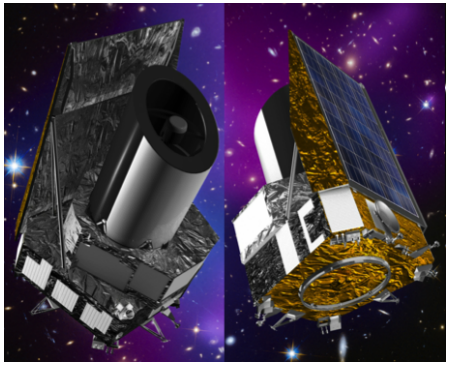
Data pipelines for Euclid



Martin Kümmel
mkuemmel@usm.lmu.de

LMU

LUDWIG-MAXIMILIANS-UNIVERSTITÄT MÜNCHEN

Faculty of Physics
Chair for Cosmology and Structure Formation

Outline

- Euclid:
  - Who / Why / What?
- The Euclid Science Ground Segment:
  - Structure: OU's, SDC's, SWG's
- What pipelines exist?
- A pipeline example: MER PF
- Hardware and IT
- Science Challenges

Who: Euclid Consortium (Jan '20)

- 16 countries (14 European + USA + Canada);
- 225 institutes/labs (astrophysics, cosmology, theoretical physics, high energy, particle physics and space science);
- 1545 full members (>900 researchers);
- Long term project (already 200 alumni)
- Yearly Euclid Mission Conference with ~400 participants;
- EC is likely the largest existing astronomy consortium!

Why: Science Goals of Euclid

- Determine the expansion of the Universe at various cosmic ages;
- Understand the nature of the expanding Universe;
- Explore the nature and properties of dark energy, dark matter and gravity
- **Two experiments:**
  - **Weak lensing analysis (Baryonic Acoustic Oscillations);**
  - **Galaxy clustering (Redshift Space Distortion);**
- Redshifts are required:
  - From NISP slitless spectroscopy;
  - From photo-z;
- (Legacy science);

What: Euclid project

- Project to achieve some science goals;
- Reaching the science goal needs massive data reduction;
- Lots of data is necessary;
- Data comes from:
  - Ground based surveys (EXTernal data);
  - The Euclid satellite;
- Euclid satellite is central to the project;

Euclid Satellite

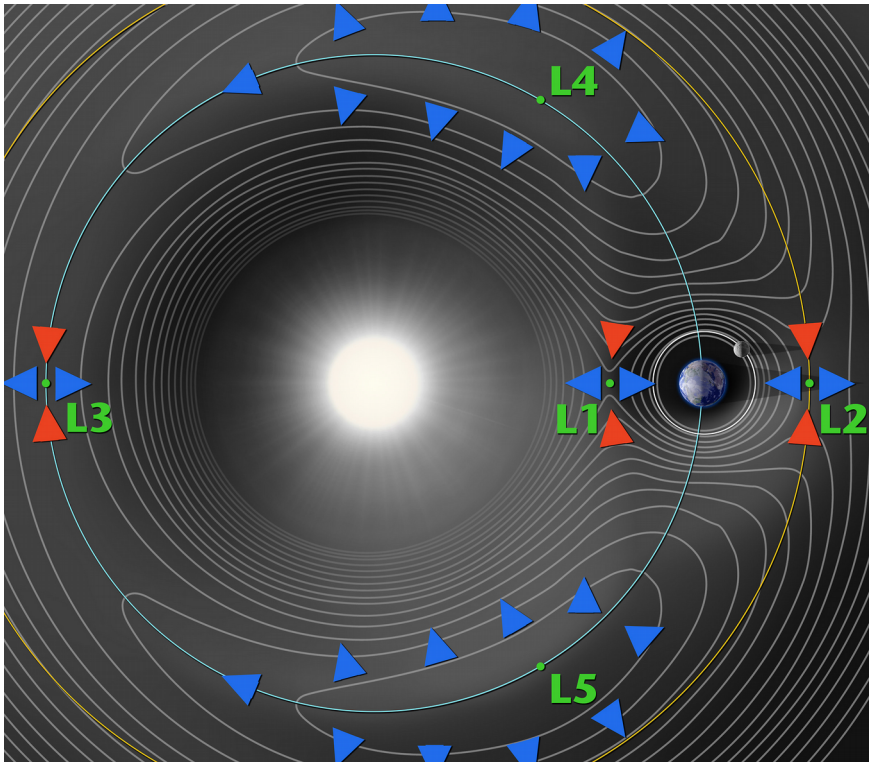
Euclid Consortium

Telescope	1.2m Korsch, 3 mirror anastigmat., f=24.5m				
Instrument	**VIS**	**NISP**			
Field-of-View	0.787x0.709 deg^2	0.763x0.722 deg^2			
Capability	Visual Imaging	NIR Imaging Photometry			NIR Spectroscopy
Wavelength range	550-900 nm	Y	J	H	1100-2000 nm
Detector Technology	36 arrays 4k x 4k CCD	16 arrays 2k x 2k NIR sensitive HgCdTe detectors			
Pixel Size/FWHM	0.1" / 0.2”	0.3 " / 0.3”			0.3 "/
Spectr. Res.	-	-			R=250

Launch: June 2022

Then: travel to L2

Euclid at L2



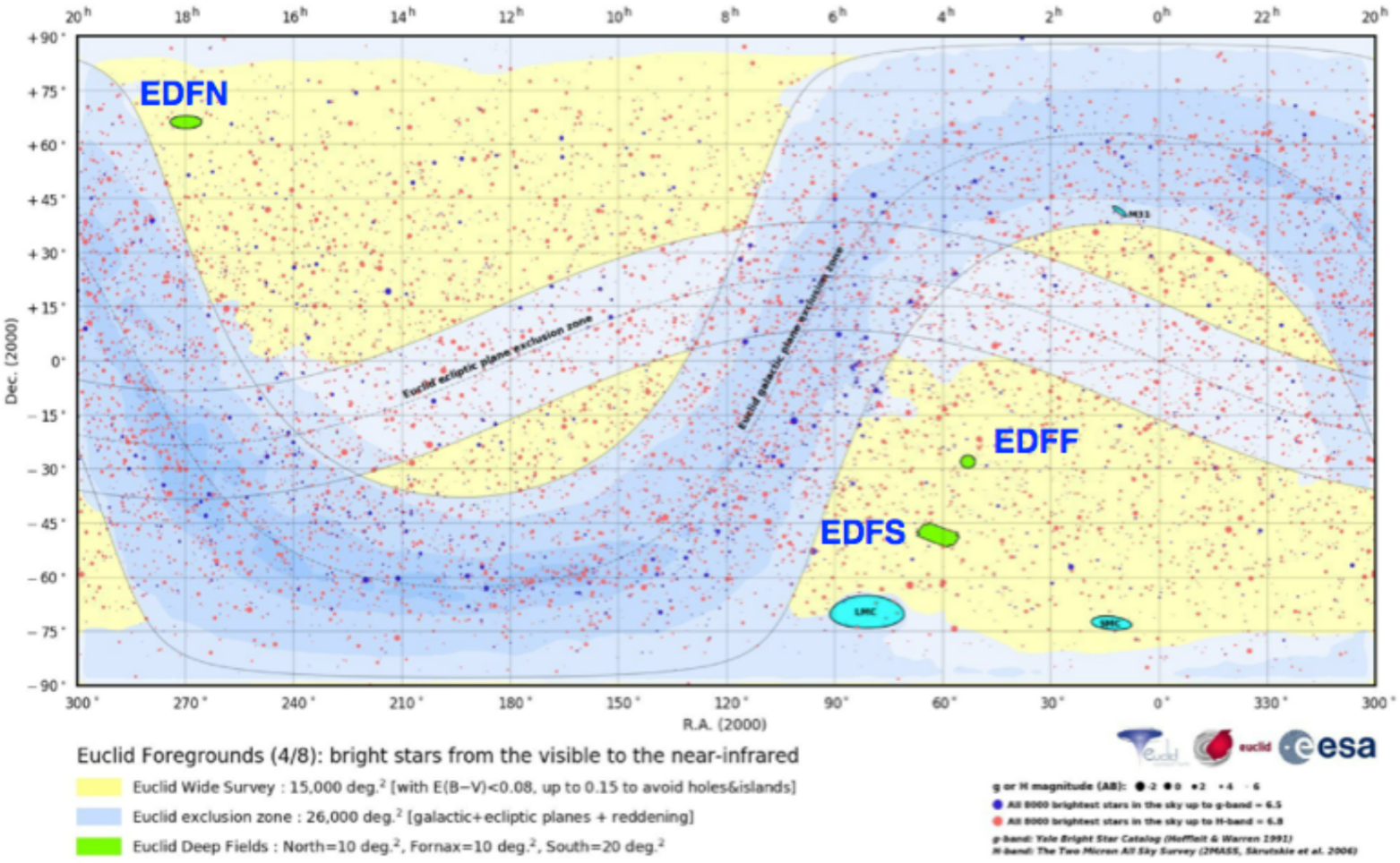
- L2 Sun-Earth Lagrangian point;
- Unstable → orbit;
- Propellant necessary → limited lifetime;
- **6 year nominal lifetime;**

Euclid Survey(s)

- Wide Survey:
  - 15,000 deg^2 (we are running out of sky...)
  - Survey specs:
    - VIS: 24.5 mag (10σ)
    - Y/J/H: 24.0 mag (5σ)
    - *g/r/i/z* = 25.2/24.8/24.0/24.0mag (10σ)
    - Spectroscopy down to $3x10^{-16}$erg cm $^{-2}s^{-1}$
- Deep Survey:
  - 40deg^2: (EDF-North and EDF-South):
    - 10 deg^2 EDF-North (NEP);
    - 20 deg^2 EDF-South (Stadion);
    - 10 deg^2 EDF-Fornax;
  - Survey specs:
    - VIS: 26.5 mag (10σ)
    - Y/J/H: 26.0 mag (5σ)
    - *g/r/i/z* = 27.2/26.8/26.0/26.0mag (10σ)
- NIR self-calibration field;
- Photo-z calibration field;

Euclid Surveys

Euclid Consortium



Jean-Charles Cuillandre

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

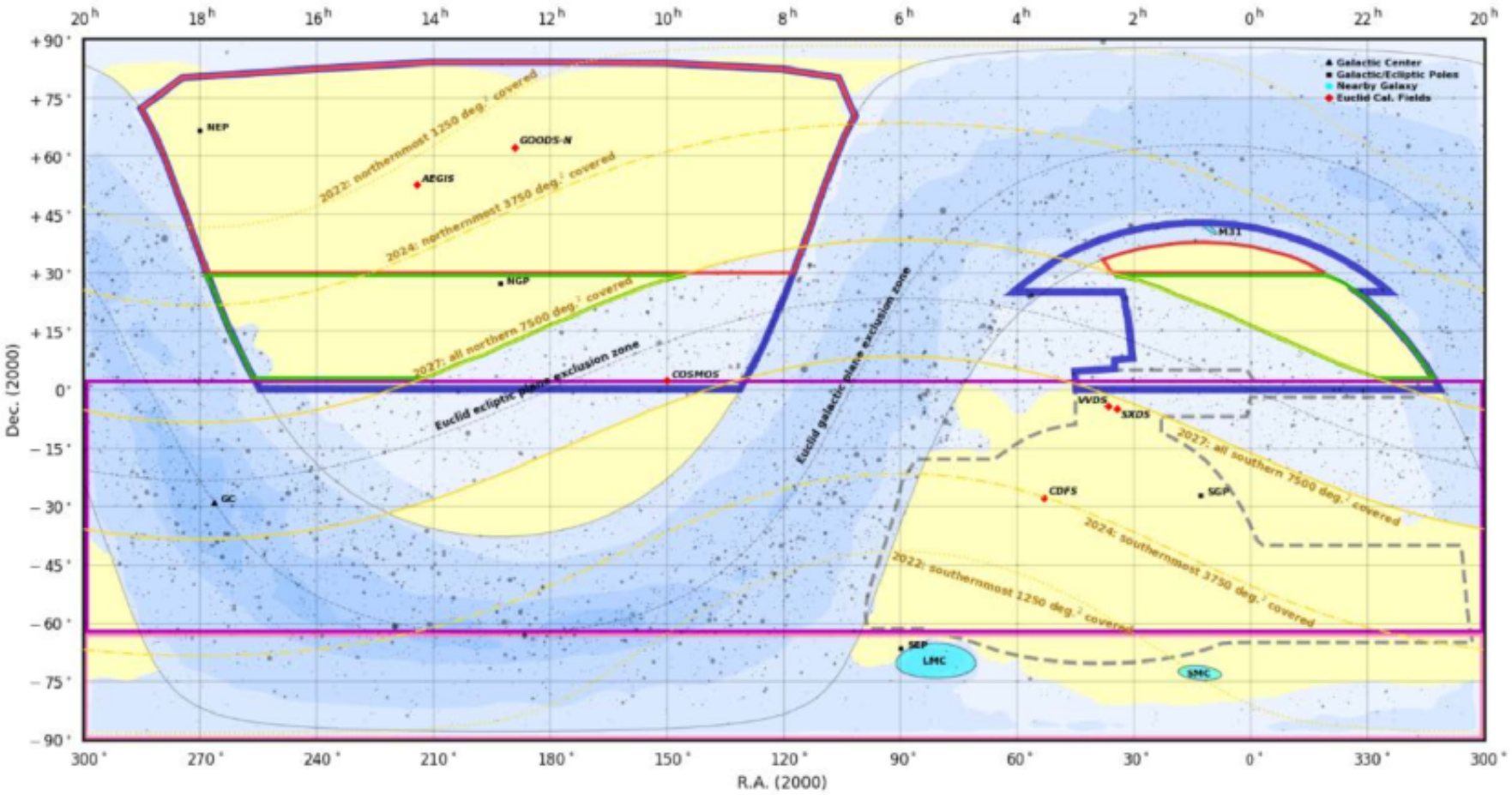
External Surveys: photometric

- Dark Energy Survey (**DES**): *griz,* south;
- Kilo Degree Survey (**KiDS**): *ugriz,* south;
- CFHT survey **CFIS**: *ur*, north
- **JPAS** survey: *g*, north;
- **PAN-STARRS 1 / 2**: *iz*, north;
- **LSST**: *griz*, south and north (special extension);

For photometric redshifts!

External Surveys: photometric

Euclid Consortium



Expected ground-based coverage of the Euclid Wide Survey DR2/DR3 (2026/2029) (origin/bands/overlap)

Euclid Wide Survey : 15,000 deg.² [with E(B−V)<0.08]

Euclid exclusion zone : 26,000 deg.² [galactic+ecliptic planes]

→ Ecliptic isolines track the space survey started at the ecliptic poles

DES−griz : 4500 deg.²

CFIS−u : 7300 deg.²

CFIS−r/JEDIS−g/Pan−STARRS−iz : 4800 deg.²

LSST main survey, ugriz : 7000 deg.²

LSST south extension, ugriz : 1000 deg.²

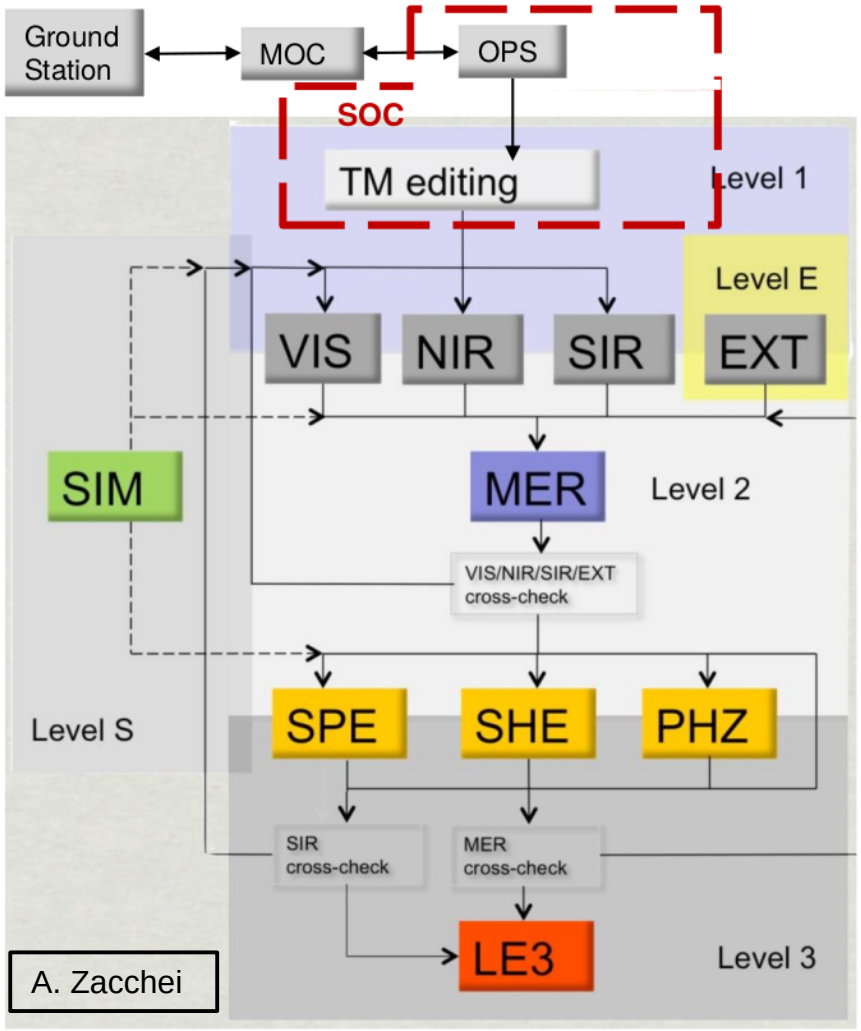
LSST north extension, griz : 3000 deg.²

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Science Ground Segment (SGS)

- Organizational Units (OU's):
  - Find data reduction methods;
  - Develop (at least) prototype code;
  - Design and implement the Processing Function (→ Data Pipeline);
- Science Data Centers (SDC's):
  - Euclidize the Data Pipelines;
  - Optimize software to production level;
  - Run the Pipeline Functions;
- Science Working Groups (SWG's):
  - Cover Weak Lensing/Cluster/Galaxy Clustering/Strong Lensing/Cosmology
  - Set the requirements for the data reduction;
  - Do the scientific analysis;
  - Write papers;
- Euclid Archive System (EAS):
  - Hosts the data and metadata;
  - Controls the processing;
  - Data delivery;

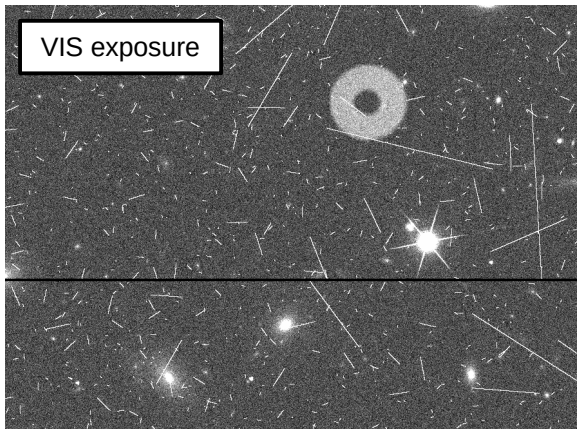
OU-PF → Euclid PF

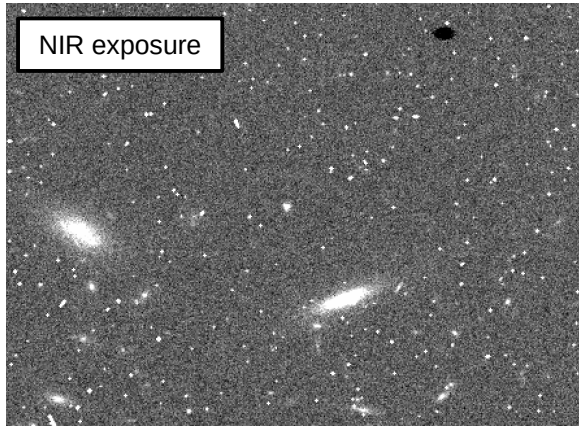


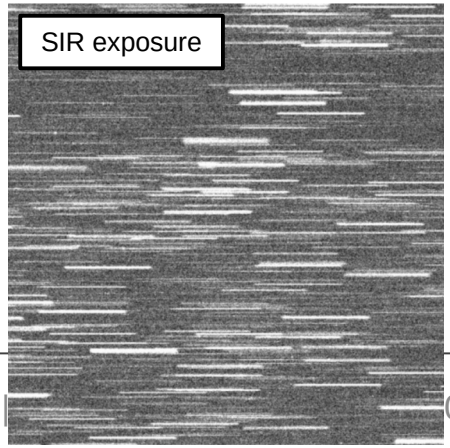
- VIS, NIR, EXT: production of fully calibrated photometric exposures from Euclid and ground-based surveys
- SIR: production of fully calibrated 1D spectra extracted from the NISP spectroscopic exposures.
- MER: production of a source catalog containing consistent photometric and spectroscopic measurements.
- PHZ: production of the photometric redshift for all catalogued sources.
- SPE: production of spectroscopic redshifts for all sources with spectra.
- SHE: measurements of galaxy shapes.
- LE3: production of all high-level science products.
- SIM: production of all the simulated data necessary to validate the data processing stages, and to calibrate observational or method biases.

OU-VIS, OU-NIR, OU-SIR pipelines

- Instrument (Euclid) oriented OU's;
- Base task: production of fully calibrated pixel data!
- OU-VIS:
  - Cosmic-ray identification;
  - Determine the Point-Spread-Function (variable);
  - Ghost masking;
  - Correct the CTE (Charge Transfer Efficiency) effect!;
- OU-NIR:
  - Determine the Point-Spread-Function (variable);
  - Non-linear sensitivity;
  - Persistence!
- OU-SIR:
  - Instrumental effects as OU-NIR;
  - Slitless spectroscopy specific effects:
    - "Object finding";
    - Contamination;
    - Spectral orders;

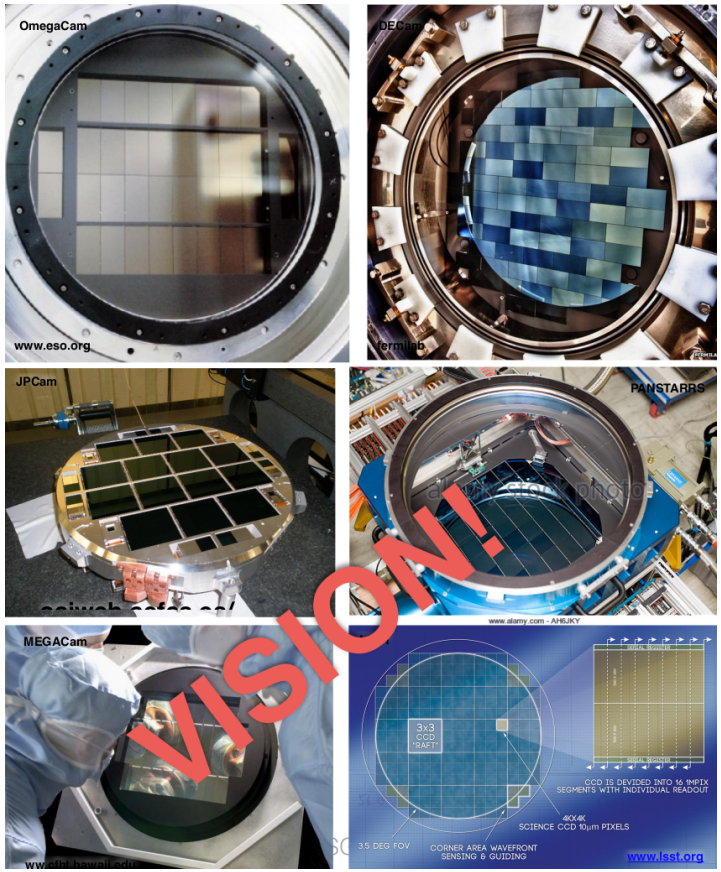






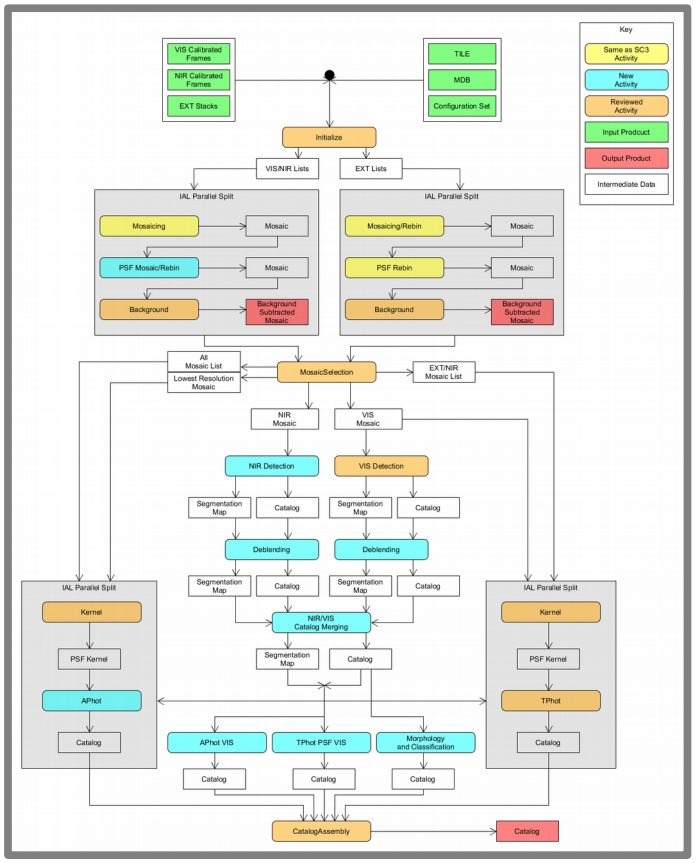
OU-EXT pipeleine

- Data from DES, KiDS, CFIS, JPAS, PAN-STARRS, LSST;
- Survey oriented OU(s);
- Stage 1 pipeline:
  - Survey specific code!
  - May contain legacy code;
  - Generate fully calibrated science exposures!
- Stage 2 pipeline:
  - Identical for all surveys!
  - Exposures → output products;
  - Image co-addition;
  - Psf processing



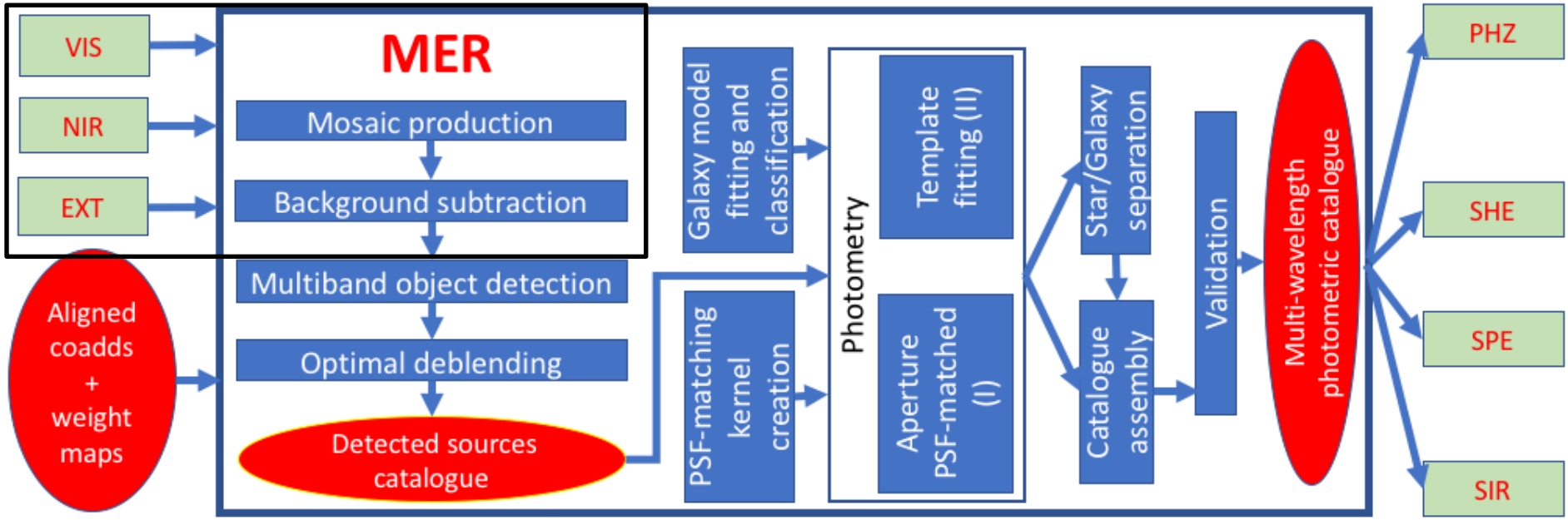
OU-MER pipeline

- Object detection;
- Photometry;
- Object classification;
- Object morphology;



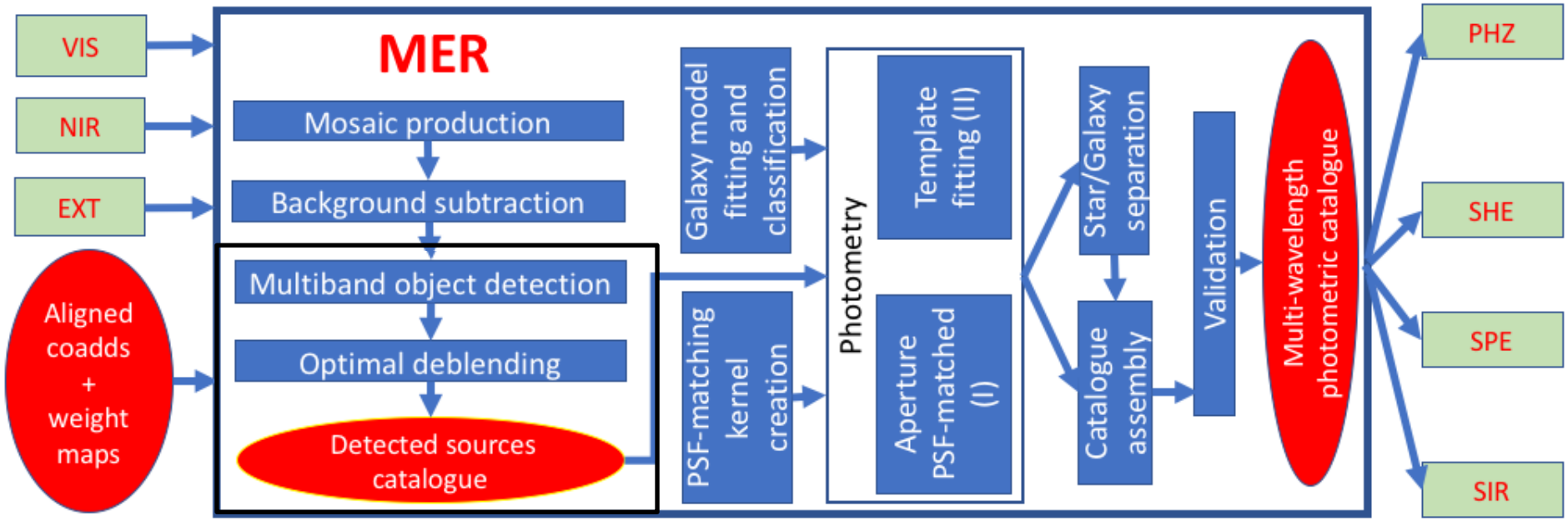
Data flow diagram

The OU-MER pipeline



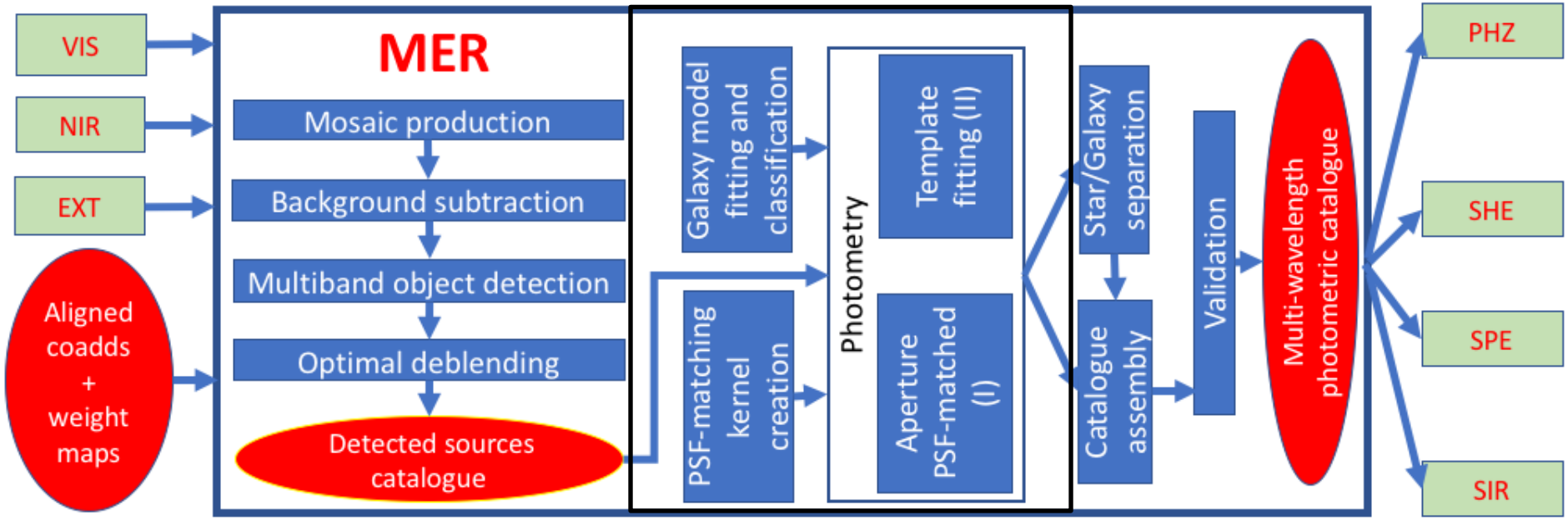
- **<u>Mosaic production</u>:** input images are combined in order to produce the final detection frame and appropriate measurement frames on pre-defined sky tiles;
- **<u>Background subtraction</u>:** any residual background is subtracted.

The OU-MER pipeline



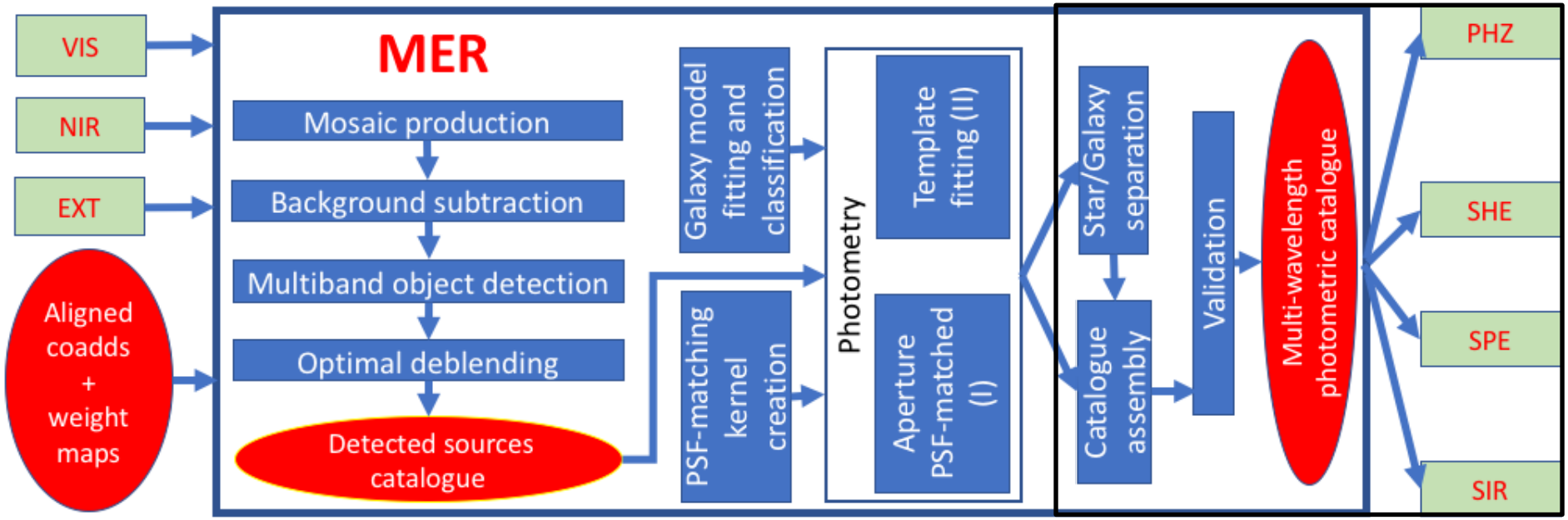
- **Detection**: all “significant” pixels of the VIS and NIR mosaic are individuated and flagged in segmentation map.
- **Deblending**: an algorithm is applied to separate blended sources and produce final VIS and NIR source lists and segmentation. Source lists are then combined.

The OU-MER pipeline



- **PSF matching:** convolution kernel to match the different PSFs are computed and all PSF-matched images needed are produced;
- **Photometry**: images are analysed by the photometric module(s) and a multiwavelength photometric catalogue of all sources is extracted.

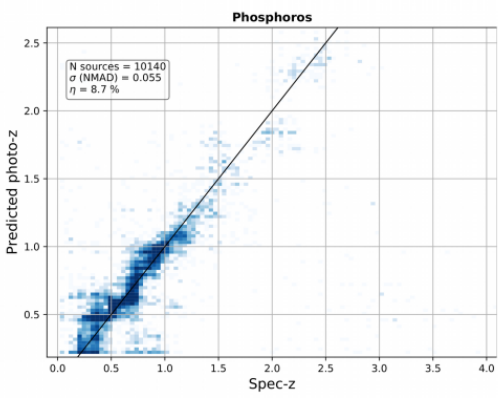
The OU-MER pipeline

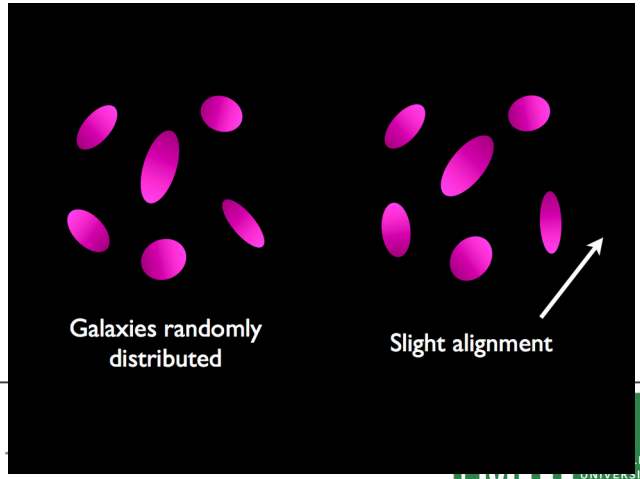


- **Morphological parameters** and other feature of the detected objects are computed, point-like and extended sources are separated.
- **Catalog Assembly** puts together all the information in the Euclid catalog and make final adjustments (unit conversion, unique ID assignment, etc.)

OU-SPE, OU-PHZ, OU-SHE pipelines

- “Science oriented” OU’s
- OU-SPE pipeline:
  - Takes the 1-d (slitless) spectra from SIR;
  - Determine the redshifts;
  - Identify and measure emission lines;
- OU-PHZ pipeline:
  - Takes the photometric information from MER;
  - Determines the photometric redhsifts!
  - Strong requirements for accuracy and failure rate!
- OU-SHE pipeline:
  - Measures the morpholgy of (suitable) galaxies;
  - No suitable algorithm existed at project start!
  - Three competing algorithms!

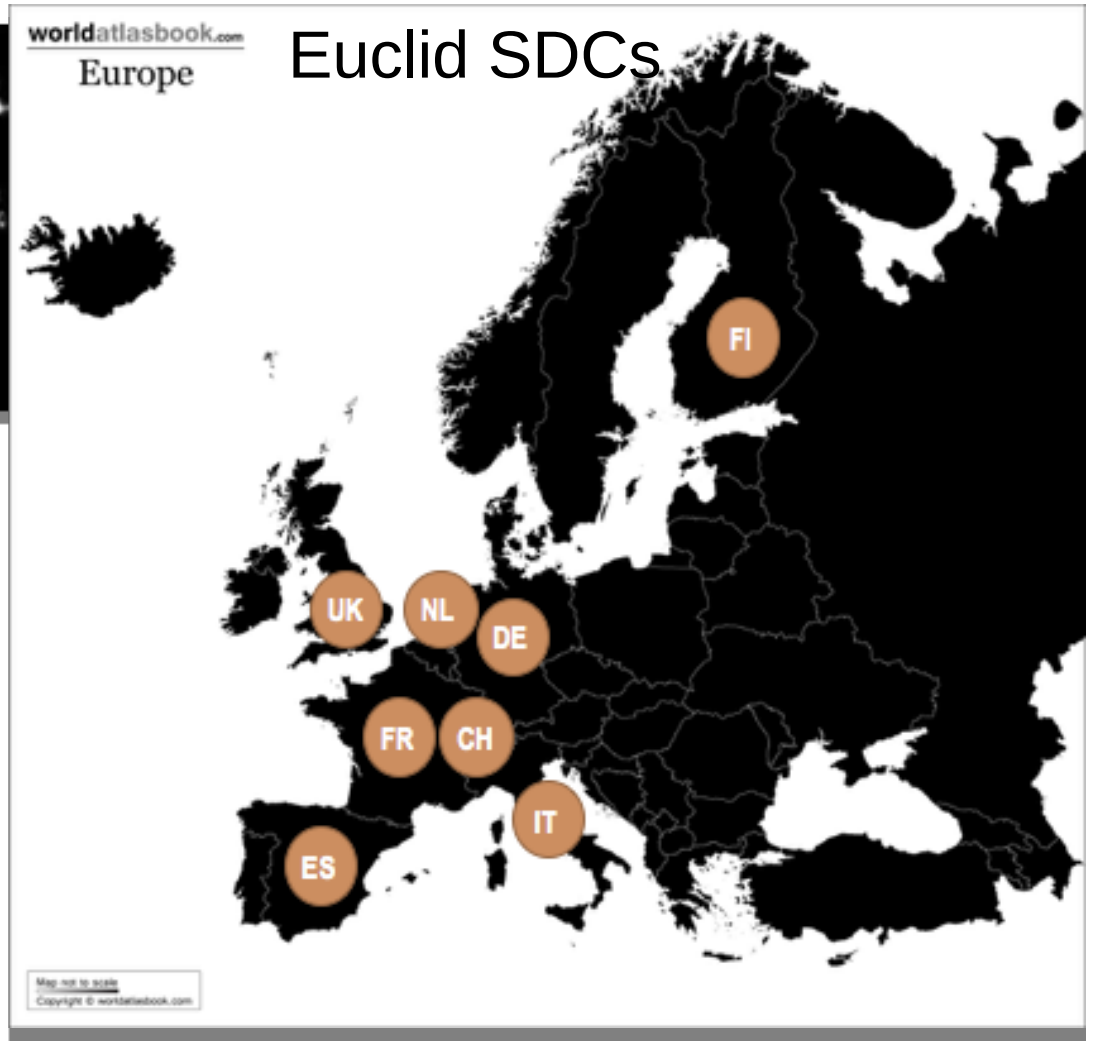




OU-LE3 pipeline

- 250 members: the largest OU!
- 47 define Processing Functions (→ Pipelines):
  - 12 with Priority 1 → Primary Core Science;
  - 4 with Priority 2 → Secondary Core Science (depending on available human resources);
  - 31 with Priority 3 / 4 → Related to Legacy Science (implemented if possible);
- Purpose of the Pipelines:
  - Preparation and characterization of various catalogs of objects;
  - Analysis of spatial distribution and weak lensing of galaxies;
  - Analysis of galaxy clusters;
  - Analysis of nearby galaxies and Milky Way objects;

Science Data Centers



Science Data Centers (cont.)

- Each major contributing country has one SDC;
- Different organizations (dedicated hardware, general computing center, …)
- Classical server hardware;
- No GPU's (due to heterogeneity);
- Every SDC must be able to run every Euclid code (… except EXT legacy code);
- SDC-DE:
  - Part of the Max-Planck-Gesellschaft computing center;
  - Dedicated hardware;
  - 1st generation cluster with 648 cores already retired;
  - 2nd generation cluster with ~300 cores, 600 TB storage running;
  - During survey processing: 6000 cores;

Processing / Software development

- Processing:
  - In Virtual Machines (VM), CENTOS7 based;
  - Defined set of libraries available;
  - Latest VM with CVFMS (→ automatic upgrading);
  - Continuous software deployment;
- Software development:
  - C++11 and python3;
  - Few (non-C++/python3) legacy code;
  - CMake based build system;
  - Git/github;
  - Not many fundamental libraries;

Science Challenges (SC)

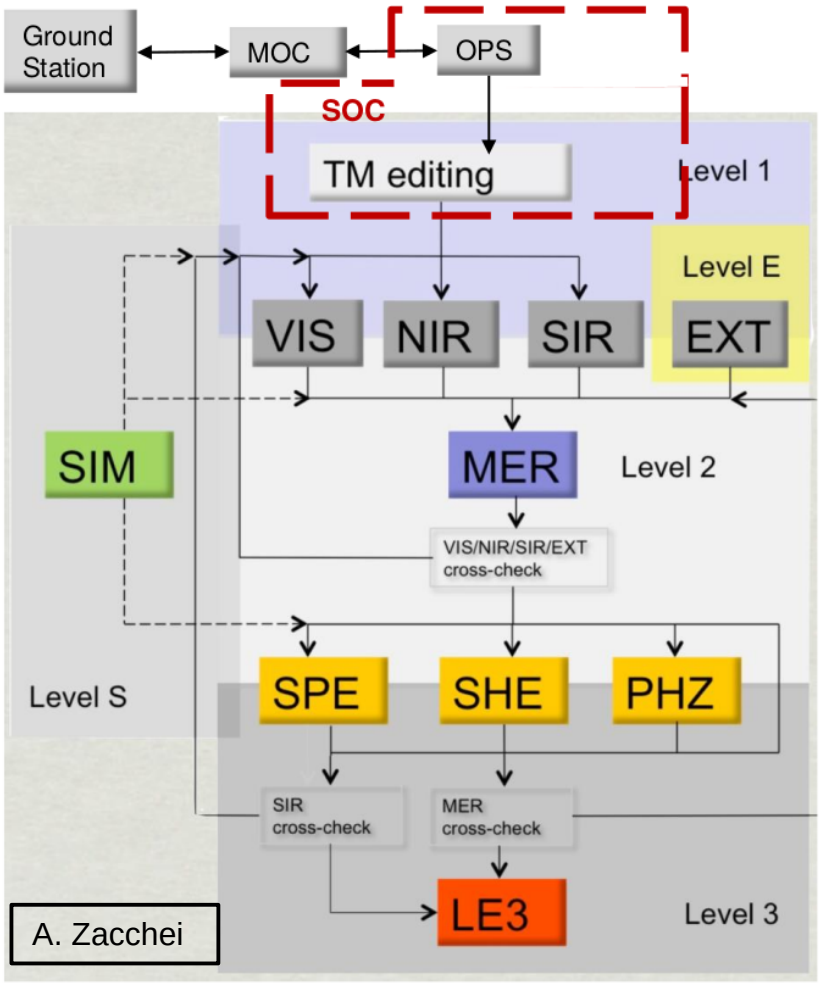
- Project started in 2011, first data comes 2023 → **What are we all doing?**
- **We use simulated data** (→ OU-SIM pipeline)!
- Science Challenge: pipelines reduce simulated data!
- Science Challenge runs in the production environment;
- Science Challenges sequence:
  - Increase in area;
  - Increase in processing depth;
  - Increase in level of detail;
- Goal: end-to-end processing over entire survey area!
- Central for pipeline validation and performance validation

OU-SIM pipeline

- Simulations for all instruments:
  - VIS;
  - NIR (photometry and spectroscopy);
  - EXT (all flavours DES, KiDS, LSST, ...);
- **All** (CTE, persistence, PSF's+variability, ...) instrumental effects need to be simulated;
- Requires tons of input:
  - True universe from theoretical simulations;
  - Stellar model (for the Galactic stellar populations);
  - Galaxy and solar system input (extinction and sky background);
  - Model parameters for the various instrumental effects;
- The quality of the reduction pipelines depends on the quality of the simulation pipeline!

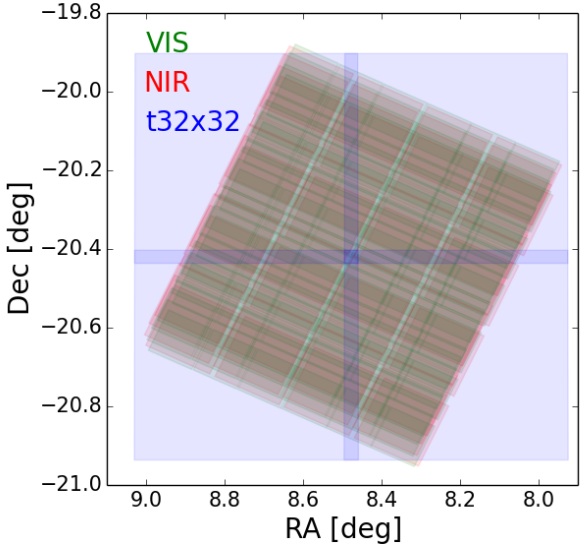
Science Challenge Sequence

- SC1: SIM (VIS/NIR/SIR data)
- SC2: SIM1 + data model
- SC3: SC3 + VIS, NIR, SIR, EXT, MER
- SC4/5/6: SC3 + SPE, SHE, PHZ
- SC7: SC4/5/6 (all!) + partially LE3
- SC8: everything!



Science Challenge 3 data

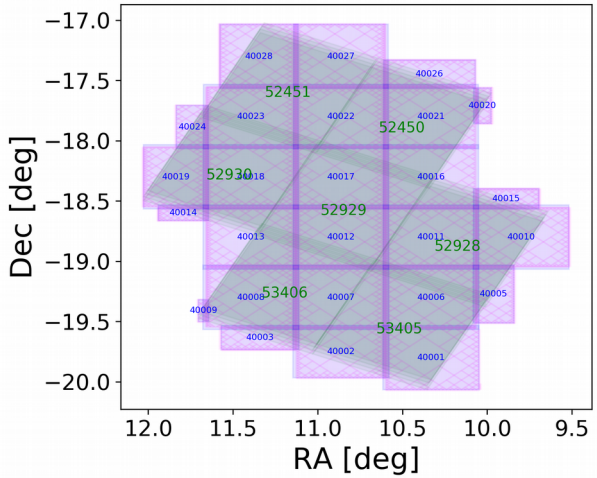
- 1 Euclid FOV
- 4 pipeline fields
- Took more than 12 months!

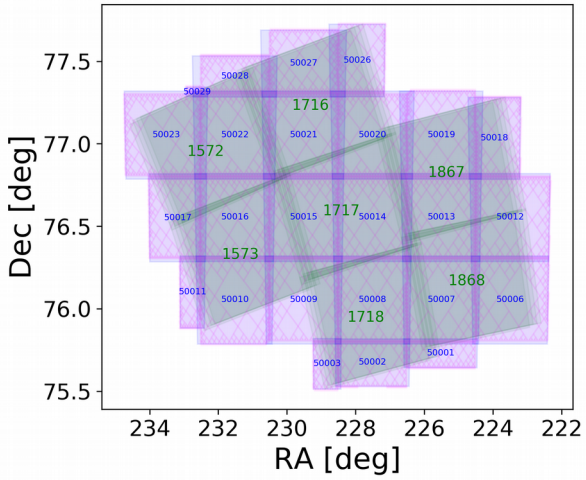


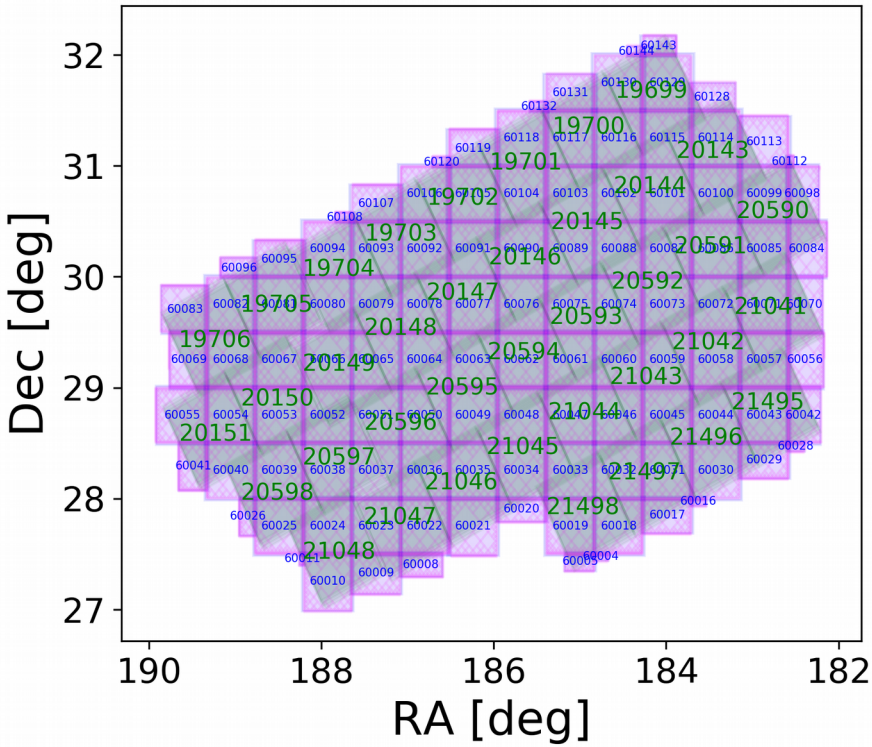
SC3 data

Science Challenge 4/5/6 data

- 60 Euclid FOV
- ~150 pipeline fields
- Took about 22 months!

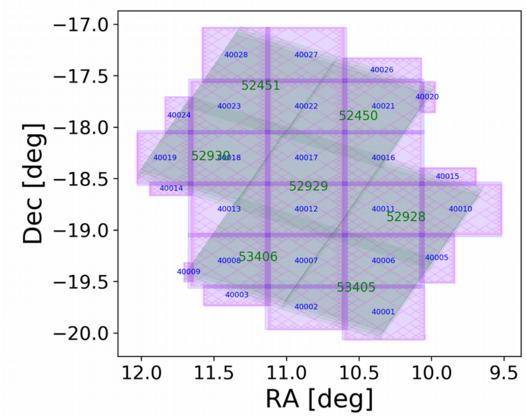


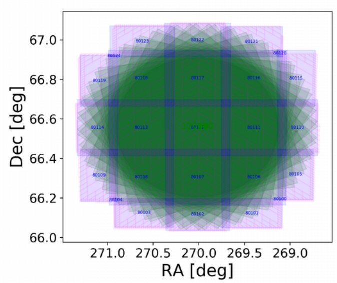


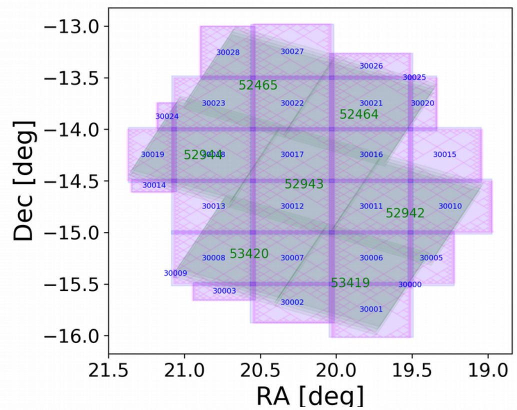


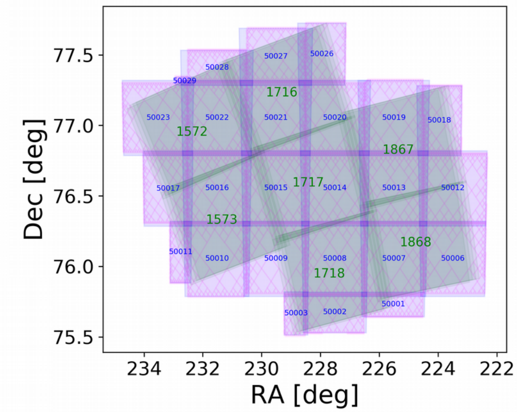
Science Challenge 7 data

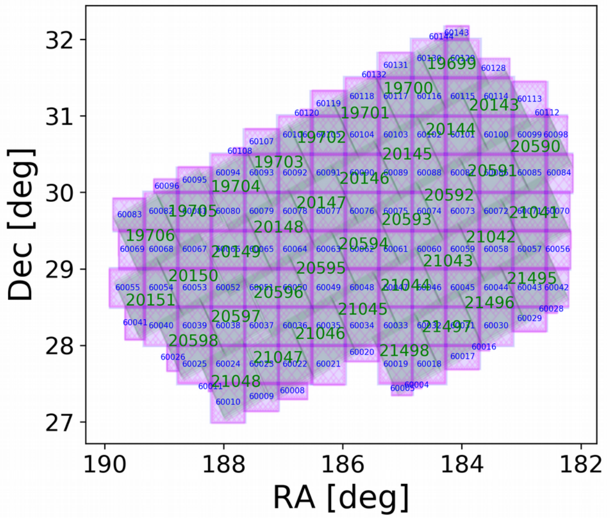
- 120 Euclid FOV + Deep Field
- ~300 pipeline fields
- Time frame 3 months!

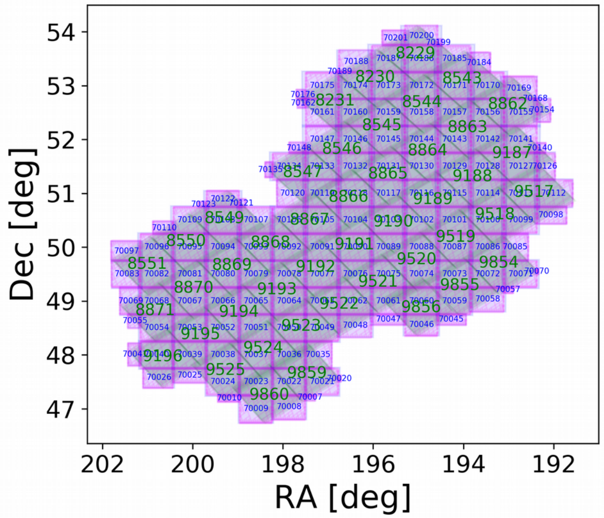


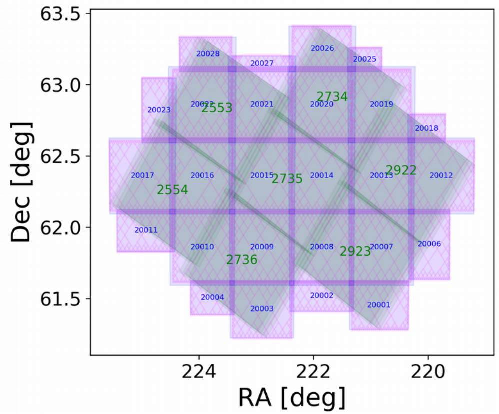




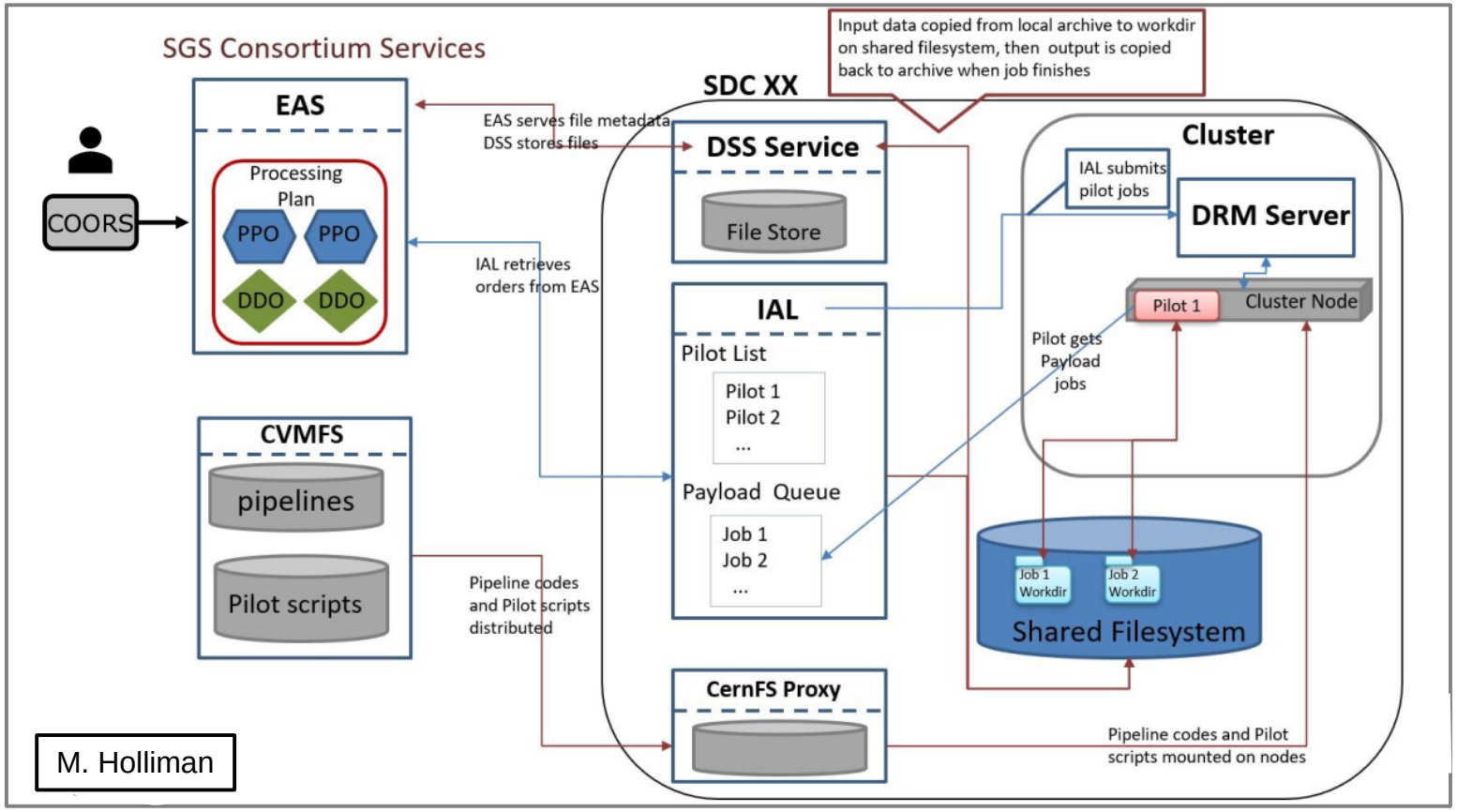








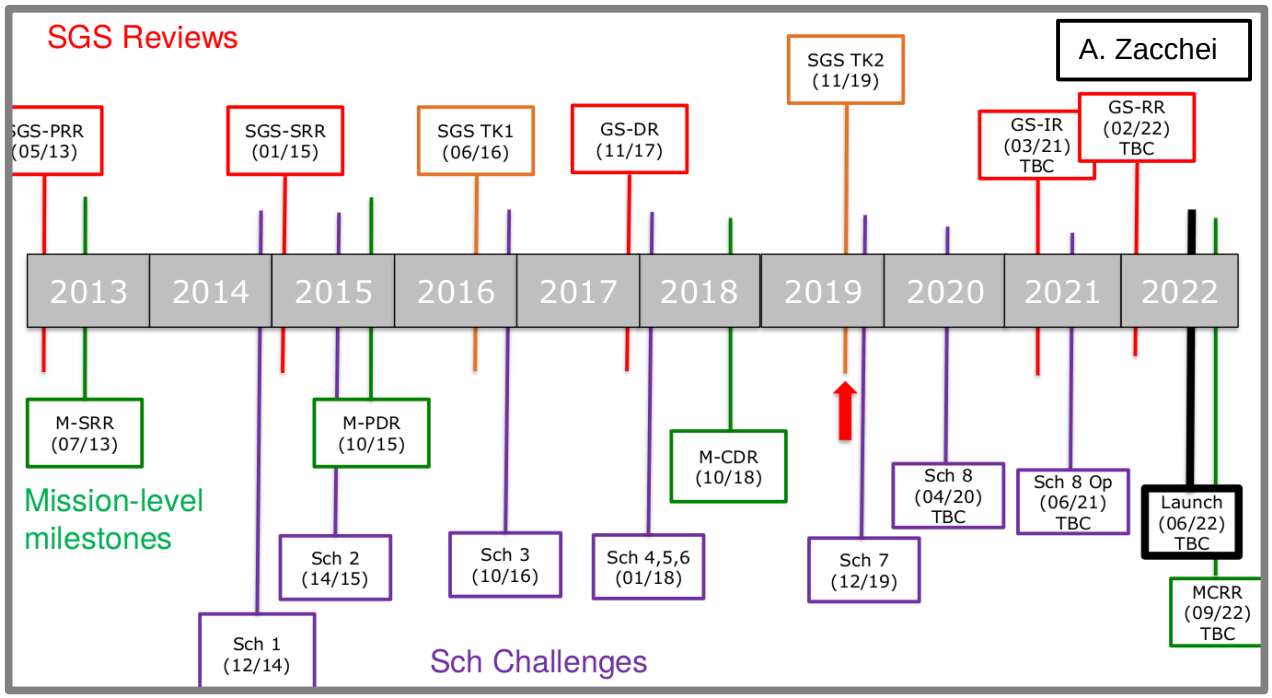
Pipeline infrastructure



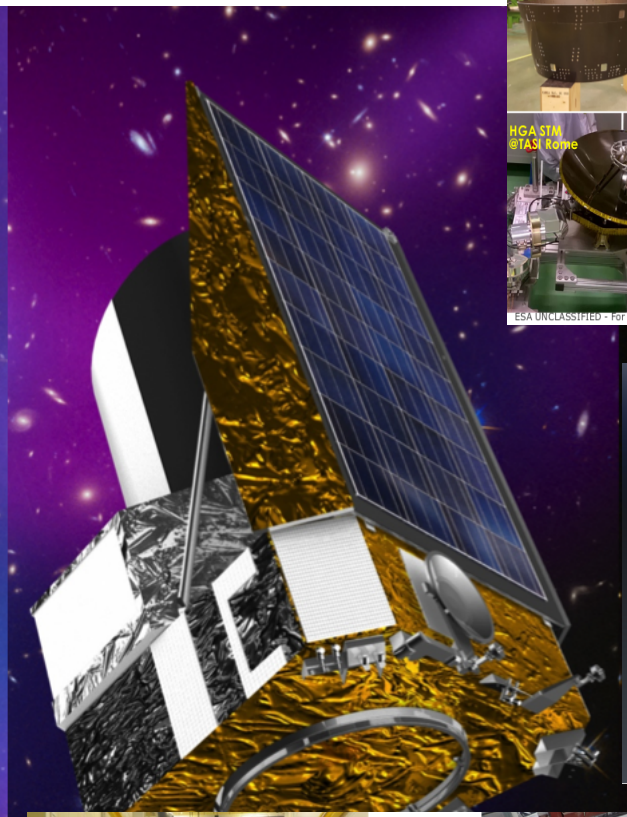
Lessons learned (personal)

1) Interfaces (IT) matter!
2) Interfaces (sociological) matter!
3) Interfaces (background) matter!
4) Project management helps!
5) It is not “**only software**”!
6) People matter!

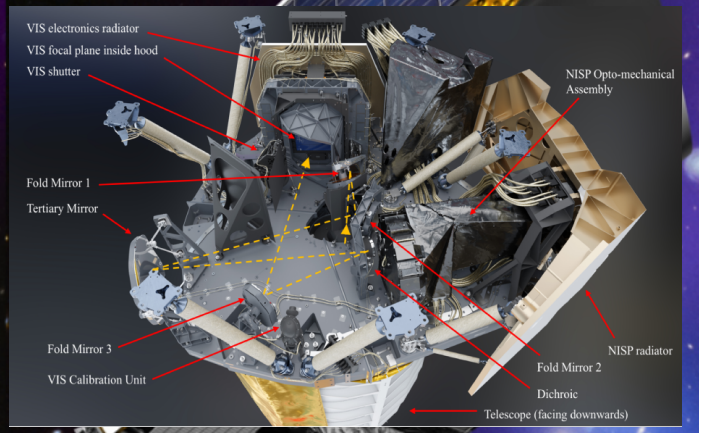
Project Status

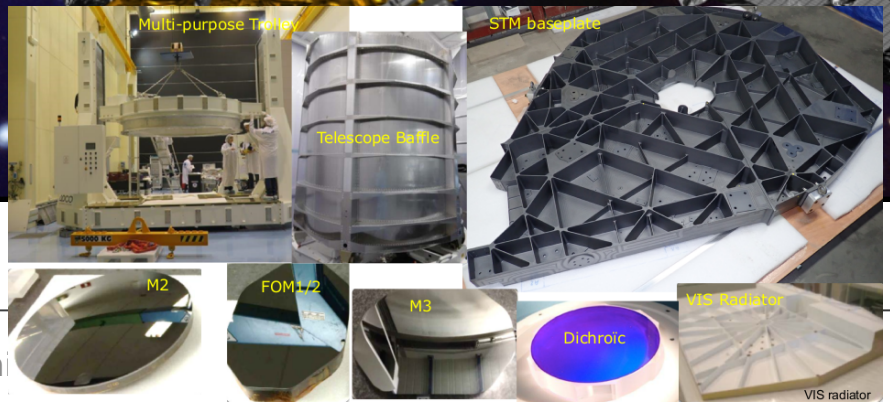


- The launch is approaching!
- Still lots of things to do!
- Data pipelines are in a reasonable state;
- Cautiously looking forward!




SVM Hardware
CFRP cone & final assembly jig @ADS CASA
SSH STM @RUAG
SVM AVM : CDMU & HRS @Torino
HGA STM @TAS Rome
VIS EM on PLM AVM @TAS-I
TIC EM panel @Madrid
ESA UNCLASSIFIED - For Official Use


VIS electronics radiator
VIS focal plane inside hood
VIS shutter
NISP Opto-mechanical Assembly
Fold Mirror 1
Tertiary Mirror
Fold Mirror 3
VIS Calibration Unit
NISP radiator
Fold Mirror 2
Dichroic
Telescope (facing downwards)


Multi-purpose Trolley
Telescope Baffle
STM baseplate
M2
FOM1/2
M3
Dichroic
VIS Radiator
VIS radiator

Further reading

- Euclid consortium page:

  https://www.euclid-ec.org/

- Euclid@ESA:
  - https://sci.esa.int/web/euclid
  - http://www.esa.int/Science_Exploration/Space_Science/Euclid_overview