Euclid Science Ground Segment

Structure & organization
(A personal point of view...)

Martin Vannier, PM Euclid@Lagrange
Sophie Maurogordato, co-lead of «Clusters» activities in Euclid SGS

« Lessons learned from other space missions »

Towards LISA catalogs, 12 june 2022

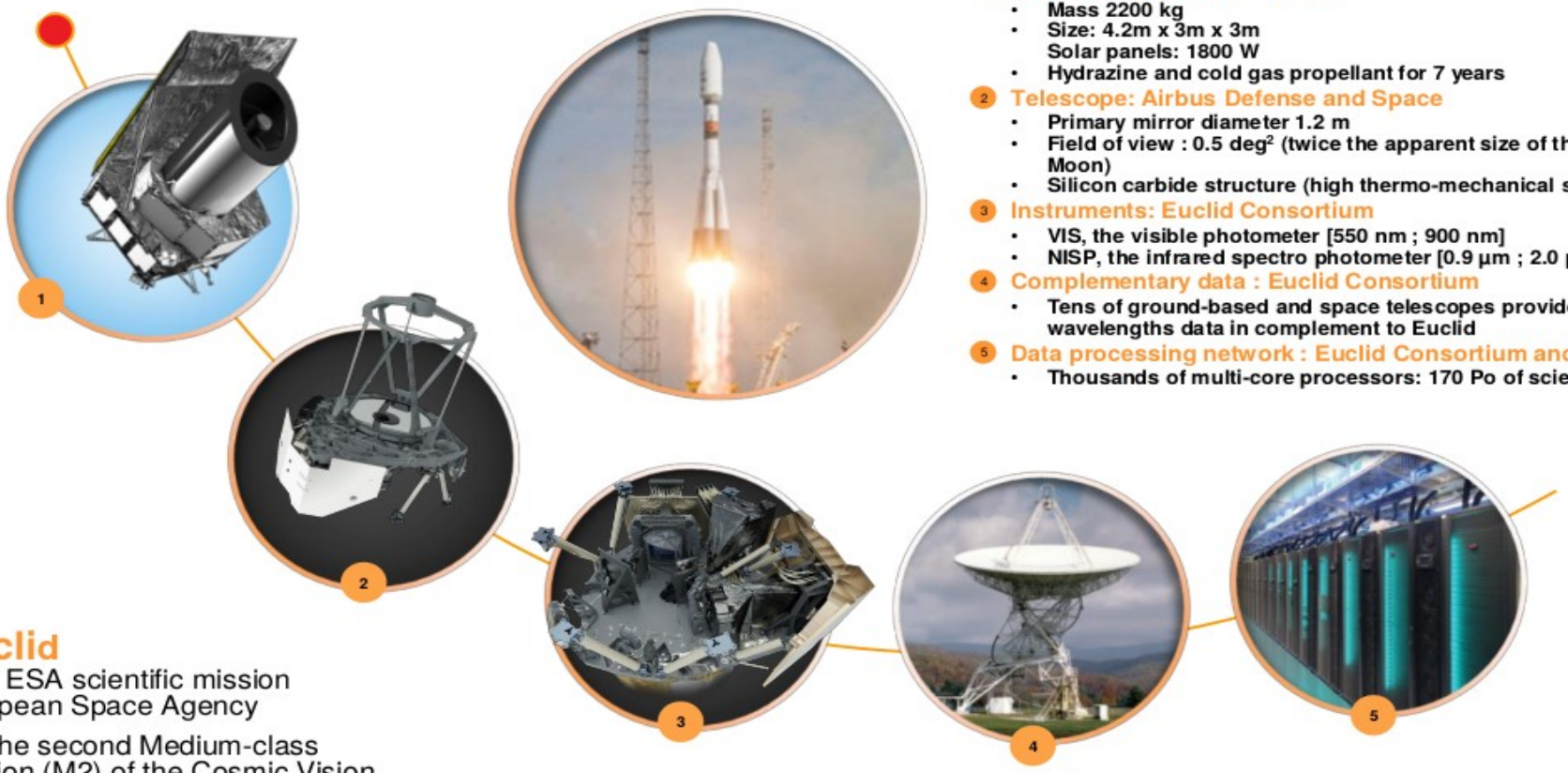
Euclid Ground Segment

- Euclid mission in 2 slides
- Euclid SGS :
  - Structure & organization
  - Interactions, development & maturity assessments
  - A few lessons learnt (partial and subjective!)
  - Local contributions and common topics with LISA L2→ L3 task

The ESA Euclid space mission

Launch foreseen in july 2023 (SpaceX, Cap Canaveral)

- 6-year mission around the Sun / Earth L2 Lagrange point,
- 6-month commissioning and performance readiness preparation,



1. **Satellite: Thales Alenia Space**
   - Mass 2200 kg
   - Size: 4.2m x 3m x 3m
     Solar panels: 1800 W
   - Hydrazine and cold gas propellant for 7 years
2. **Telescope: Airbus Defense and Space**
   - Primary mirror diameter 1.2 m
   - Field of view : 0.5 deg^2 (twice the apparent size of the full Moon)
   - Silicon carbide structure (high thermo-mechanical stability).
3. **Instruments: Euclid Consortium**
   - VIS, the visible photometer [550 nm ; 900 nm]
   - NISP, the infrared spectro photometer [0.9 µm ; 2.0 µm].
4. **Complementary data : Euclid Consortium**
   - Tens of ground-based and space telescopes provide multi-wavelengths data in complement to Euclid
5. **Data processing network : Euclid Consortium and ESA**
   - Thousands of multi-core processors: 170 Po of science data.

Euclid
is an ESA scientific mission
European Space Agency

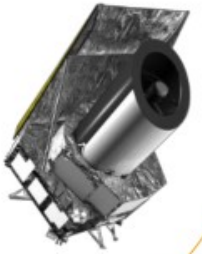
It is the second Medium-class mission (M2) of the Cosmic Vision programme

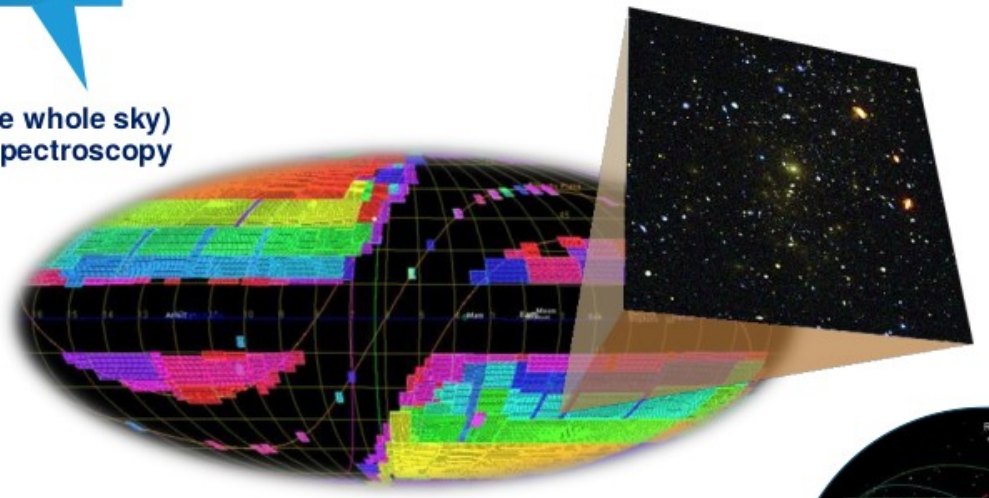
The Euclid survey

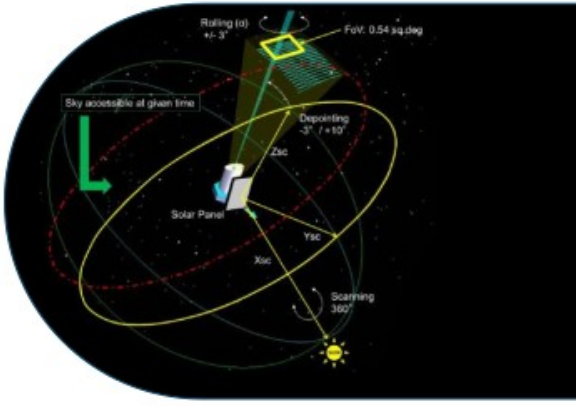
10 billion galaxies spread over 10 billion light-years

Euclid is a statistical "tracker" (a cosmological probe) of the history of the cosmic expansion and the formation of structures in the Universe

Survey of 15 000 deg² (1/3 of the whole sky)
Visible an infrared, imaging and spectroscopy

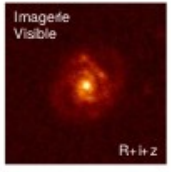


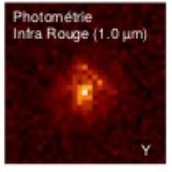


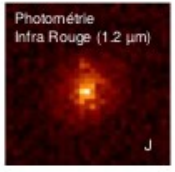


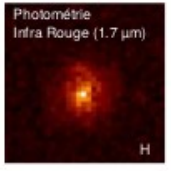
The Euclid survey will be built from observations of:
30 000 different fields of 0.5 deg² each on the sky
10 000 calibration fields

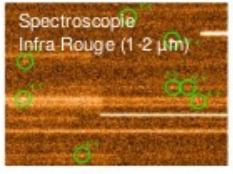
with:
Selection of the best areas of the sky (minimizing zodiacal light and stars contaminations)
Complex constraints on the satellite to minimize thermal perturbations







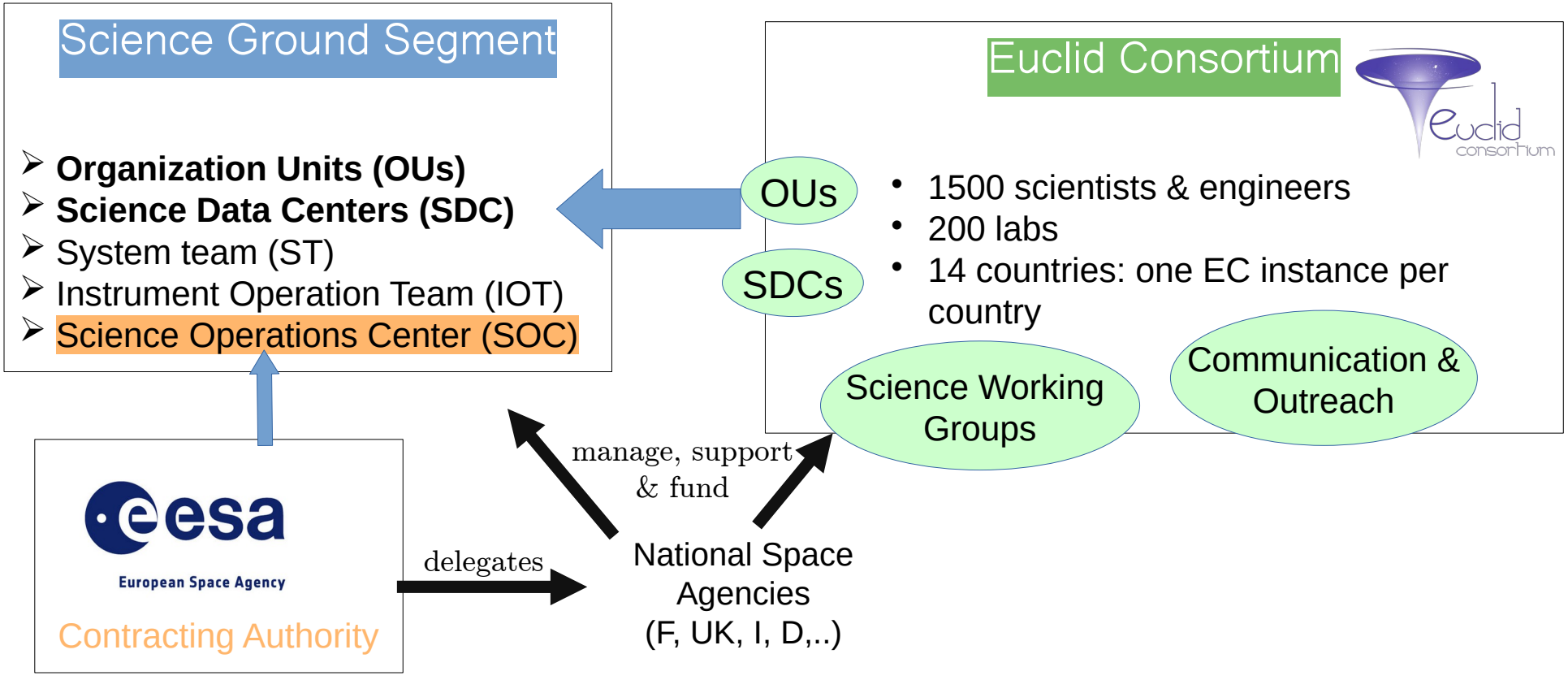




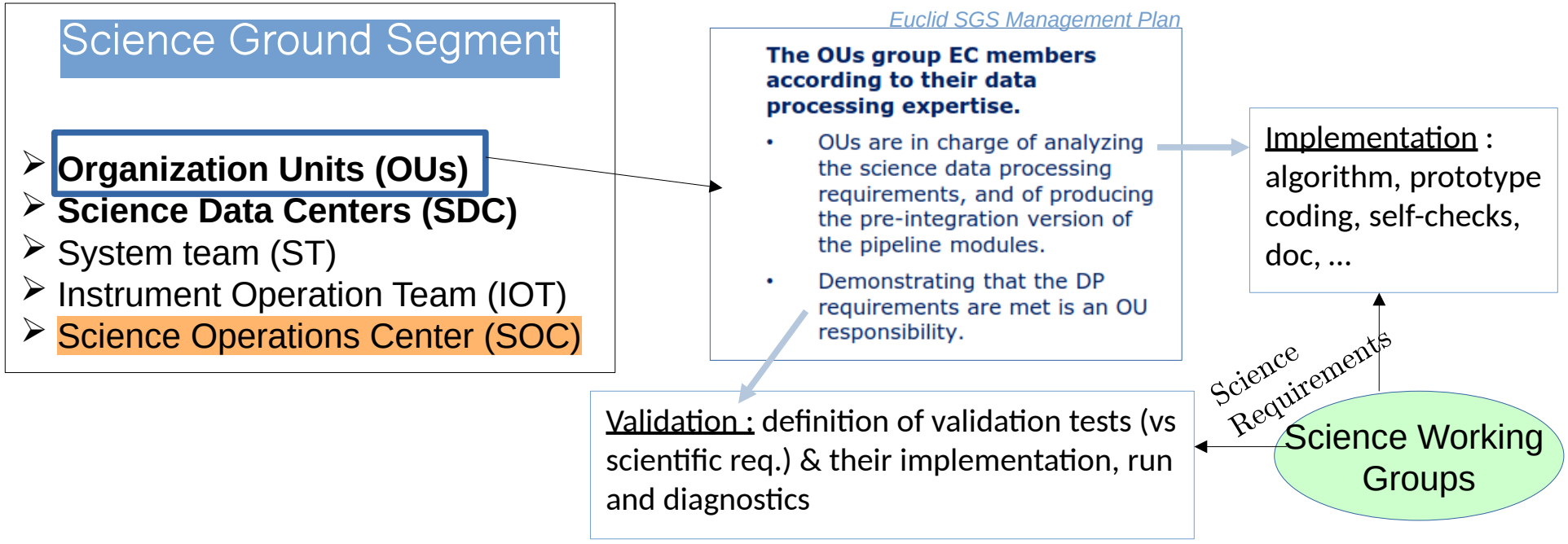
10 billions of galaxies observed
in visible and infrared photometry

30 millions of
infrared spectra

Euclid SGS & Consortium : the big picture



Le Segment Sol (Science Ground Segment)



This scheme applies to any « **Processing Function** » :
- well defined funtional perimeter
- own lead and dedicated team, own development cycle

- *Detection of Galaxy Clusters*
- *Selection Function of the detected Galaxy Clusters*

Le Segment Sol (Science Ground Segment)

Science Ground Segment

- **Organization Units (OUs)**
- **Science Data Centers (SDC)**
- System team (ST)
- Instrument Operation Team (IOT)
- Science Operations Center (SOC)

Euclid SGS Management Plan

The SDCs are built around national HPC facilities, they gather IT support as well as developer expertise.

- SDCs are in charge of the pipeline development (software and support architecture).
- Production software will run in the SDCs.
- Act as focal points for instrument operations tasks.

Before the launch :

- Technical implementation (integration,deployment, tests, ...) of OU codes according to SGS standards (see « PF maturity progress »)
- Processing Functions chained in pipelines using proper Data Model

Before the launch :

- Production/exploitation of these processing pipelines
- Archive of intermediate and finite data products

Le Segment Sol (Science Ground Segment)

SGS Structure

The SGS consists of the following entities:

- The Organization Units (OUs)
- The Science Data Centers (SDC)
- The Science Operations Centre (SOC)
- The System Team (ST)
- The Instrument Operations Team (IOT)

The System Team has the responsability to build the Common Software infrastructure and define the Data Model

The IO Team has the responsability to operate the instrument during nominal operations

Organisation Units (OU)

OUs are organized by :

- Data flow level (1 to 3)
- Technical scope

Each OU is divided in work-packages

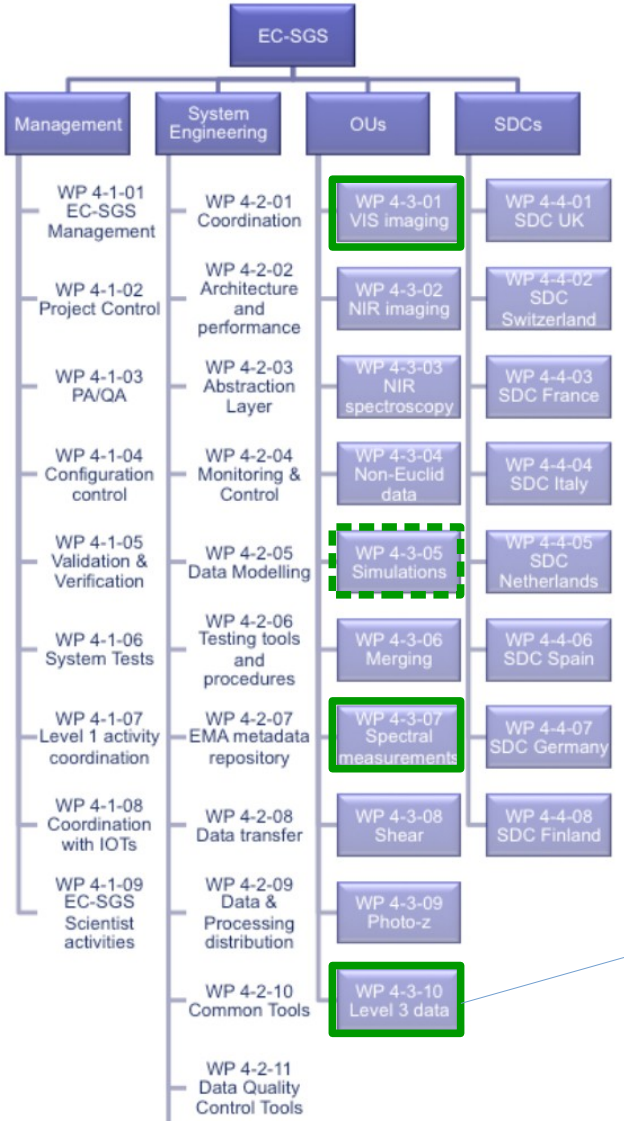
LE1

LE2

LE3 Work packages :

- Internal Data Handling
- External Data Handling
- Galaxy Clustering
- Clusters of Galaxies
- Lensing
- Time Domain
- Milky Way and Nearby Galaxy

Co-lead : S. Maurogordato (OCA)
(Implementation)



Science Working Group
« Clusters of Galaxies »

Science req.

Processing Functions « Clusters of Galaxies »
(**co-lead @ Lagrange**) :

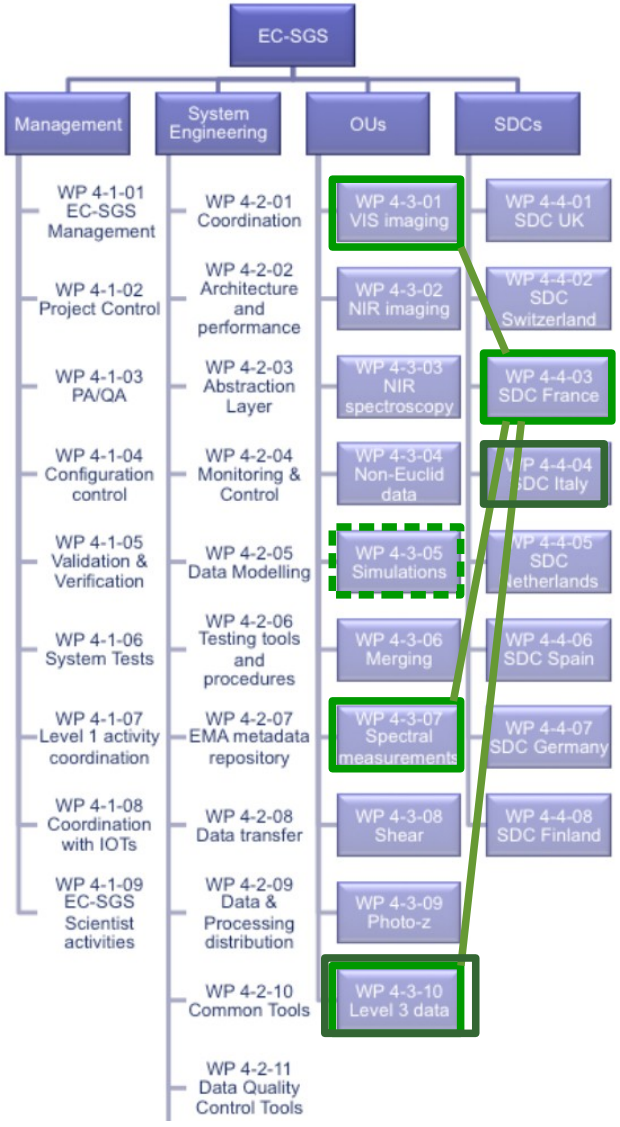
- *DET-CL: Detection of clusters of galaxies from the L2 catalogue*
- *RICH-CL : Richness and membership probabilities of the clusters*
- *PROF-CL : Radial profile of clusters*
- *COMB-CL : Mass estimate of clusters using weak-lensing*
- *SEL-CL: Selection Function of the detected galaxy clusters*
- *CAT-CL : Merge and cross-match of the detection catalogs and their parameters*
- ...+ a few more in development.

OU-LE3 Work packages :

- Internal Data Handling
- External Data Handling
- Galaxy Clustering
- Clusters of Galaxies
- Lensing
- Time Domain
- Milky Way and Nearby Galaxy

Implementation : algorithm, coding, self-checks, doc, ...

Validation : definition of validation tests (scientific req.), implementation, run and diagnostics



Science Data Centers (SDC)

- SDCs are organized by country
- Each is (co-)lead for some OUs, in charge of :

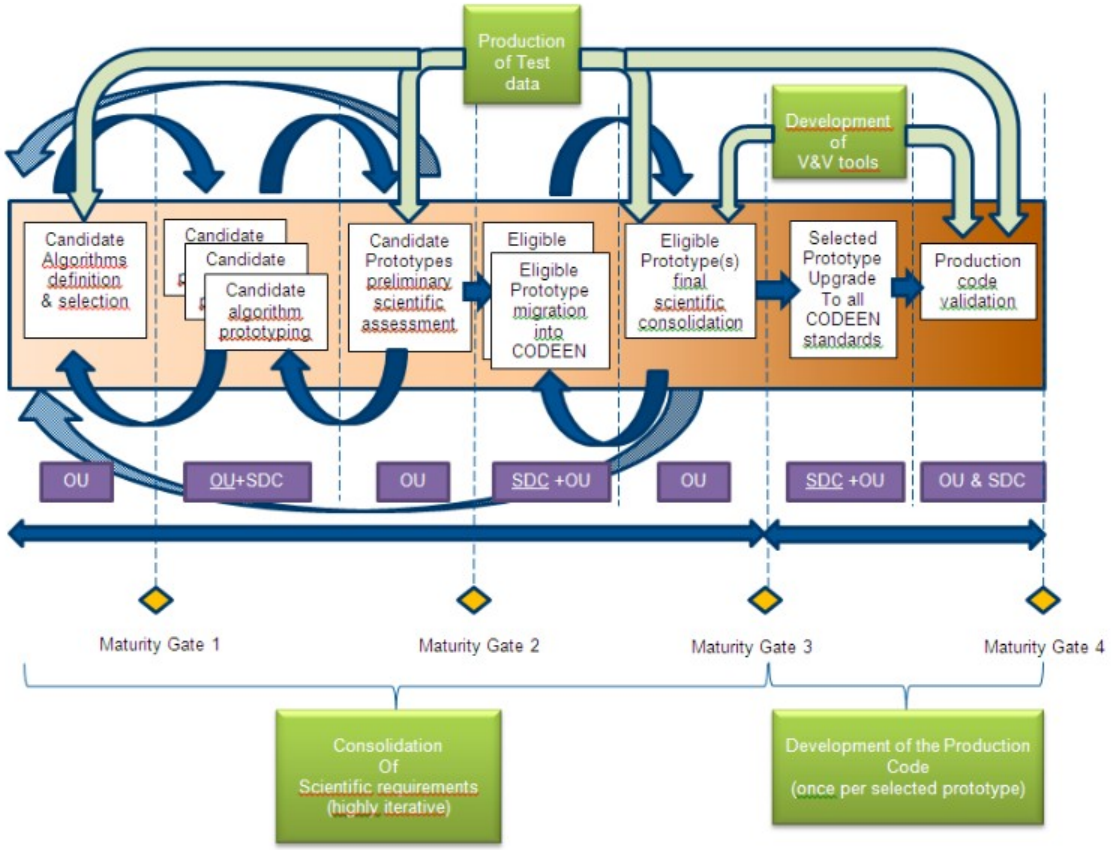
→ Technical implementation (integration, deployment, tests, ...) of OU codes to strict SGS standards

→ Production/exploitation of these processing pipelines

→ Archive of intermediate and finite data products

In France : 14 labs (INSU, CEA, IN2P3)

Development & Maturity Levels



- Each Processing Function undergo Maturity Level Assessments as it progresses.

- Should meet requirements on :
  - Science (definition and validation of specifications)
  - Interface (Data Model, etc...)
  - Technical (algorithm, code implementation, integration)
  - PA/QA (code quality, test coverage, doc...)

- ➔ Formal reviews
- ➔ Dedicated tools and metrics (Jenkins, SonarQube, ...)
- ➔ Continuous iterations between OU (science specs. and validation, new features, ...) and SDC (technical implementation)

Organisation

- Classical phases for project management (currently D/E1) and associated milestones (« Readiness Review », etc...)
- Contributions to SGS supervised and funded at national level.
- (CNES) : Commitments (delivrables, planning) and needs (missions, short-term engineers,...) described in « Technical and Financial Proposal » for each dev. phase.
- Typically , 1 FTE → + 1.5k€/y hardware

Communication

- Regular telecon and meetings and different levels (OU, PF, SGS-Fr) and for each scientific topic (OU+SWG « Clusters »)
- Intensive use of **Redmine** (wiki, file repo, « issues », etc...=) : + 120 Euclid projects, and gitlab for (prototype or official) code sharing and versioning

Computation choices

- Originally : C++ and Python as « glue language »
- In practice, C++ and/or Python for any application (except specific exec. Time constraints)
- Most of the standard sci/astro/io libraries included

What works well… and not so well

(a subjective, partial, non-echaustive view)

- CNES very supportive, at all levels
- Incentives for formal management (reports, planning), documentation and PA/QA (code, tests, etc...) are a pain, but **a very necessary one.** Formalizing things (up to a certain limit) takes time but brings order.
- Technical choices were globally good (things work !) : integrated environment, including plenty of usefull common tools...

- Formal division between some groups may increase inertia or miscommunication, although they might have important mutual dependencies

SWG <=> OU, for definition and detailed analysis of scientific requirement.

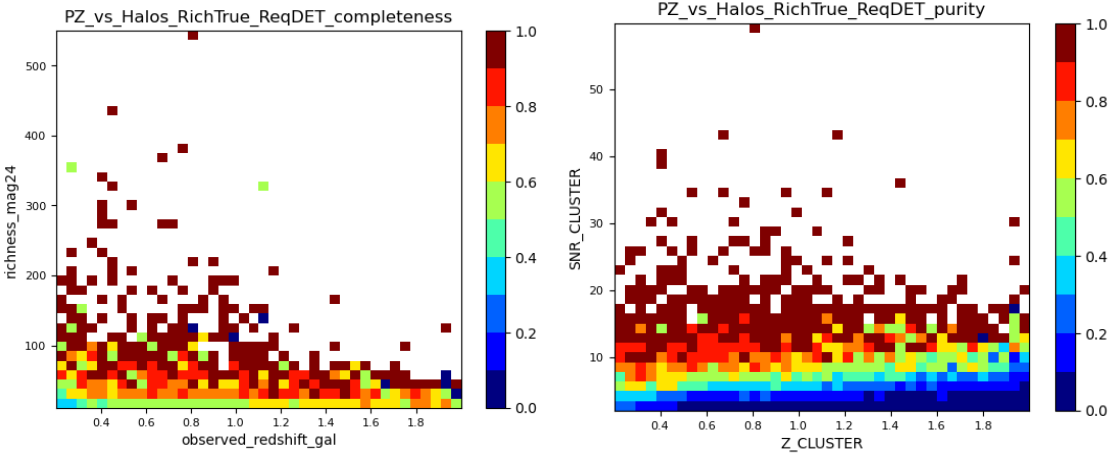
- Steep learning curve on the technical aspects (SDC) make it difficult for most newcomers to put their hands in actual development process.

→ OU scientists often require SDC engineers to upgrade the code (or even to run it), which slows down the interactions. (Would have been much worse with C++ as default langage)

- Difficult to produce (and to get) large end-to-end simulations. Custom « LE3 » simulations very usefull, but idealistic

Some similarities between our local contribution on Euclid and the « L3 catalog » LISA workpackage

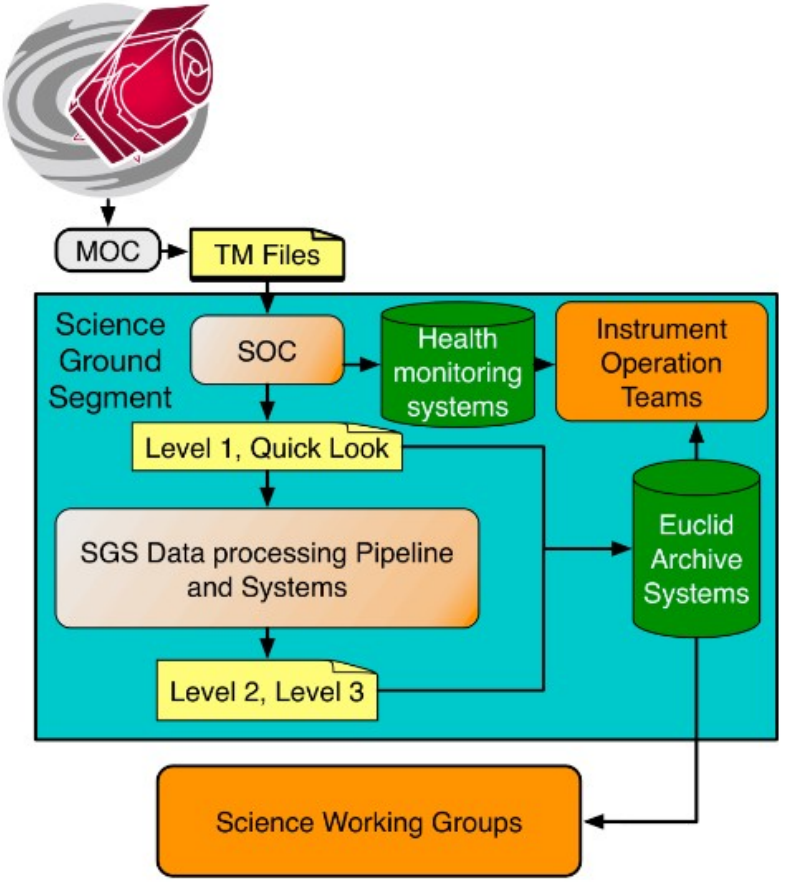
- Our Euclid task : Detect and characterize the « clusters of galaxies » using LE2 outputs
- Four « cluster finder » challenges using simulated data (increasing complexity and scale)
- A « matching » is performed between detection and « truth table » . The results of a finder are estimated in terms of Completeness and Purity
-



<u>But :</u>

- What is a « real match » ? How to define exclusive areas in the parameters space ? How to assess fragmented or merged detections ? Different matching techniques...
- Representativity of simulations (scientific and technical complexity) may be dubious, especially at low SNR, high redshift, etc. where it becomes interesting !
- How to compare, without truth table, different detections sets (each with its biases, limits,..) ?

The SGS perimeter and operational role



- This is an overview of the SGS system for **Euclid operations**.
- In **grey**: systems performing operations, in **orange** teams performing actions and taking decisions.
  - SGS and SOC are both!
- Calibration activities (not represented) occur both at SGS data processing level and at instrument operations level (scope is different).
- **EAS is in fact an active system**, it provides data management and a transfer system for the SGS.