SDC-IT Status

M. Frailis

on behalf of the SDC-IT team

The Italian Science Data Centre (SDC-IT)

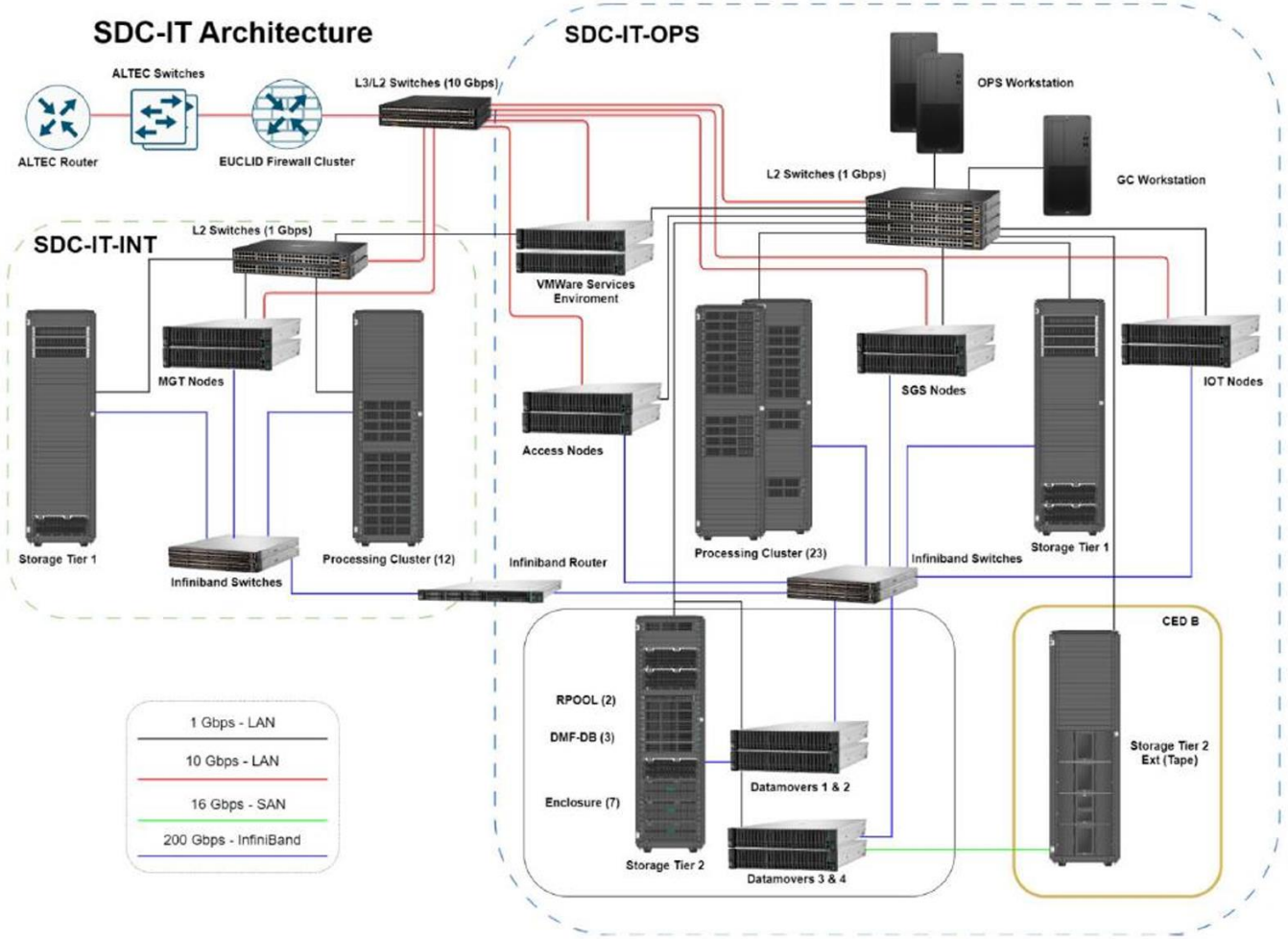
SDC-IT-DEV at INAF-OATs

- Pipelines **development**, **integration**, **testing** and release (LE1, NIR, SIR, MER, LE3, SPE)
- **Data Quality Checks pipelines (DQC)** development and integration
- **Planning** and **management** of the production infrastructure (with ASI and ALTEC)
- Support to the NISP Instrument Operations Team
- **Computing and storage infrastructure for development**:
  - PLEIADI computing system at INAF-OATs: 40 nodes, 2848 cores
  - 2 dedicated BeeGFS storage servers (650 TB)
  - InfiniBand + OmniPath network fabric

SDC-IT-PROD at ALTEC SpA

- Computing infrastructure dedicated to the Euclid SGS **operations**
- Providing 25% of the total SGS mission resources:
  - **Computing and Storage resources** for integration and production (see next slide)
- Support to the **development** activity: LE1 NIS Processor, AstrOmatic sw re-engineering, SIR and NIR code validation common software
- IOT **dedicated workstations**: ICR Tool and IODA software developed by the INAF team
- Operating the DQC pipelines for NIR and SIR
- Access nodes and tools available to the SDC-IT Dev team (and the OU) for verification activities on operational data

SDC-IT-PROD architecture (ALTEC)



SGS Cluster: data exchange and processing management.

Processing Cluster: 23 nodes OPS (**3648 cores**); 12 nodes INT (896 core).

Tier 1: high performance; Lustre Appliance **1.5 PB**.

InfiniBand: Storage, Processing Cluster and SGS Cluster connectors (**200 Gbps**).

Access Nodes: data access and analysis.

IOT Nodes: instrument data analysis and ICR (Instrument Command Requests)

SHARED

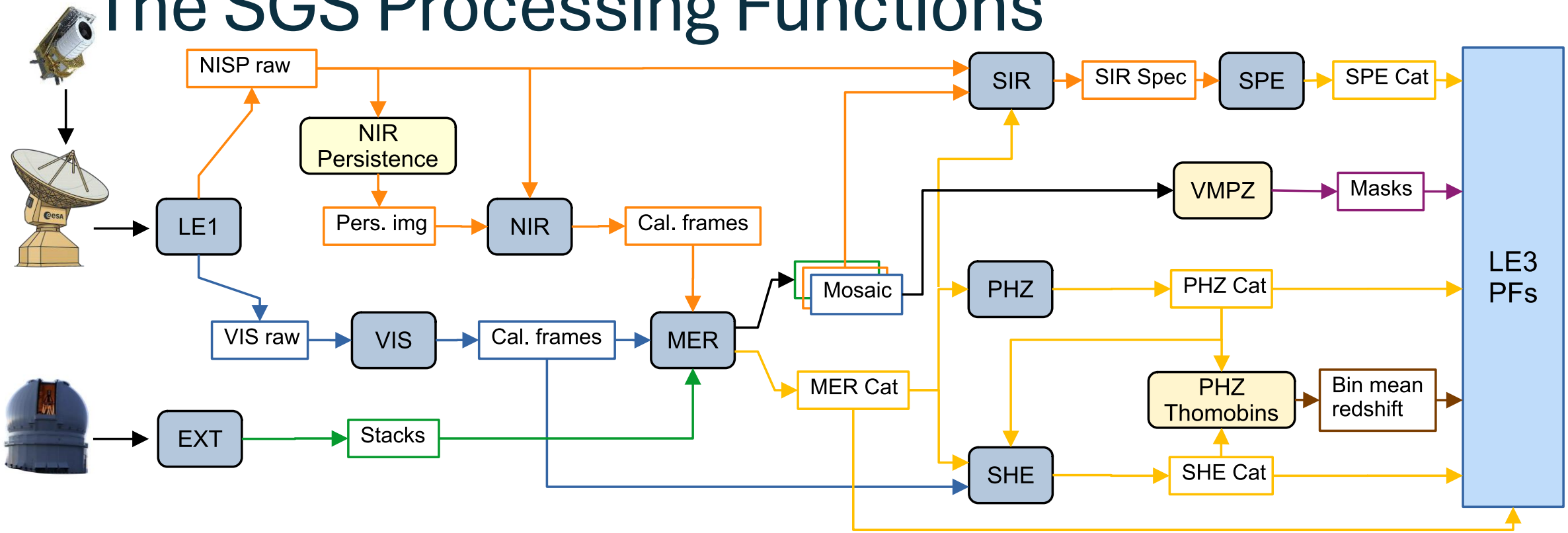
External link: **10 Gbps**.

Tier 2: high capacity; **4.4 PB** HPE ZeroWatt Storage & Data Management Framework.

Tape Library: backup; MSL6480 6 drive and 4 modules with 280 LTO 9 Tape.

Management Cluster: services for operability and center management: cluster Vmware.

The SGS Processing Functions



LE1: S/C telemetry editing
EXT: External data
NIR: Near-infrared
VIS: Visible
MER: Merging
SIR: Spectroscopy Infrared

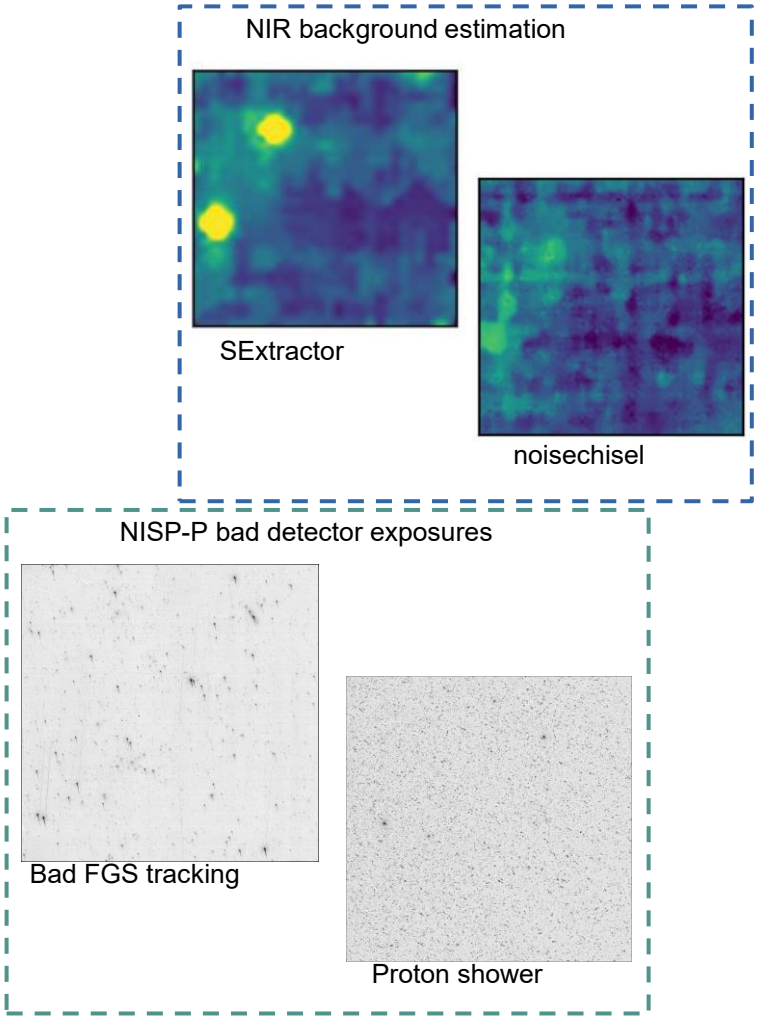
SPE: Spectral measurements
PHZ: Photometric redshift
SHE: Cosmic Shear
VMPZ: Visibility Mask PhotoZ
LE3: Galaxy Clustering,
Cluster of Galaxies,
External data, etc.

Agenzia Spaziale Italiana

INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

Pipeline integration activities

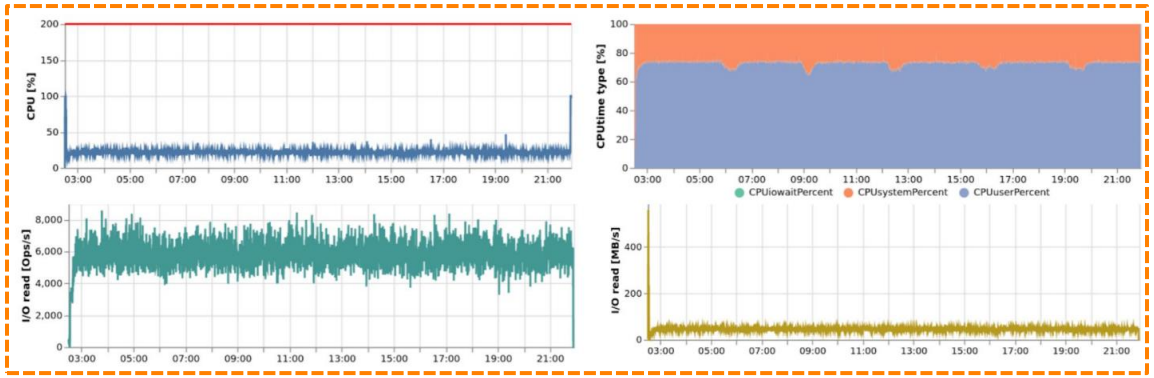
- **Interface** issues (Data Model)
- Tasks **resources** estimation and **optimization** (RAM, I/O)
- Software **re-engineering**
- **Tuning** of algorithms parameters
  - Several end-to-end pipeline runs (impact on downstream steps)
- **Comparison** of alternative analysis methods
  - e.g. background estimation: sextractor vs noisechisel
- Analysis issues identified by downstream pipelines
  - **Feedback loops** to identify the involved step and test the solution
- Automatic handling of **bad data**
  - e.g. Fine Guidance Sensor (FGS) failures, solar flares (cosmic-rays)
  - Products to be created and flagged as invalid by the pipeline



Task resources monitoring and optimization

- Monitoring RAM and I/O resources
  - at the level of the full Pipeline Processing Order
  - at pipeline task level
- I/O bottleneck
  - Storage at all SDCs is based on a network file system (Lustre, BeeGFS, etc.)
  - Some tasks can perform thousands of small I/O operations per second
- Mitigation
  - Check file I/O patterns in the code
    - process data objects in batches
  - Use of HDF5 file format for intermediate steps instead of FITS
  - Use of fast SSD temporary storage available in the SDCs
    - e.g. MER Mosaicing

Pipeline execution: resources dashboard

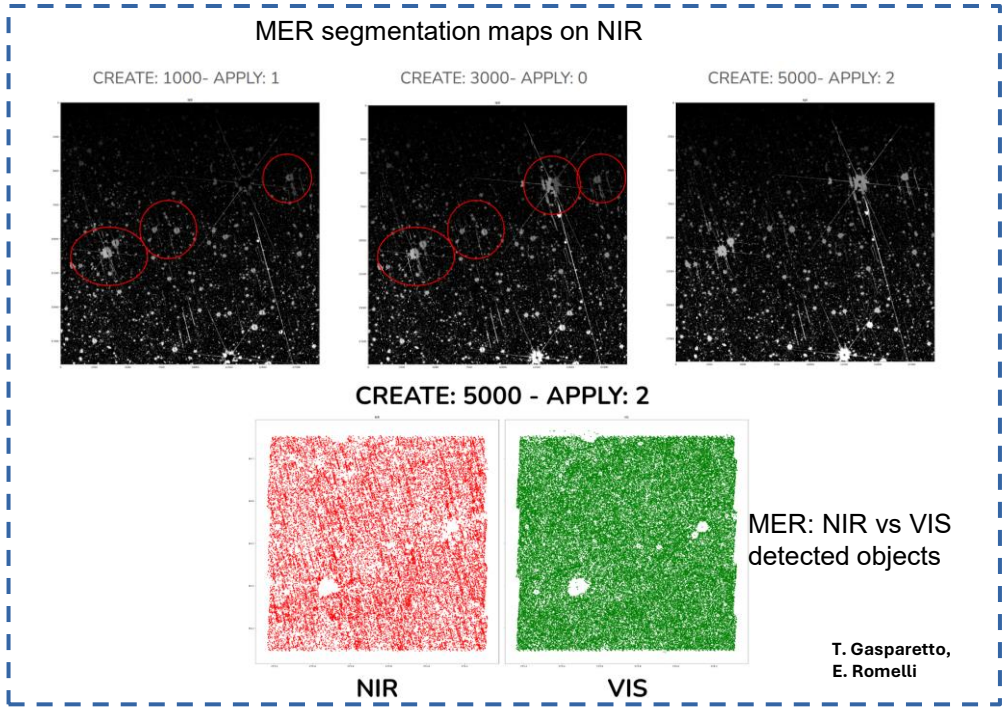


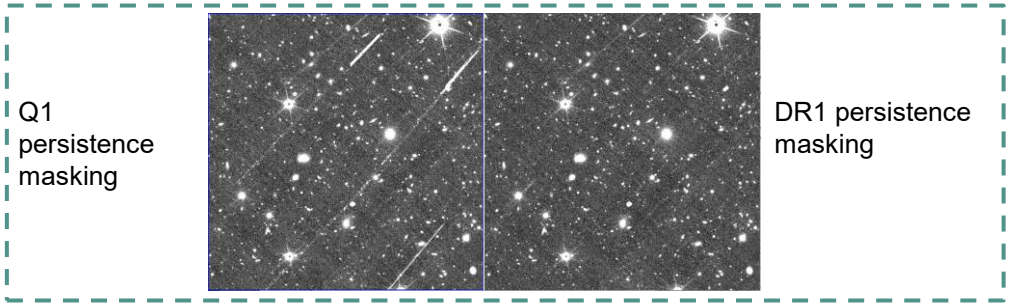
Bad task I/O example

Feedback loops between PFs: NIR and MER

NIR persistence masking

- Before Quick-release 1
  - based on ground persistence model
  - Several configuration tests
  - MER source deblending affected
- For Data Release 1
  - New masking based on power law persistence model computed in-flight
  - Persistence charge image based on 5 hours history
  - Improved spectra traces masking

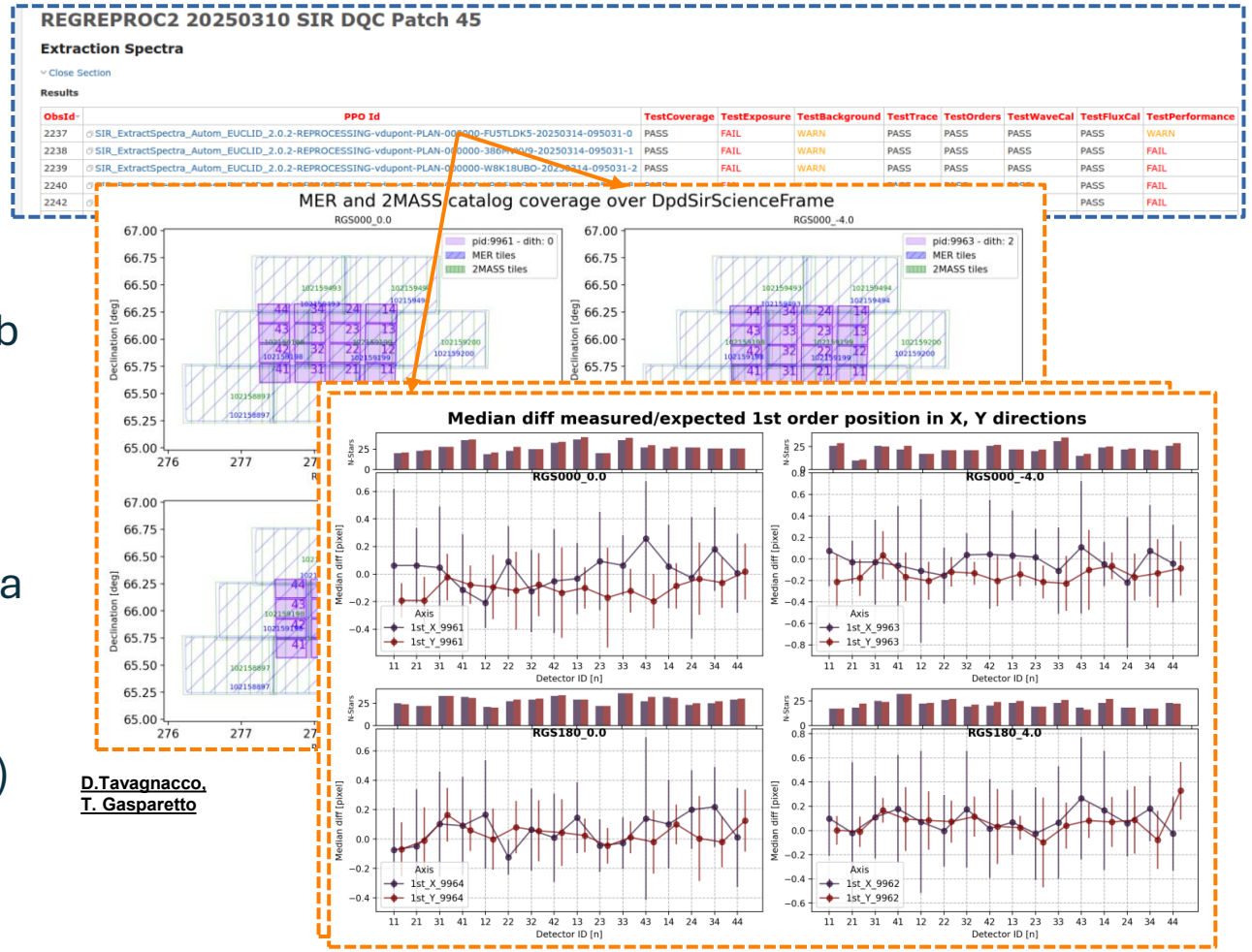




Data products validation

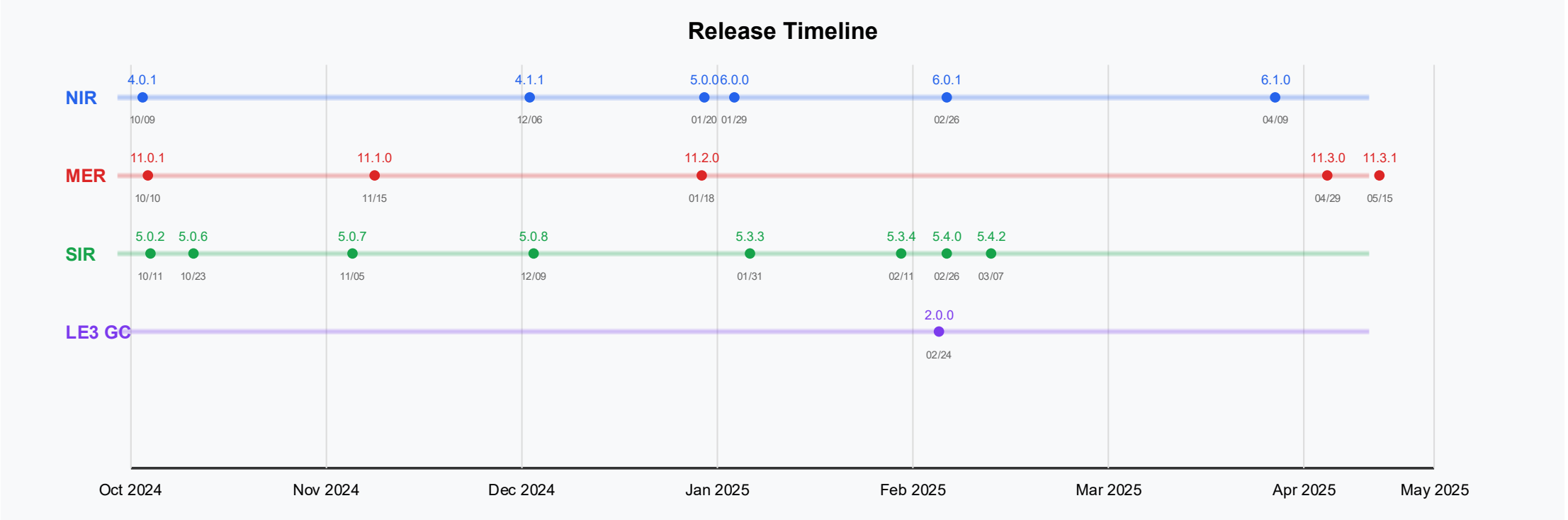
- Survey data too large to be checked manually

  Data Release 1:
  - 1900 deg2 (wide)
  - **~10.000** MER catalogs (3200 columns)
  - **~45.000** NIR calib. frames, ~22.000 VIS calib frames
  - **~500M** combined spectra
- In-pipeline computation of quality measures
  - stored as metadata information in each data product
- A separate **Data Quality Check pipeline** (DQC)
  - developed for each PF (by the **SDCs teams**)
  - producing quality reports
- All data products are validated by the OUs



D.Tavagnacco, T. Gasparetto

PF releases up to DR1 (after Q1)



DR1 production started for **NIR** and **MER**

DR1 on-going development

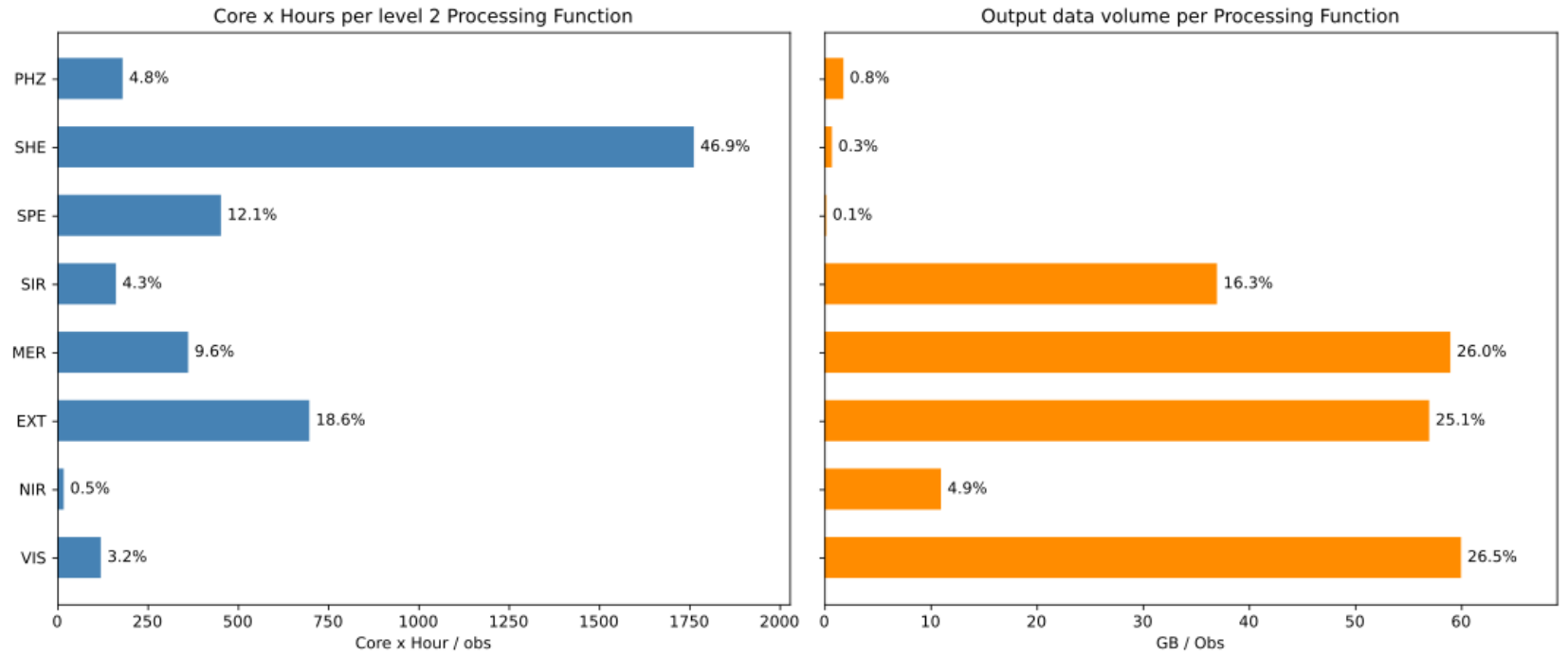
- SIR tests and integration
  - SIR updated features after RR2
  - Many production runs for each feature and configuration
- LE3 GC
  - Version for RR2 released
  - New version in preparation to adapt to SEL_ID DM changes
- LE3 CL
  - Runs performed on RR2 data
  - LE3_TILING_CL, LE3_DET_CL and Merging
  - On-going analysis on:
    - galaxy selection and masks
    - P(z) handling strategies

Name	SIR	MER Catalog	ROS	Padding	Clip	H Depth
REGREPROC2_R1	5.4	RR2 wide	RR2	NA	4	22.5
Zeroth-orders						
TEST_ZERO_ORDERS	5.5	RR1 + deep	1-17	3	4	22.5
TEST_ZERO_ORDERS_R2	5.5	RR2 + deep	1	3	4	22.5
TEST_ZERO_ORDERS_R3	5.5	RR2 + deep	1	0	3	22.5
TEST_ZERO_ORDERS_R3_WIDE	5.5	RR2 + wide	1	0	4	22.5
TEST_ZERO_ORDERS_R3_clip4	5.5	RR2 + deep	1	0	4	22.5
Combination						
TEST_STRICTER_CLIP	5.5	RR2 + deep	1-4	3	3	22.5
TEST_STRICTER_CLIP_R3	5.5	RR2 + deep	1	0	3	22.5
MER catalog						
TEST_MAG_23_5	5.5	RR2 + deep	1	3?	4	23.5
TEST_MAG_23_5_WIDE	5.5	RR2 + wide	1	0	4	23.5
TEST_MAG_23_5_WIDE_NOSP	5.5	RR2 + wide	1	0	4	23.5
Persistence						
TEST_MAG_23_5_WIDE_NOSP_MASK30	5.5	RR2 + wide	1	0	4	23.5
Optimal extraction						
TEST_MAG_23_5_WIDE_NOSP_OPT	5.5	RR2 + wide	1	0	4	23.5

SIR PF test runs for DR1 (completed by D. Tavagnacco)

SGS Technical budget

- Estimate of resources needed by level 2 Processing Functions
- Based on recent SGS regression testing campaigns



The SDC-IT team

INAF team

- Marco Frailis (SDC Lead and DEV Lead for NIR)
- Deborah Busonero (SDC Validation team)
- Samuele Galeotta (SDC-DEV Lead for MER and LE3)
- **Thomas Gasparetto** (**former** SDC-DEV team: NIR; now OU-MER)
- **Marius Lepinzan** (SDC-DEV: MER, OU-MER)
- **Gianmarco Maggio** (SDC Infra Sys Admin)
- Davide Maino (SDC Scientific Coordinator, SDC-DEV: SIR)
- **Erik Romelli** (OU-MER Co-Lead, SDC-DEV team: MER , LE3, ICR Tool)
- **Federico Rizzo** (SDC-DEV team: LE3 and NIR)
- Giuliano Taffoni (SDC Infra Manager)
- Daniele Tavagnacco (SDC-DEV Lead for SIR, LE3, ICR Tool)
- **Roberta Giusteri** (IOT Deputy Coordinator)
- Emiliano Munari (OU-LE3, CL Implementation Lead)
- **Thomas Vassallo** (**former** SDC-DEV team: MER and LE3; now OU-EXT)
- Claudio Vuerli (SGS Project Office, SDC PA/QA Lead)

ALTEC team

- Filomena Solitro (Program Manager)
- Lorenzo Bramante (System Engineer, IV&V Test Engineer)
- Rosario Messineo (Senior Data Processing Engineer)
- Federico Farinetto (Software Engineer, IV&V Test Engineer)
- Lucio Vincenzo Costa (Software Engineer)
- Alberto Alessio (Software Engineer)
- Luigi Antonio Squillante (Operations)
- Alessandra Cavallari (Operations)
- Stefano Lanza (CM, PA/QA)
- Matteo Del Giudice (Infrastructure Engineer)
- Stefano GIannuzzi (Infrastructure Engineer)
- Marco Richichi (Infrastructure Engineer)

Fixed-term contract (post-doc, TD, PhD)