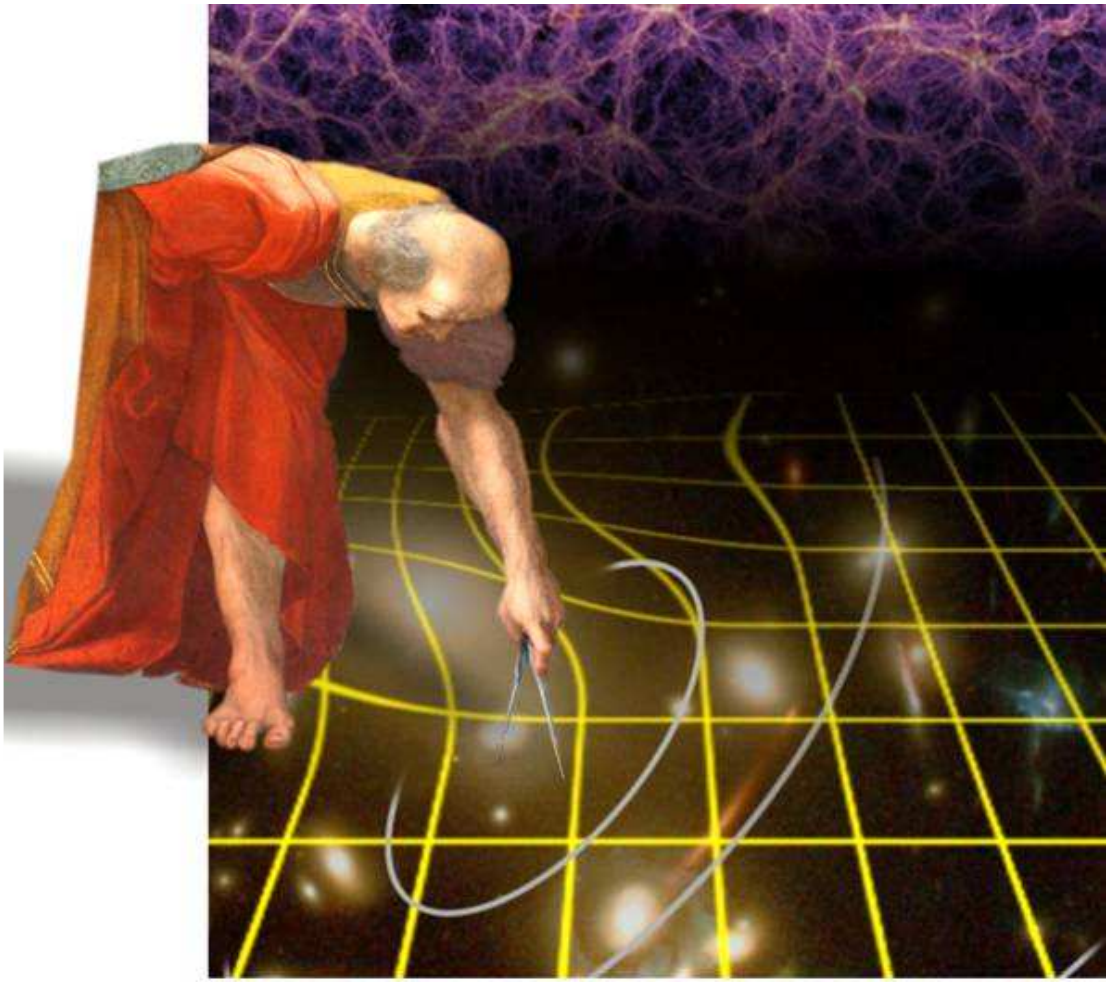


Euclid

The challenge of Euclid SGS (Science Ground Segment) Architecture

M. Poncet (CNES)
On behalf of
Euclid SGS System Team

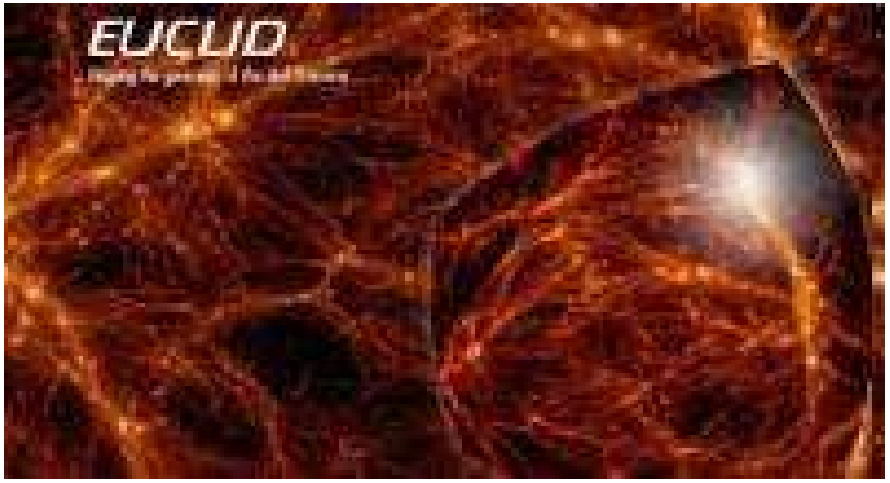
http://www.euclid-ec.org

Euclid - Agenda

- ✓ **Euclid Mission**
- ✓ **Space Segment**
- ✓ **Ground segment & Organization**
- ✓ **Data & processing Challenges**
- ✓ **Design principles**
- ✓ **SGS Overall Architecture**
- ✓ **Current Status & Schedule**
- ✓ **Conclusion**

Euclid - Context

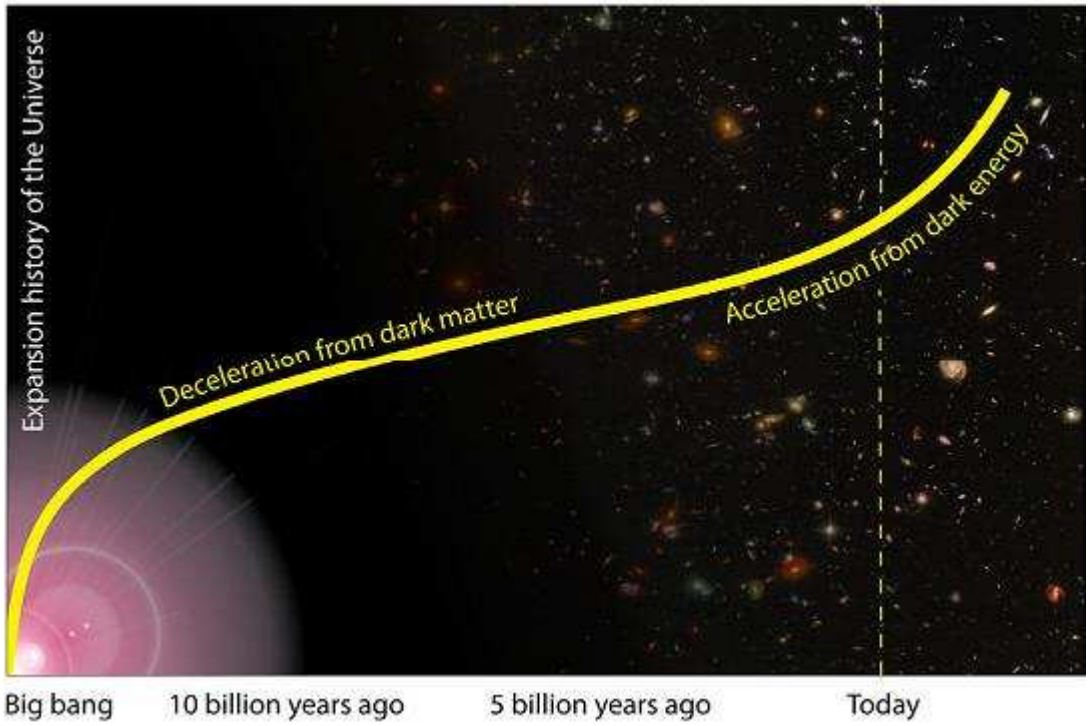
- M2 mission in the framework of **ESA Cosmic Vision Program**
- Euclid mission objective is to map the geometry of the **dark Universe**.
- Mission endorsed by **ESA** and a **European Consortium**



- For more information see :
  - http://sci.esa.int/science-e/www/area/index.cfm?fareaid=102
  - http://www.euclid-ec.org
  - http://smsc.cnes.fr/EUCLID/index.htm

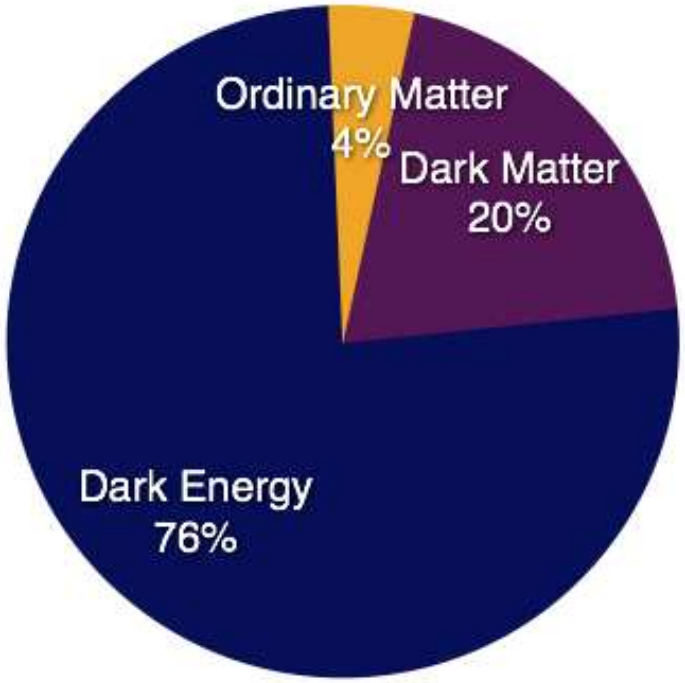
Euclid – The big Question

Euclid Consortium



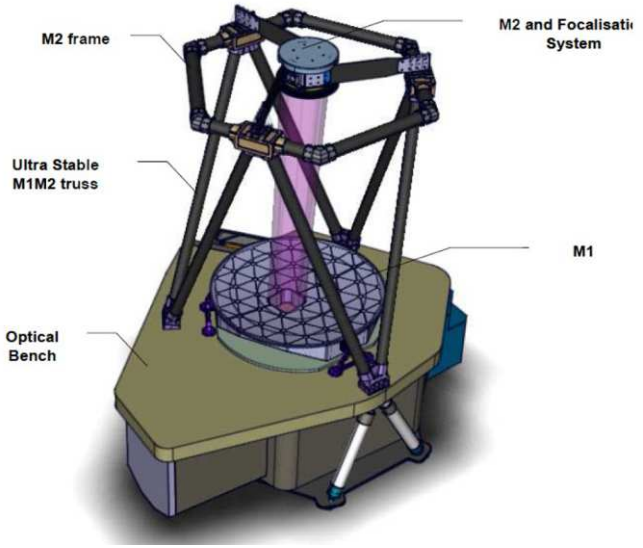
The acceleration of the Universe is produced by a new component called « Dark Energy »

The expansion of the Universe is accelerating !



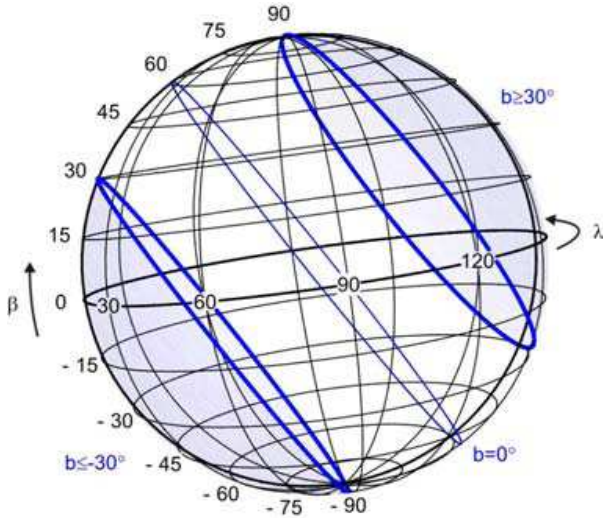
Euclid – Objectives of the mission

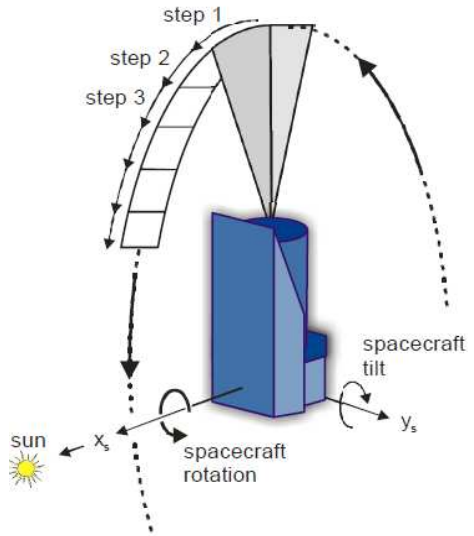
- Understand the **origin** of the Universe's accelerating expansion
- Probe the **properties** and **nature** of dark energy, dark matter, gravity, and
- **Distinguish** their **effects** decisively by:
  - Using at least 2 independent but complementary probes
  - Tracking their (very weak) observational signatures on the
    - geometry of the universe: Weak Lensing (WL) and Galaxy Clustering (GC)
    - cosmic history of structure formation: WL, Redshift-Space Distortion (RSD), clusters of galaxies (CL)
  - Controlling systematic residuals to an unprecedented level of accuracy.



Euclid - Mission = Survey

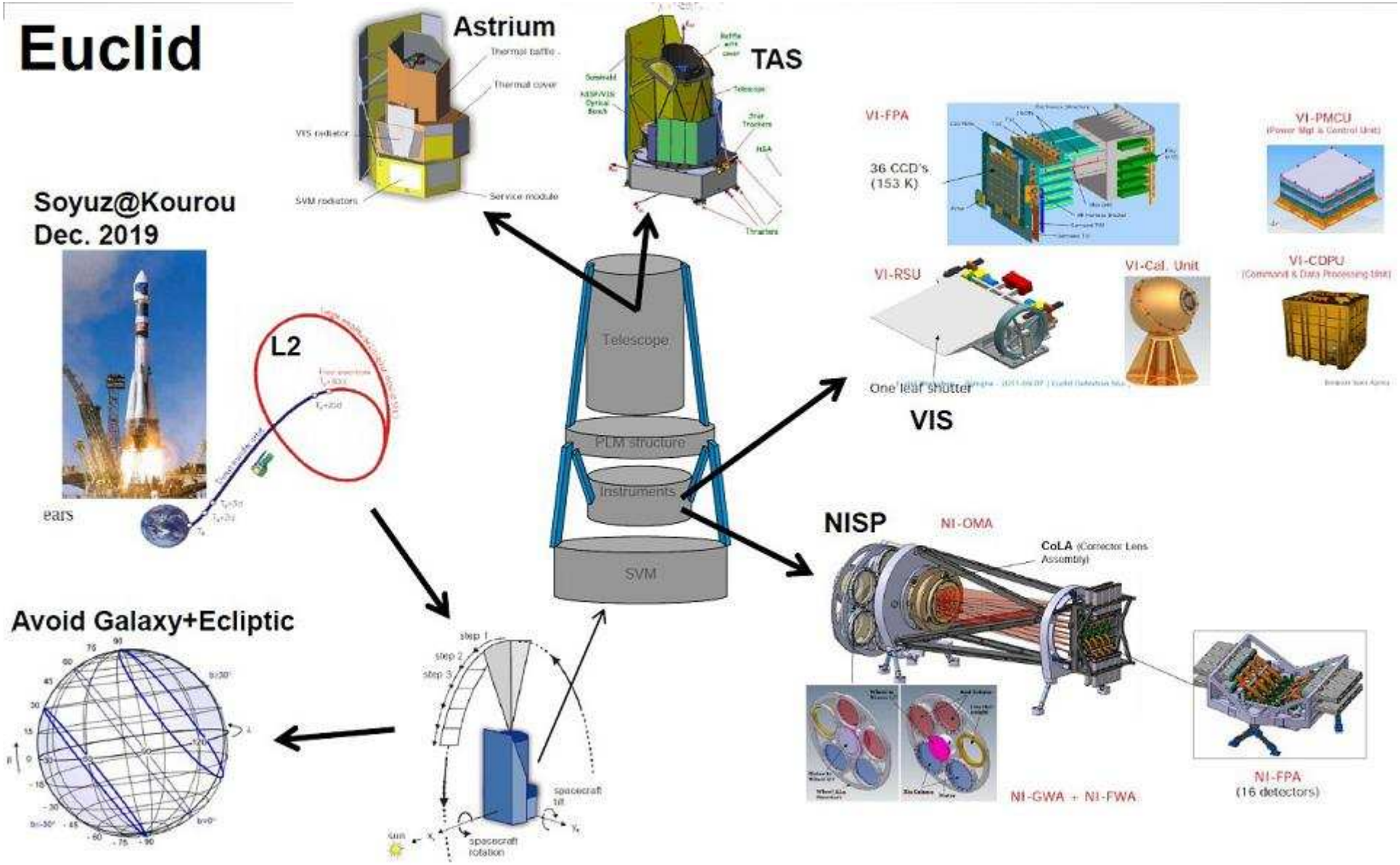
- Survey **mission** with **6 years** nominal science operation duration.
- The wide **extragalactic** sky **survey** covers 15 000 deg2, and about 1.5 10^9 galaxies
- The **deep survey** covers 40 deg2 around ecliptic poles, and about 10,000 galaxies
- The 3 axis stabilized spacecraft is operated in step and stare mode (around the S/C sun axis) to observe galactic latitudes > 30 degrees.)





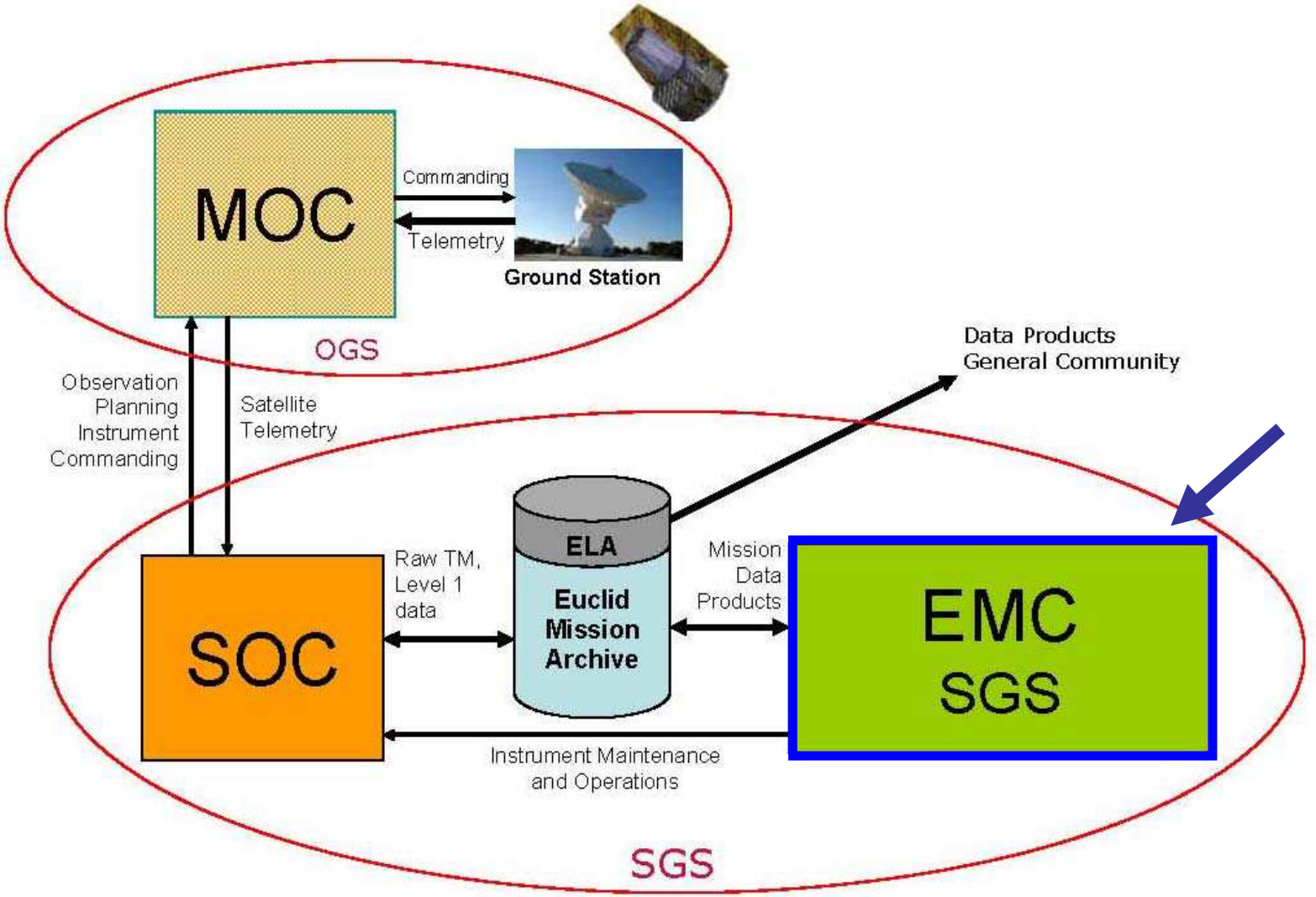
Euclid – Space Segment at a glance

Euclid Consortium



Euclid - Ground Segment (1/2)

Euclid Consortium



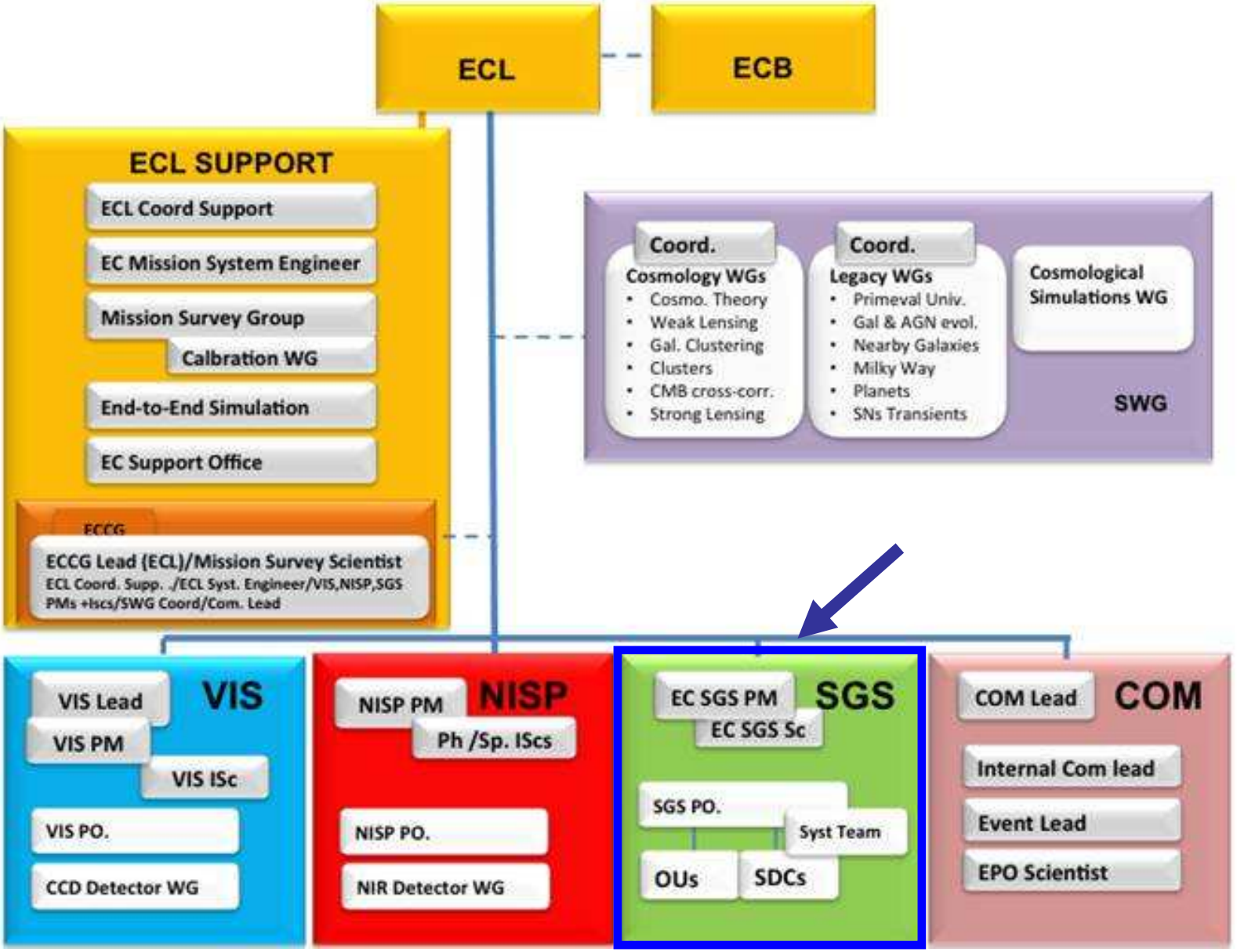
Euclid - Project Organization (2/2)

Euclid Consortium

- Euclid Mission Consortium (**EMC**) is constituted by national space agencies and scientific labs.
- **13 countries** contributing to the EC : Austria, Finland, UK, Denmark, Germany, France, Netherlands, Italy, Spain, Norway, Switzerland, Romania, Portugal + Berkeley labs;
- About **900 persons** in EMC, 500 researchers from about 109 labs/departments
- **ESA** is in particular responsible for the satellite, the telescope, the Operation Ground Segment (Mission Operation Center and Ground Stations) and for Science Operation Center.
- **EMC** is in particular responsible for the instruments (VIS and NISP) and the Scientific Ground Segment (science and processing).
- => **Close cooperation** between ESA and Euclid Consortium.

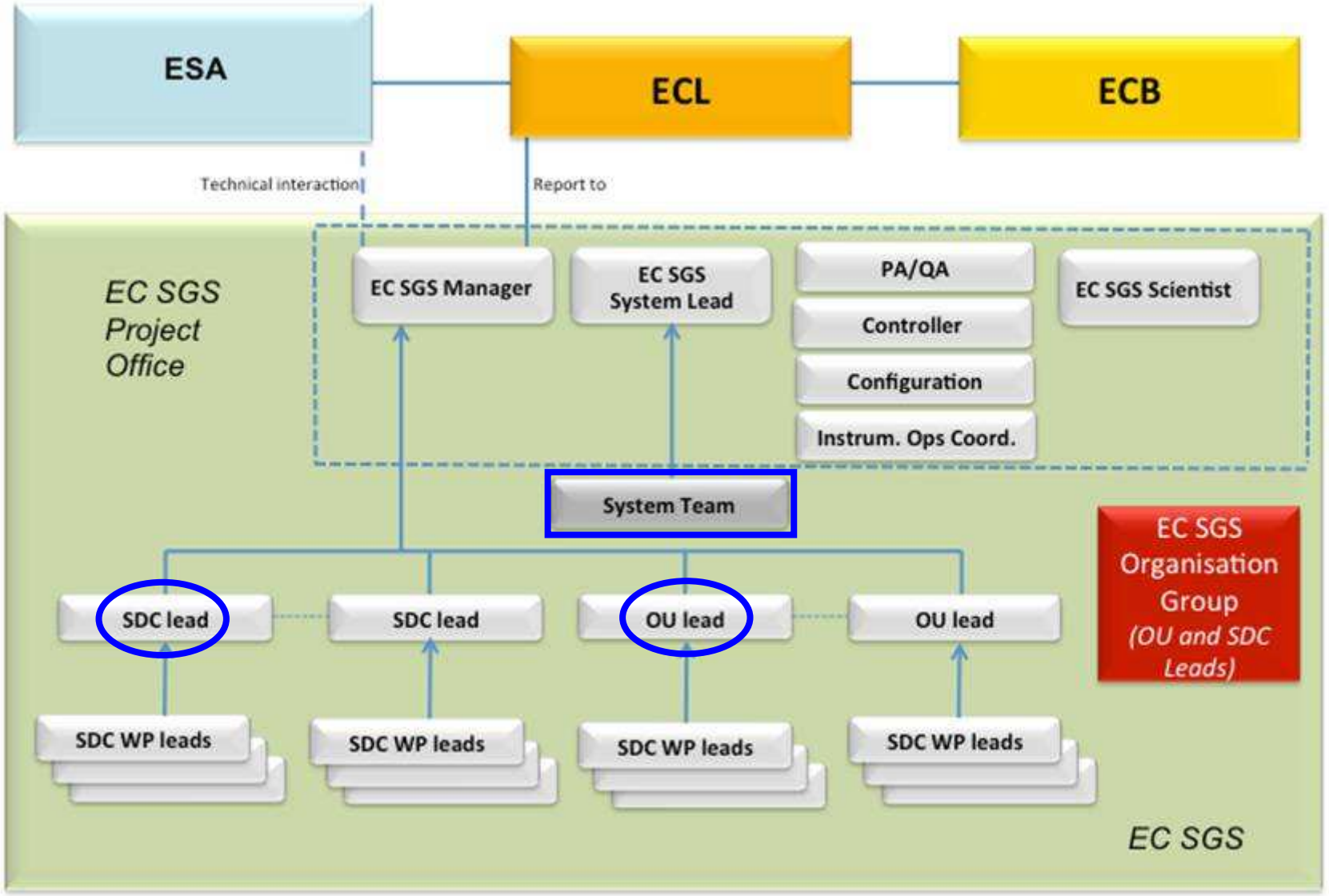
Euclid - Consortium Organization

Euclid Consortium



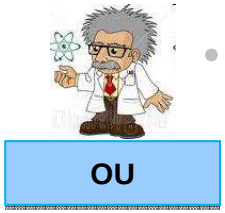
EMC – EC SGS Organization

Euclid Consortium



Euclid – OUs / SDCs (1/2)

- Euclid overall processing from raw data up to level 3 is divided into "**pipelines"**, according to levels, sub-steps or thematic of processing.

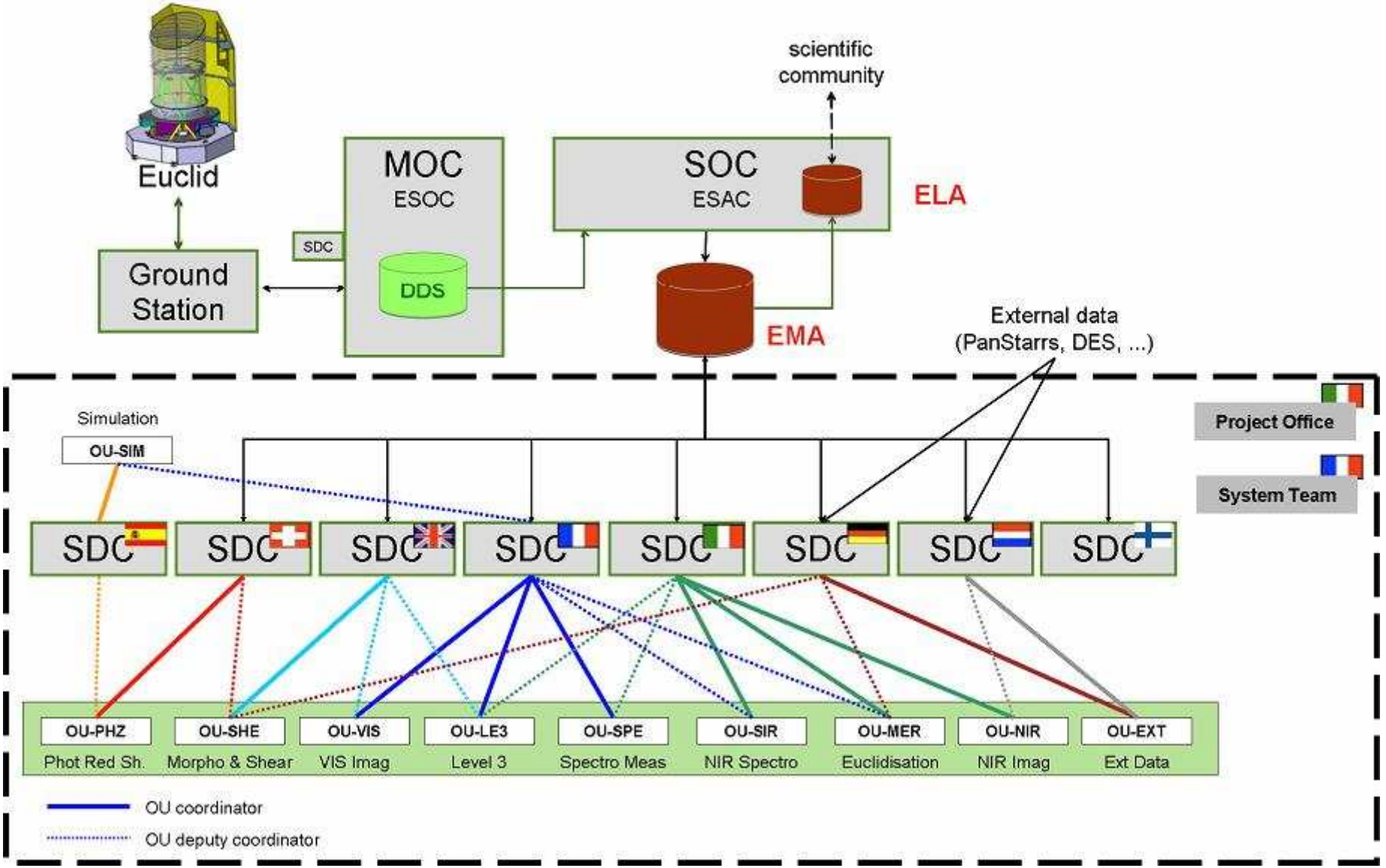


- **OUs** (Organization Unit) :
  - Responsible for the definition and prototyping of a given pipeline
  - Responsible for the validation of a given pipeline

- **SDCs** (Science Data Centers) :
  - Responsible for the S/W development of pipelines
  - Responsible for the H/W processing infrastructure
  - Responsible for the pipeline processing operations

Euclid – OU / SDCs (2/2)

Euclid Consortium



Euclid - SGS Challenges

Key challenges

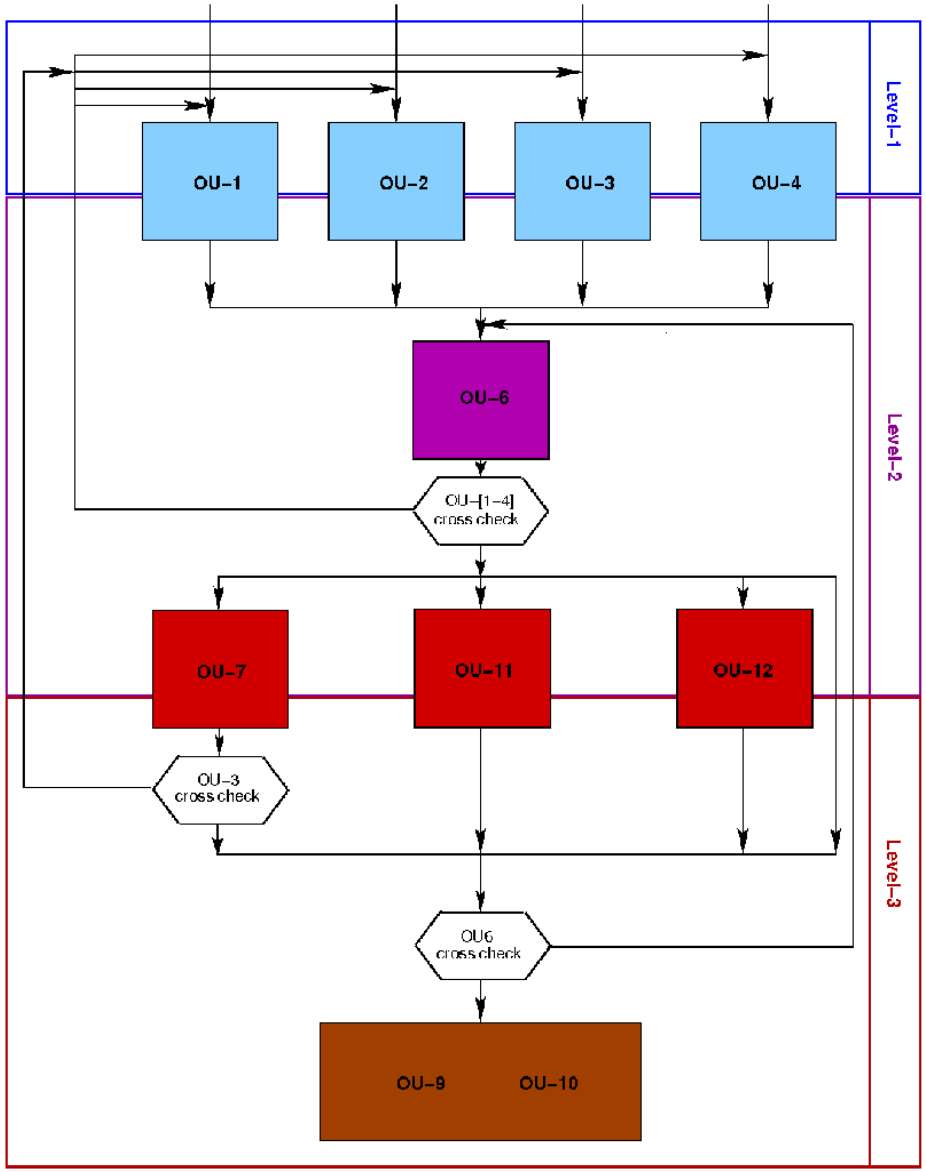
- Heavy **simulations** before the mission,
- heavy (re)**processing** needed from raw data to science products (e.g. on Planck volume multiplied by dozens),
- amount of **data** that the mission will generate per yearly release :
  - 26 PBytes of data => storage !
  - 1.10^{10} objects => database !
- **accuracy and quality** control required at every step.

=> Data Centric

Euclid - Processing Challenge

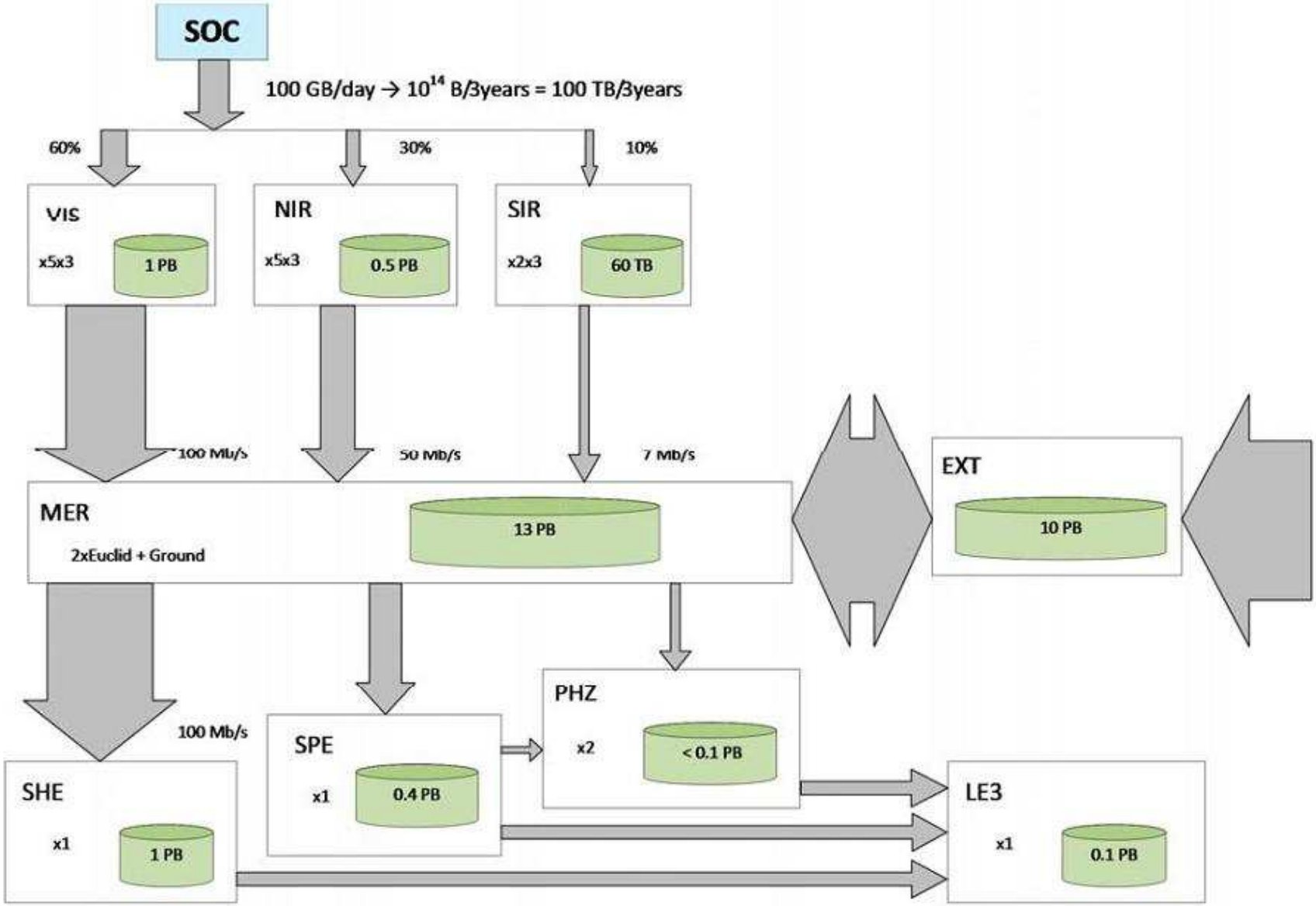
Euclid Consortium

OU-1	OU-VIS	Visible imaging
OU-2	OU-NIR	NIR imaging
OU-3	OU-SIR	NIR spectro
OU-4	OU-EXT	External data
OU-5	OU-SIM	Simulation
OU-6	OU-MER	Merging
OU-7	OU-SPE	Spectral measur.
OU-11	OU-SHE	Shear
OU-12	OU-PHZ	Photo-z
OU9+10	OU-LE3	Level3 products



Euclid - Data Challenge

Euclid Consortium



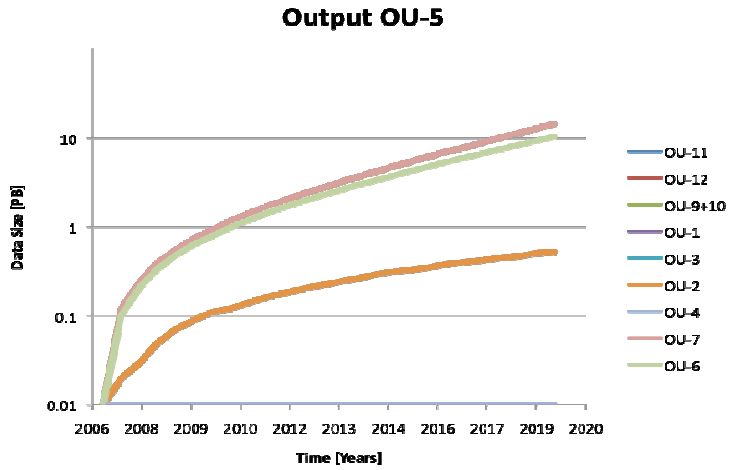
Euclid - Technology bottleneck ? (1/2)

To understand how those **data size** evolutions constraints the design we have to compare them with **hardware** evolution.

For a first estimation, we use the next laws:

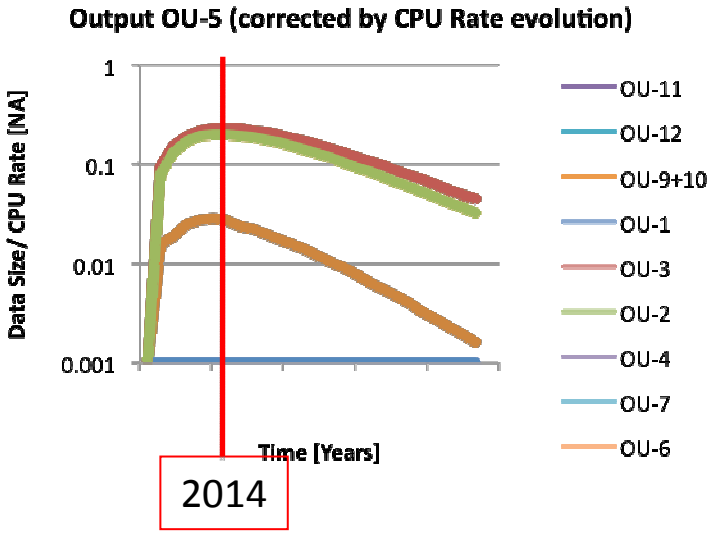
- Moore's Law : CPU double every 18 month.
- Kryder's Law : Disk size double every 12 month.
- Disk rate increase by 10% every 12 month.
- Nielsen’s Law Network bandwidth double every 21 month.

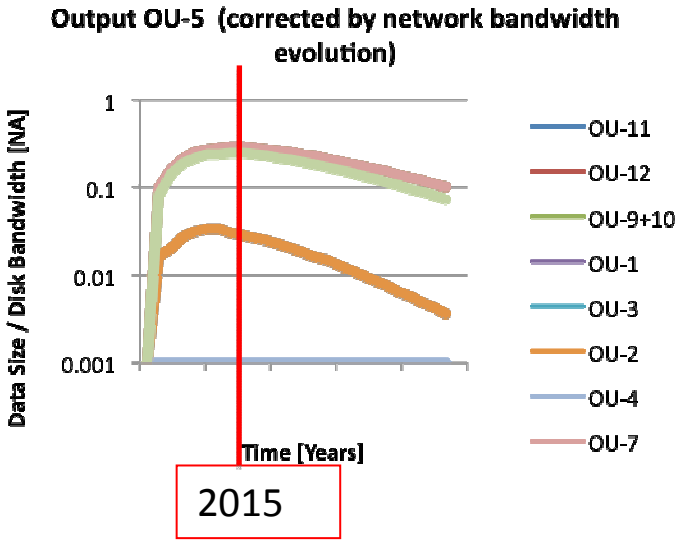
toward a given pipeline dataflow estimation :

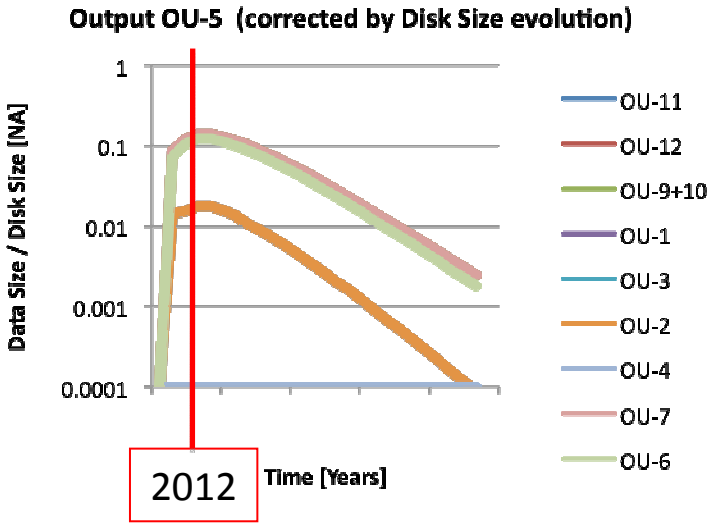


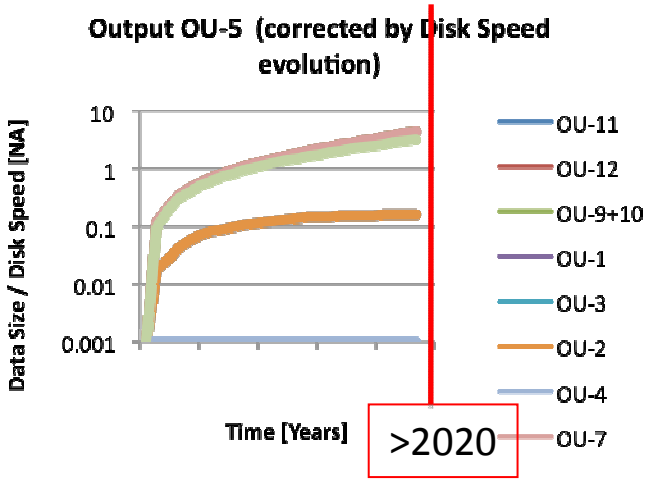
Euclid - Technology bottleneck ? (2/2)

Euclid Consortium









Euclid – Architecture Key Drivers

Euclid Consortium

- **Disk Size**: Data increase slowly than storage increasing. Could be a driven requirement for really first cycles.
- **CPU Rate** : CPU increasing much faster than data to be treated. Could be a problem at the beginning, but naturally disappear.
- **Network Bandwidth**: Before the launch, this is a real problem. May be in 2020, it will be easier to exchange data, but could be very costly.
- **Disk Rate**: It is also a critical driver for the design.

=> Data flows optimization is a key driver

- **decrease** as much as possible **data access**;
- **decrease** as much as possible **data exchange**;
- keep a very **thin infrastructure** between software modules and data access to avoid overheads while
- **accessing data locally** inside each SDC infrastructure;
- allow the inclusion of new national SDCs, as needed;
- **simplify** as much as possible the system design.

Euclid - SGS Design Principles

- **Low coupling** between components
- As **asynchronous** as possible
- **Service** Oriented / **Event** Driven
- **Aggregation** of services and **parameterization** rather than "hard coding"
- **Agreed** (and respected) Conventions & **Interfaces** rather than complex APIs

- Don't reinvent the wheel or over constrain
- Occam's razor: the law of "parsimony, economy or succinctness"

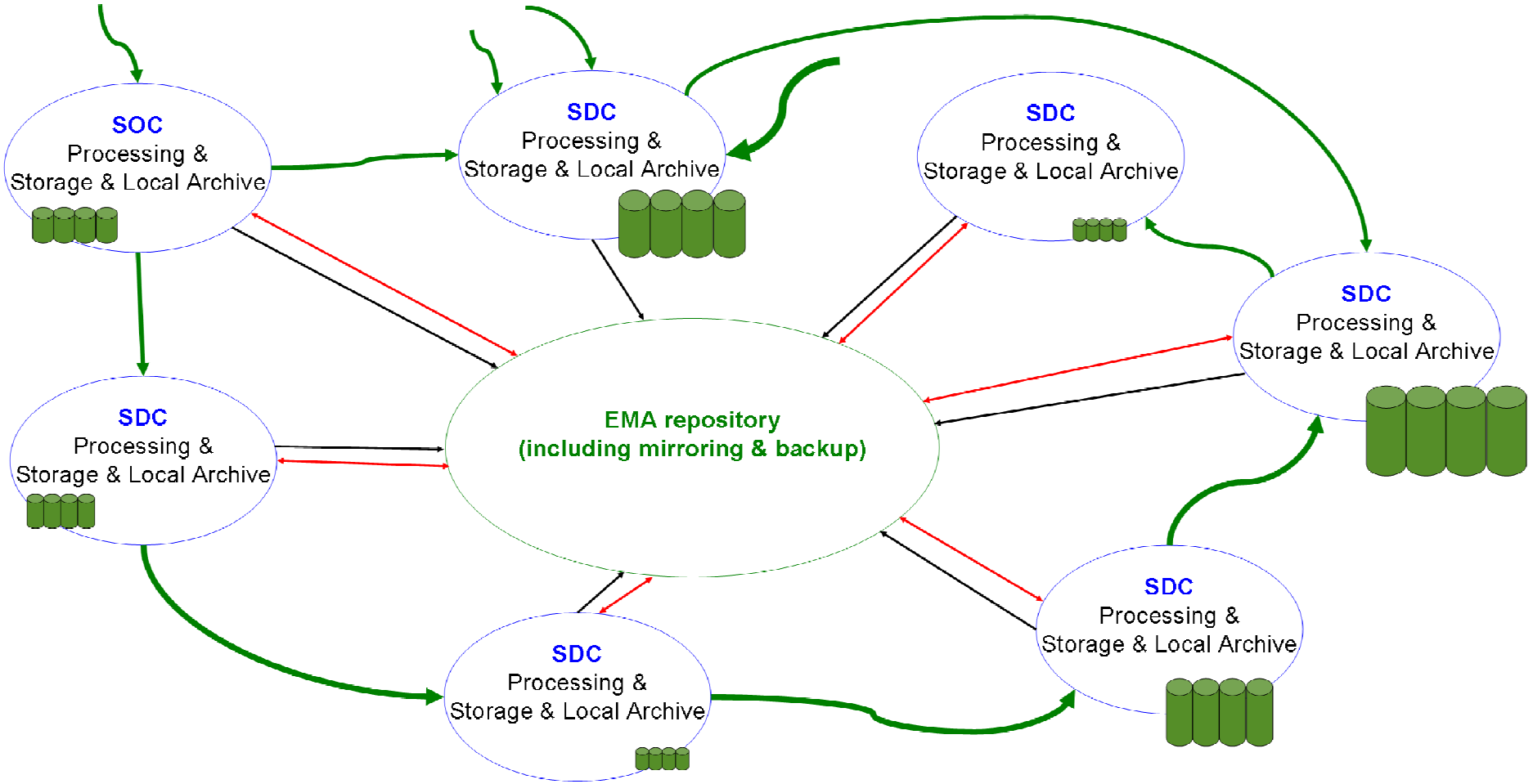
Euclid - SGS Logical Architecture (1/2)

Euclid Consortium

- A **single metadata repository** which inventories, indexes and localizes the huge amount of distributed data,
- A **distributed storage** of the data over the SDCs (ensuring the best compromise between data availability and data transfers),
- A **set of services** (SOA) which allows a low coupling between SGS components : e.g. metadata query and access, data localization and transfer, data processing M&C, …
- An **Infrastructure Abstraction Layer** (IAL) allowing the data processing software to run on any SDC independently of the underlying IT infrastructure, and simplifying the development of the processing software itself,
- A common **Decentralized Processing Control**, data and event driven, deployed on each SDC.

Euclid - SGS Logical Architecture (2/2)

Euclid Consortium

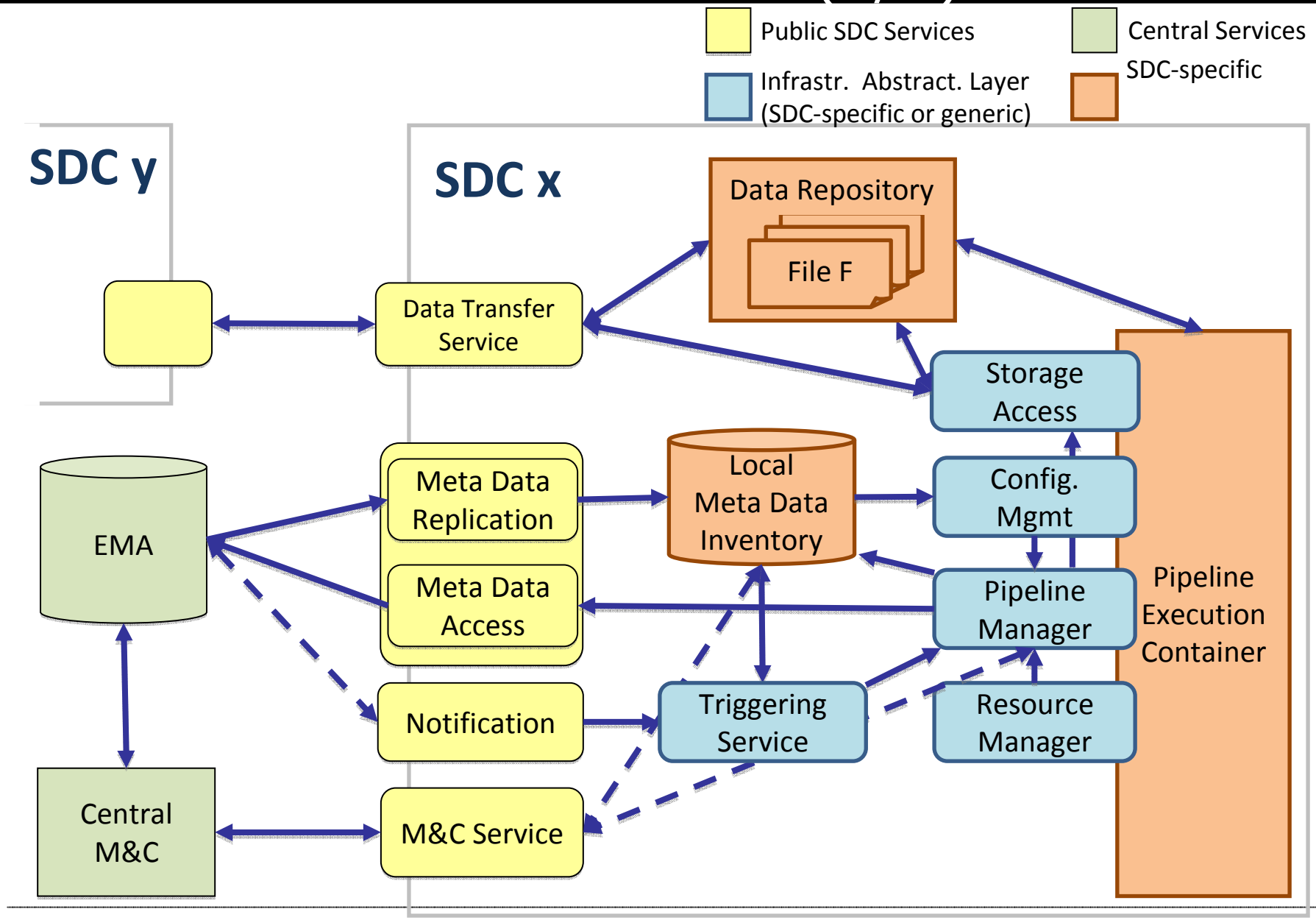


Euclid – SGS Main Services (1/2)

- The Euclid logical Architecture is based on functions/services (also seen as Work Packages for the system team) :
  - Monitoring and Control
  - Data Transfers
  - Infrastructure Abstraction Layer (IAL)
  - Metadata repository and data localization (EMA)
  - Data and Processing distribution/orchestration
  - Common tools

Euclid – SGS Main Services (2/2)

Euclid Consortium



Euclid – EMA principles

- Euclid Mission Archive (EMA) :
  - Data **Inventory**
  - Data **Localization**
  - **Metadata** Repository (generic header + specific metadata, incl. data quality and lineage)
  - **Query** Interface (DBMS agnostic)
  - **Publication** to ELA (data access for the science community)
  - Version control
  - Data access **rights** management

Euclid – Data Storage principles

Euclid Consortium

- Data Storage :
  - Data are **stored** at SDC level
  - Data are **distributed** among SDC
  - Data are **replicated** : at least a primary storage and a secondary storage
  - Data **distribution policy** should minimize the data transfers
    - By data processing level
    - By sky area
    - …

Euclid – Processing principles

Euclid Consortium

- Processing are **distributed** on the SDCs
- => The SDCs are by nature heterogeneous (cluster, grid, cloud, …) : implies kind of mix federation/integration : "**fedegration"**
- Processing should occur where the data is in order to minimize the data transfers
- Any processing pipeline should be able to run on any SDC (at least on two : primary and secondary)
- => needs of minimal set of commonality
- => needs for **infrastructure abstraction** : IAL (Infra. Abstraction Layer), virtualization

Euclid – IAL Principles

- **Infrastructure Abstraction Layer** (IAL) isolates the pipeline from the underlying infrastructure :
  - **Pre processing** step :
    - Queries and retrieves the input data
    - Takes care of the resources needed by the pipeline
    - Creates the working context for the pipeline
  - **Running step**
    - The pipeline runs in a "sandbox" and knows only about it (no external access)
    - IAL manages pipeline control and data flow
    - A Set of minimal and basic pipeline interfaces : inputs/outputs, M&C, parameterization
  - **Post processing** step :
    - inventory and storage of outputs (metadata + data),
    - notification

Euclid – Orchestration Principles

- **Inversion of control** : strong centralized M&C (“big brother”) would be too complex, thus SDCs should have some autonomy and gives in return visibility on their activity and status.
- Two kinds of decentralized orchestration are foreseen:
  - **Polling Mode** (Pull Mode): check periodically if new data is available or if there is something to do.
  - **EDA** (Event Driven Architecture) Mode (Push mode): subscribe for notification on given events and be notified when any previously subscribed event occurs (new data available, …), then triggers the corresponding action(s).

Euclid – Common Tools

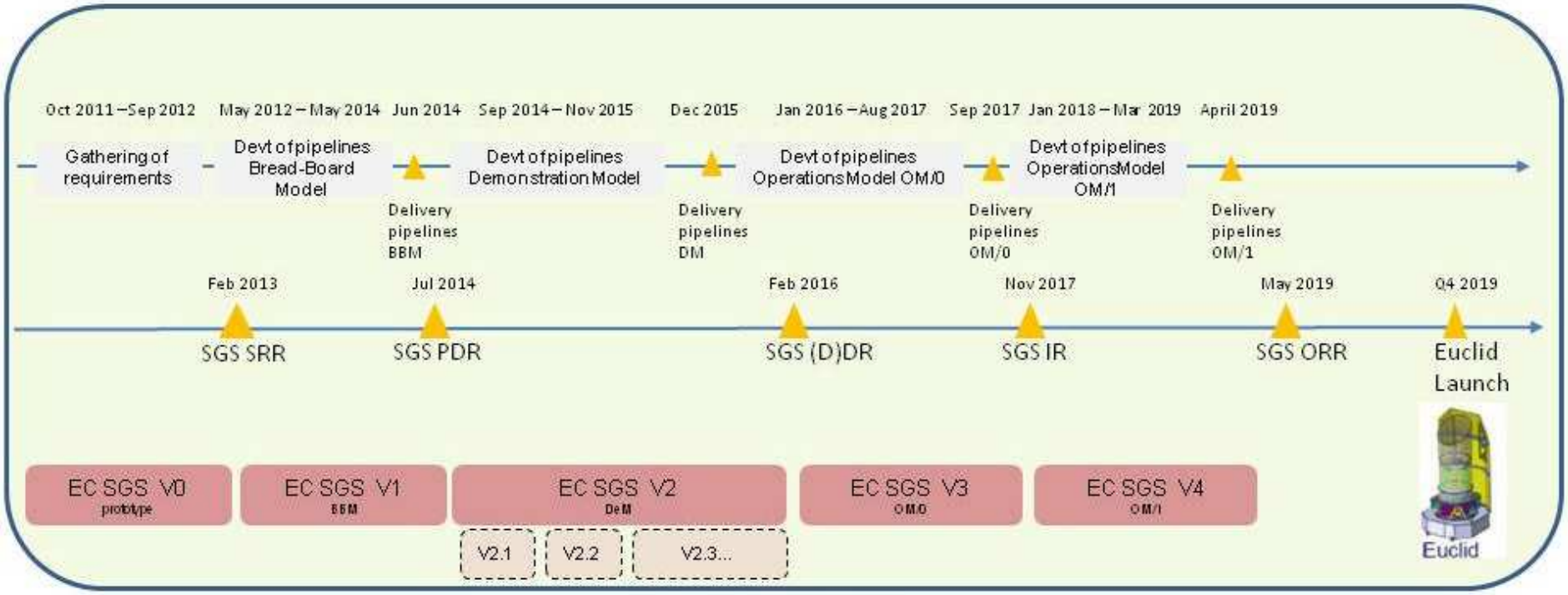
- **Common** scientific libraries.
- **Development** approach :
  - Agile methodology
  - Collaborative development platform
  - Continuous Integration
  - Test Based Development
- Linux O/S
- C/C++,Python, Java languages (TBC)

Euclid – EC SGS next steps

- Euclid SGS Architecture **Mockup** foreseen in next months :
  - Kind of Proof of Concept (POC) : architecture & tools
  - Flagship OUs (SIM, VIS)
  - Flagship SDCs
- Euclid SGS KOM meeting : 5th-9th Mars 2012 at Bologna, IT
- Hard work at System Team level in close cooperation with OUs and SDCs in order to translate the **logical** architecture into a **physical** one.

Euclid - SGS Schedule

Euclid Consortium



Euclid - Conclusions

Euclid Consortium

- The SGS part of the Euclid project is at **early definition phase**…
- But there are already **active working groups** which are motivated to :
  - Share their experience acquired during previous projects (Gaia, Planck, Herschel, …),
  - Try innovative ideas,
  - Build the SGS from prototype to the target system by iterations and increments.

- There is a lot to do but there is still about 5000 days to achieve it…

Acknowledgments

- Yannick Mellier (IAP), Fabio Pasian (INAF), Oriana Mansutti (INAF), Claudio Vuerli (INAF), Marco Frailis (INAF), Andrea Zacchei (INAF), Marc Sauvage (CEA), John Hoar (ESAC), Christophe Dabin (CNES), Keith Noddle (UoE), Jean-Marc Delouis (IAP), Laurent Vibert (IAS), Rees Williams (RuG), Christian Neissner (PIC), Johannes Koppenhoefer (MPG), Joseph Mohr (LMU), Christian Surace (LAM), Pierre Dubath (Unige), Stéphane Paltani (unige), Martin Melchior (ETH), Stefan Muller (ETH), Elina Keihanen (FIN), Massimo Brescia (INAF), Luigi Paioro (INAF), …

- Which are part of the about 900 people already involved in this fantastic project !

Thank you

for your attention

Any questions ?

Please speak slowly

I'm French ☺