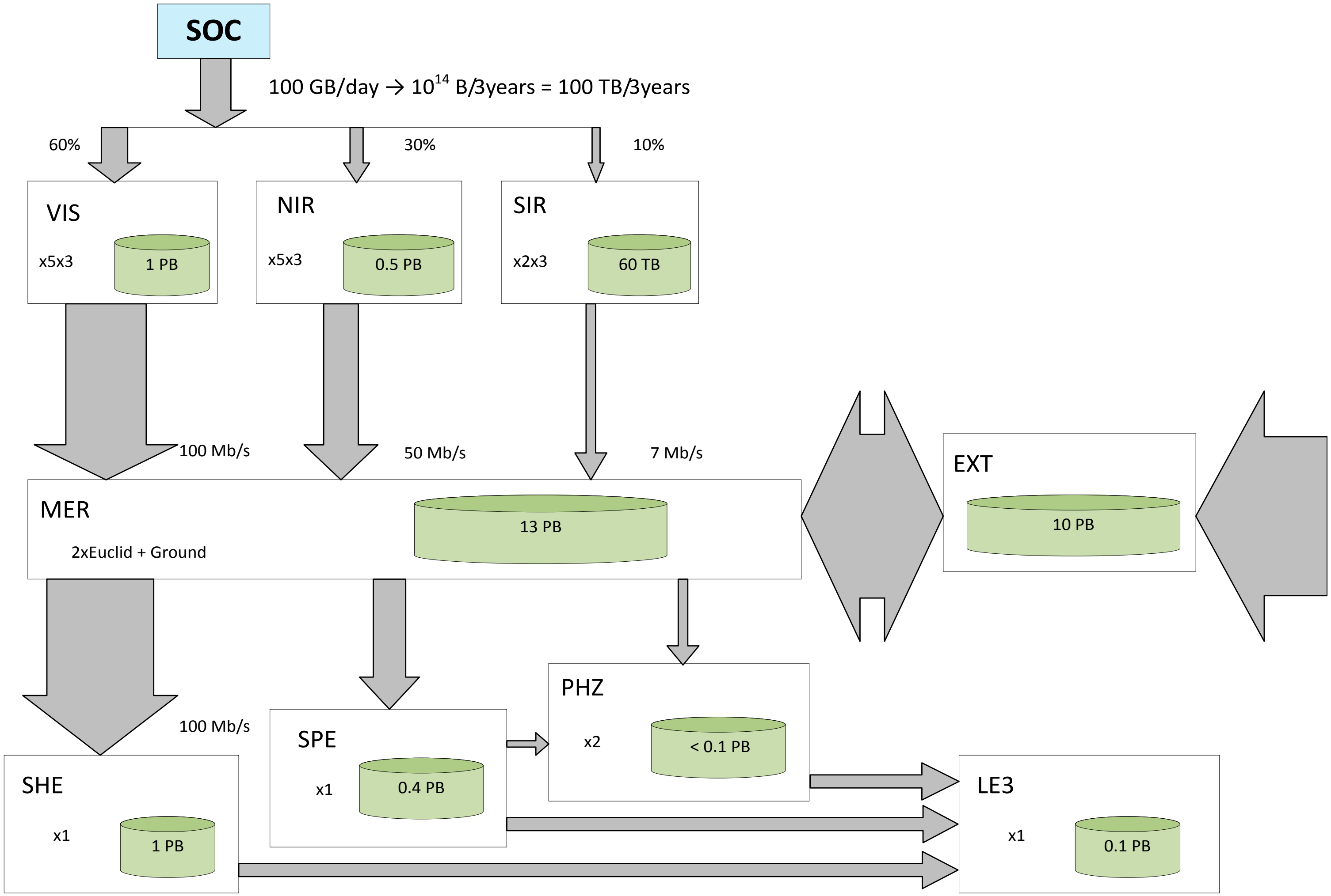
Processing Levels and Processing Functions

The contents of the Processing Functions are defined by the OUs and implemented by the SDCs.

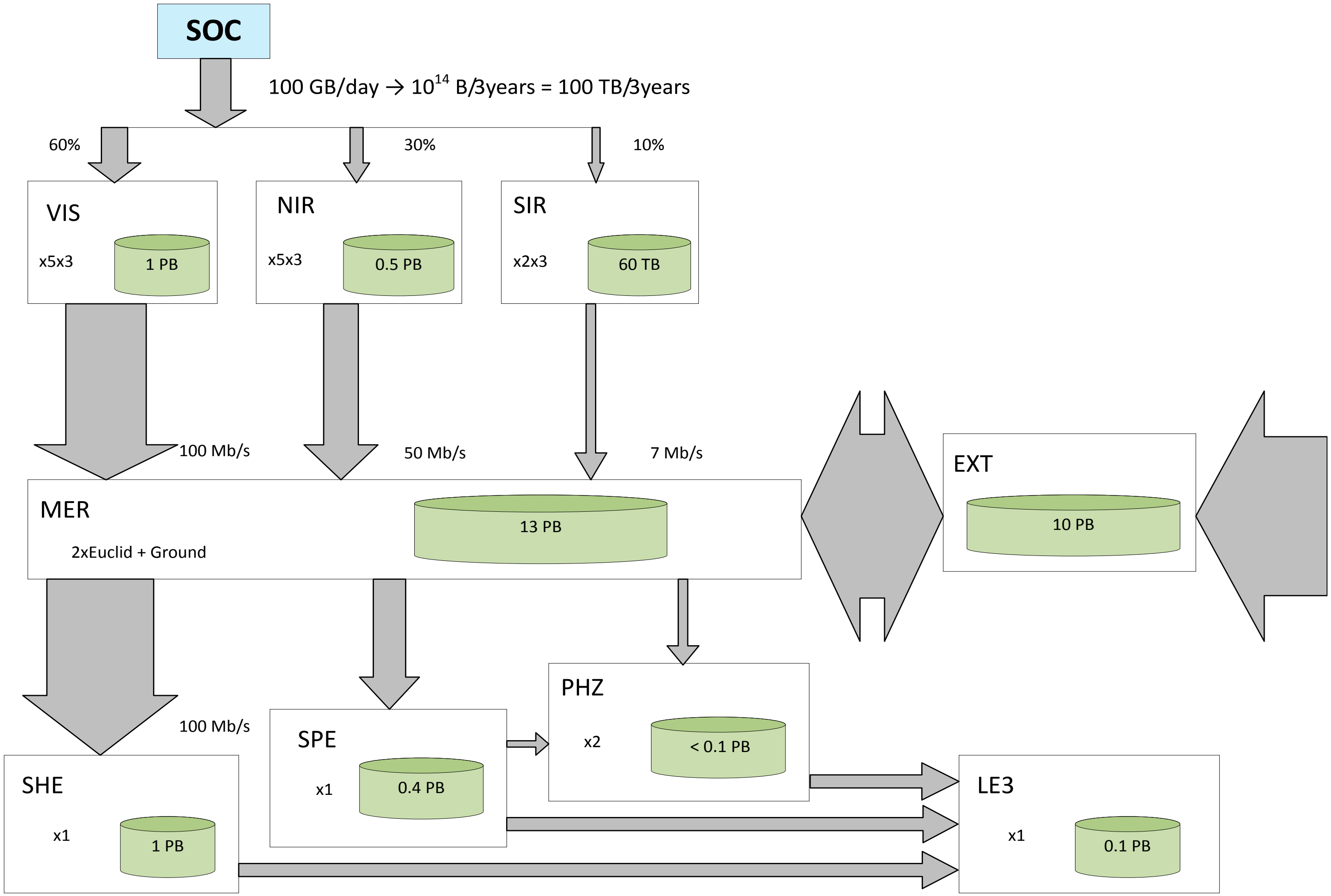
The OU teams

VIS	Process VIS tiles to calibrated images	H. McCracken, C. Grenet, K. Benson
NIR	Process NISP tiles to calibrated images	A. Grazian, R. Bouwens
SIR	Process NISP tiles to calibrated spectra	M. Scodeggio, C. Surace
EXT	Provide external data (photometric images and catalogs, spectroscopic samples)	G. Verdoes-Kleijn, J. Mohr
SIM	Provide instrument simulators and survey simulations	S. Serrano, A. Ealet
MER	Merge all data to produce the Euclid source catalog(s)	A. Fontana, M. Kuemmel, M. Douspis
PHZ	Compute photometric redshifts	S. Paltani, F. Castander
SPE	Extract spectroscopic redshifts	O. Le Fèvre, M. Mignoli
SHE	Compute galaxy ellipticities	A. Taylor, T. Schrabback, S. Courbin
LE3	High-level catalog processing and more	J.-L. Starck, F. Abdalla, S. Borgani

Data flow estimation

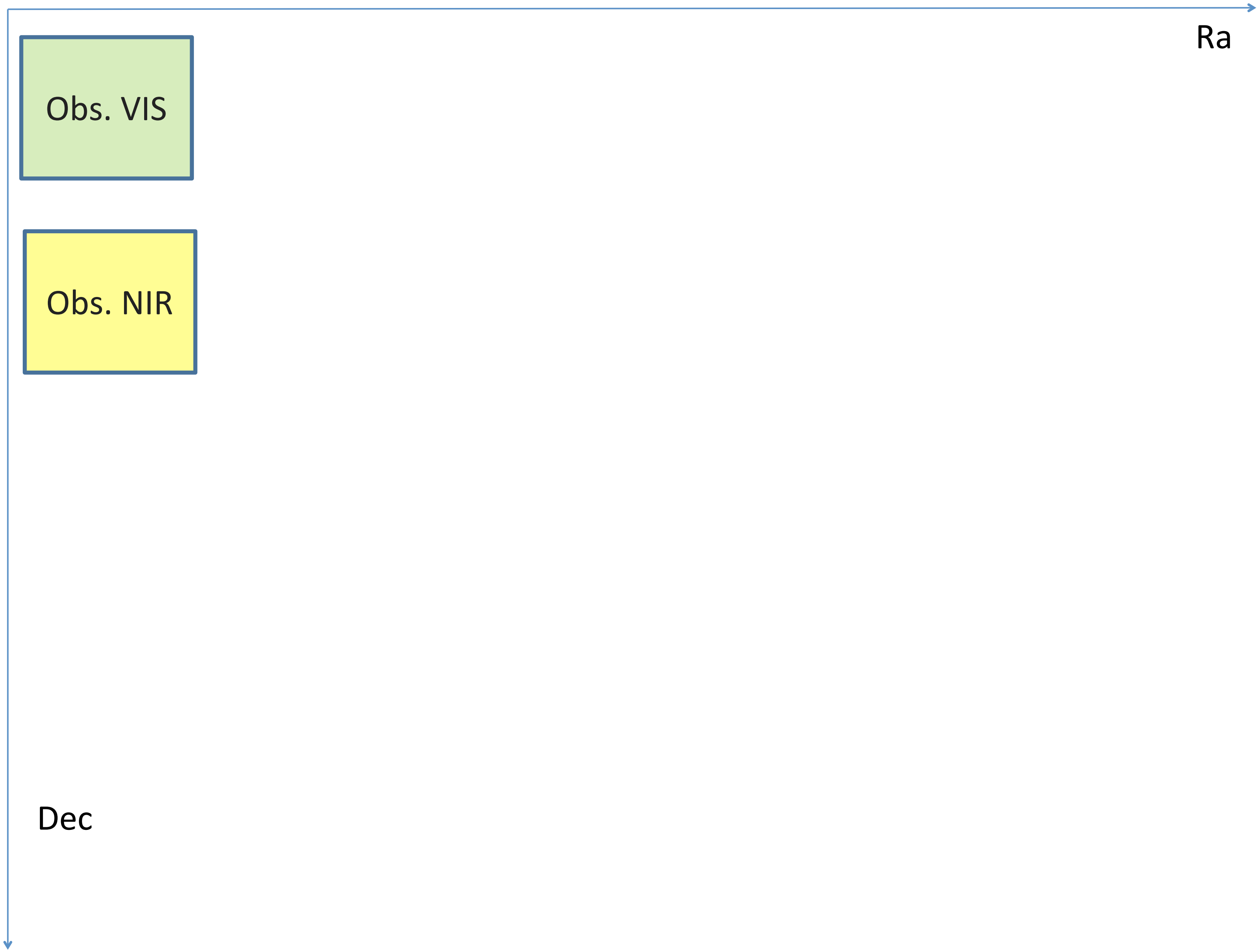


Data flow estimation

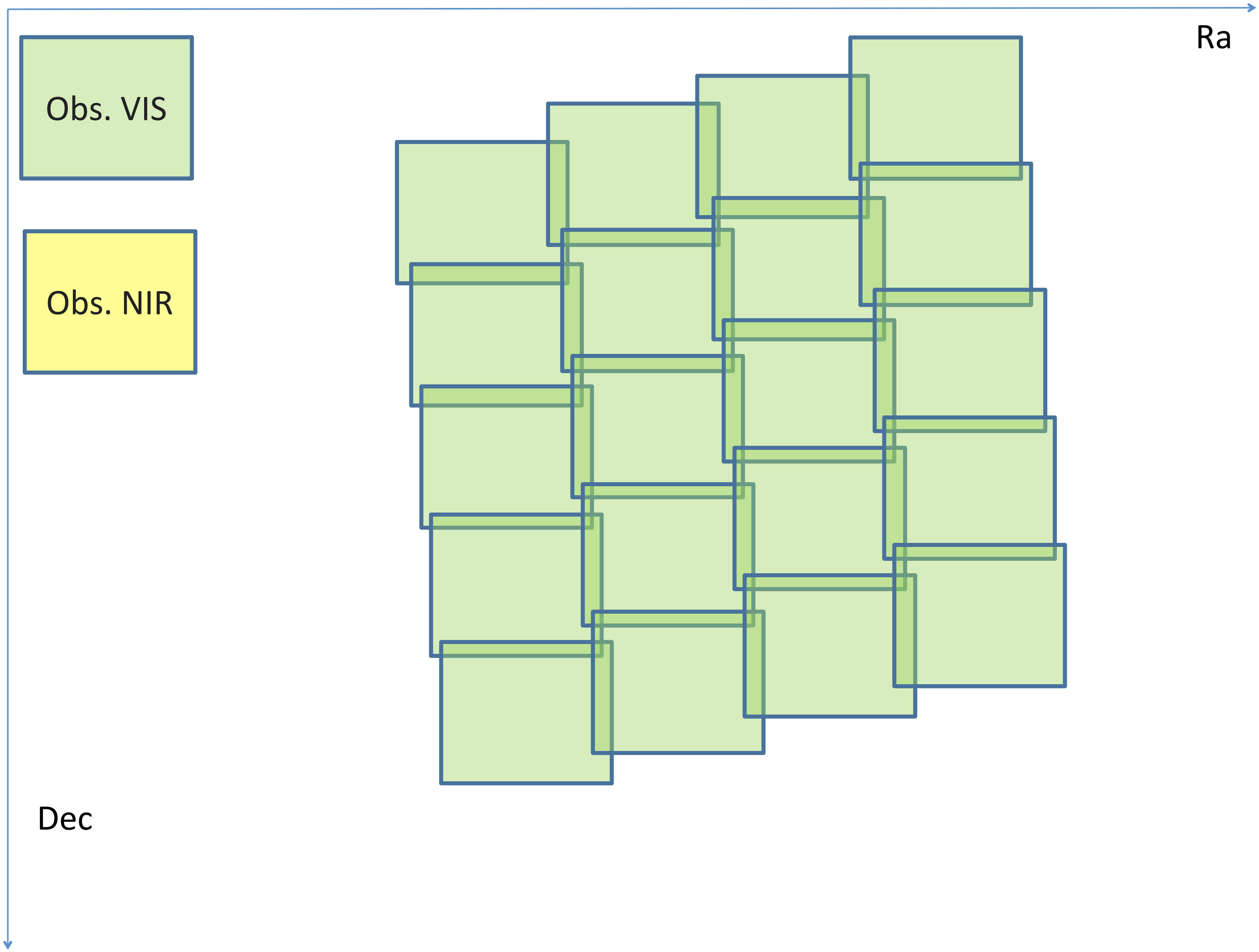


- "Brute force" SDC-SDC bulk data transfer is inefficient → need to move the processing software, not the data.

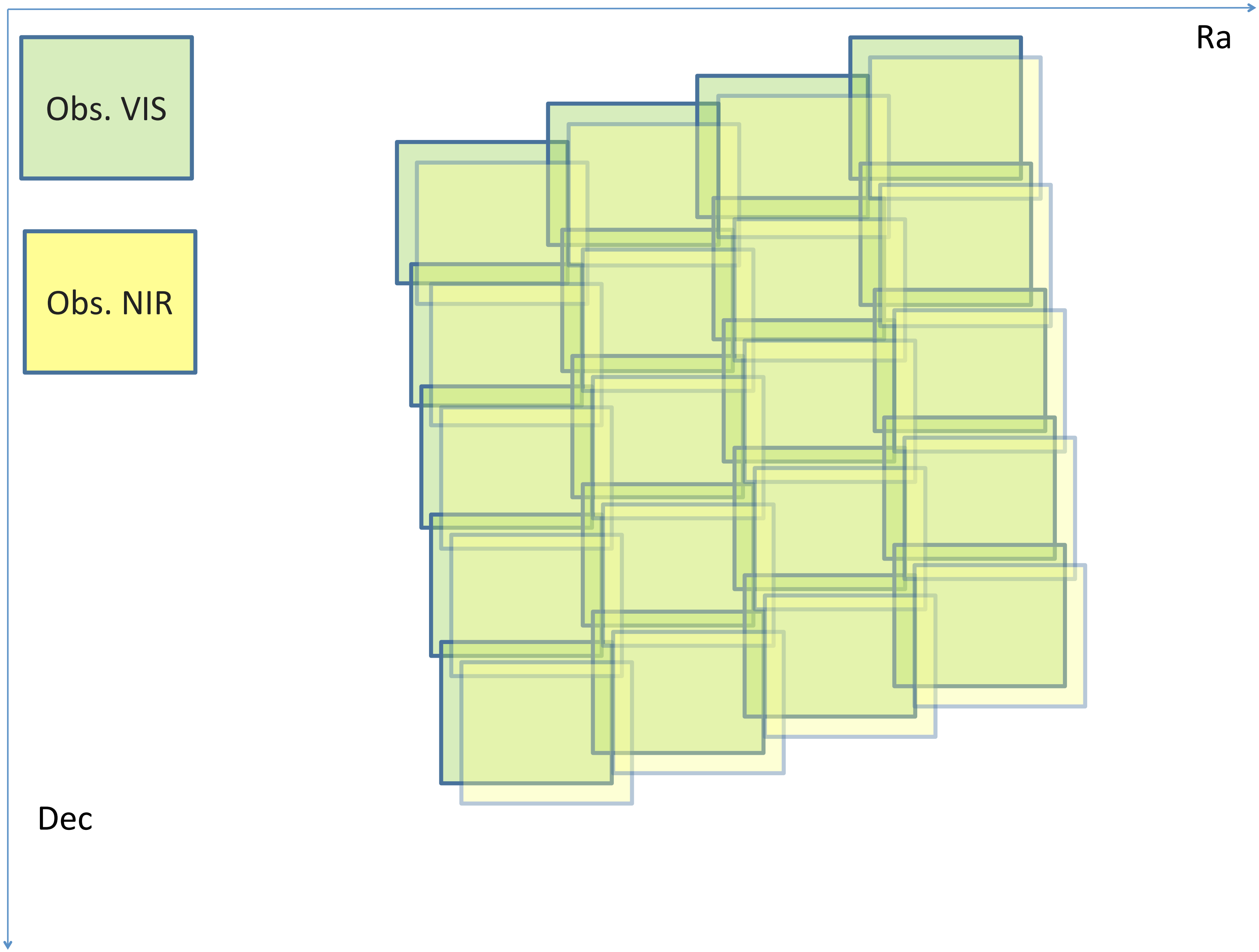
The survey will/should proceed by sky patches (50-100 sq. degrees)



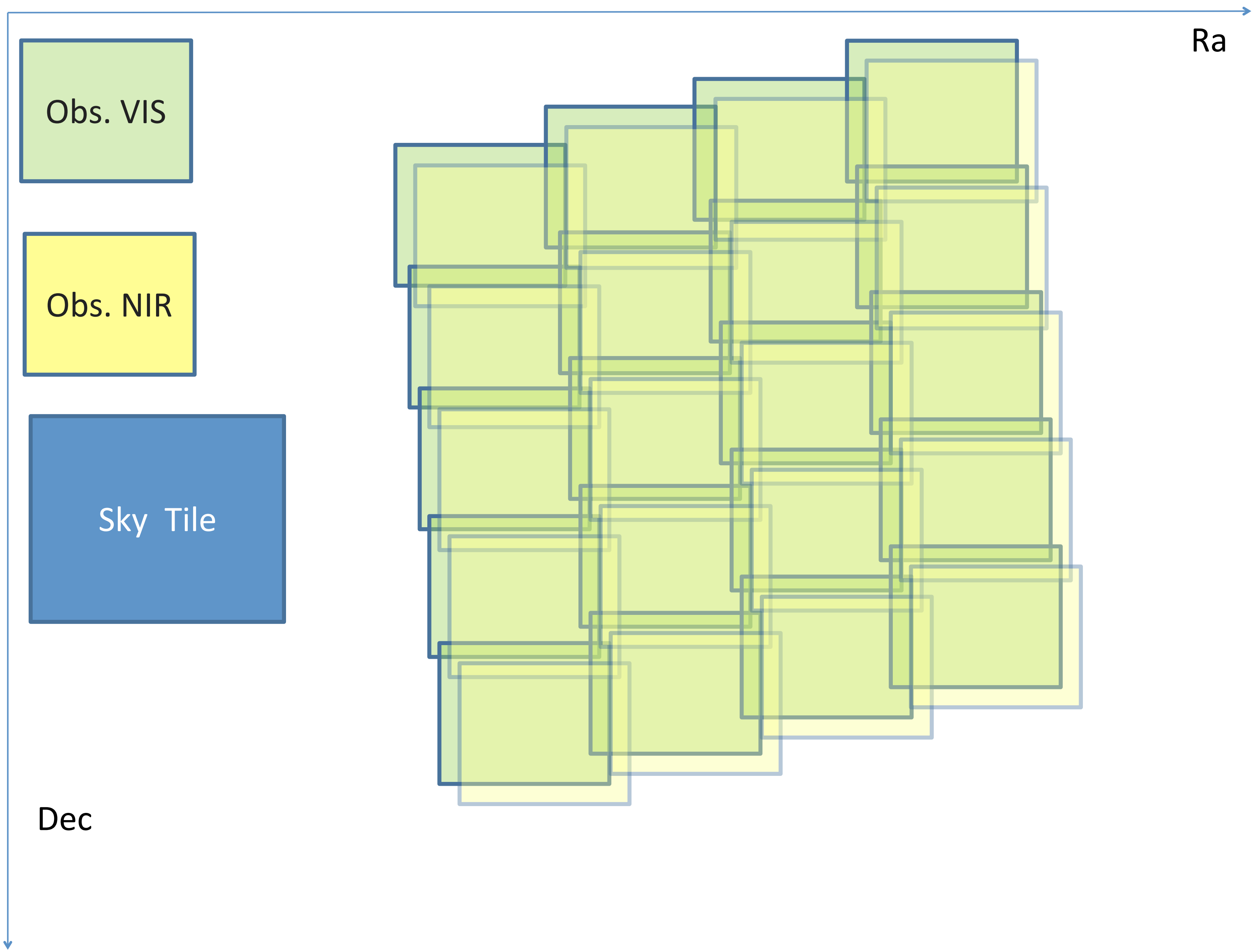
The survey will/should proceed by sky patches (50-100 sq. degrees)



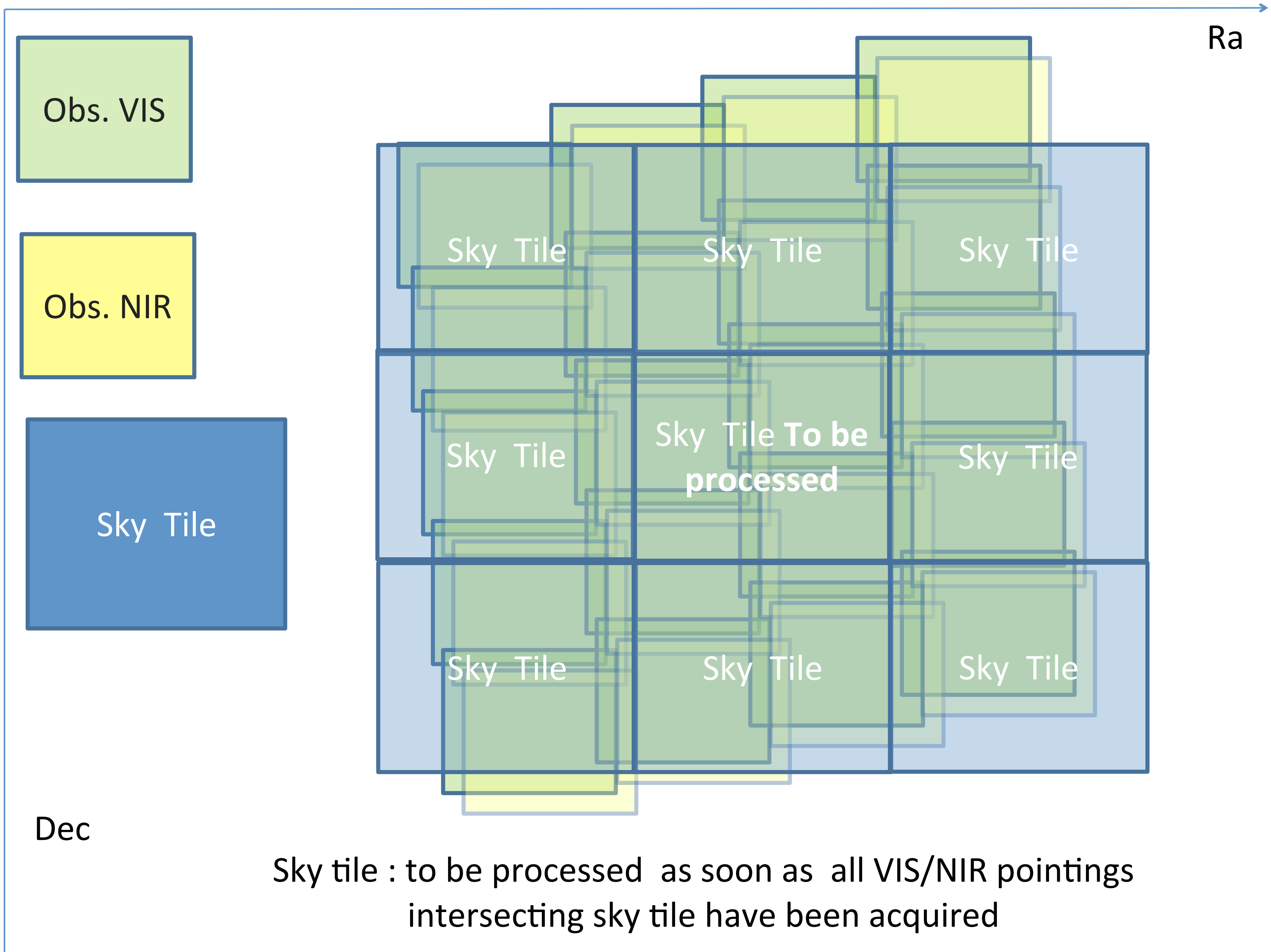
The survey will/should proceed by sky patches (50-100 sq. degrees)



The survey will/should proceed by sky patches (50-100 sq. degrees)

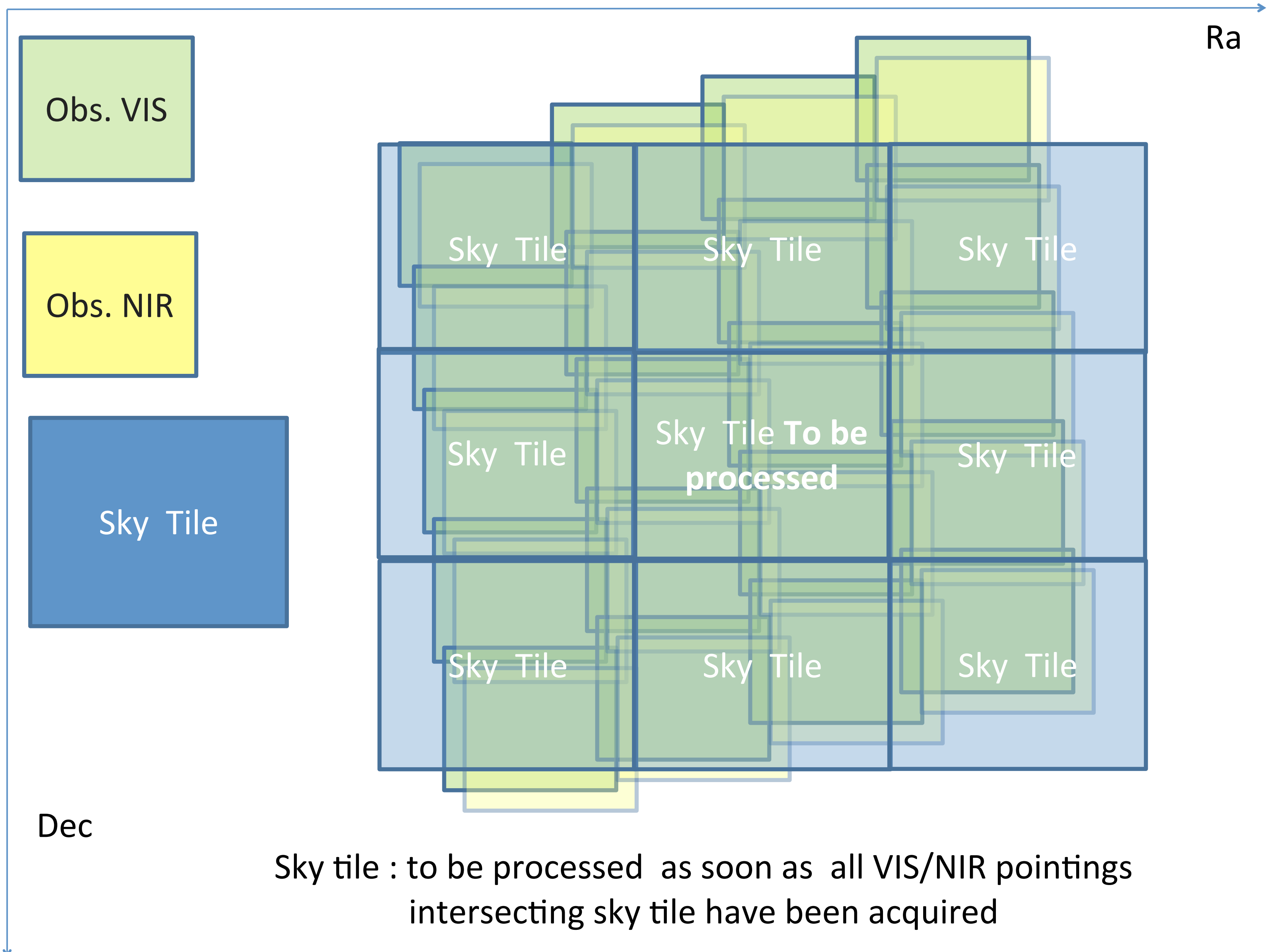


The survey will/should proceed by sky patches (50-100 sq. degrees)



Sky tile : to be processed as soon as all VIS/NIR pointings intersecting sky tile have been acquired

The survey will/should proceed by sky patches (50-100 sq. degrees)



Sky tile : to be processed as soon as all VIS/NIR pointings intersecting sky tile have been acquired

The processing strategy/orchestration is heavily dependent on the survey strategy

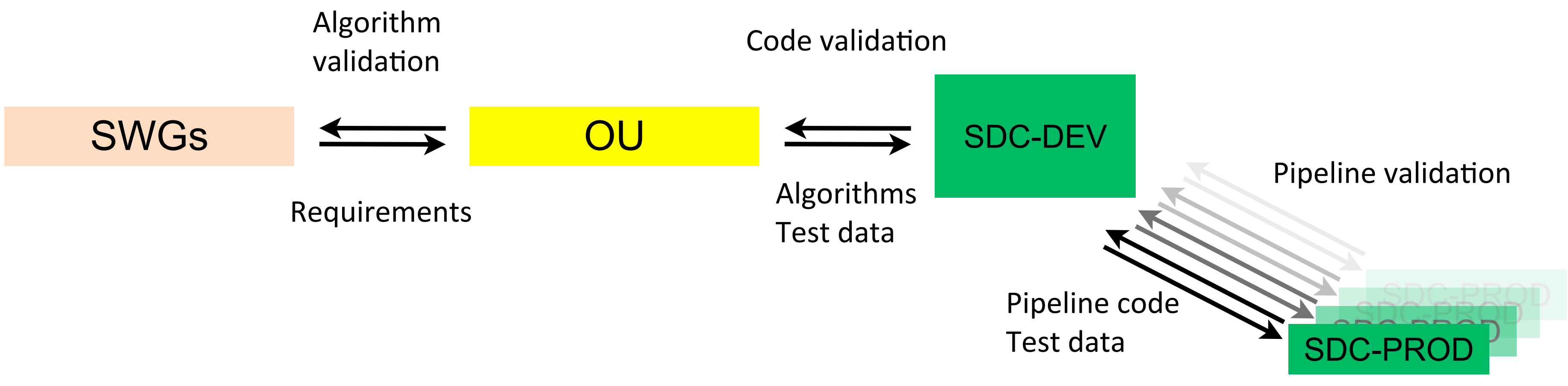
Science Working Groups

- The SGS development is performed by the Organization Units (OUs) assisted by the SDC developpers.
  - They are responsible for the algorithmic research and development.
- The scientific overview and responsibilities rests within the Science Working Groups.
  - They are organized by Cosmological probes and science topic
  - They provide the requirements for each analysis step.
  - They verify and validate the algorithms created by the OU

Science Working Groups

- The SGS development is performed by the Organization Units (OUs) assisted by the SDC developpers.
  - They are responsible for the algorithmic research and development.
- The scientific overview and responsibilities rests within the Science Working Groups.
  - They are organized by Cosmological probes and science topic
  - They provide the requirements for each analysis step.
  - They verify and validate the algorithms created by the OU

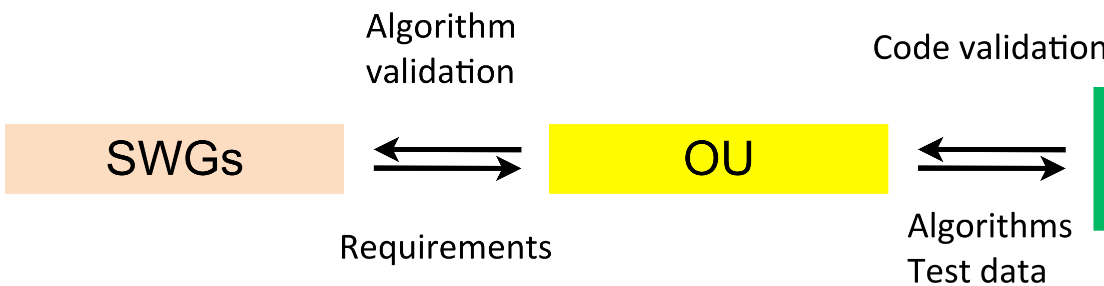
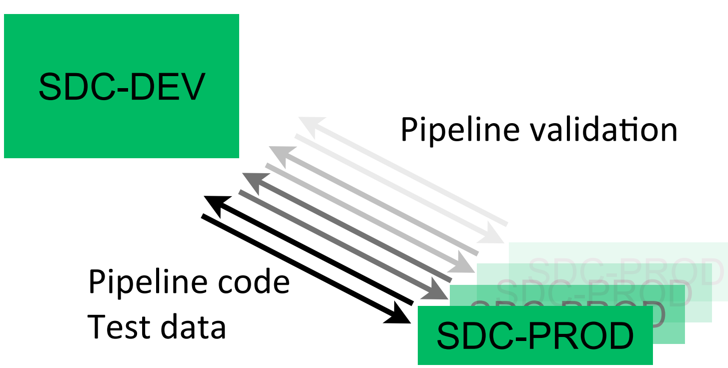
For every step of the science processing



Science Working Groups

- The SGS development is performed by the Organization Units (OUs) assisted by the SDC developpers.
  - They are responsible for the algorithmic research and development.
- The scientific overview and responsibilities rests within the Science Working Groups.
  - They are organized by Cosmological probes and science topic
  - They provide the requirements for each analysis step.
  - They verify and validate the algorithms created by the OU

For every step of the science processing

The SWG-OU interface is one of the most important interface in the EC.

A central element: the Euclid archive

Core concepts:

- Data distributed across SDCs and SOC (data, e.g. images, spectra)
    - Each data set resides in at least two SDCs, but not all the data is available everywhere.
- Meta-data located in central repository at SOC
    - i.e. physical location of data files, catalogues, quality information, lineage, etc…
    - SDCs have local copy for processing.
    - Mirror database envisioned.
- We need an orchestration system that distributes the data as a function of:
    - processing readiness (all data needed for processing availalble).
    - computing resources availability.
- The system must support the EC in its activities and the wider scientific community once data is made public.

The Euclid Archive is implemented by the **Euclid Archive System** (EAS)

The Euclid Archive System

The EAS consists of:

- Euclid Archive Core System (EACS): the meta-data repository.
- Advanced applications: archive access tools needed by the EC to achieve the science goals
- Storage systems at the SDCs/SOC for the distributed data
- Data transfer tools: managing the data on and transfer between SDC/SOC storage facilities
- Management tools: access control, reporting, etc.

The EAS is collaboratively developed within the SGS System Team by ESA and the EC

- The Core System will be developed by ESA/ESAC – this allows ESA to meet its objectives for long-term data preservation
- Other elements of the EAS will be developed by the EC – support to the specific requirements of the EC SGS SDCs, OUs and the SWGs.

Understanding the needs of the EC thus is vital…

Where are we now?

- The SGS has gone through the Preliminary Requirements review.
  - Reasonable list of actions on the plans and organisation to be closed by end 2013.
  - SIM OU leading the way, with instrument simulators being integrated in the "proto" EAS.
  - First data challenges starting in different parts of the science processing system (shear measurements, PSF modelling, merging and cataloguing).
  - Definition of the Euclid Archive requirements, and use-cases resting on the legacy from previous large surveys (e.g. SDSS).

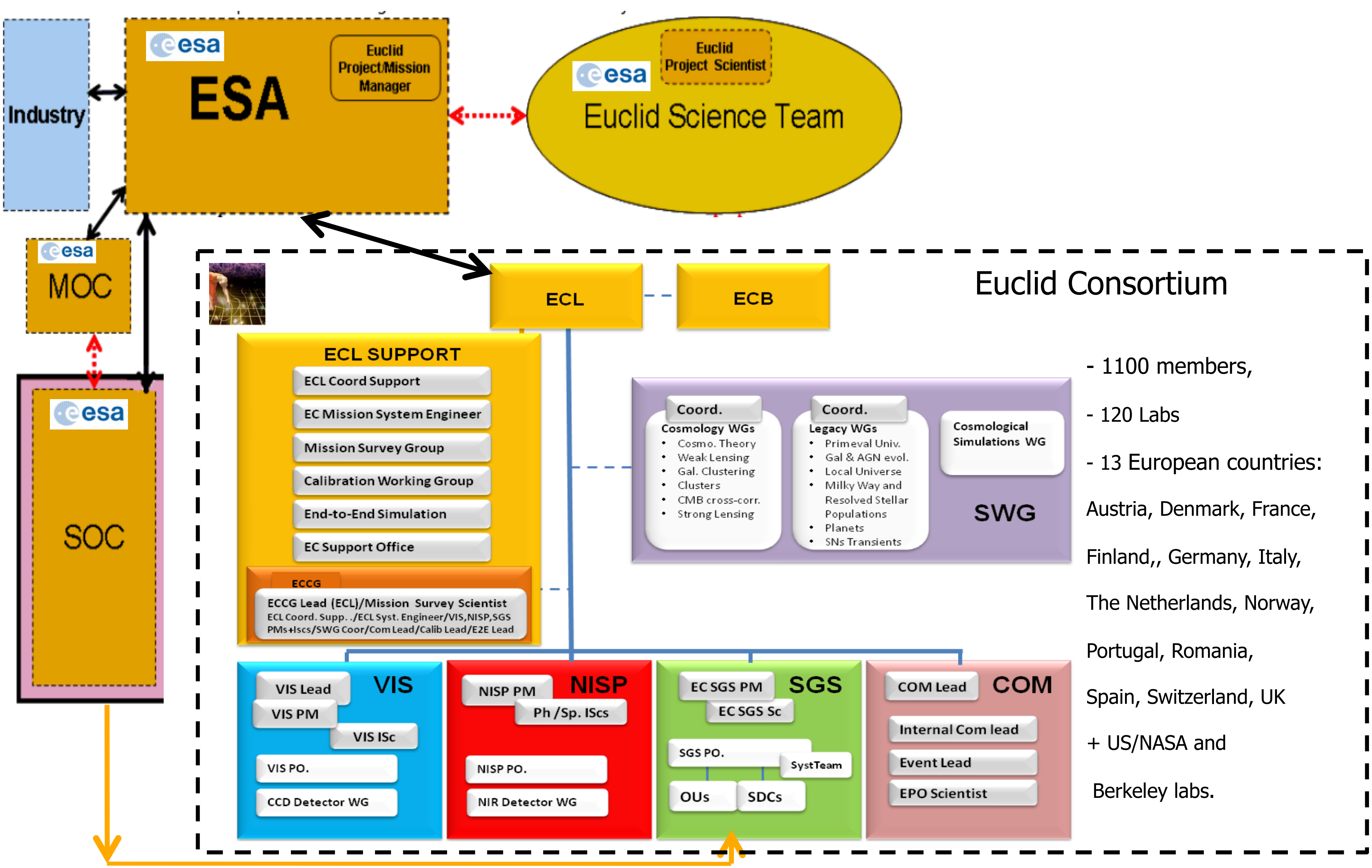
- For an in-depth view of the data processing, one can consult the Euclid Data and Processing Flow Document.

Marc Sauvage – EC SGS Scientist

THANK YOU!

Supplementary material

Euclid: organisation



Complexity at all levels

- The accuracy requirements for Euclid data processing are impressive...
- But the implementation challenges are as well:
  - 8+ data centers participating in the processing.
  - 10 main sectors of data processing.
  - Data volume reaching the pB scale.
  - Multiple processing of the whole dataset to be expected.
  - Quality control at every stage.
- Guidelines to achieve a conceptual vision of this flow:
  - The basis is the Ground Data Processing Requirement Document (GDPRD 2.0), capturing the flow down from the science requirements
  - Augmented by the Legacy Requirement Document (not formally binding).
  - The Science Implementation Requirement (SIRD 0.7) contains requirements that affect the SGS in a general way, but strictly not for its data processing activity.
- At present, concepts are captured in the Data and Processing Flow Document.

Euclid Legacy: Euclid data=a gold mine for astronomy for 2020-2040

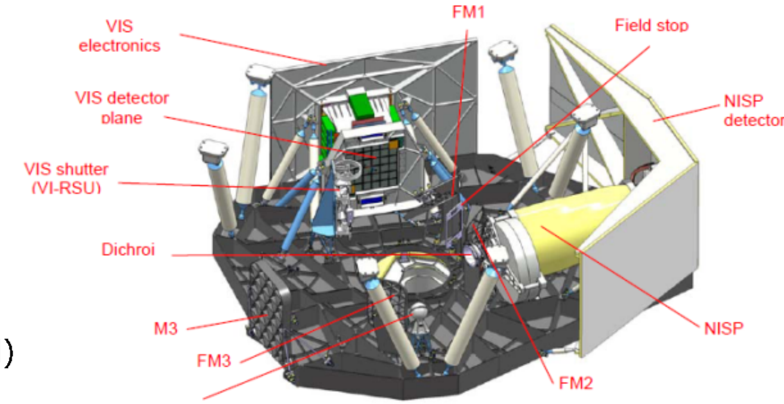
Objects	Euclid	Before Euclid
Galaxies at 1<z<3 with precise mass measurment	~2x10^8	~5x10^6
Massive galaxies (1<z<3))	Few hundreds	Few tenss
Hα Emitters with metal abundance measurements at z~2-3	~4x10^7/10^4	~10^4/~10^2 ?
Galaxies in clusters of galaxies at z>1	~2x10^4	~10^3 ?
Active Galactic Nuclei galaxies (0.7<z<2)	~10^4	<10^3
Dwarf galaxies	~10^5	
T_{eff} ~400K Y dwarfs	~few 10^2	<10
Lensing galaxies with arc and rings	~300,000	~10-100
Quasars at z > 8	~30	None

Operations

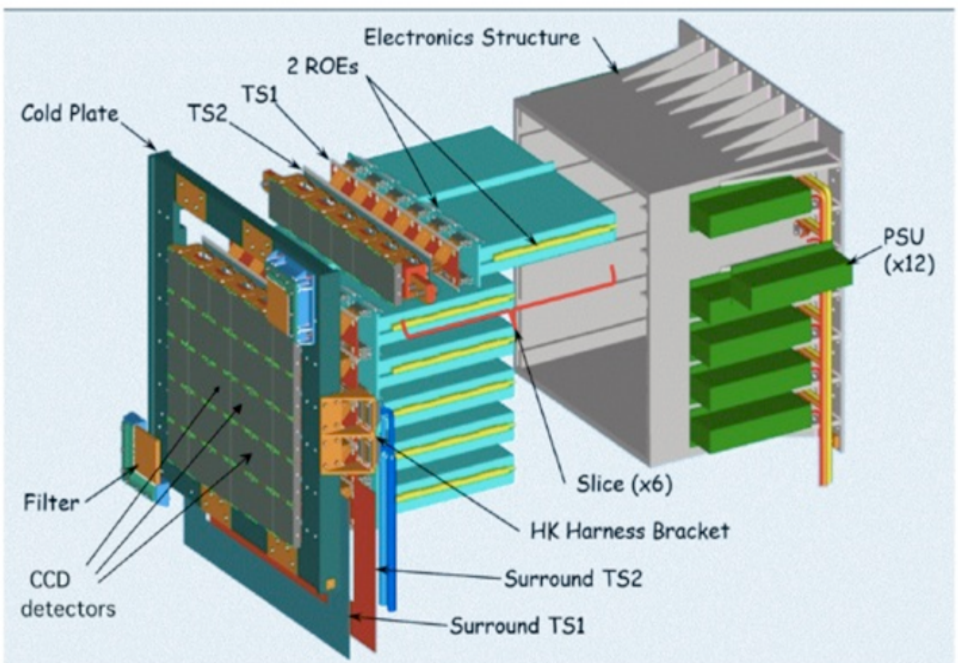
- MOC: provides planning information, including a predicted orbit, planned events at s/c and GS level, and tools/data to correctly plan the spacecraft pointing
- PS: provides survey strategy (EST recommendations)
- SOC: implements survey strategy: generates long term plan, derives a series of daily/weekly/monthly observation sequences
- IOTs (VIS, NISP):
  - maintain a routine calibration plan which is delivered to the SOC for execution; submit unplanned calibration requests to SOC as observation requests
  - single contact point to ESA: IOTs coordination person is part of the EC SGS PO
  - build upon instrument and data processing expertise

The VIS instrument

- large area imager – a ‘shape measurement machine’
- 36 4kx4k CCDs with 12 micron pixels
- 0.1 arcsec pixels on sky
- bandpass 550-900 nm – narrow band channel
- limiting magnitude for wide survey of magAB = 24.5 for 10σ (extended)
- data volume – 520Gbit/day

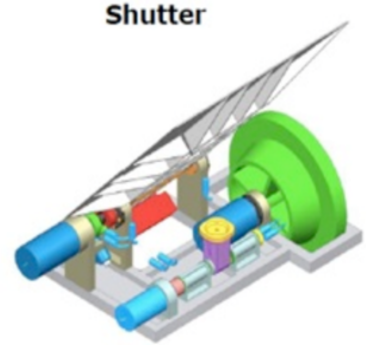


COLD

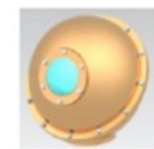


Focal Plane Assembly

Shutter

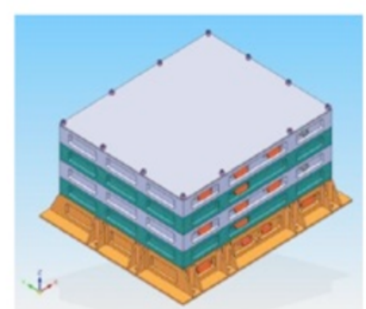


Cal Unit



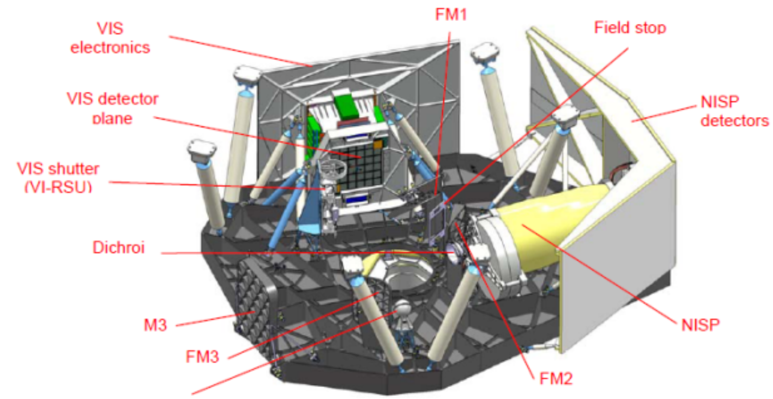
WARM

Power and Mechanisms Control Unit

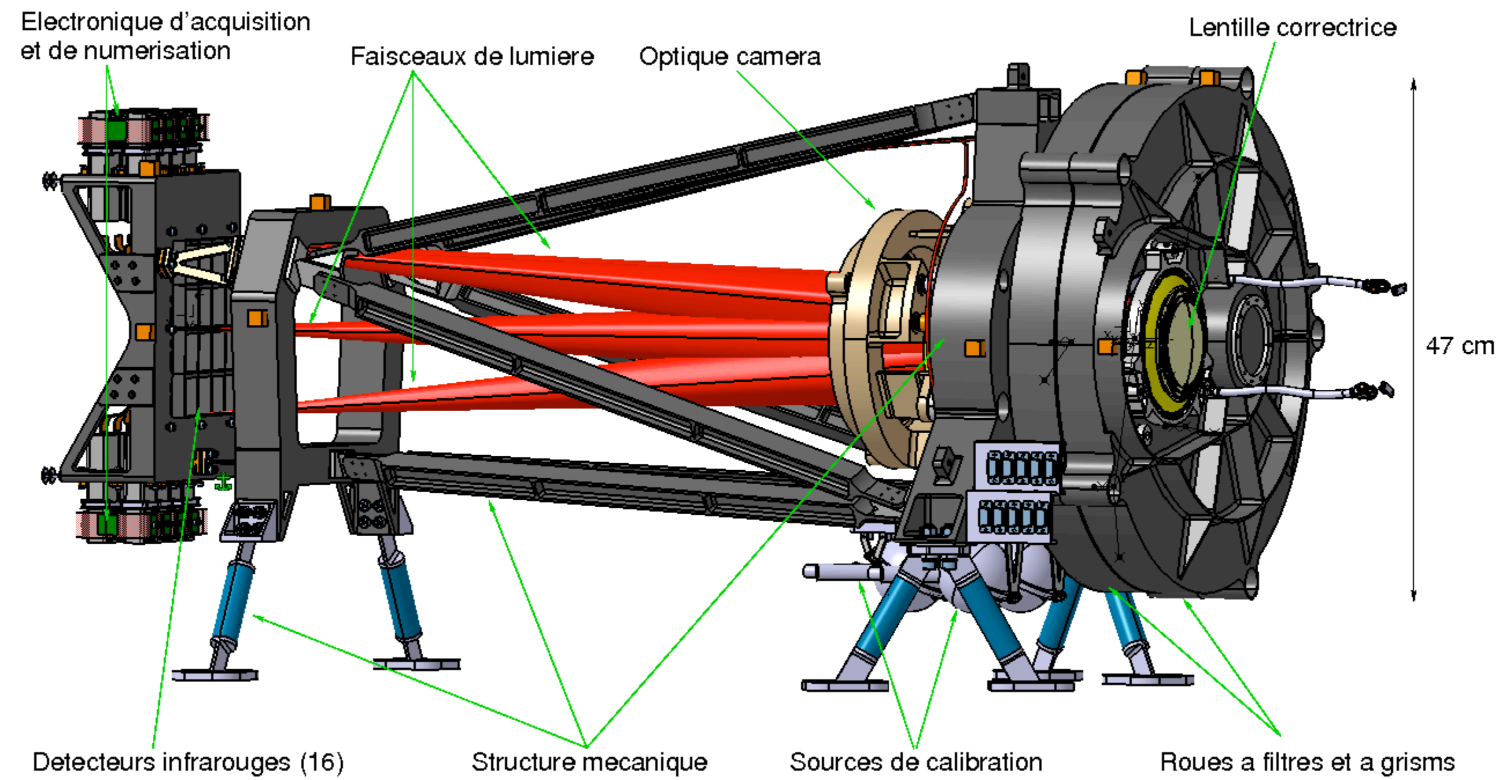


Command and Data Processing Unit

The NISP instrument



The NISP instrument: Near Infrared Spectrometer and Photometer



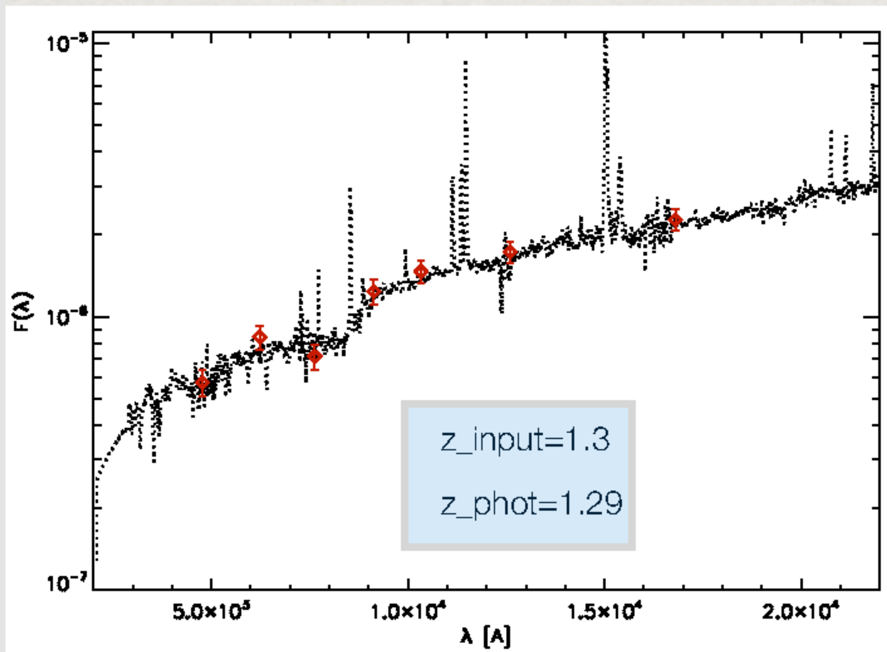
16 NIR detectors 2kx2k H2RG - 0.3 arcsec/pixel

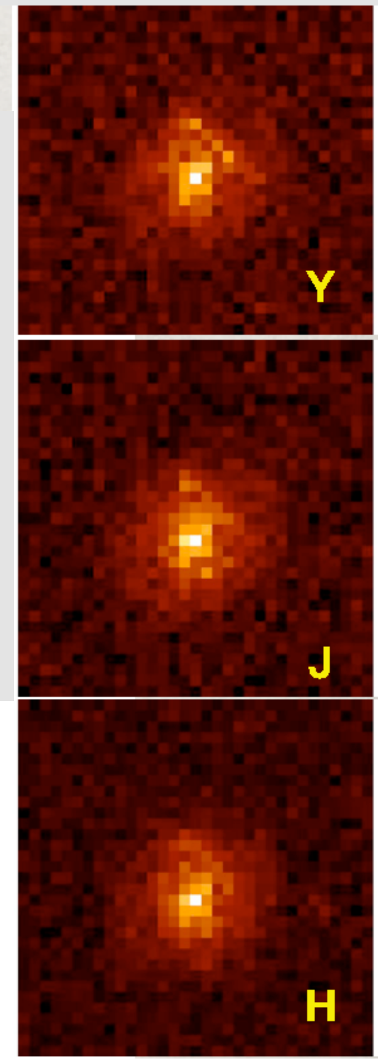
4 Grisms (2 B, 2 R, rotated by 90 deg.) ; 3 filters NIR: Y, J H - Telemetry= 180 Gbit/d

Data analysis of Euclid space+ground data

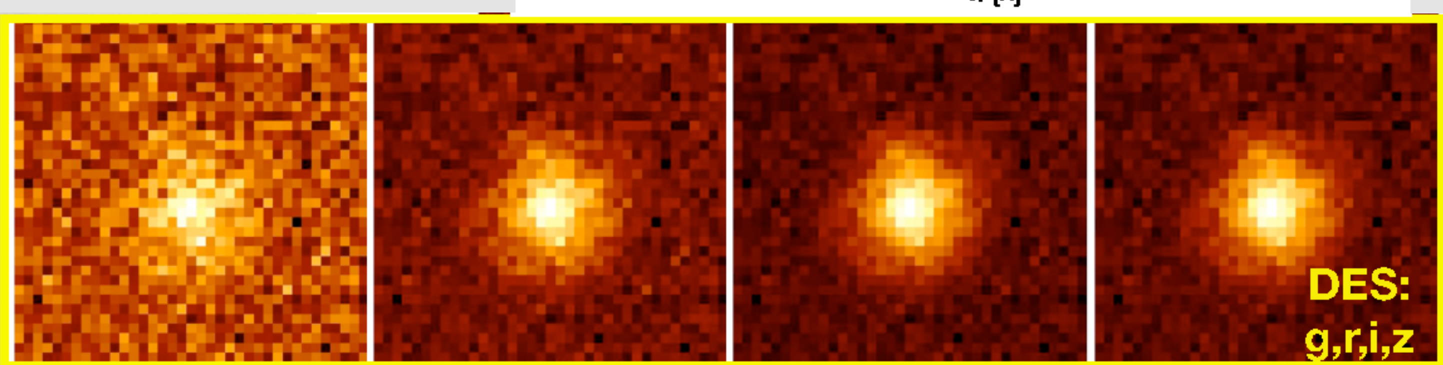
Photometric redshifts with Euclid

Nearinfrared data from Euclid NIR images.

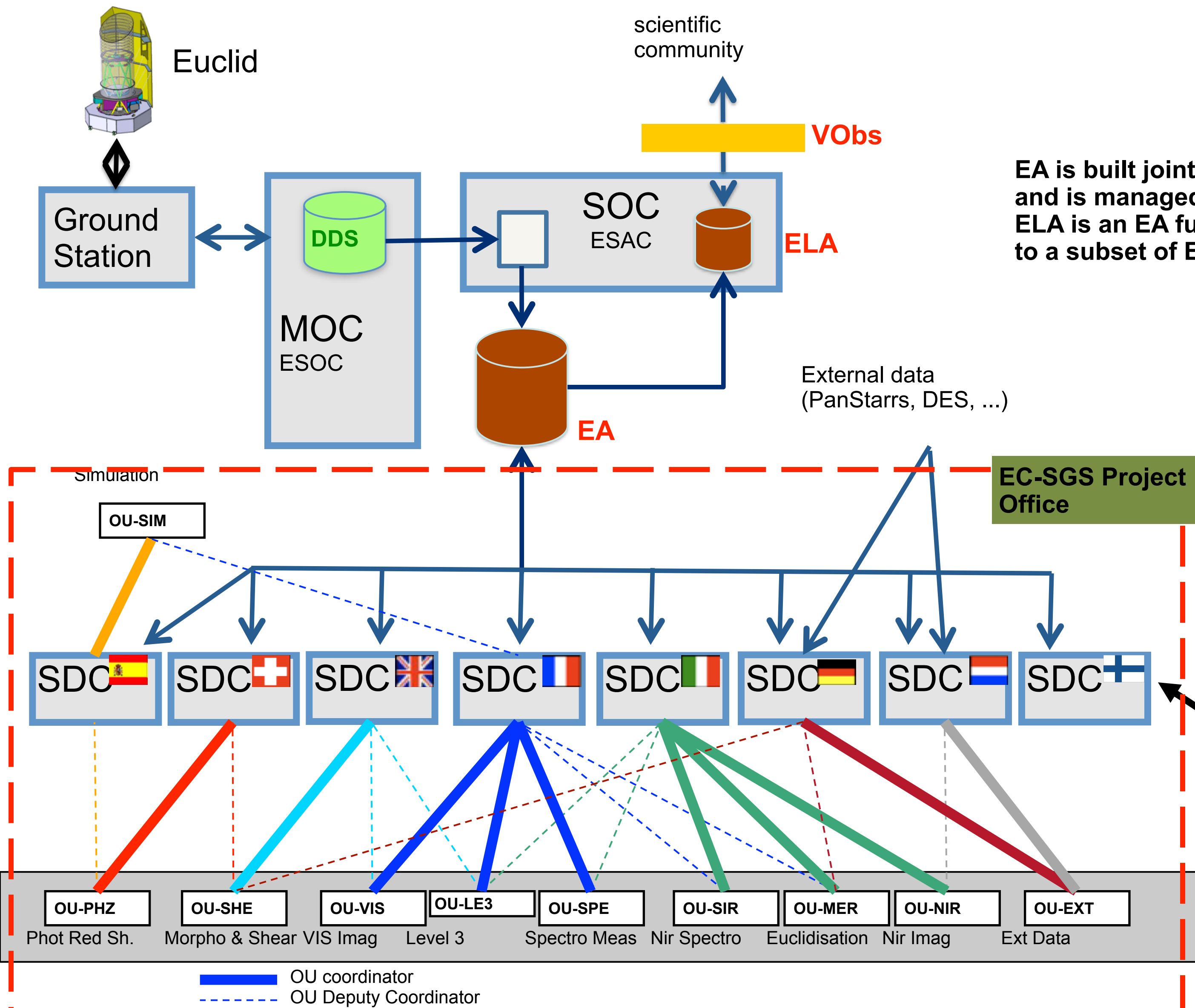




Visible data obtained from ground based telescopes



GS: overall view



EA is built jointly by EC and SOC, and is managed by SOC
ELA is an EA function, allowing public access to a subset of EA data

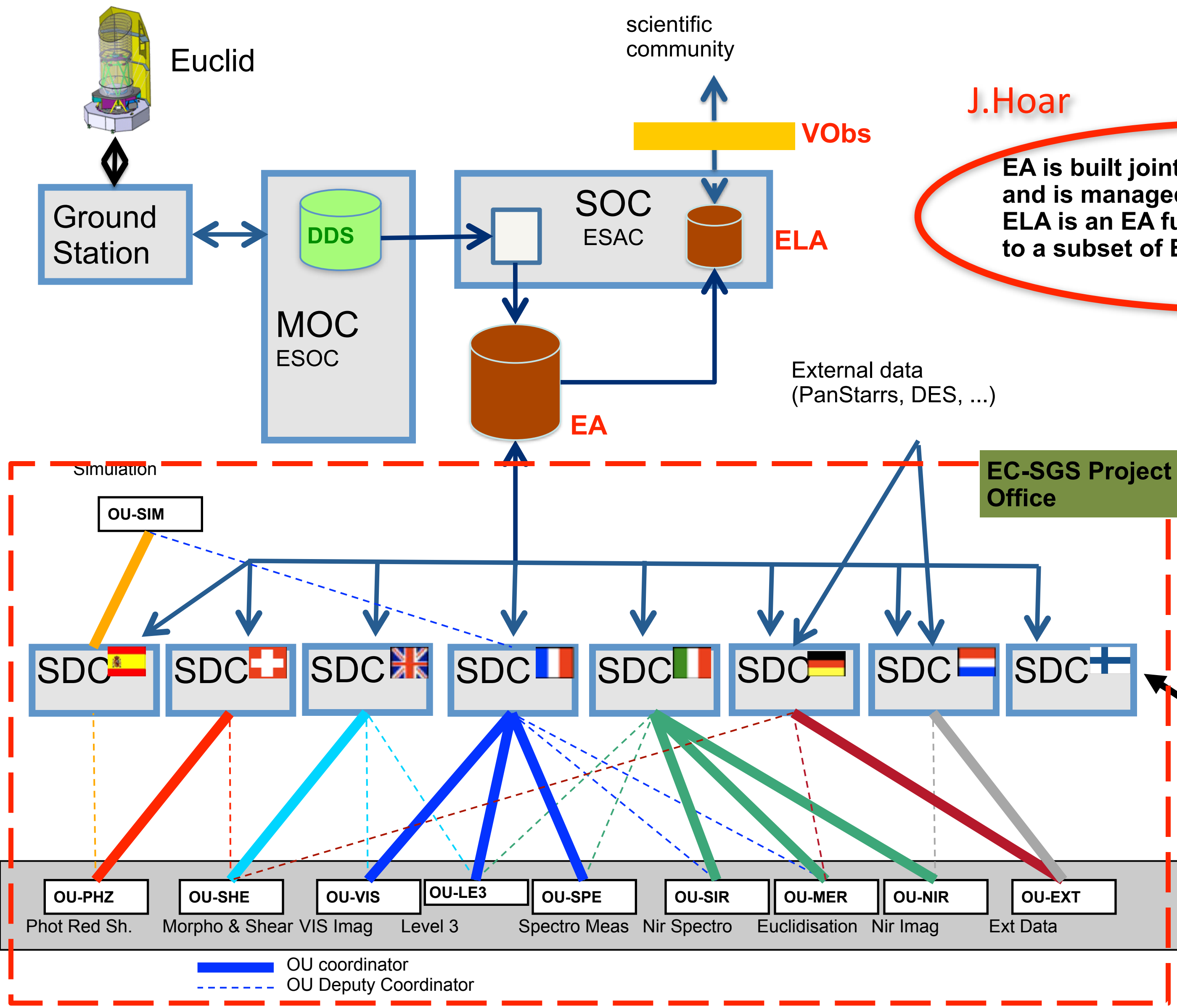
An organization based on the decomposition in Organization Units (OU), corresponding to a subset of overall EUCLID Data Processing.

Each OU produces algorithms which are integrated and executed in an SDC (Science Data Center)

The SGS System Team provides support and tools for the whole of the SGS (SOC + EC-SGS)

OUs are transnational

GS: overall view



An organization based on the decomposition in Organization Units (OU), corresponding to a subset of overall EUCLID Data Processing.

Each OU produces algorithms which are integrated and executed in an SDC (Science Data Center)

The SGS System Team provides support and tools for the whole of the SGS (SOC + EC-SGS)

OUs are transnational