# Addressing Computing Challenges for Imaging Air Cherenkov Telescopes

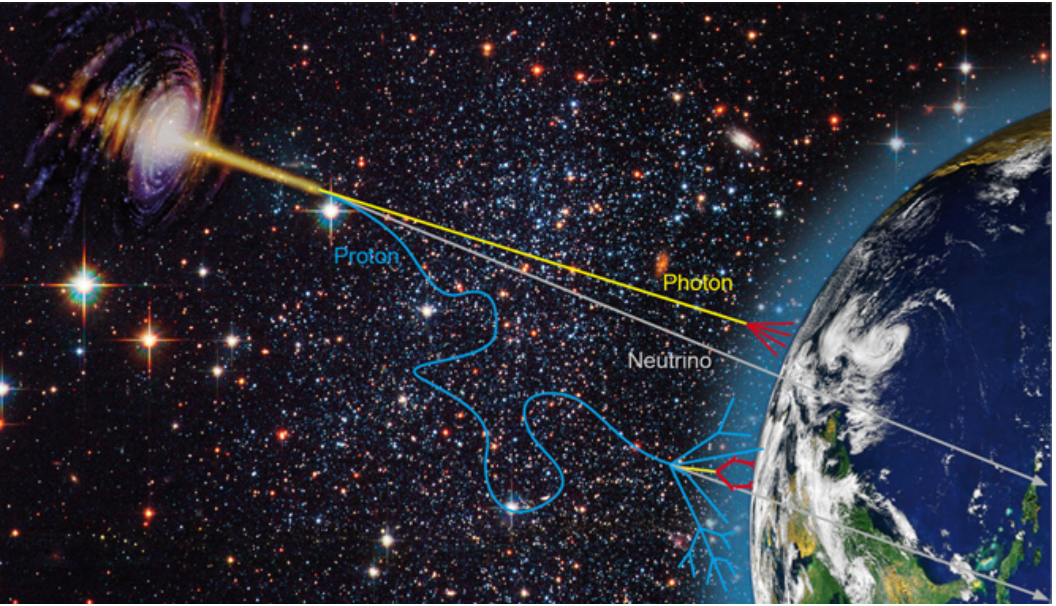
INAF USCVIII - Calcolo Critico

**Ciro Bigongiari INAF-OAR & ASI-SSDC**

# Outline

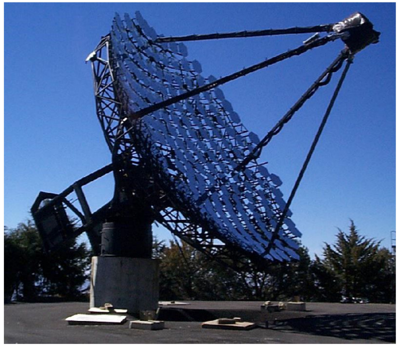
- Gamma-ray astronomy from ground
- IACT technique in a nutshell
- CTA & ASTRI
- CTA Data flow – Data flux
- CTA Computing model
- ASTRI Data flow
- ASTRI Computing resources
- ASTRI & CTA Simulations

# Gamma-ray astronomy from ground




**VHE – Very High Energy (E ≳ 30 GeV)**

**VHE Gamma-rays** produced in cosmic objects (AGNs,SNRs,GRBs...) can reach the Earth unscrambled by magnetic fields.
They interact with Earth atmosphere producing light that can be detected by **IACT (Imaging Air Cherenkov Telescopes).**

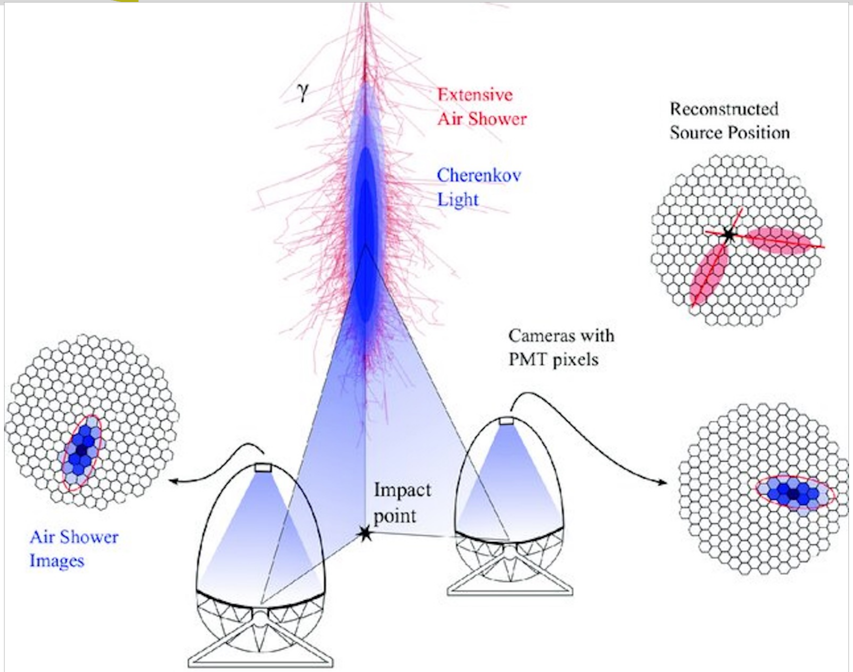


Whipple pioneering Cherenkov telescope

Array of Cherenkov telescopes outperform single telescopes

IACT arrays in operation

# IACT Technique in a nutshell




Courtesy of MAGIC Collaboration

**IACT – Imaging Air Cherenkov Telescopes**

- VHE cosmic particles interacts with atmospheric nuclei
- A cascade of subnuclear particles develops
- Ultra-relativistic charged particles emit Cherenkov light
- Cherenkov light is focused onto PMT/SiPM cameras
- Pictures of Cherenkov flash are taken
- From the shape, size and orientation of the images the nature (gamma or hadron) of the primary particle, its incoming direction and its energy can be estimated

**IACTs takes pictures as many other telescopes do, but**

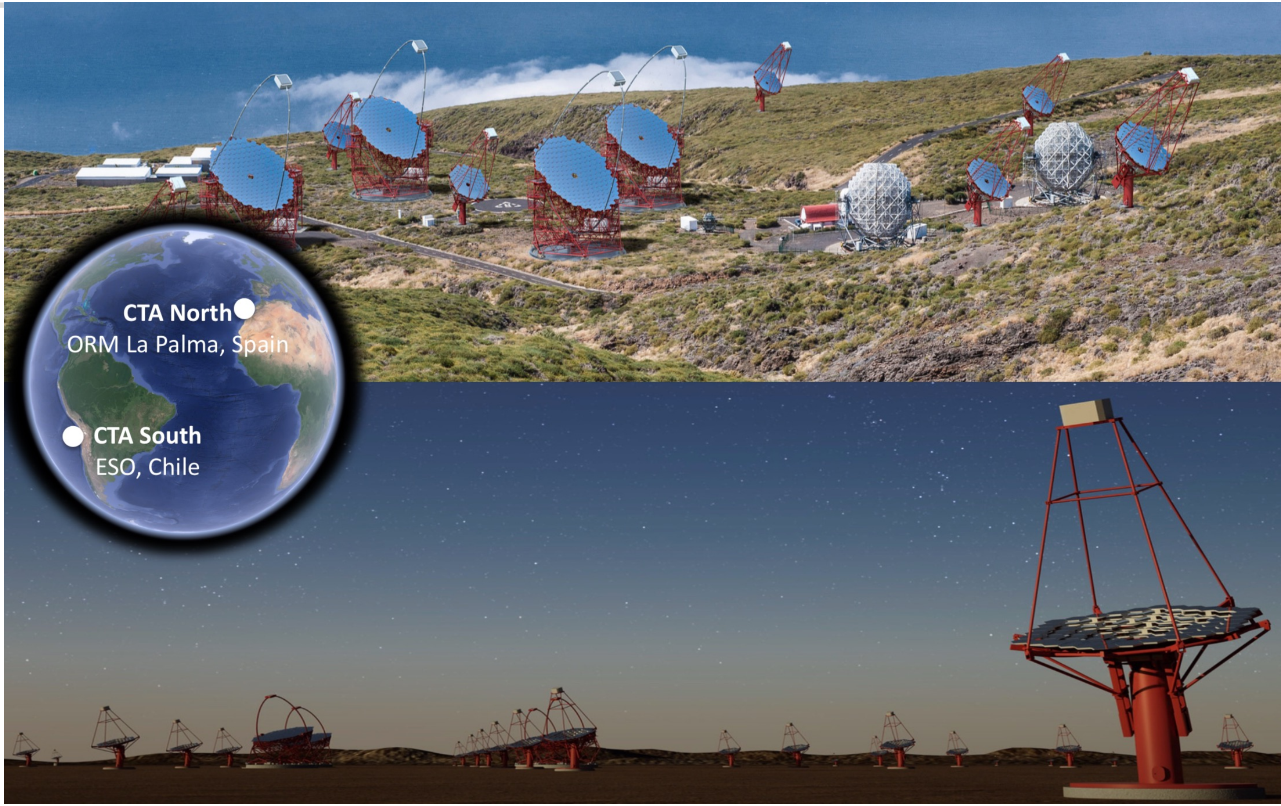
- Resolution is much lower ~2000 pixels
- Acquisition rate can be much higher ~ 1KHz

# CTA Observatory

**CTAO Northern Array**
- **4LSTs+9MSTs**
- 0,25 km2 footprint
- focus on extra-Galactic science

**CTAO Southern Array**
- **14 MSTs + 37 SSTs (+2 LSTs PNRR)**
- 3 km2 footprint
- focus on Galactic science

**Two arrays -> full sky coverage**



## ASTRI-Horn

INAF-led Project funded by Italian Ministry of Research

**Prototype of the 4-m class telescopes** developed in the framework of CTA Observatory; installed and operational since 2018 on Mount Etna (Sicily, Italy)

**First detection of a gamma-ray source** (Crab Nebula) **with a dual-mirror, Schwarzschild-Couder Cherenkov telescope** (Lombardi et al., 2020)

Dedicated to **Guido Horn D'Arturo**, precursor of the segmented astronomical mirrors technique

## ASTRI mini-array: array of 9 ASTRI telescopes

INAF-led Project with international partners:
Univ. of Sao Paulo/FPESP (Brazil), NWU (S. Africa), IAC (Spain), FGG, ASI/SSDC, Univ. of Padova, Perugia and INFN)

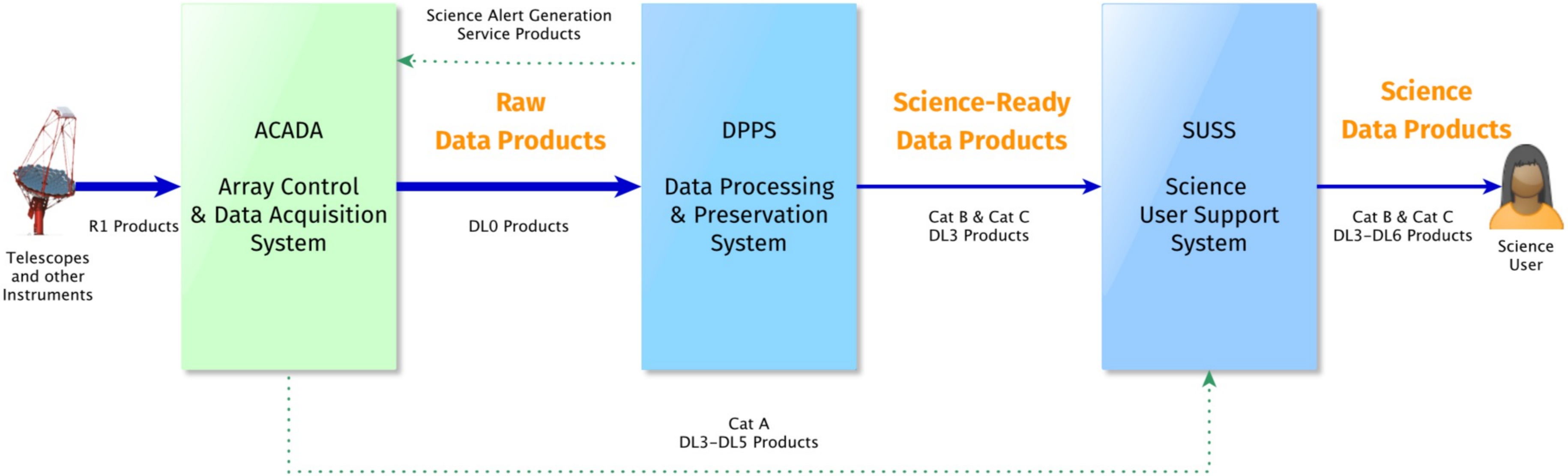
**Being deployed at the *Observatorio del Teide*** (Tenerife, Spain)
First telescope installed; first camera expected for July 2023
Two more telescopes by the end of 2023. Two more camera in spring 2024

**Fully operational in 2025**

**First 4 years → *Core Science projects***, following 4 → *Observatory*

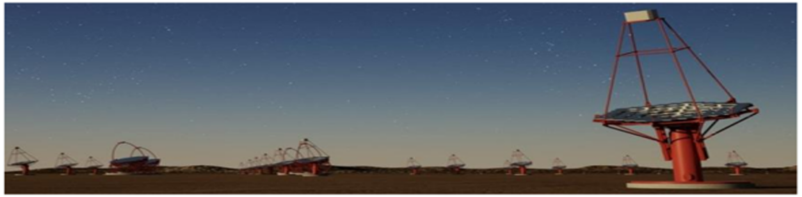
First telescope of the ASTRI mini-array installed at Teide Observatory, Tenerife

# CTAO Core Data flow




Where data are stored: **On-site Storage** | **Bulk Archive** | **Science Archive**

# CTAO – Data flux



**CTAO-North, Canarian island**

- 4 Large Size Telescopes
- 9 Mid Size Telescopes

**CTAO-South - Chile**

- 14 Mid Size Telescopes
- 37 Small Size Telescopes

Not including 2 PNRR LSTs !!!

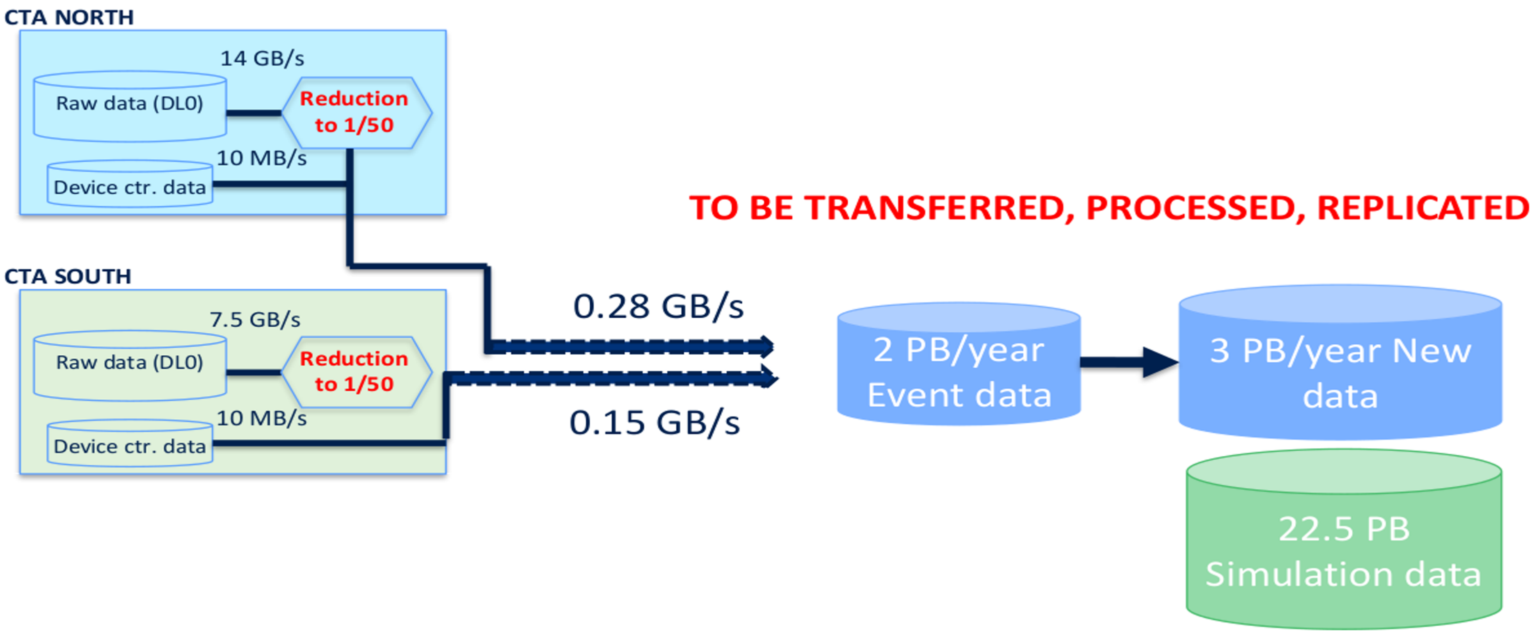
- ❑ 1314 observation hours per year
- ❑ 20% monitoring and service data

14 GB/s → 67 PB/year

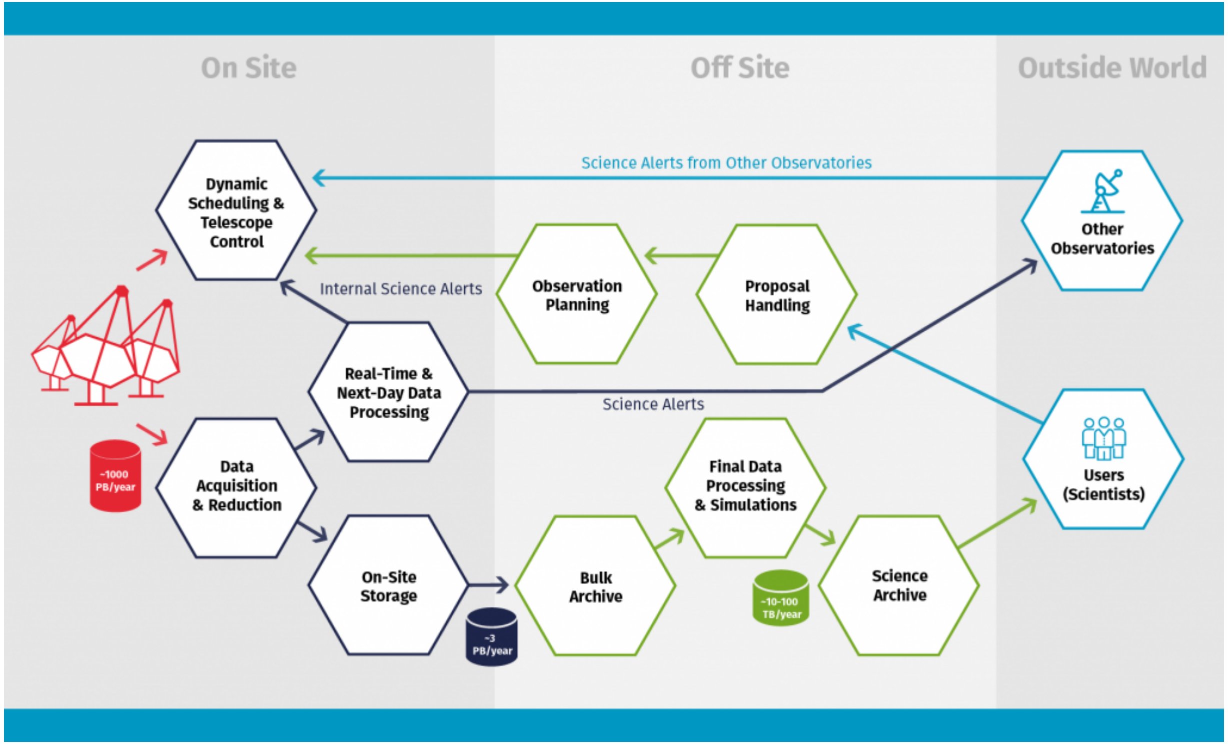
7.5 GB/s → 36 PB/year

# DATA Compression

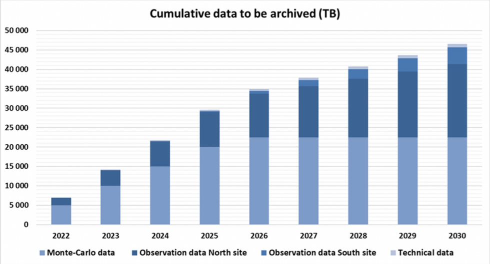
## Required Data Volume Reduction ratio: 50



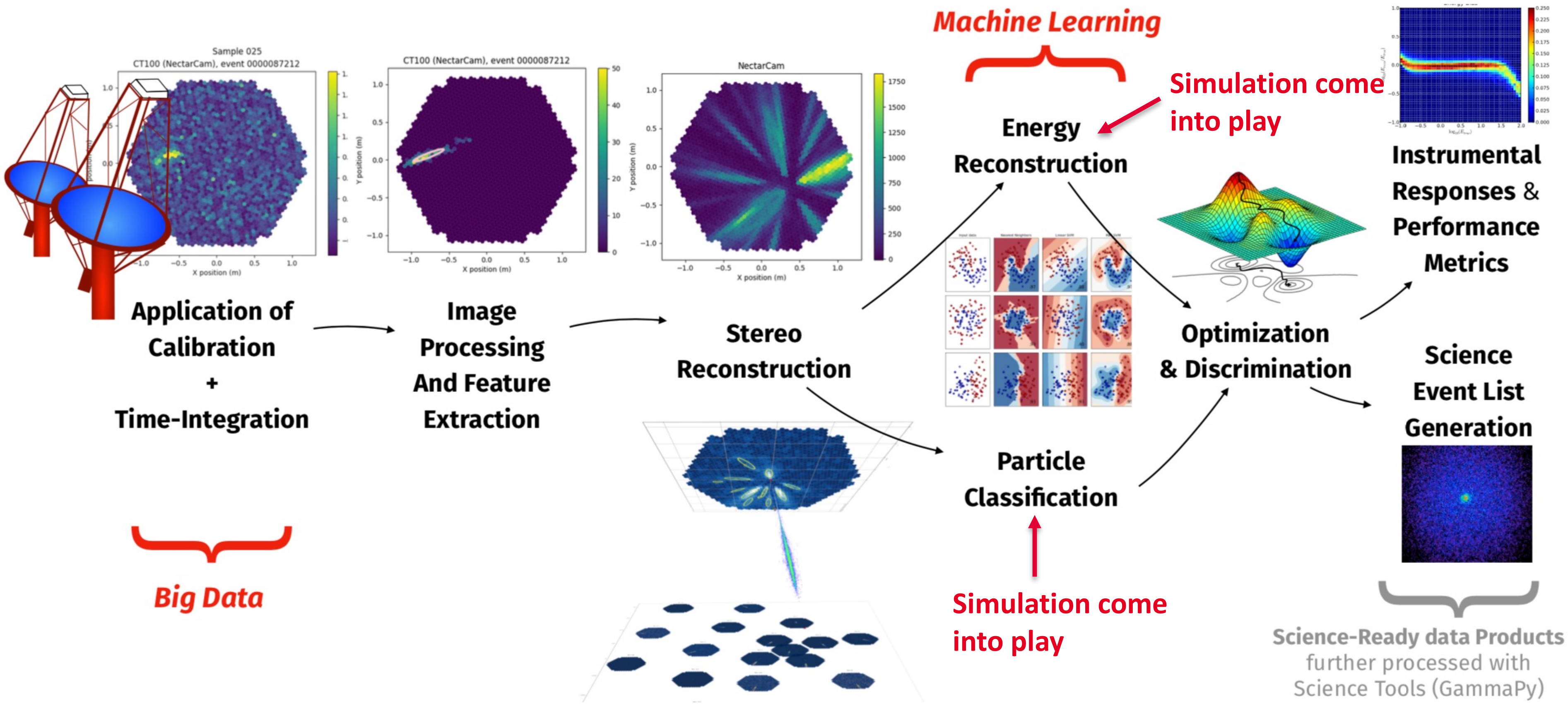
# CTAO Data flow



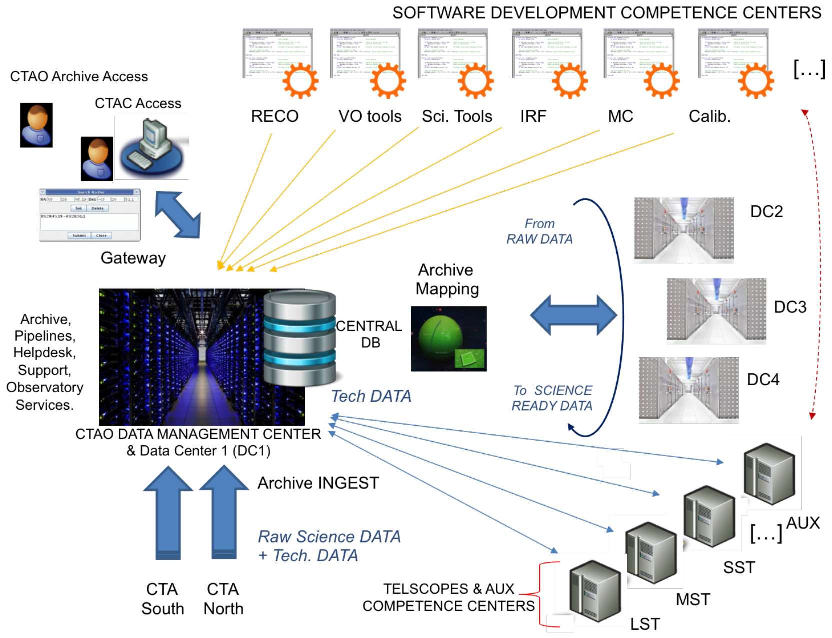

The telescopes on the two CTA array sites will produce **hundreds of petabytes (PB) per year of** raw data. They will be written after compression to a few PB per year to the off-site data centres for further processing and storage. Additionally, **a few tens of PB of simulated data** will be produced and processed.



# Data reduction & analysis

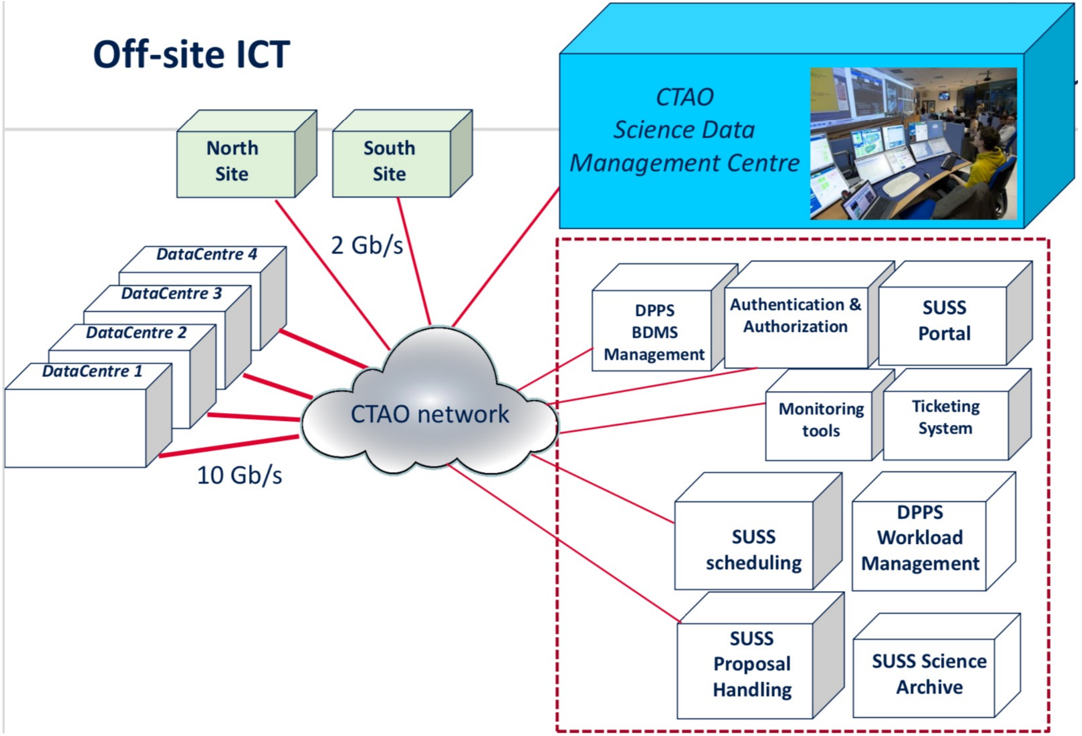


# CTAO – Computing model




- A distributed ICT infrastructure.
- Based on the data centres' IT equipment, hosting the CTA distributed applications based on DPPS, SUSS and SOSS software.
- Data Centres interconnected by Wide Area Networks (WANs) in a redundant topology (10 Gbps bandwidth).
- Connected to the two CTA array sites (CTA-North, CTA-South) for data transfer (Minimum of 2 Gbps bandwidth).
- CTA applications and WAN networks centrally controlled and monitored by the SDMC team.

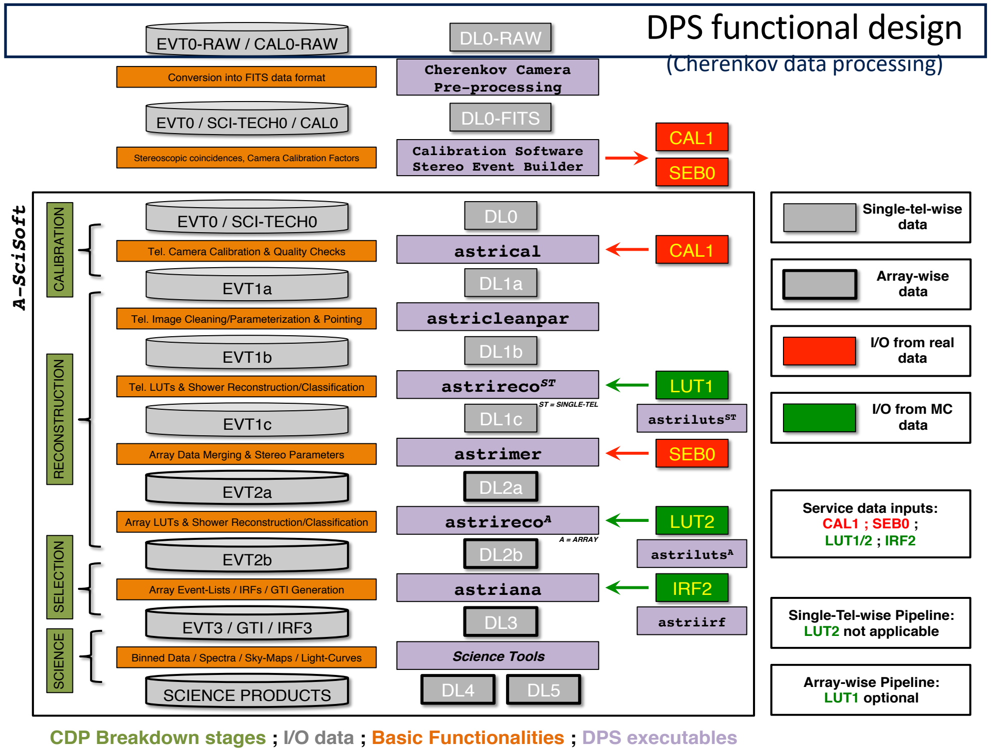
# CTAO offsite ICT infrastructure




Data pre-processed onsite will be transferred to 4 Data Centers where they will be further reduced and stored

- **PIC** in Barcelona, Spain
- **DESY** in Zeuthen, Germany
- Swiss National Supercomputing Centre (**CSCS**) in Lugano, Switzerland
- INAF/INFN in **Frascati**, Italy
- **SDMC in Zeuthen, Germany**

# ASTRI Data Flow & Data levels




- **Cherenkov Camera Pre-Processing:** all Cherenkov raw data converted in FITS data format (adopted as common I/O data format for each DPS subsystem);
- **Calibration Software:** generation of suitable calibration factors for Cherenkov data calibration;
- **Stereo Event Builder:** identification of stereoscopic Cherenkov events induced by the same extensive air shower;
- **Cherenkov Data Pipeline:** end-to-end Cherenkov data reduction (`A-SciSoft` software package) up to high-level science-ready data (DL3) and standard scientific products (DL4/5)
  - Calibration (DL0 → DL1)
  - Reconstruction (DL1 → DL2)
  - Selection (DL2 → DL3)
  - Science* (DL3 → DL4/5)
    - *external Science Tools: `Gammapy` and `ctools`

- 2 end-to-end scientific data processing levels:
- **short-term** pipeline (for scientific quick-look)
- **long-term** pipeline (for consolidated products)

# ASTRI-MA – Data flux

| Archive/DB Units | GB/day MAX | GB/day AVG | TOT MAX (AVG) [TB/yr] |
|---|---|---|---|
| Bulk Archive (only RAW) | 5117[2] | 558[3] | 604 (91)[4] |
| Bulk Archive (DL0 FITS + pipe products)[5] | 15680 | 1710 | 1853 (278) |
| Science Archive | 250 | 200 | ~35 (~25) |
| Swap-tmp Loc.Repo[6] | | | 200 |
| Simulation Archive (MC)[7] | | | 100 |
| Quality Archive | 33 | 24 | 3.9 (3.9) |
| Log / Monitor / Alarm Archive | 54 | 27 | ~20 (~10) |
| System Configuration DB | 5 | 4 | ~0.6 (~0.5) |
| CALDB | | | 0.2-0.4 (TBD) |
| Performance DB | | | 0.5-1.0 (TBD) |
| Interferometry Instrument (SI3) | | | 1200 (??) |
| **hot-storage TOTALS:** | **5405** | **786** | **~640 (~120)** |
| **cold-storage Backup** | **~16000** | **1737** | **~1873 (~290)** |

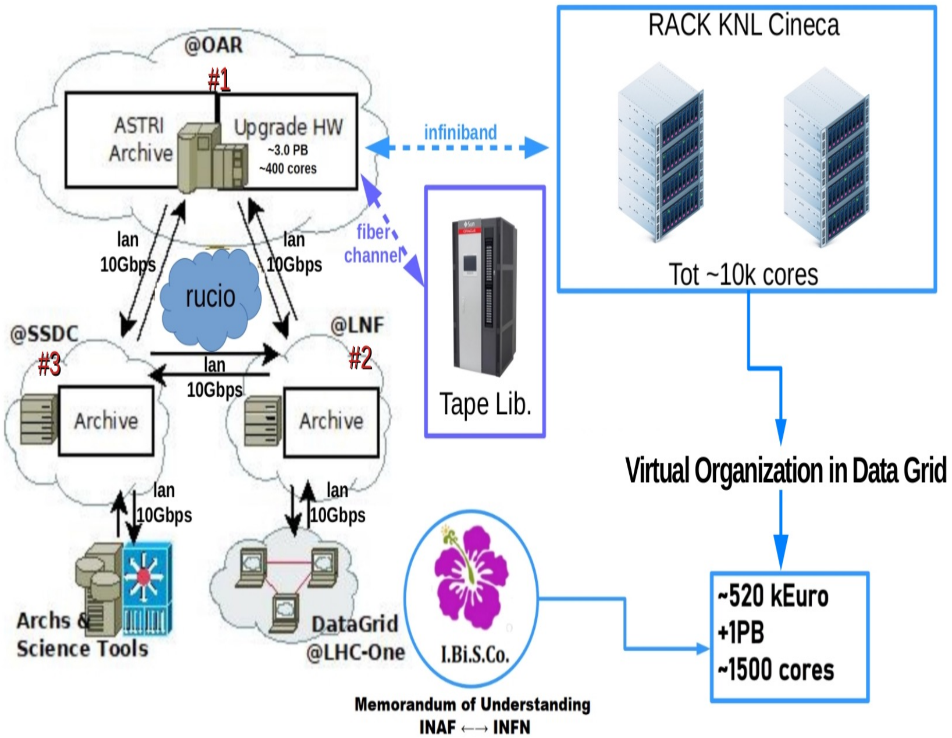
Considering:

- Packet dimension **13kB x 9 telescopes**
- **Worst Case**:
  - **1.0 kHz** trigger rate
  - **11hr** acquisition/dd
- **Average Case**:
  - **150kHz** trigger rate
  - **8hr** acquisition/dd

Optimal **HD Space** → ~**0.75 PB** (hot+MC )/y
Optimal **Tape Space** → ~**1.15 PB** (cold + hot+MC )/y

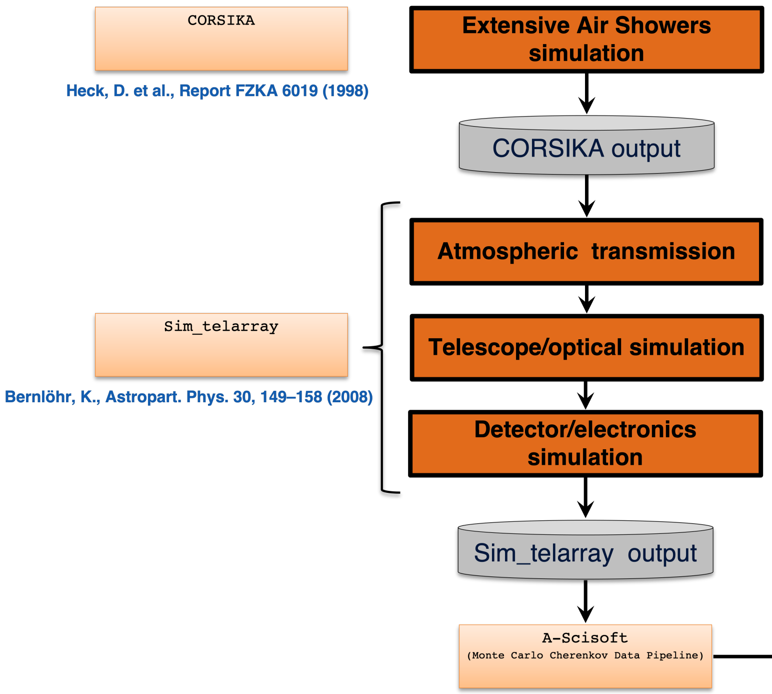
Not including interferometry data !

# ASTRI – Computing resources




- Storage:
  - 3PB @ OAR
  - 1PB @ LNF
  - Tape library for cold storage
- CPU
  - 400 cores @ OAR Ready
  - 1500 cores @ LNF ~Ready
  - ~10000 cores

# ASTRI – Simulation Chain

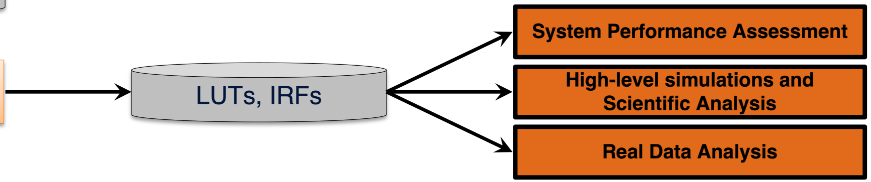



**Simulation of:**
- gamma and background (p, $e^-$, He,…) species
- Zenith and Azimuth angles
- Telescope positions in the array
- Geomagnetic field (@ array site)
- Atmospheric density profile

- Atmospheric extinction (🡐 MODTRAN®)

- Ray Tracing (🡐 Mirror reflectivity)

- SiPM response (🡐 PDE Single Photo Electron Response)
- Camera Electronics (🡐 SlowShaper)
- Trigger (🡐 FastShaper)

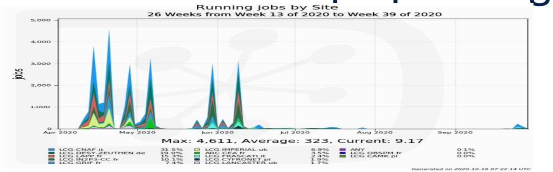


# ASTRI-MA simulation

## Second massive production of simulated events for ASTRI-MA: ASTRI-MA-Prod2

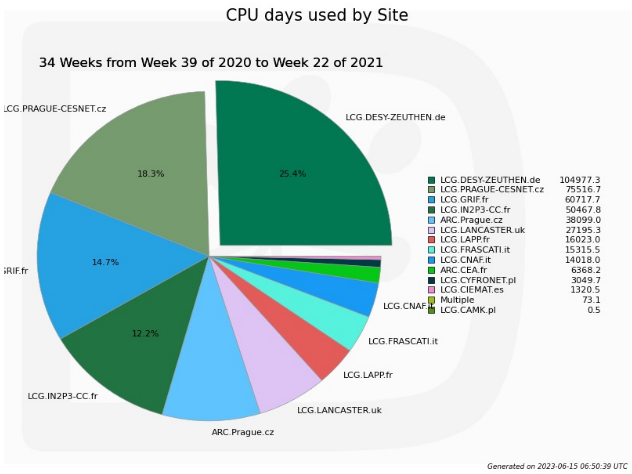
| PARTICLE | Emin [TeV] | Emax [TeV] | Spectral Slope | IPmax [m] | View Cone [deg] | Zenith [deg] | Azimuth [deg] | Number of Simulated Runs | Number of Simulated Showers per Run | Number of Simulated Showers |
|---|---|---|---|---|---|---|---|---|---|---|
| GammaPointlike | 0.1 | 330 | -1.5 | 2000 | 0 | 20 | 180 | 2 × 2000 | 10000 | $4 \times 10^7$ |
| GammaDiffuse | 0.1 | 330 | -1.5 | 2400 | 10 | 20 | 180 | 2 × 10000 | 20000 | $4 \times 10^8$ |
| Electron | 0.1 | 330 | -1.5 | 2400 | 10 | 20 | 180 | 2 × 5000 | 20000 | $2 \times 10^8$ |
| Proton | 0.1 | 600 | -1.5 | 2400 | 10 | 20 | 180 | 2 × 50000 | 20000 | $2 \times 10^9$ |
| **TOTAL** | | | | | | | | **134000** | | **$2.64 \times 10^9$** |

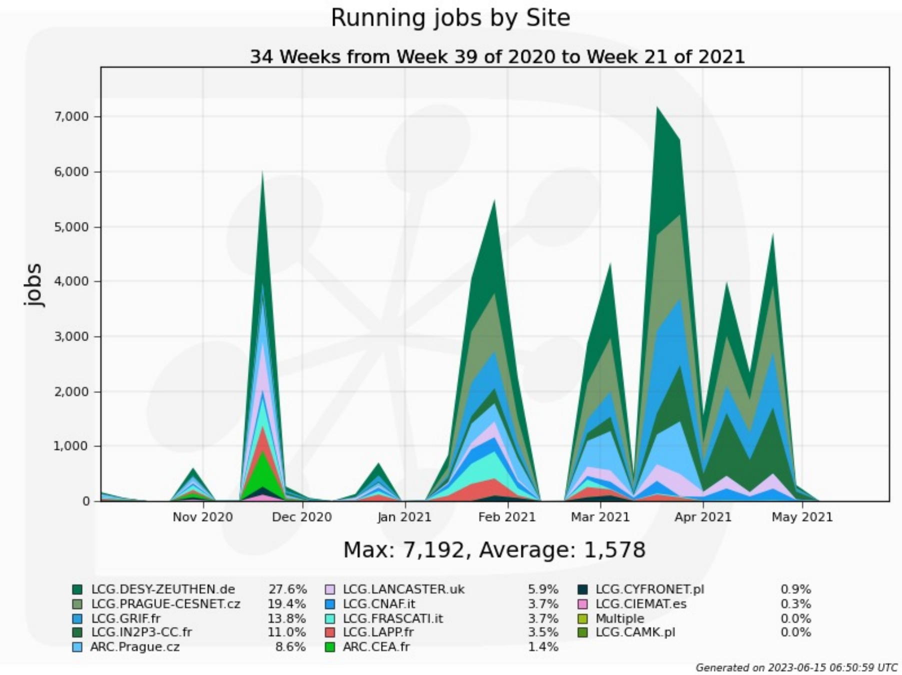
Just one telescope pointing




| Production | CORSIKA storage [TB] | CORSIKA CPU [$10^6$ HS06hours]$^2$ | Sim_telarray storage [TB] | Sim_telarray CPU [$10^6$HS06hours] |
|---|---|---|---|---|
| ASTRI-MA-Prod1 | 34 | 16.3 | 1.0 | 1.6 |
| ASTRI-MA-Prod2 | 62 | 31.2 | 1.9 | 3.0 |

# CTA – Simulations Prod5b




## Prod5b – last massive CTA production

CPU = 127 M HS06 hours
Storage = 1700 TB.

HS06 = HEP-Spec06 Benchmark

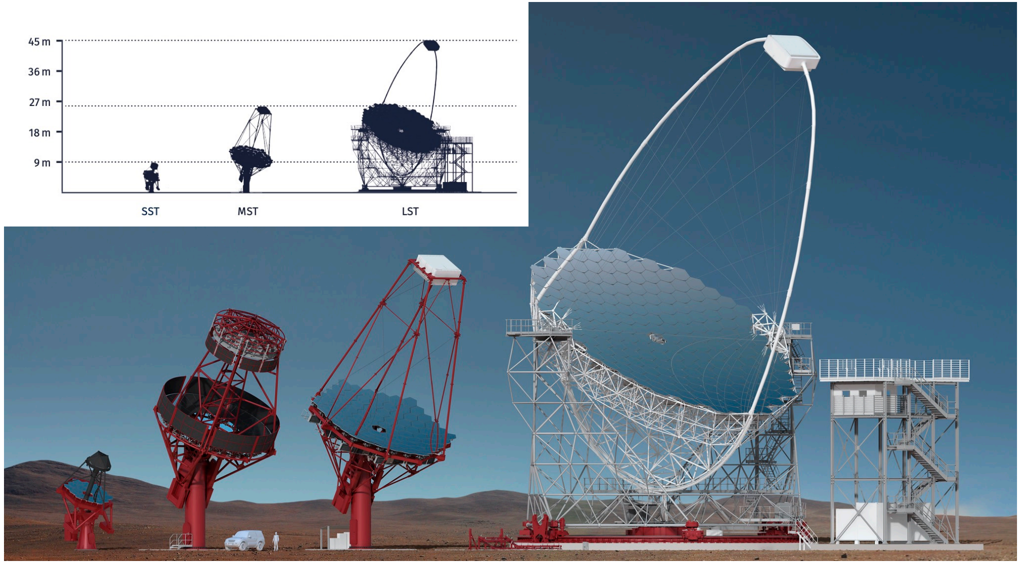
# Take home messages

- New generation IACT arrays are ready to go
  - LST1 taking data and performing science at CTAO-North site
  - First ASTRI-MA telescope deployed at Teide site
  - CTAO ERIC ready in September 2023
  - ASTRI-MA fully operational in 2025
- Expected data flows ~100PB/1PB per year for CTA/ASTRI
  - To be processed
  - To be stored
- INAF deeply involved in both projects

# Backup slides

BACK-UP

# Characteristics of CTA telescopes




Three different-size telescopes, large size telescopes LST, medium size telescopes MST, and small size telescopes SST. Two different kinds of MST are under development.

| | Large-Sized Telescope (LST) | Medium-Sized Telescope (MST) | | | Small-Sized Telescope (SST) |
|---|---|---|---|---|---|
| | | FlashCam | NectarCam | SCT | |
| Required energy range | 20 GeV – 3 TeV | 80 GeV – 50 TeV | | | 1 TeV – 300 TeV |
| Energy range (in which subsystem provides full system sensitivity) | 20 GeV – 150 GeV | 150 GeV – 5 TeV | | | 5 TeV – 300 TeV |
| Number of telescopes (alpha configuration) | 0 (South)<br>4 (North) | 14 (South) | 9 (North) | | 37 (South)<br>0 (North) |
| Optical design | Parabolic | Modified Davies-Cotton | | Schwarzschild-Couder | Schwarzschild-Couder |
| Primary reflector diameter | 23.0 m | 11.5 m | | 9.7 m | 4.3 m |
| Secondary reflector diameter | -- | -- | | 5.4 m | 1.8 m |
| Effective mirror area (including shadowing) | 370 $m^2$ | 88 $m^2$ | | 41 $m^2$ | >5 $m^2$ |
| Focal length | 28 m | 16 m | | 5.6 m | 2.15 m |
| Total weight | 114 t | 82 t | | 80 t | 17.5 t |
| Field of view | 4.3 deg | 7.7 deg | 7.9 deg | 7.6 deg | 8.8 deg |
| Number of pixels in Cherenkov camera | 1855 | 1758 | 1855 | 11328 | 2048 |
| Pixel size (imaging) | 0.1 deg | 0.18 deg | 0.18 deg | 0.067 deg | 0.16 deg |
| Photodetector type | PMT | PMT | PMT | SiPM | SiPM |
| Telescope readout event rate (before array trigger for MSTs and SSTs) | >7.0 kHz (after LST array trigger) | >6 kHz | >7.0 kHz | >3.5 kHz | >0.6 kHz |
| Positioning time to any point in the sky (>30° elevation) | 20 s | 90 s | | | 90 s |
| Pointing precision | <14 arcseconds | <7 arcseconds | | <10 arcseconds | <7 arcseconds |
| Observable sky | Any astrophysical object with elevation > 24 degrees | | | | |

# Expected Data acquisition rates

According to detailed Monte Carlo simulations the expected event rates for CTAO telescopes are:

- LSTs **15 KHz**
- MSTs 9 KHz
- SSTs 0.6 KHz

~2000-pixel cameras
Readout of roughly 60-100 ns with 0.25-1 GHz sampling

ASTRI-MA telescopes 0.15 KHz
2368-pixels cameras
Just two float per pixel

# Data levels description

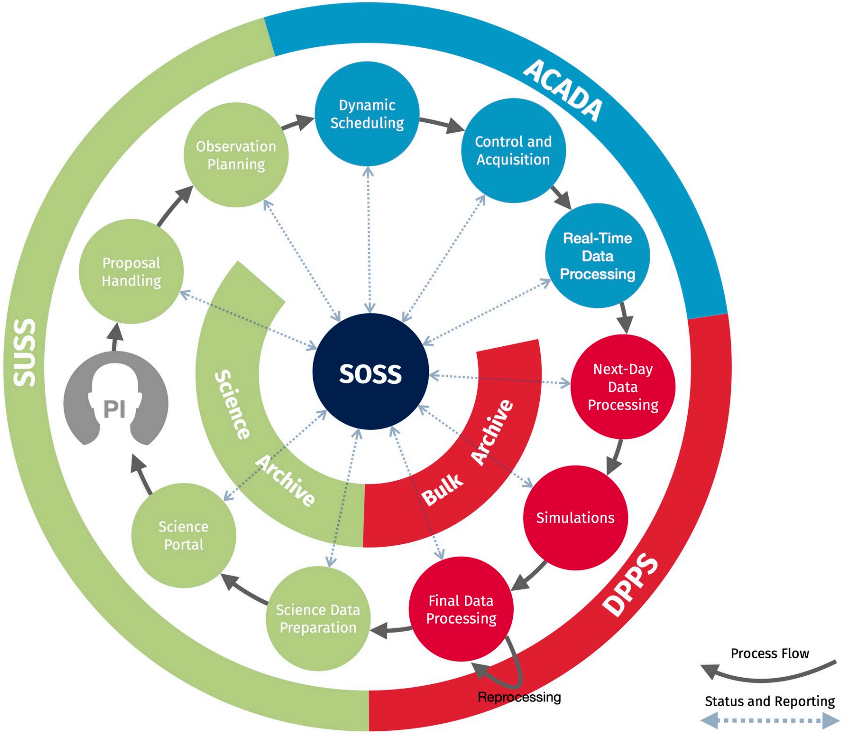
- **Level 0 (DL0)**: raw data from the hardware/software data acquisition components that are permanently archived;
- **Level 1 (DL1)**: telescope-wise reconstructed data (*reconstructed shower parameters per telescope*). Specific to ASTRI data model, the following sub-data levels are defined:
  - Level 1a (DL1a): telescope-wise calibrated data;
  - Level 1b (DL1b): telescope-wise cleaned and parameterized data (*telescope-wise image parameters*);
  - Level 1c (DL1c): telescope-wise fully reconstructed data (*telescope-wise energy, arrival direction, particle identity discrimination parameters per telescope*)
- **Level 2 (DL2)**: array-wise reconstructed data (*reconstructed shower parameters per event*). Specific to the ASTRI data model, the following sub-data levels are defined:
  - Level 2a (DL2a): array-wise merged data (*array-wise event parameters*);
  - Level 2b (DL2b): array-wise fully reconstructed data (*array-wise energy, arrival direction, particle identity discrimination parameters per event*)
- **Level 3 (DL3)**: reduced data (*selected list of events plus corresponding instrument response functions*);
- **Level 4 (DL4)**: science data (*high-level scientific data products*);
- **Level 5 (DL5)**: observatory data (*legacy observatory data and catalogs*).

# CTAO Data Processing Categories

| Data Product Category | Name | Purpose | Results | Systematic Errors | Sensitivity | Responsible (DL0→DL3) | Responsible (DL3→DL5) |
|---|---|---|---|---|---|---|---|
| Category-A | Realtime | Fast alerts and operator feedback | In seconds | High | Low | ACADA | SUSS |
| **Category-B** | **Next-Day** | Proposal Monitoring | By next evening | Medium | Medium | **DPPS** | SUSS |
| **Category-C** | **Final** | Publication | By next month | Low | High | **DPPS** | SUSS |
| **Category-C+** | **Reprocessed** | Publication | A most once per year | Lower | Higher | **DPPS** | SUSS |

CTA will have at least 3 data categories, likely 4

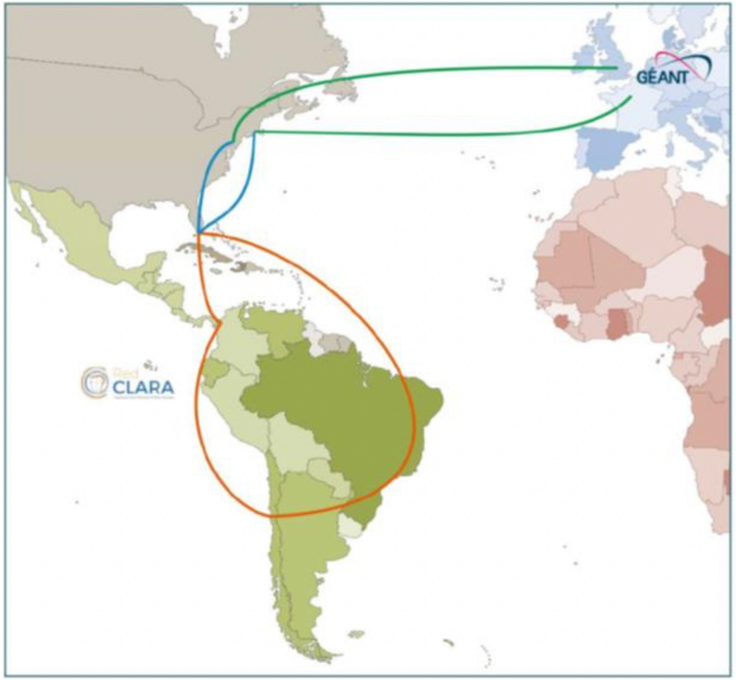
# CTAO – Archives




- Archives are in the centre of science operations from begin to end
- Archiving of data at different levels (DL0-DL3, DL5, DL6)
- Archiving of metadata linked to the different levels, including provenance information
- Additional information:
  – Proposals and Schedules
  – Status information
  – Monitoring and engineering data

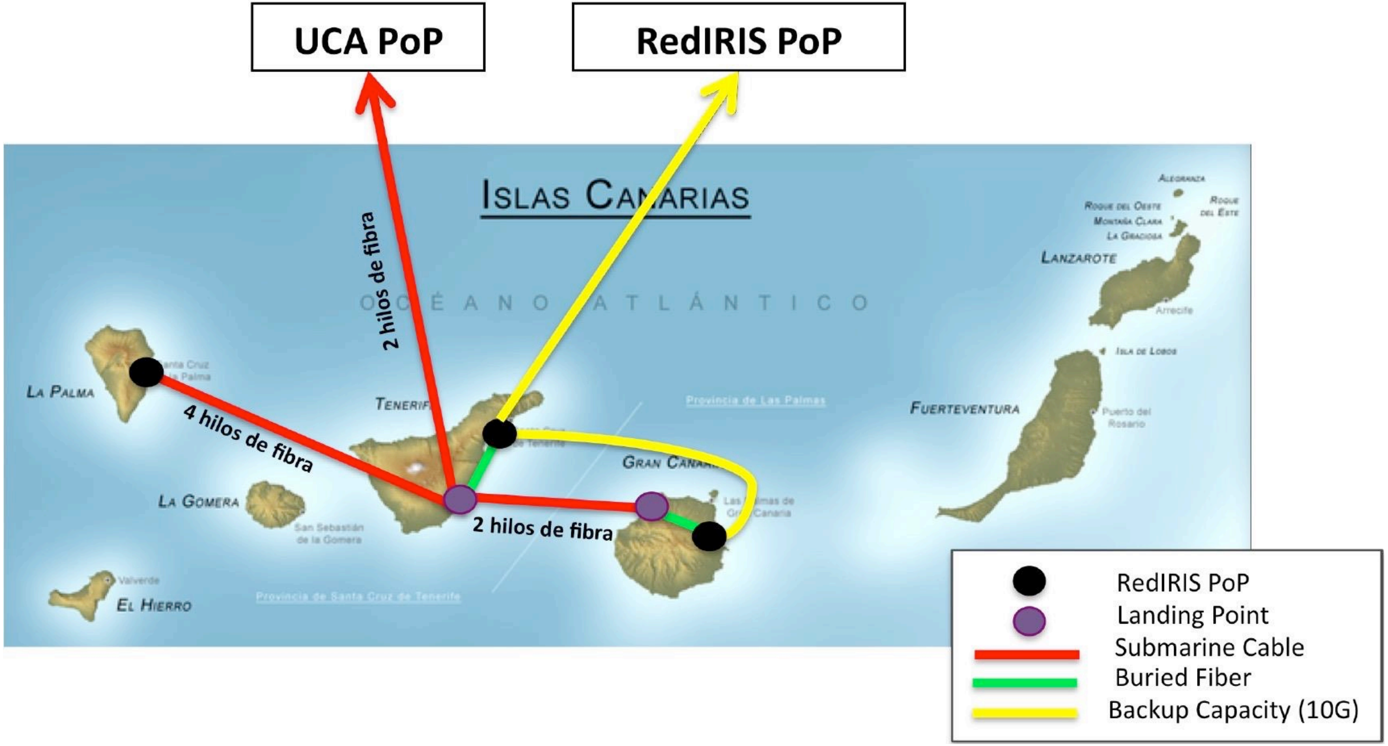
# CTAO South connection to Europe

Current route at 10 Gbps via USA (will stay as redundant route)




Future route at 100 Gbps (BELLA-S project originally planned end of 2020) to GEANT in Europe

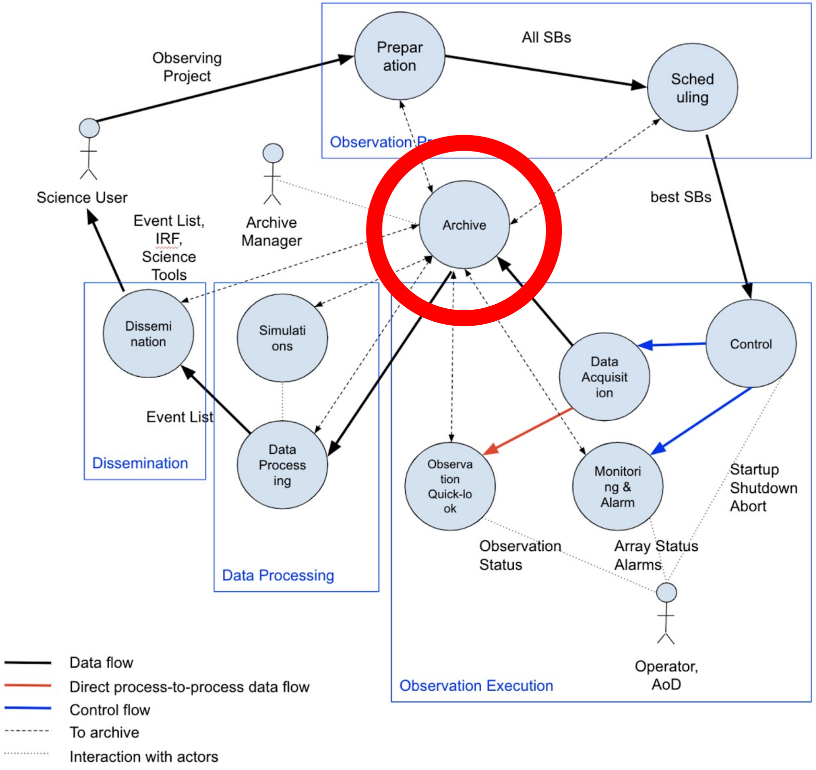
# CTAO North connection



# ASTRI-SSDC

- The ASI-SSDC (Space Science Data Center):
  - Wide experience as MWL data center, both for low-level data products (AGILE data center, Fermi-LAT/SWIFT/... data mirror center) and high-level data, data products and catalogs.
  - Data and data products integrated in a fully MWL environment (MMIA: Multi-Mission Interactive Archive).
  - Possibility to perform cross-catalog searches between resident and external catalogs.
  - Powerful tools to extract SED of sources and modelisation.
  - VHE catalog products from literature already integrated in the TeVGeV Catalogue.
  - Proposed to host also DL4 data from CTA

# ASTRI - Archives




## ASTRI Mini-array Archive System

- plays a **central role** in the whole observing life-cycle of the array
- shall guarantee **long-term data preservation and access**
- shall manage: *observation plans; science data; monitoring/alarm/logging data; system configurations*

**2 separated physical units** (10 Gbit/s connected):

***On-Site Archive System:***

- @ Teide Observatory (Tenerife, Spain)
- Archive System for temporary storage (~1 week) and services
- On-site ICT

***Off-Site Archive System:***

- @ Rome (Italy)
- Archive System for long-term data storage and services
- Off-site ICT ➔ **ASTRI Data Center**