PLATO 2.0

P.M. Marrese on behalf of PLATO Italia

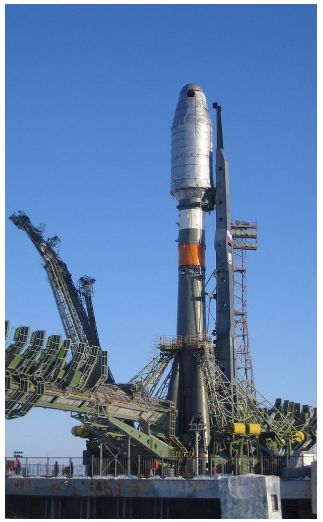
ICT Workshop 2017 Bologna 28 Novembre 2017

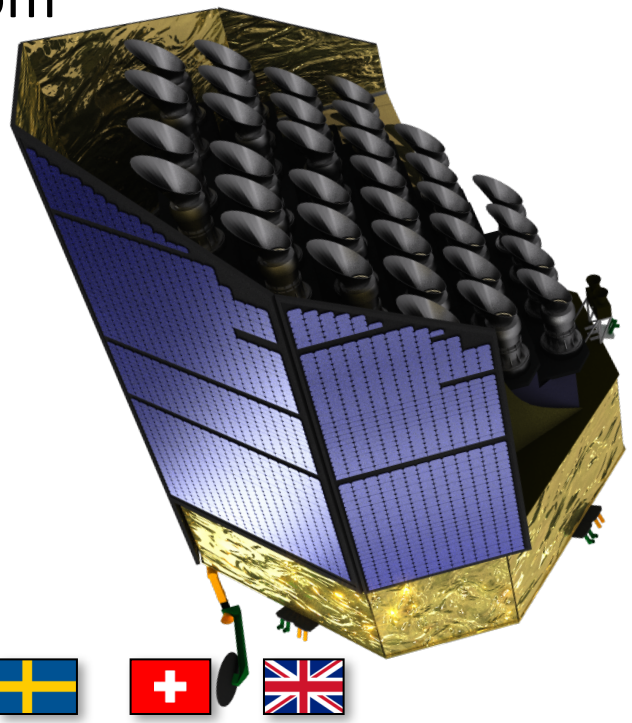
PLATO PLAnetary Transits & Oscillations of Stars

AIM : detect and characterize (density, age) terrestrial planets around solar-like stars up to the habitable zone

FoV
~48°.5 x 48°.5

- M class mission (M3)
- Budget envelope ~ 650 M€ (≤ 500 M€ from ESA)
- Launch: 2026 – Launcher Soyuz Fregat from Kourou
- Operation: 4.25 (+2) yrs
  built for 8 yrs





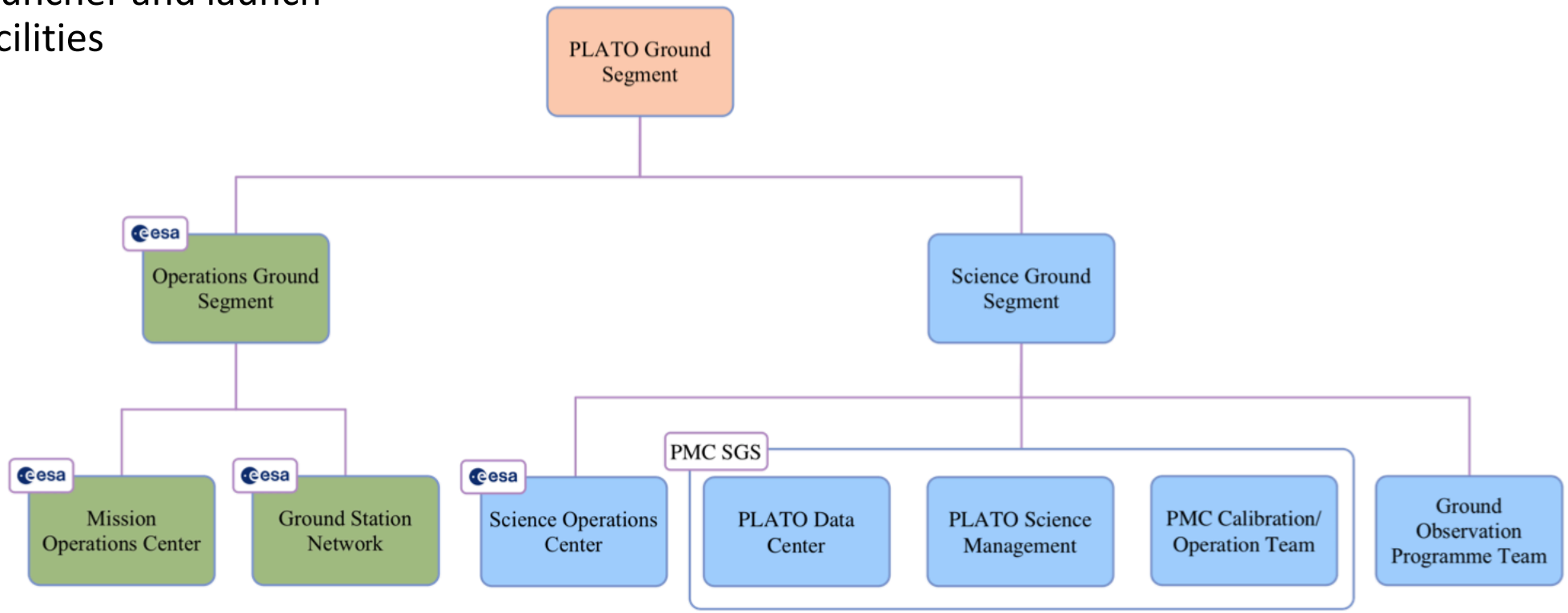
PLATO

ESA

- Satellite
- Mission Operation Center (MOC)
- Science Operation Center (SOC)
- Launcher and launch facilities

PLATO Mission Consortium:

- Payload
- Data Center (PDC)
- Plato Science Management (PSM)



PLATO Italian Contributions

- Payload
  - Telescope Optical Units
  - Instrument Control Unit
- Science Preparation
  - Input catalogue definition
  - Main Contributions to:
    - the specification and validation of methods for light-curve processing and derivation of final data products;
    - the system architecture of the PLATO Data Centre;
    - The preparatory and follow-up database management.
- PDC
  - Input catalogue implementation

✓ **INAF**
- **OA Catania** (Science, Payload)
- **OA Padova** (Science, Payload)
- **OA Brera** (Science, Payload)
- **IAPS-Roma** (Science, Payload)
- **FGG** (Payload)
- **OA Palermo** (Science)
- **OA Torino** (Science)
- **OA Capodimonte** (Science)
- **OA Roma (+Teramo)** (Science)
- **OA Arcetri** (Science)

✓ **Padua University, Physics & Astronomy Dep.** (Science)

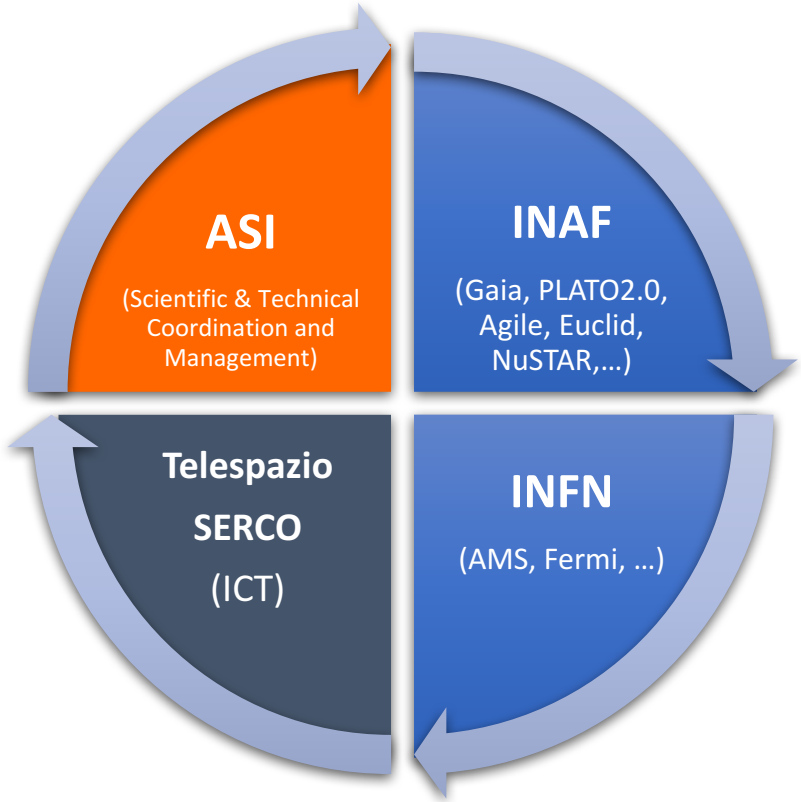
✓ **ASI-SSDC** (PDC, Science)

SSDC
SPACE SCIENCE DATA CENTER

MAIN GOAL

acquire, manage, process and distribute data from (mainly) space based mission adopting the FAIR (Findable, Accessible, Interoperable, Reusable) principles.

SSDC adopts international standards ensuring both the long term preservation of archives and the interoperability with other data centers.



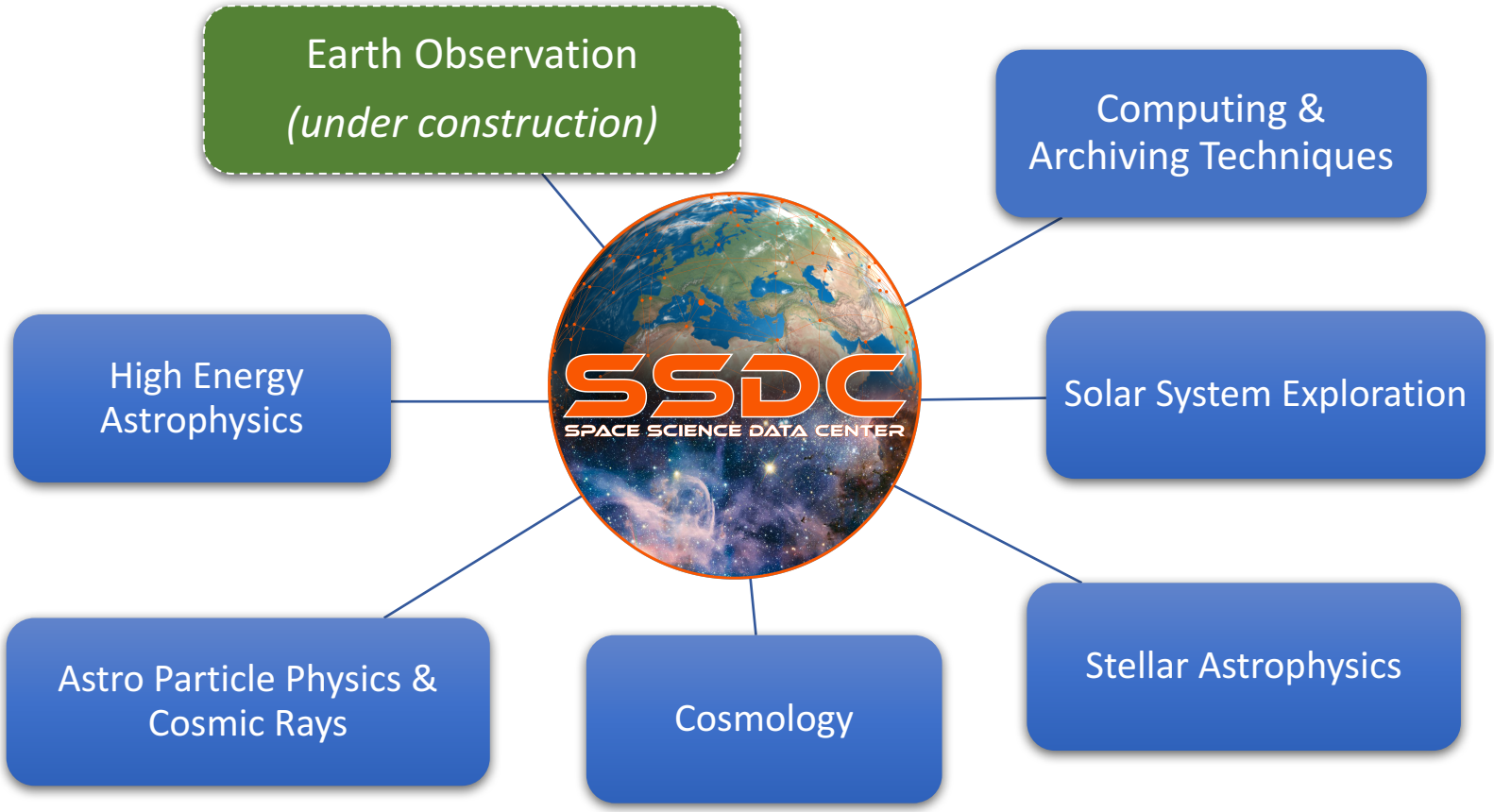
SSDC – UNIVERSE OBSERVATIONS management and organization involves several Research Institutes:

- **ASI** – Italian Space Agency
- **INAF** – National Institute for Astrophysics
- **INFN** – National Institute for Nuclear Physics

Industries are involved for Information and Communication Technology supports.

SSDC Scientific Expertise

At present, SSDC team involves around 40 people: scientists from ASI, INAF, INFN and SW engineers from Telespazio & SERCO, expert in different fields.



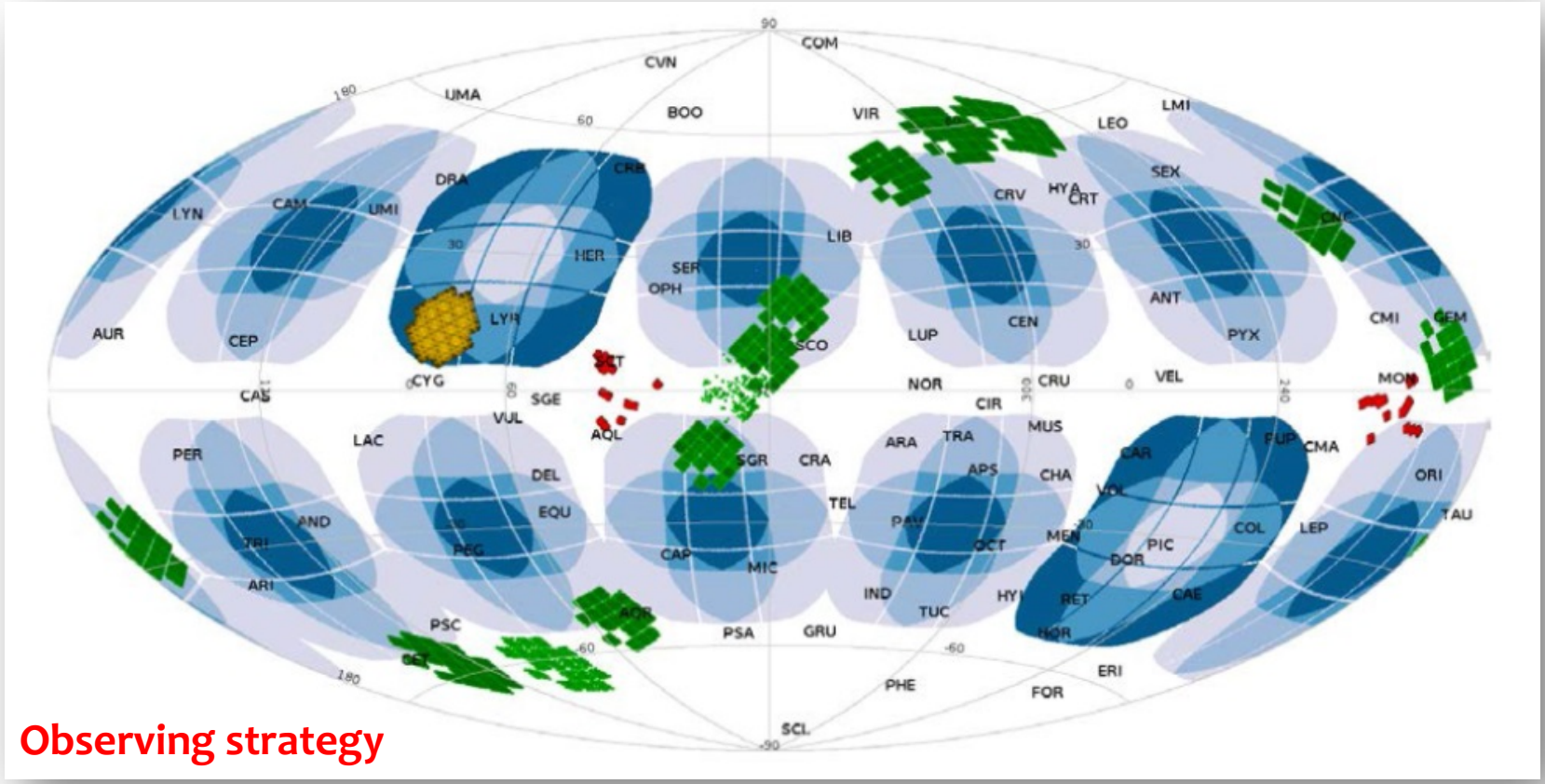
Effective approach: Developers and Users belong to same communities.

PIC - PLATO INPUT CATALOGUE

PLATO requires the pre-selection of the target stars best suited to the detection of planets, especially terrestrial planets. The success of the mission also relies on our ability to select fields that maximize the number of dwarf and subgiant stars with spectral type F5 or later for which we can acquire photometry with the required signal-to-noise ratio

P1 and P5 allsky sources V< 13.0 :

5.5 million + contaminants (PSF ~30 arcsec)



Observing strategy

- **Baseline**
  2 long pointings of 2 years
- **Alternative**
  3 years + 1 year step-and-stare phase

PIC – IMPLEMENTATION STRUCTURE

PIC IS DELIVERED BY SSDC (PDPC-A) to PDC-DB

PLATO ID + characterization

Additional source IDs

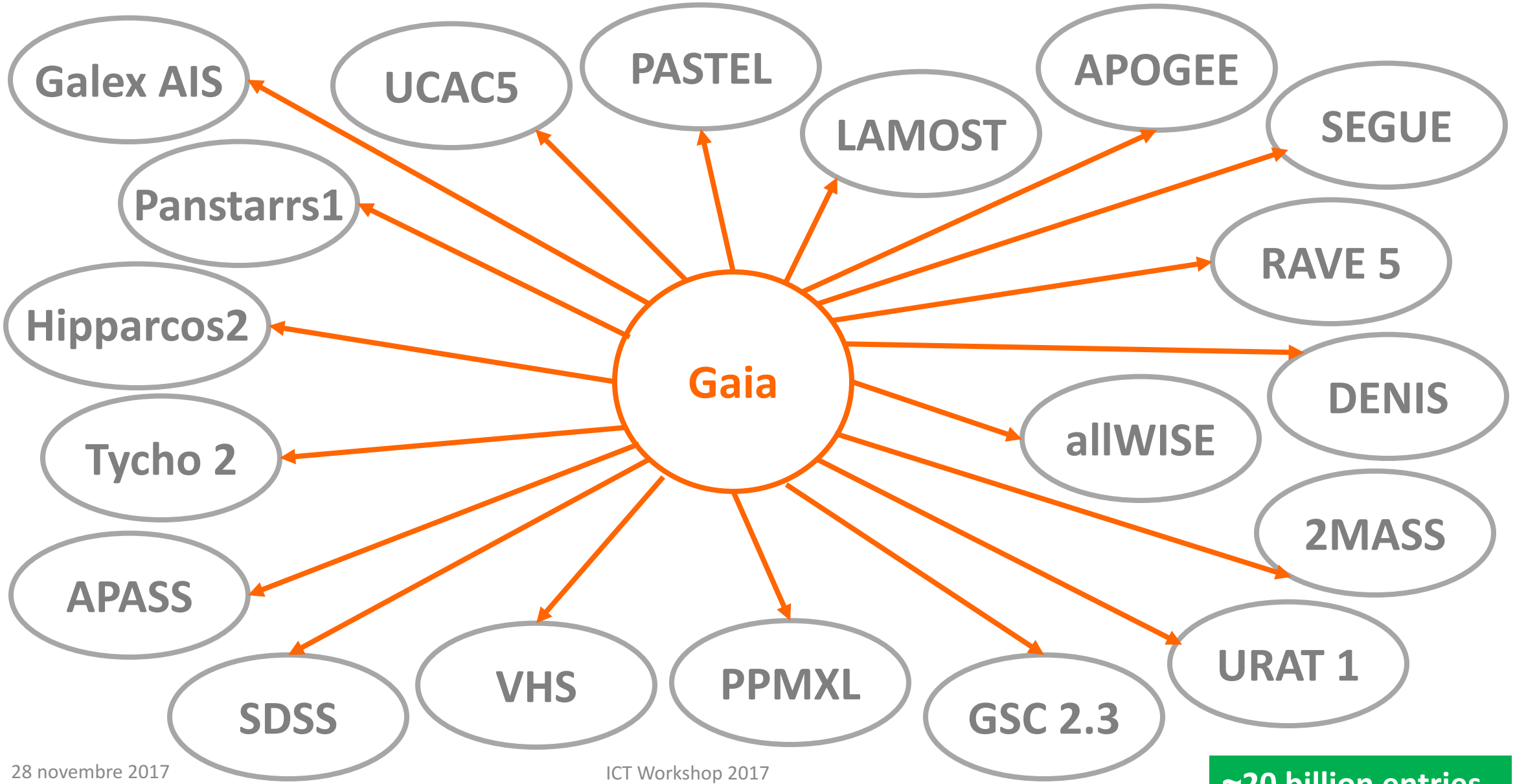
References for each target and each column

PLATOCont ID + PLATO ID + characterization

Additional source IDs

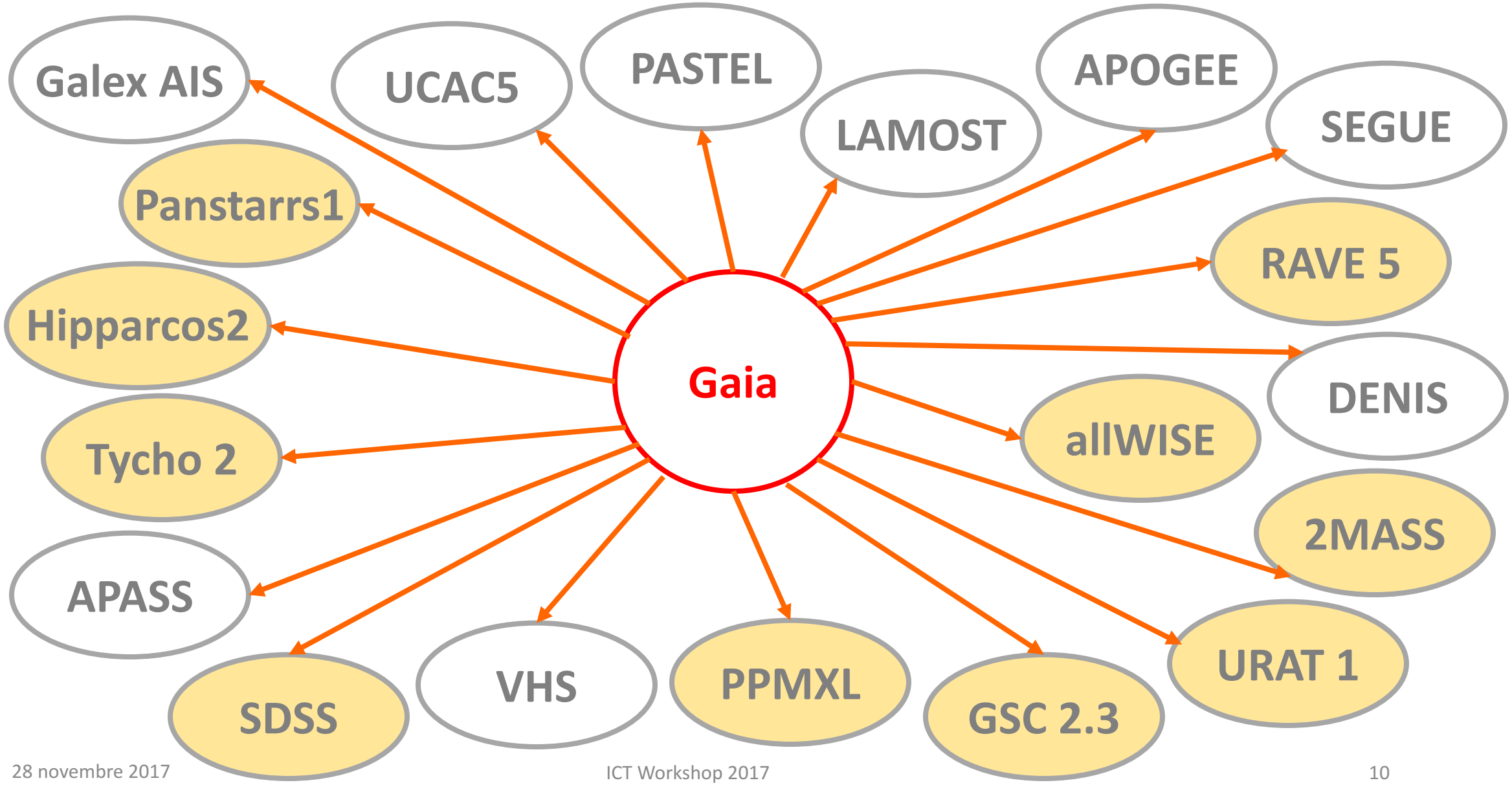
References for each contaminants and each column

TARGET CHARACTERISATION - Extended DB



~20 billion entries

Gaia – PLATO synergy



XM software for PLATO and Gaia

SW configuration control

XM BASICs

- Pre-computed
- Uses Gaia position + proper motions + parallax + errors + correlations
- Uses ext catatalogues positions + errors
- Defines mates in Gaia and neighbours in the ext Catalogue
- Defines a figure of merit (distance, convoluted errors on positions, density in second catalogue)
- No magnitudes
- Different algorithm for surveys and sparse catalogues
- Second catalogue density pre-computed and fed to the XM
- Dedicated XM for PLATO

XM implementation

- Several instances run in parallel (2-3xCPU)
- Read data once
  - Keep data in RAM
  - Perform calculations
  - Delete data not useful from RAM
  - Cycle
- 3 DB connections (up to 3 servers)

I/O limited : writing in DR1, reading in DR2

XM size

Gaia DR1 1.1 billion

GSC 2.3 0.95 billlion

Gaia DR2 1.5 billion

Panstarrs 2.2 billion

DR1

Catalogue	Time (minutes)
UCAC4	39
GSC2.3	239
PPMXL	172
SDSS DR9	56
URAT-1	26
2MASS PSC	69
allWISE	450

Architecture & DB design for data access + XM

Requirements

Data size : scalability

Data complexity :
combination of different surveys
catalogue entries + epoch data

Interrogation flexibility:
no typical query, but a variety of scientific cases

Static archive : not frequent updates (new catalogues or releases)

Access: read intensive, complex queries

XM : read intensive, but also write intensive (concurrent writing)

XM Adopted solution

- Data size reduction: BaseCat
- SQL : output needs then to be accessed
- Read : Toku-DB
  Write : XtraDB/InnoDB (no ACID compliant, no transactions, yes concurrent writing)

ACCESS Adopted solution

- Distributed scalable system
- Shared-nothing architecture : each node is independent and self-contained (nodes do not share RAM and storage)
  -> horizontal data distribution and parallel query execution
  (DB sharding, no joins across nodes)
- Relational DBMS : effective joins
- DBMS : MariaDB with engines MyISAM, XtraDB/InnoDB, TokuDB, FederatedX –> flexibility, highly configurable and customizable
- Data size reduction : normalisation
- Data redundancy : very limited (string type flags)
- Small size tables : indexes are also smaller

Sharding disadvantages :
development and management complexity, mitigated in our case by not frequent updates:

- shard preparation is easy by command replication
- replication and backup made easier by DBMS engines choice
- Automated sharding (like hash) are not easier and do not grant absence of across-modes joins

SSDC HW for Gaia-PLATO

Sistema 8 BLADE (HP Gen8)

- RAM 2 TB (256 GB per blade)
- Spazio Disco 15 TB
  - 300GB x 6 = 1.8 TB, 15K rpm SAS
  - 1.2TB x 8 = 9.6 TB, 10K rpm SAS
  - 1.8TB x 2 = 3.6 TB, 10K rpm SAS
- CPU: 8 core 16HT (2 per blade)
- Enclosure e switch

GaiaServer (HP)

- RAM 36 GB
- Spazio Disco 24 TB
  - 2TB x 12 7.2K rpm SATA
- 2 X CPU: 6 core 12HT

+

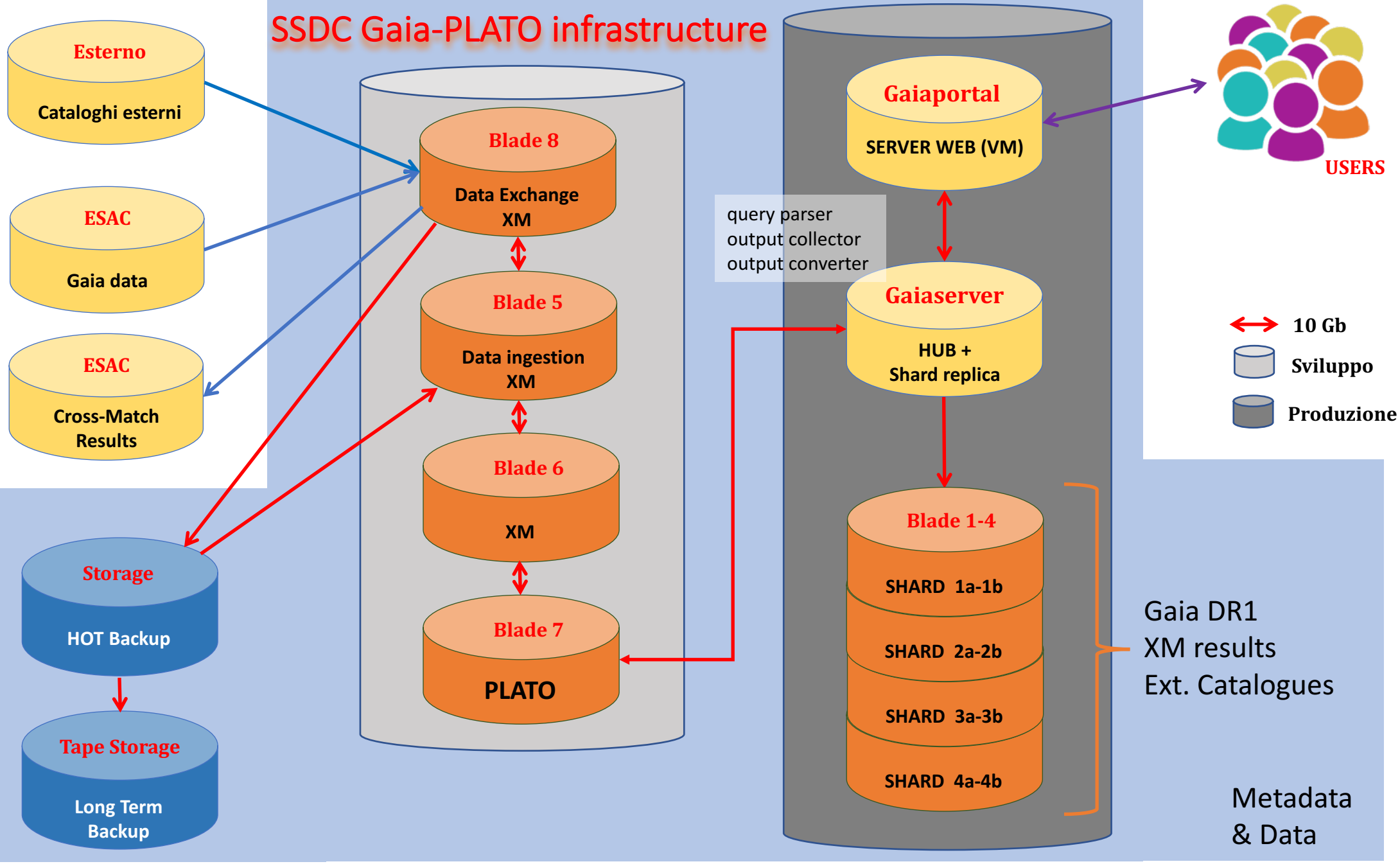
Sistema 8 BLADE (HP Gen9)

- RAM 2 TB (256 GB per blade)
- Spazio Disco 19.2TB
  - 1.2TB x 16, 10K rpm SAS

2018

- CPU: 8 core 16HT (2 per blade)
- Enclosure e switch

SSDC Gaia-PLATO infrastructure



THE FUTURE

Available HW + 2018 HW will serve
PLATO + Gaia DR2 (april 2018)
Probably also PLATO + Gaia DR3 (second half 2020)

Gaia DR4 2022 nominal end of mission,
however “The Science Programme Committee (SPC) has decided to extend the Gaia operations beyond the nominal 5 years until the end of 2020. Following the standard ESA science mission extension procedure this extension should be confirmed next year and a proposal for 2021-22 will be submitted for SPC approval.”

→ **SCALE HW + epoch data solution**