Centre de Calcul
de l'Institut National de Physique Nucléaire
et de Physique des Particules

News from LSST - Vera C. Rubin Observatory

FJPPL — Japan-France workshop on computing technologies

Gabriele Mainetti

February 18, 2025

Outline

1 Astro@CC-IN2P3

2 Rubin and LSST

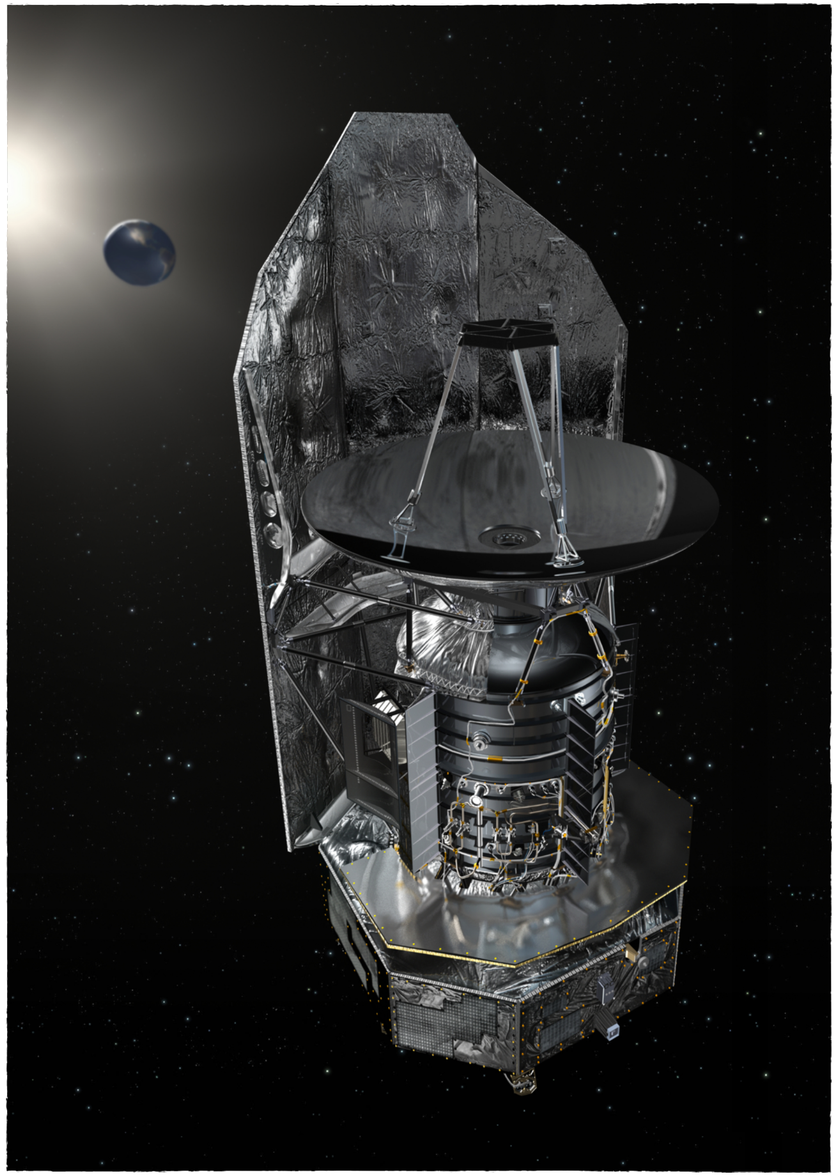
3 Rubin at CC-IN2P3

4 Conclusions

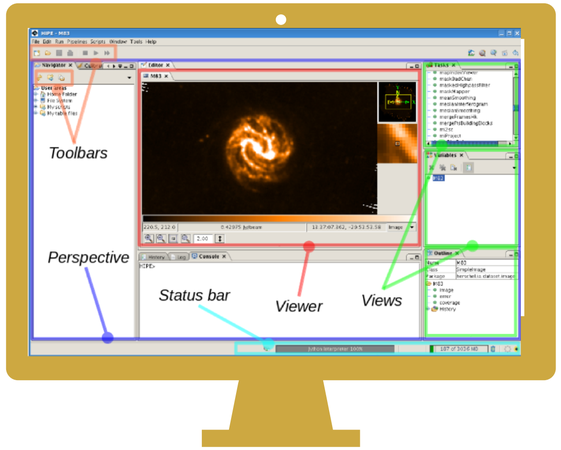
Astro @ IN2P3 Computing Center

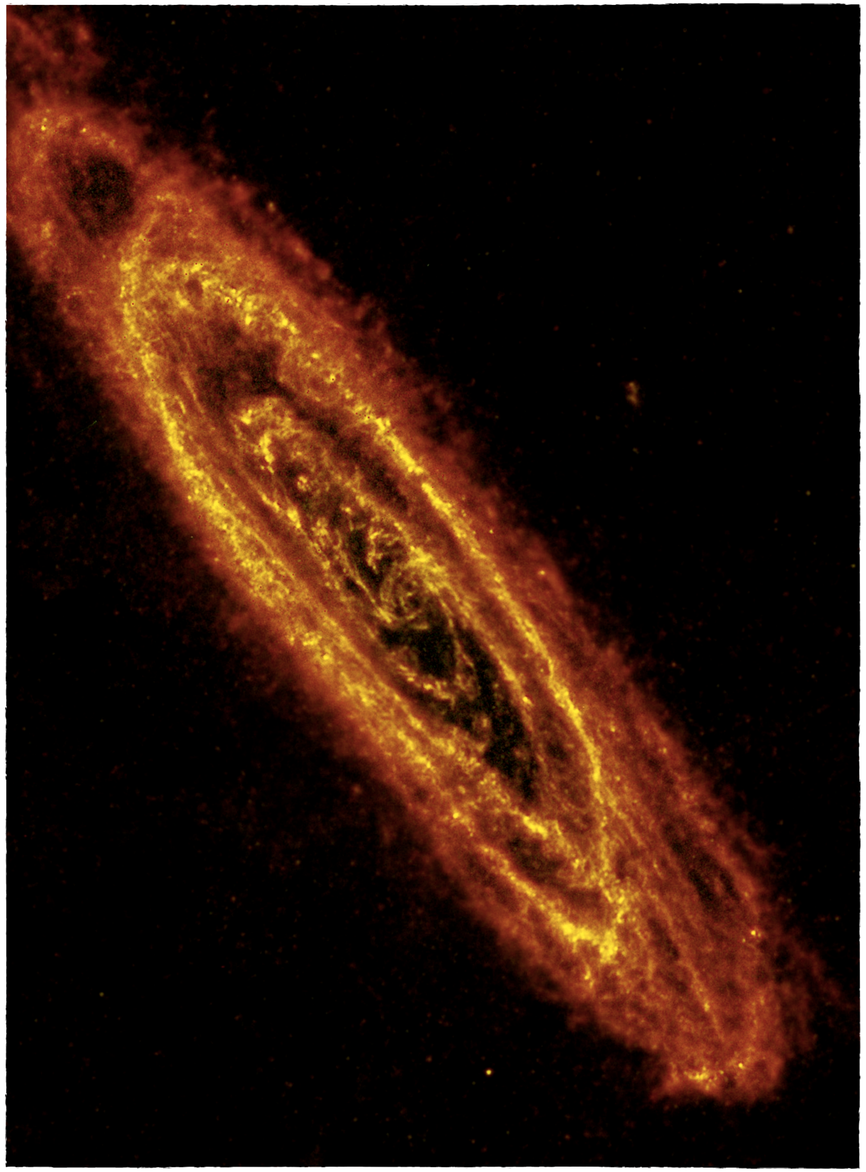
1

Astro @ CC-IN2P3



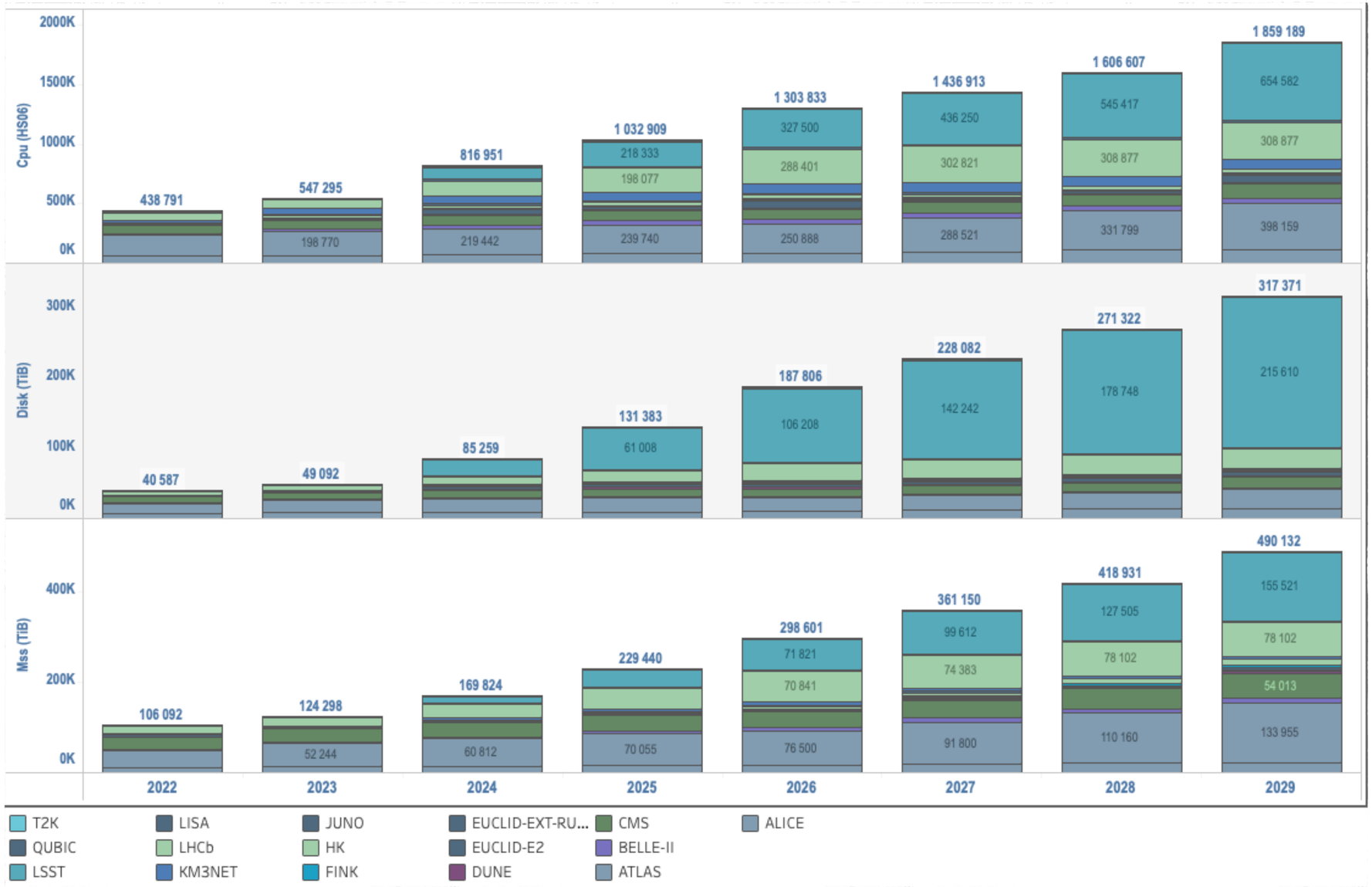
Credit: ESA/AOES Medialab





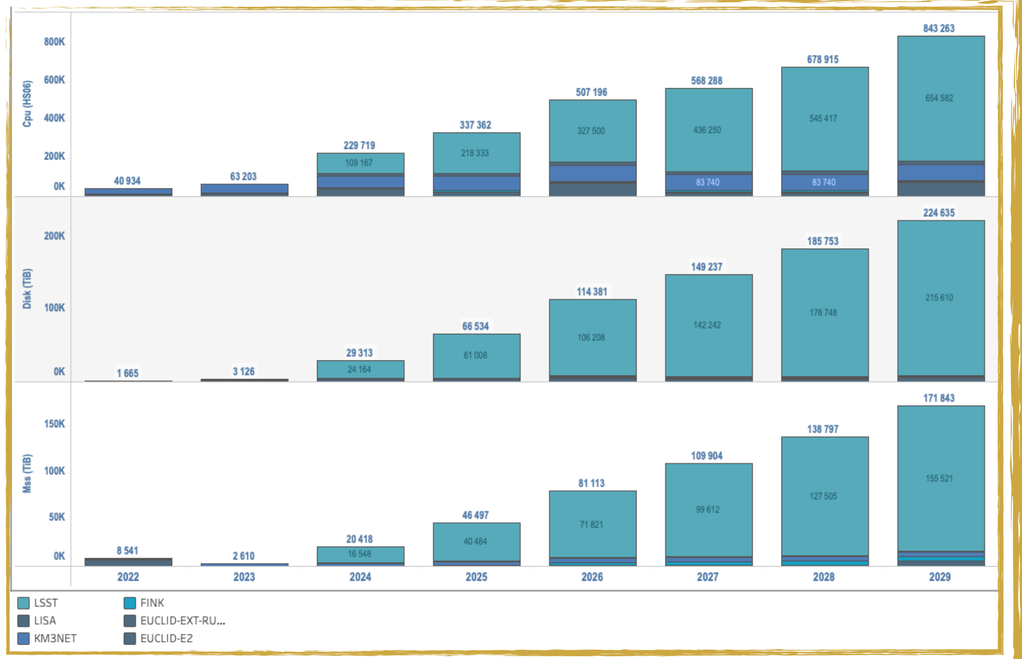
Credit: ESA/Herschel/PACS/SPIRE/J. Fritz, U. Gent

Astro @ CC-IN2P3



Astro @ CC-IN2P3

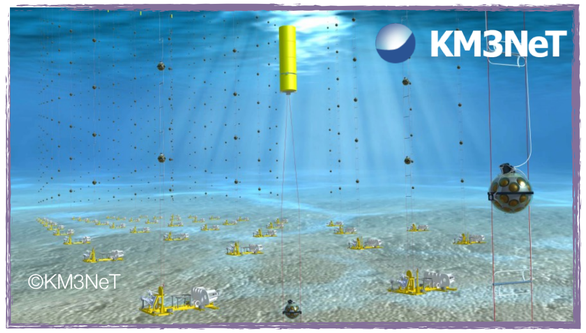
- CC-IN2P3 supports more than 40 astroparticle physics projects, from cosmology to neutrino and gravitational waves
- ~ 20% of the total CPU, 30% of GPU, 42 PB (disks and tapes) (2023)



The French Euclid Science Data Center SDC-FR

The French Rubin Data Facility - FrDF

The French KM3NeT Computing Center



The French LVK (Ligo/Virgo/Kagra) Computing Center

The Rubin Observatory and LSST

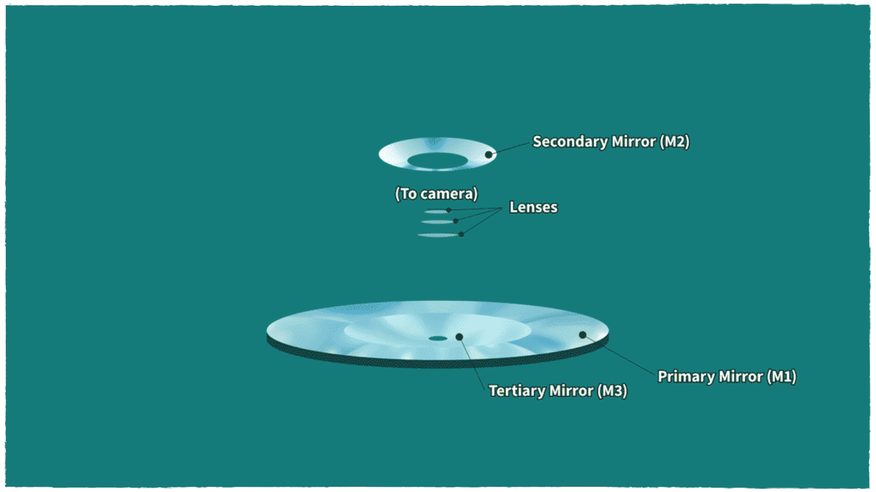
2

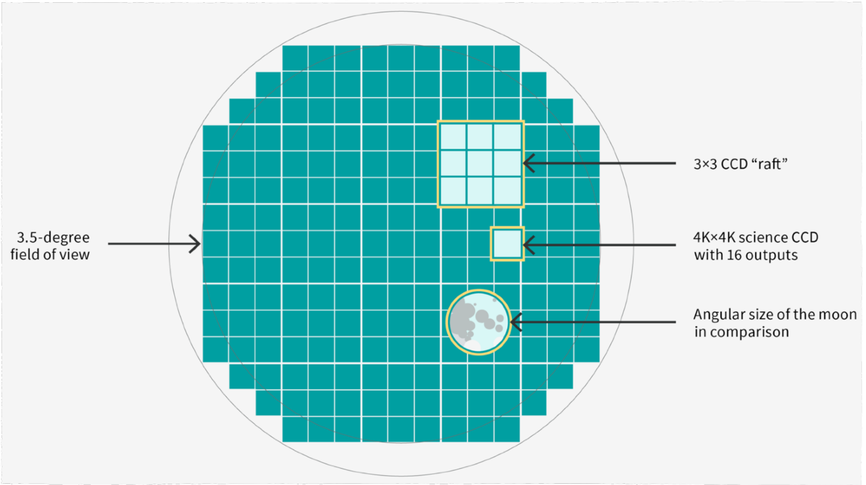
Rubin Observatory

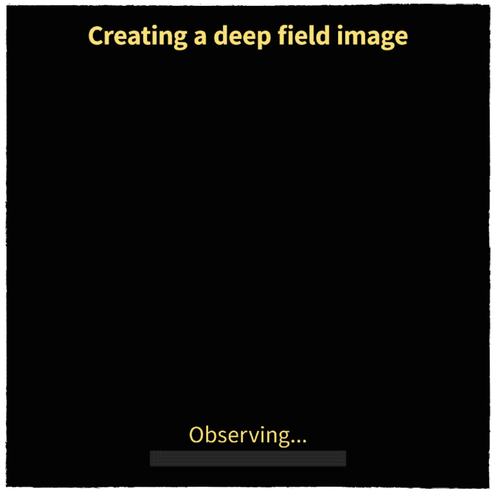
The Vera C. Rubin Observatory

- Cerro Pachón @ Chile (2647m asl)
- Main mirror **8.4m** ⌀
- **3.2 G pixels** camera
- **9.6 deg²** Field of View
- **f/1.234** aperture

Credit: Hernan Stockebrand



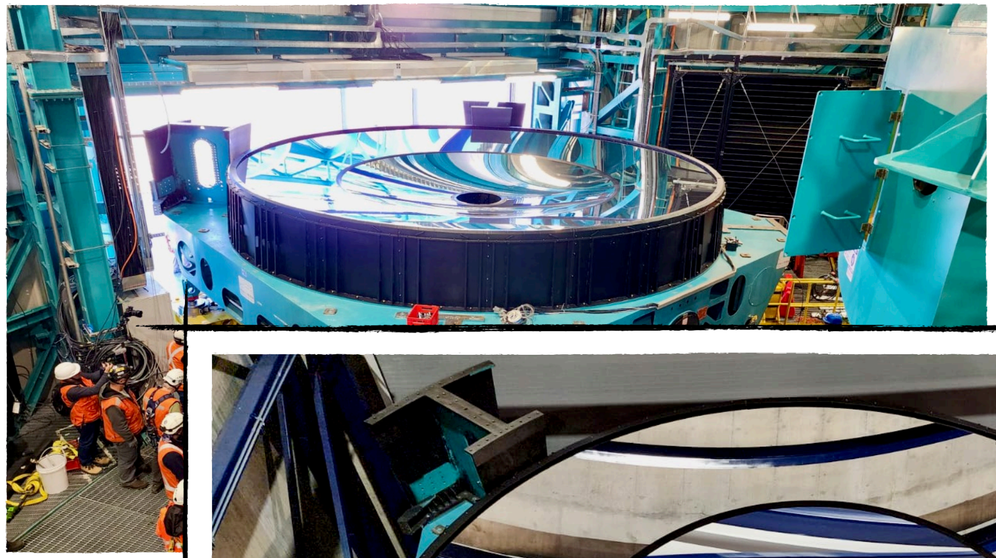


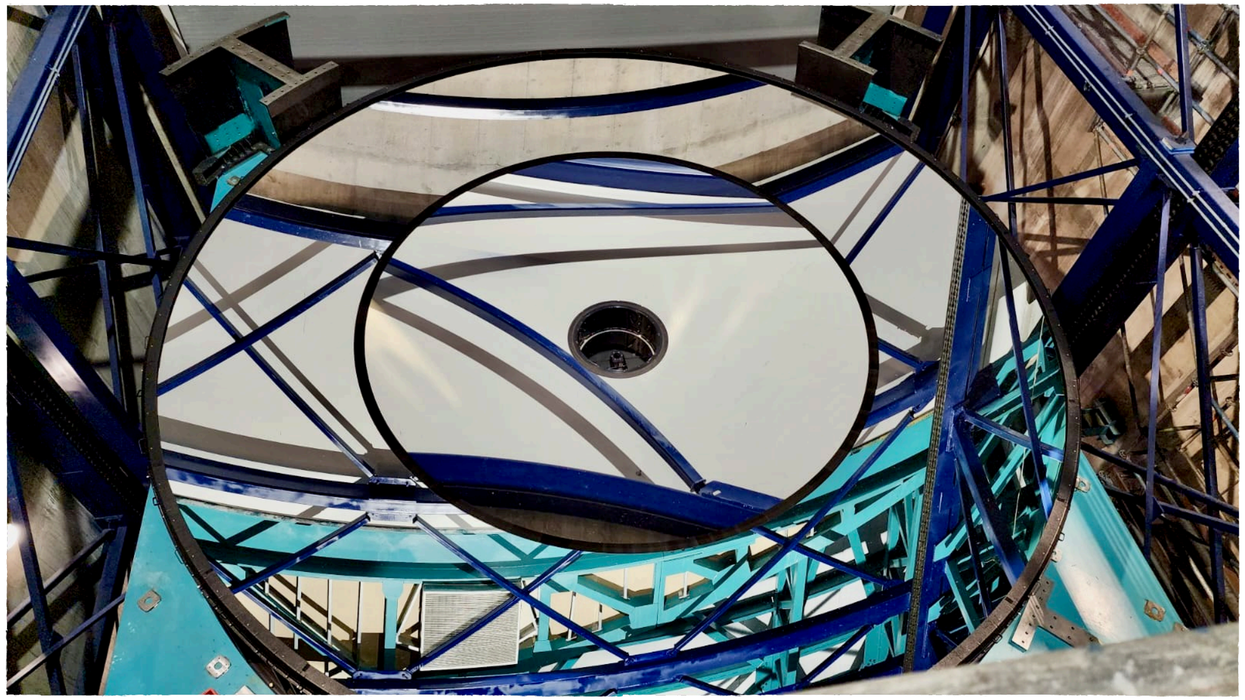


Rubin Mirrors

Credit: Rubin Observatory/NSF/AURA/T. Vičuna/S. Deppe

Rubin Mirrors



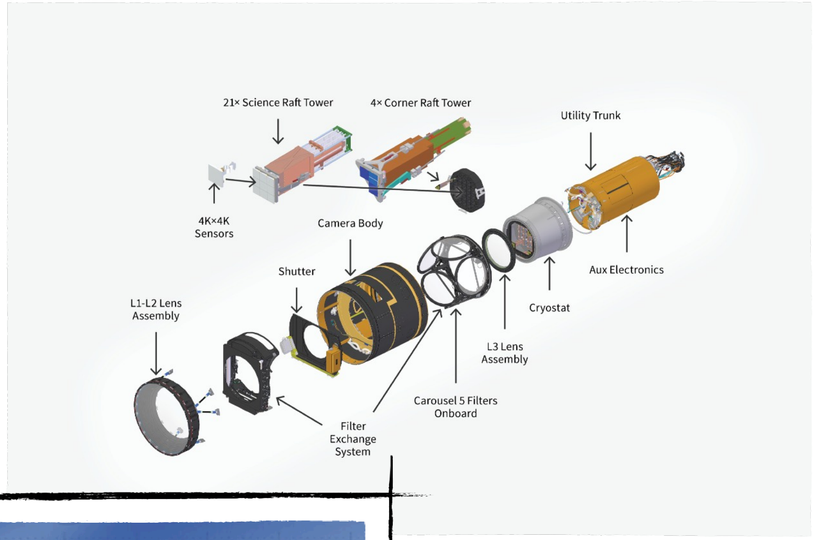


Credit: RubinObs/NOIRLab/SLAC/NSF/DOE/AURA/

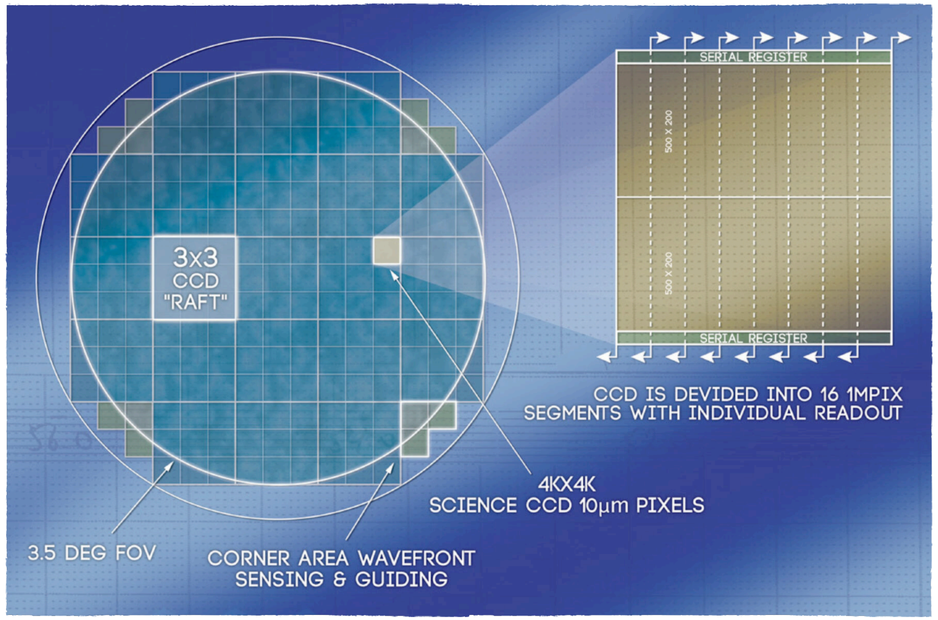
Credit: Rubin Observatory/NSF/AURA/S. Ma

Rubin Camera

Credit: Olivier Bonin/SLAC National Accelerator Laboratory



Credit: Travis Lange, SLAC National Accelerator Lab



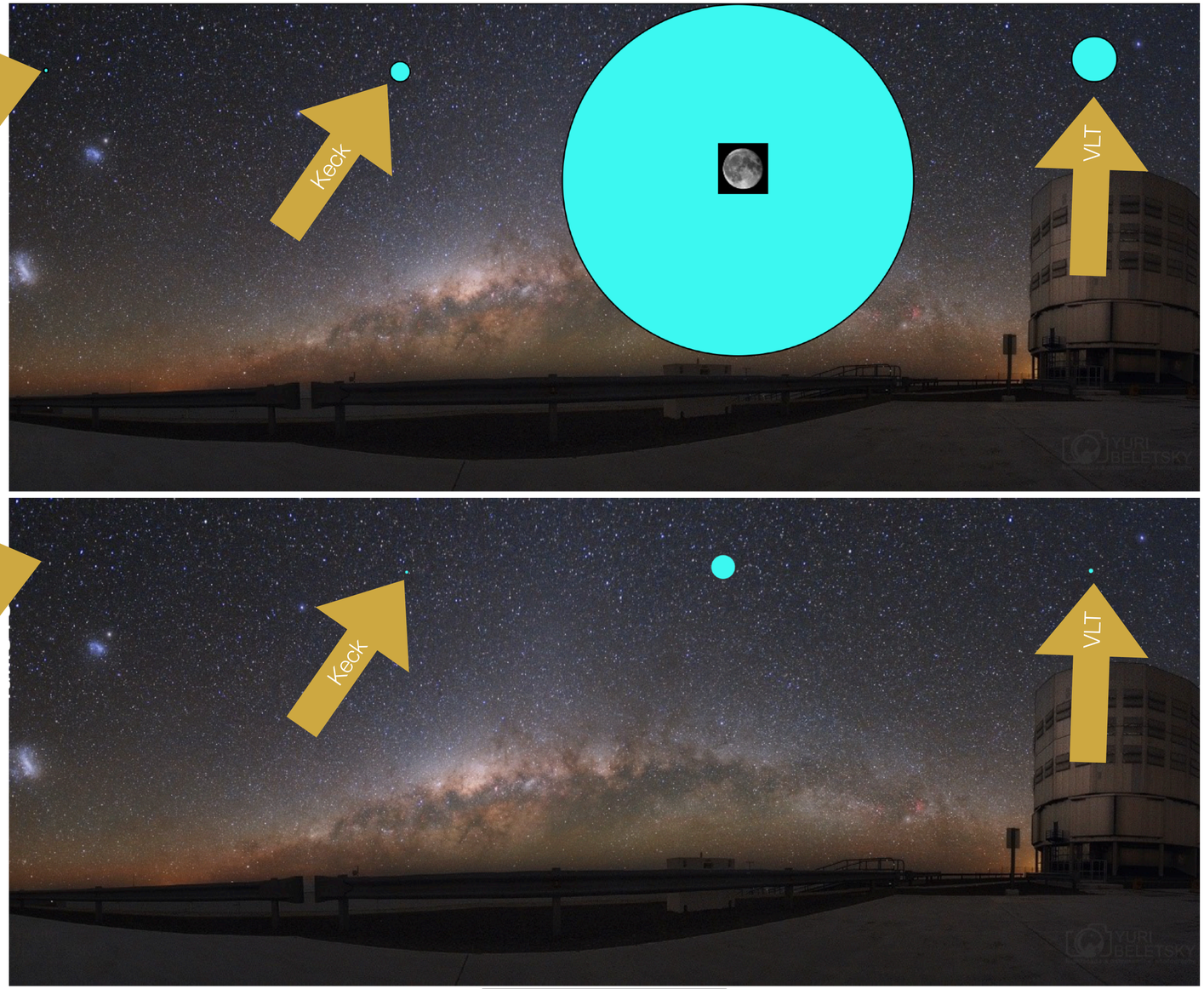
Credit: Rubin Observatory

Rubin Camera

Credit: Travis Lange, SLAC National Accelerator Lab

Rubin FoV

15x

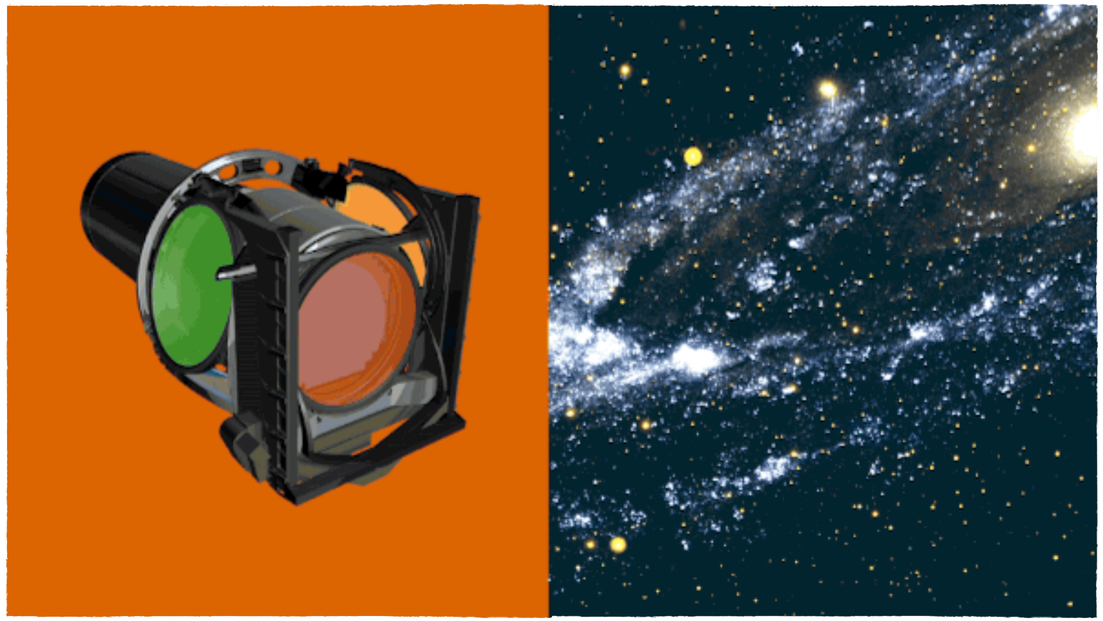


Telescope	Main mirror diameter	FoV
JWST	6.5 m	~9,7 arcmin² (NIRCam)
Hubble	2.4 m	~16 arcmin² (ACS/WFC)
Keck	10 m	~83,5 arcmin² (DEIMOS: 16,7' × 5')
Rubin	8.4 m	~9,6 deg² (LSST Camera)
VLT	8.2 m	1 deg² (OmegaCAM: 1° × 1°)

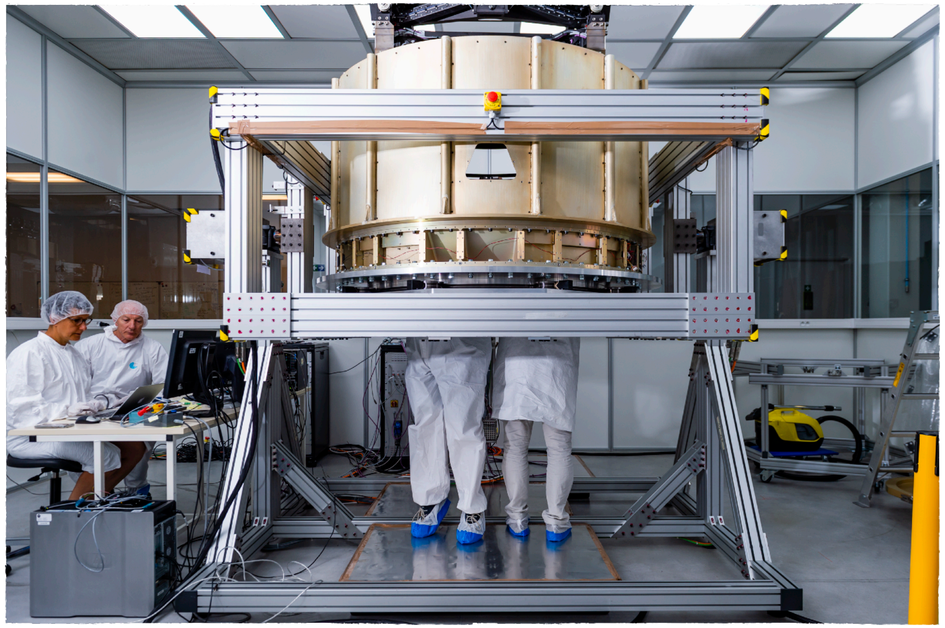
Credit: P. Antilogus

Rubin Filter Exchange

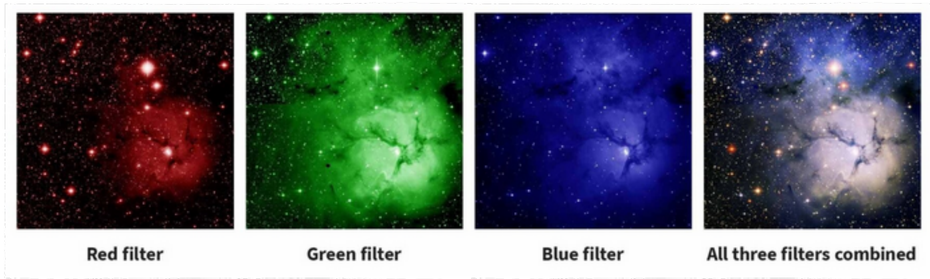
- Three subsystems built by three IN2P3 laboratories (LPC, CPPM, and LPHNE)
- The control and command system developed by the APC
- Assembled at the LPNHE



Credit: SLAC National Accelerator Laboratory



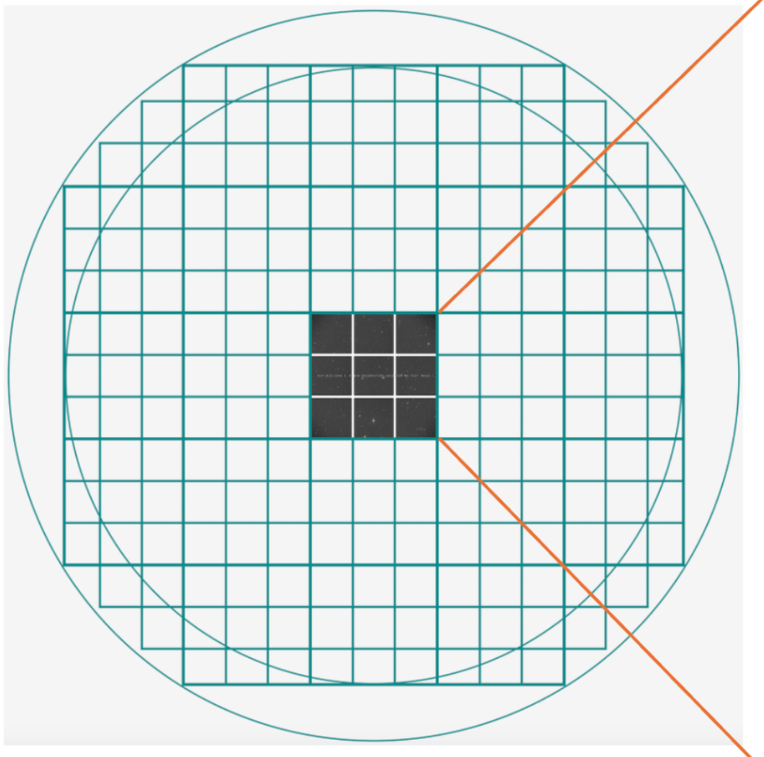
Credit: RubinObs/NSF/AURA



Credit: Olivier Bonin/SLAC National Accelerator Laboratory

Rubin Commissioning Camera

LSST Camera field of view



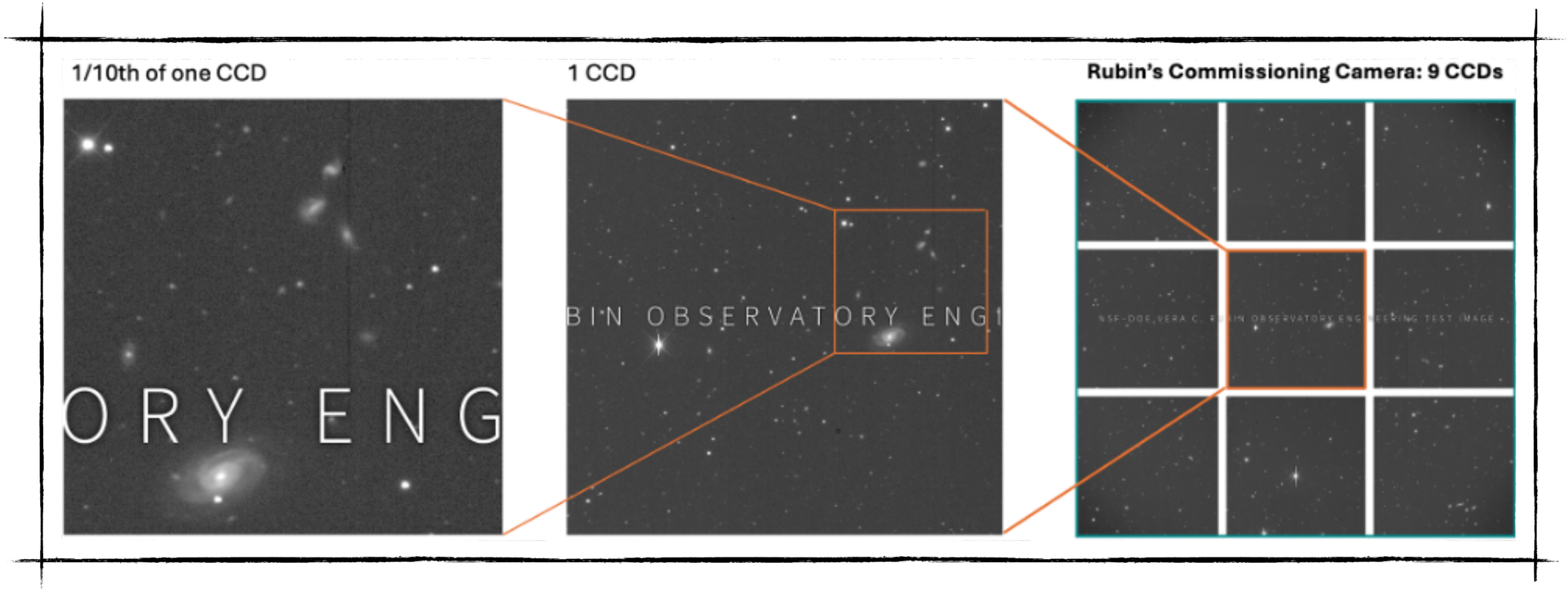
189 science 4k x 4k CCDs, 3200 Megapixels

Commissioning camera: 9 CCDs with 144 Megapixels

One of the first Rubin on-sky images from Oct 24, 2024

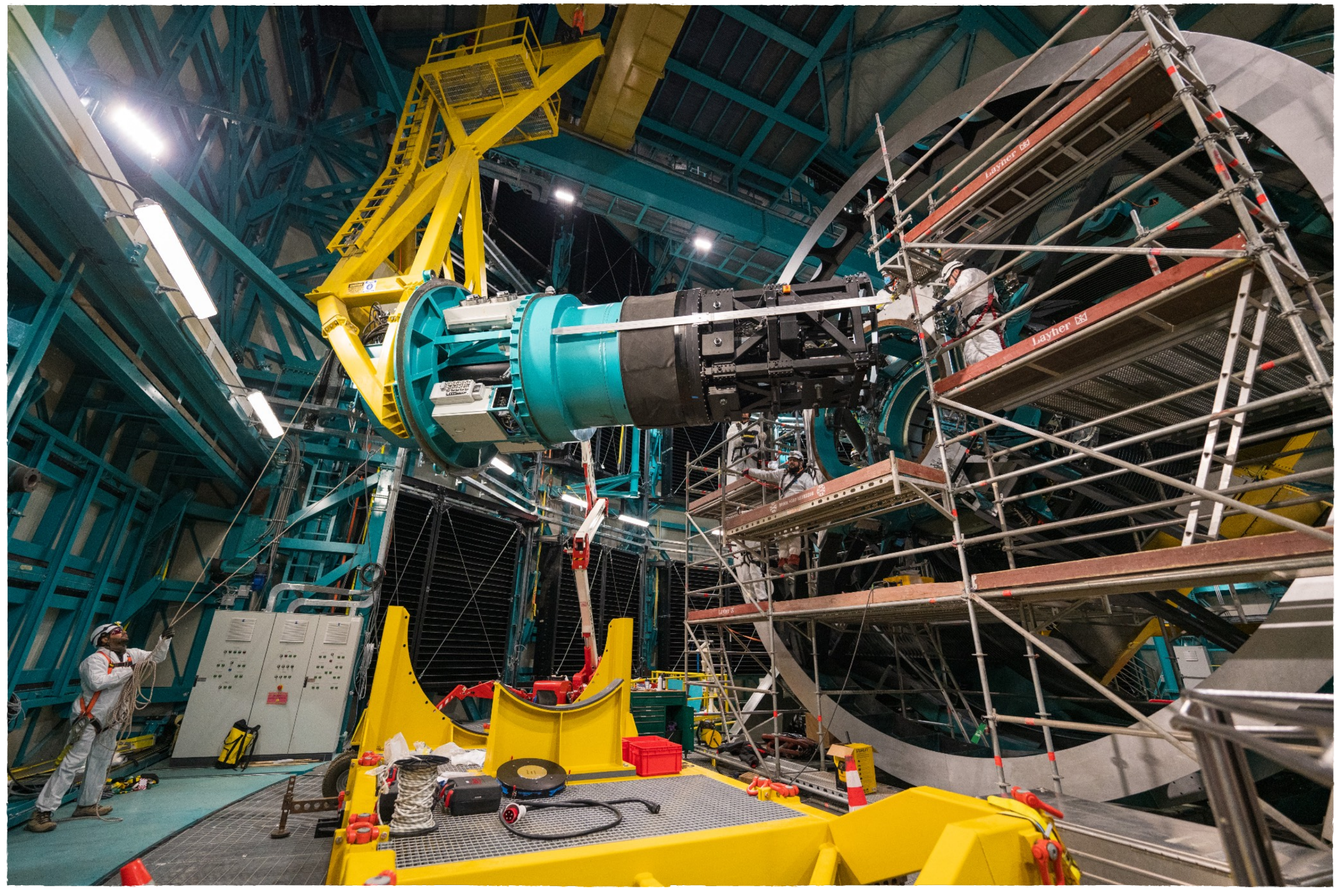
Credit: RubinObs/NOIRLab/SLAC/NSF/DOE/AURA

Rubin Commissioning Camera



Credit: RubinObs/NOIRLab/SLAC/NSF/DOE/AURA

Rubin Commissioning Camera



Credit: Rubin Observatory/NSF/AURA/H. Stockebrand

Credit: RubinObs/NOIRLab/SLAC/NSF/DOE/AURA

Rubin at CC-IN2P3

3

Rubin at CC-IN2P3

The Legacy Survey of Space and Time (LSST)

- Composition of the Universe: Dark Energy and Dark Matter
- The Solar System inventory
- The changing sky: variable and transient objects observation
- The Milky Way cartography

What CC-IN2P3 is preparing for

- 6.4 GB per exposure (compressed)
- 2500 images (science + calibration) per night (20 TB)
- 5 PB each year
- # of files: 500k per night, 150M per year, 1.5B files for the entire survey
- Up to 10 millions alerts per night
- After 10 years of operation:
  - 6 millions of raw images
  - 0.5 EB of generated data in total
  - 15 PB of catalogs in database

What CC-IN2P3 is committed to deliver to Rubin

- Annual processing of 40% of the cumulated raw data
- Long term storage of raw data and selected published data sets (on tape)

Compute Element

- ARC CE and Slurm HTC platform

Butler Repository

- PostgreSQL and CephFS → dCache

Software Distribution

- CernVM-FS

Workload Management

- PanDA (multi-site)
- Parsl (local processing)

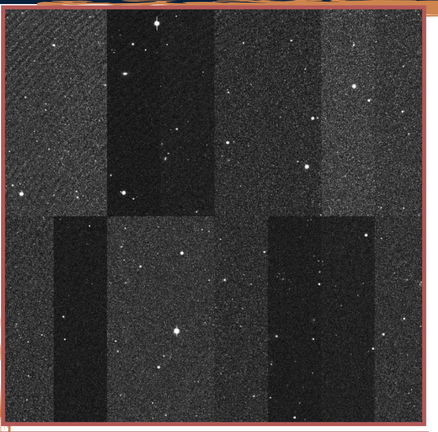
Inter-site Data Replication

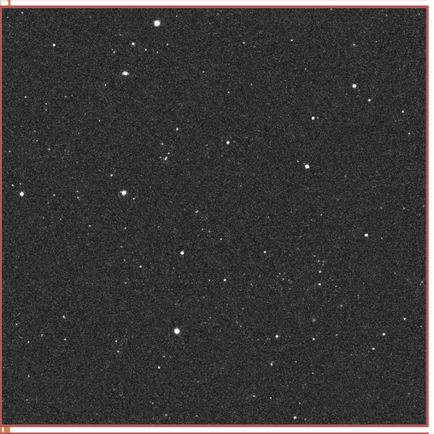
- Rucio-compatible storage element
- FTS3-compatible storage endopoint
- Kafka consumers

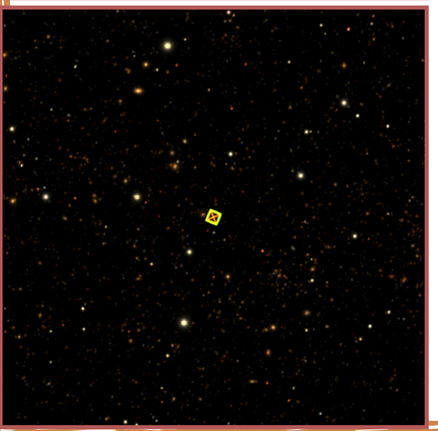
Data Access

- Catalog Database
- Analysis Platform

Rubin at CC-IN2P3





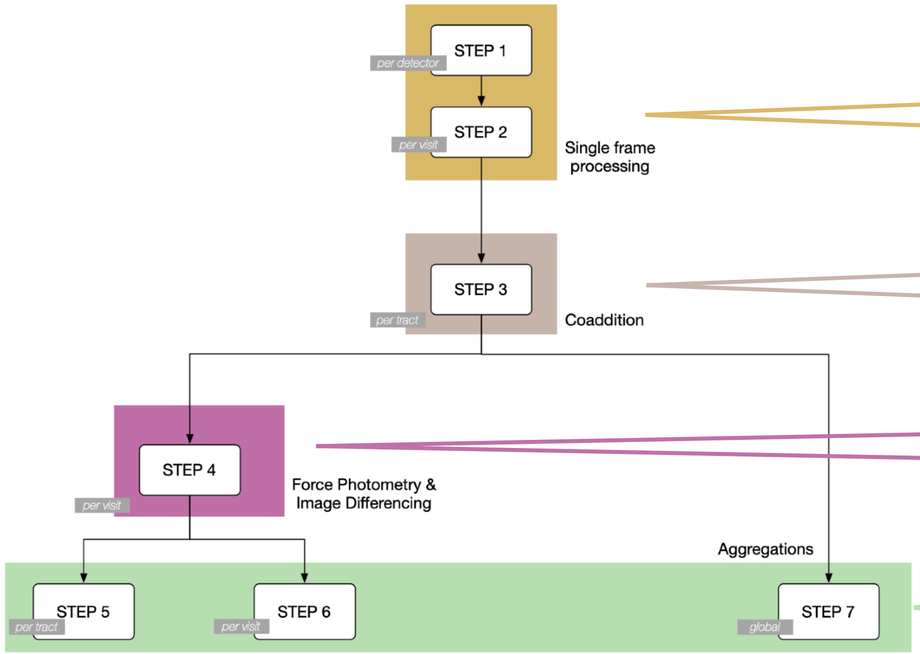


Rubin Data Preview 0.2 processing campaign in 2022

- Processing of simulated Rubin images generated by the Dark Energy Science Collaboration (LSST DESC)
- Equivalent to 5 years of data taking on 300 deg2, or 0.5% of the full survey
- 50 TB of input datasets

- ~ **60 millions** of tasks executed
- more than **2 millions CPU** hours
- **3.3 PB** of data generated
- Butler registry database of **337 GB**

- Set of ~80 kinds of pipeline tasks, grouped in 7 steps



Process raw images and calibrate them to remove any artifact from instrument

Merge images together ("coaddition") to improve signal/noise

Measure differences between images to detect changes

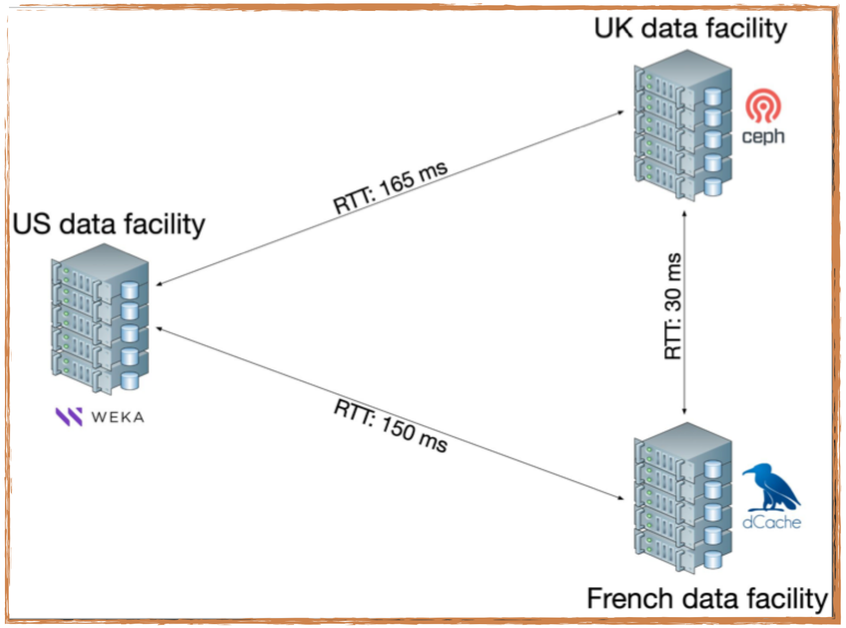
Detect objects, measure their properties and generate catalogs

- Each step corresponds to a given workflow which involves execution of many Rubin pipeline tasks

Rubin at CC-IN2P3

Rubin Data Replication

- Inter-facility Rubin data replication to use open-source software: Rucio and FTS3
  - Rucio
    - Replica catalog: Where does my data live?
    - Data policy enforcement: How many copies of the data, and where?
    - Transfer scheduling: Arranges to satisfy your policies with external services!
  - FTS3
    - Executes transfers scheduled externally on behalf of Rucio
    - Highly configurable for tuning handling of many transfers to many sites
- Kafka and Rubin-specific tools
  - To identify data which needs replication among the facilities (e.g. exclude intermediates) and ask Rucio to replicate it
  - To trigger actions at each facility to timely ingest replicated data into the local data butler repository



- Ongoing work to understand:
  - How small files (by HEP standards) impact the data transfer
  - To demonstrate our capability of routinely perform inter-facility data exchange

Rubin at CC-IN2P3

The analysis platform

- Objectives:
  - To provide researchers with a platform to easily access/analyze survey data (images and catalogs)
  - To integrate it with the CC-IN2P3 (e.g authentication, $HOME and other storage systems, …) for a smooth transition between environments
  - To deploy a scalable and resilient platform
- Technology:
  - **Kubernetes** 
  - Open source development :
    - https://github.com/lsst-sqre
    - https://www.lsst.io/
- Two main components:
  - **Qserv**: Rubin's astronomical catalog database
  - **Rubin Science Platform** (RSP): the interactive analysis platform

Qserv

- Developed by SLAC with contributions from IN2P3
- Shared-nothing Massively Parallel Processing Relational Database
- Spherical partitioning with overlap, sciSQL (UDF)
- Shared scans (concurrent query load)
- Data Replication
- 100 % Open Source

The Rubin Science Platform

- Web-based environment for interactive data analysis
- Developed by the Rubin SQuaRE Team
- An all-in-one tool allowing easy and quick data access/analysis integrating:
  - catalogs and tables viewer
  - image viewer and analyzer
  - advanced programmatic analysis with LSST Python stack via Jupyter notebooks
  - gateway to Qserv catalogs (interoperability) for IVOA tools

Conclusions

4

Conclusions

- Astrophysical, astroparticle, and cosmological projects are very important to CC-IN2P3
- The Rubin Observatory has started the commissioning phase
- A revolutionary camera with a large FoV will produce a huge amount of data in the next 10 years
- CC-IN2P3 is committed to process 40% of the cumulated raw data and to store a full copy of raw data
- HEP technologies used by Rubin for data transfer and data processing