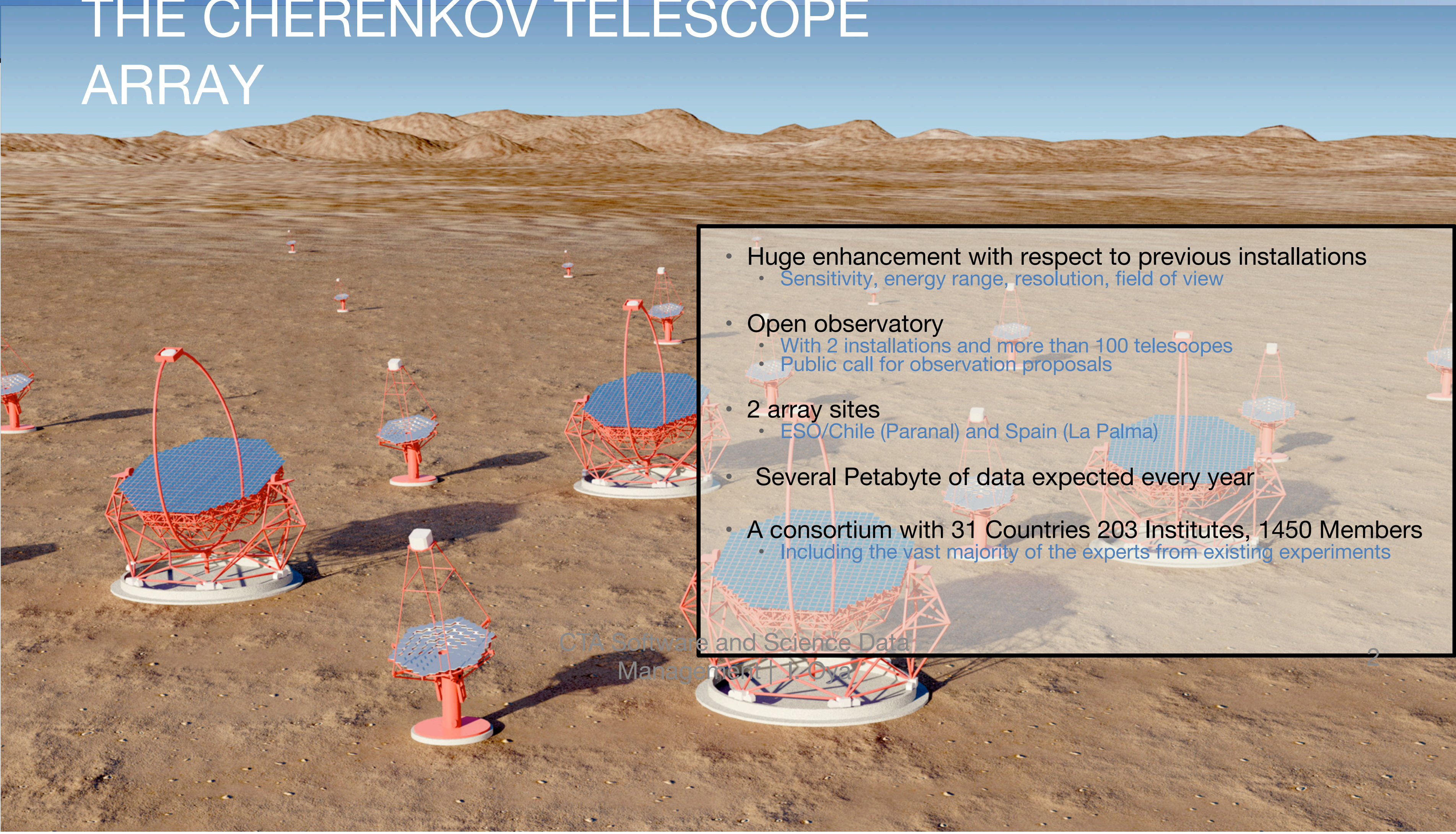
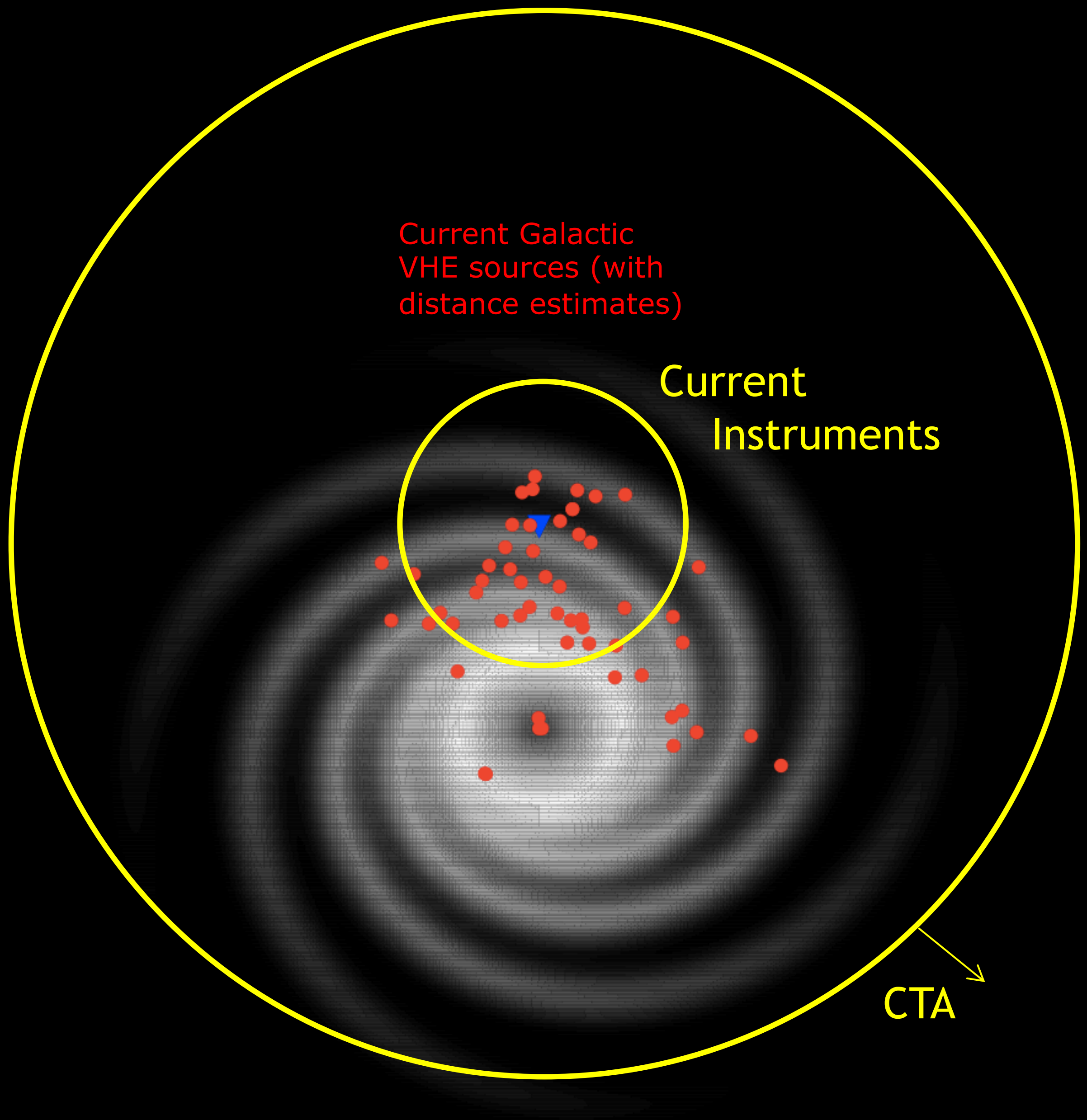
Matthias Fuessling,
CTA Observatory

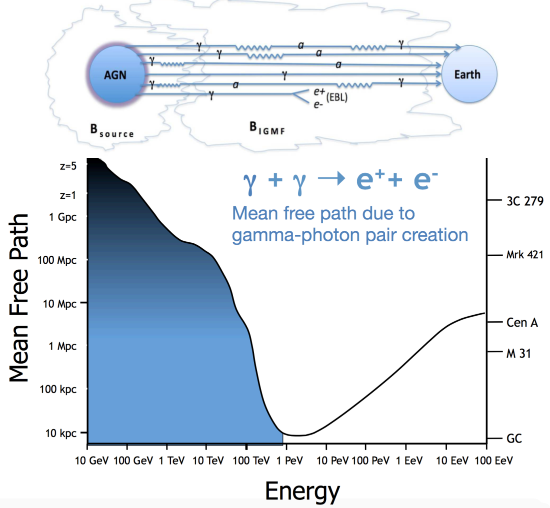
CEVO Tech Forum
2021-04

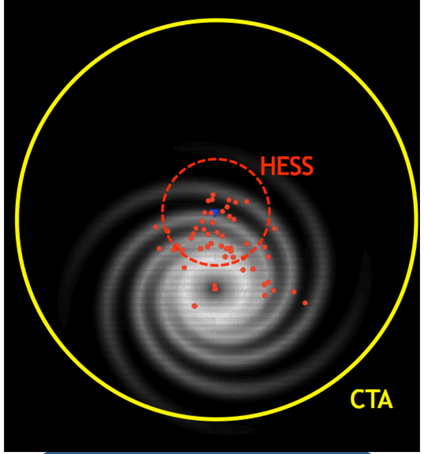
THE CHERENKOV TELESCOPE ARRAY

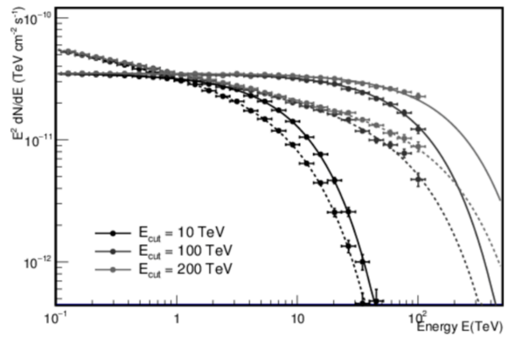


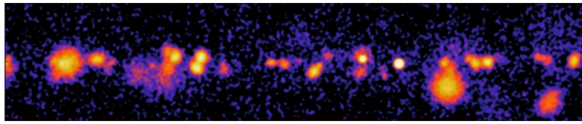
- Huge enhancement with respect to previous installations
  - Sensitivity, energy range, resolution, field of view
- Open observatory
  - With 2 installations and more than 100 telescopes
  - Public call for observation proposals
- 2 array sites
  - ESO/Chile (Paranal) and Spain (La Palma)
- Several Petabyte of data expected every year
- A consortium with 31 Countries 203 Institutes, 1450 Members
  - Including the vast majority of the experts from existing experiments

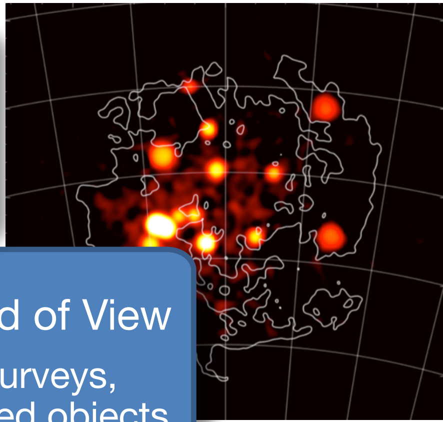

Current Galactic
VHE sources (with
distance estimates)
Current
Instruments
CTA

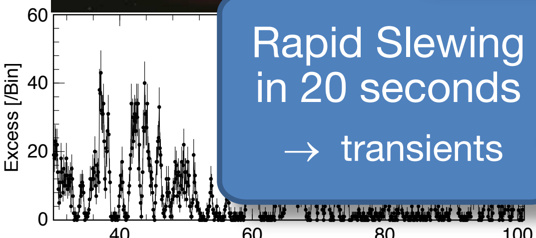


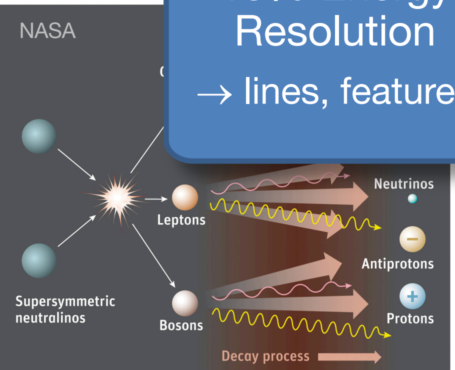


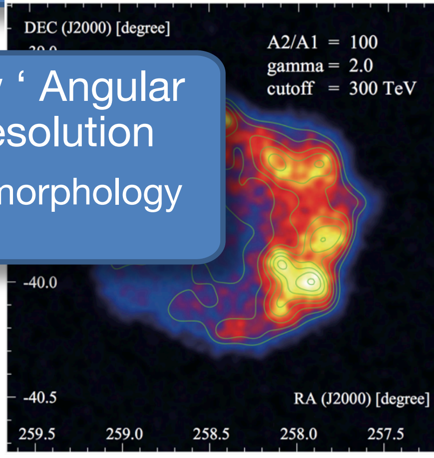












10 x Sensitivity, Large Collection Area

→ all topics

Energies down to 20 GeV

→ Cosmology++

Energies up to 300 TeV

→ Pevatrons

Rapid Slewing in 20 seconds

→ transients

8° Field of View

→ surveys, extended objects

10% Energy Resolution

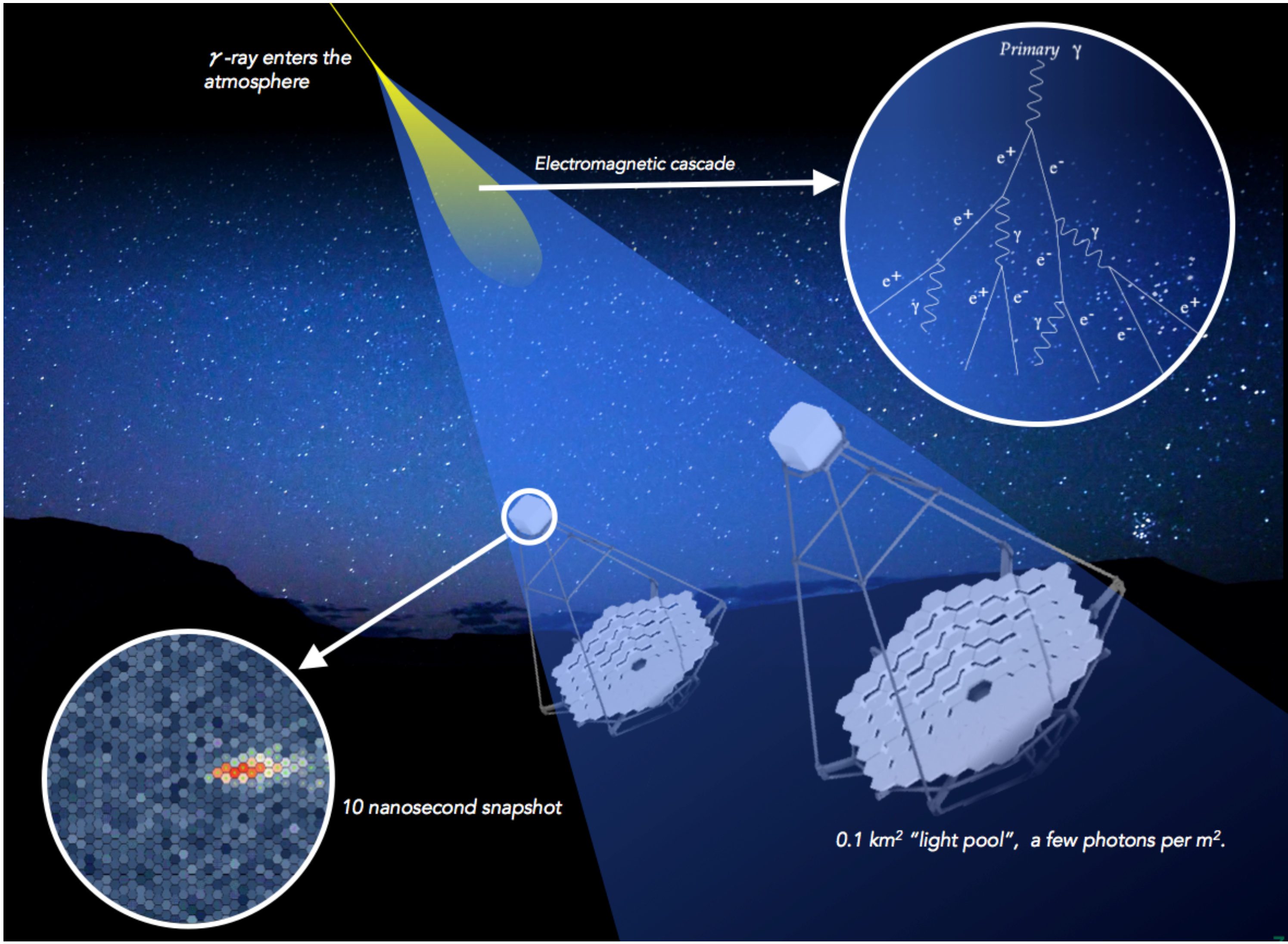
→ lines, features

Few ‘ Angular Resolution

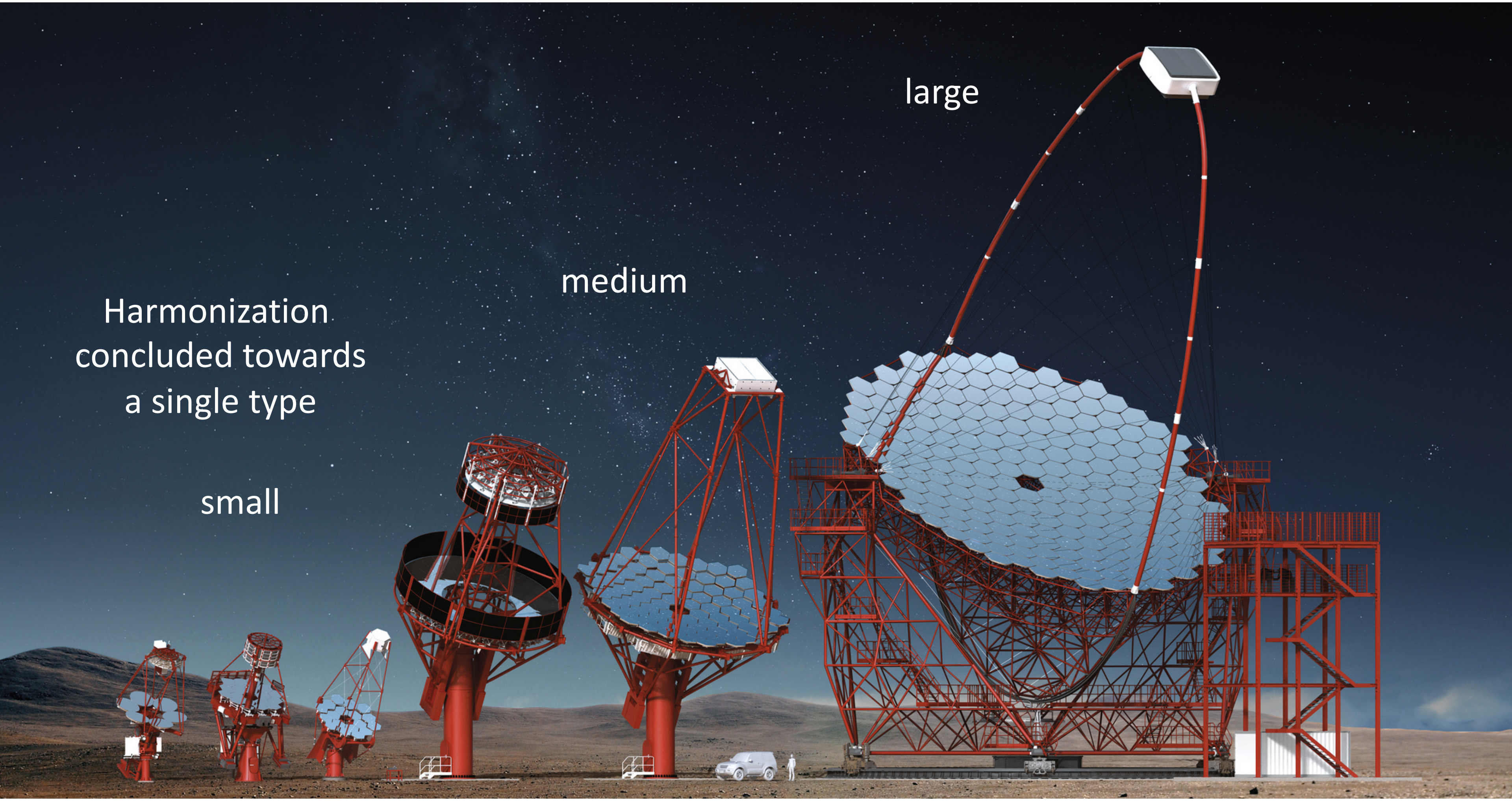
→ morphology

cta
cherenkov telescope array

DETECTION PRINCIPLE



CTA TELESCOPE PROTOTYPES

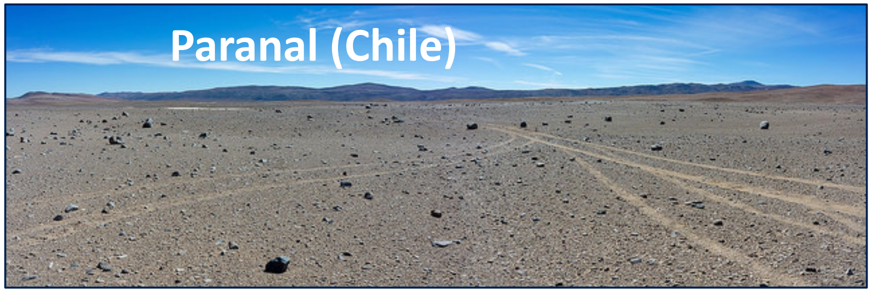


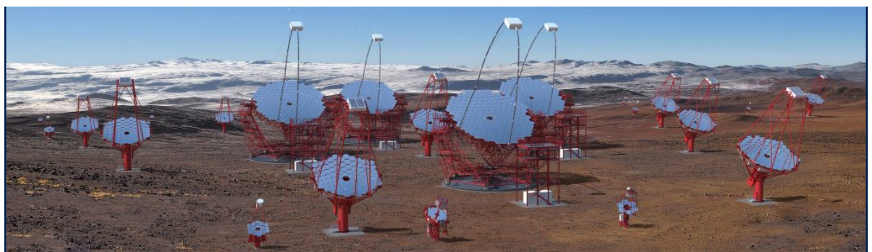
CHERENKOV TELESCOPE ARRAY

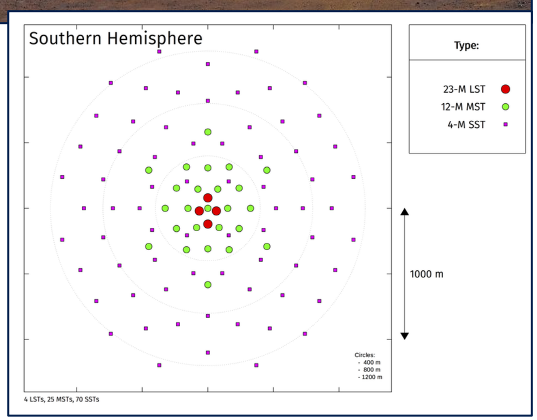
10 GeV | 100 GeV | 1 TeV | 10 TeV | 100 TeV

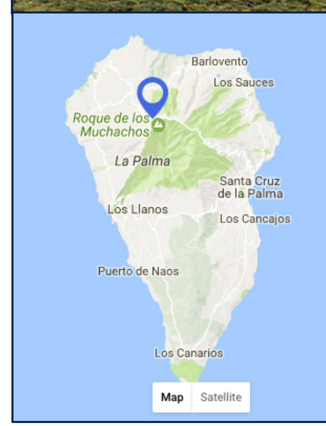
CTA SITES: ARRAYS, HEADQUARTERS, DATA CENTER

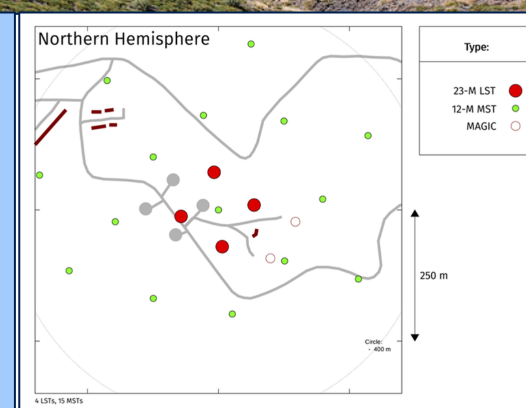
CTA AT PARANAL AND LA PALMA











CTA AS AN OBSERVATORY

- A Guest Observer Facility
  - For the **first time** in this waveband
    - Existing instruments are run as experiments
  - Annual cycles, TAC ranking, long-term schedule
  - Proposal preparation support, tracking, helpdesk +
  - Public science data archive
    - After proprietary period
- Two Telescope Arrays – one Observatory
  - Inter-site coordination
  - Uniform approach to science operations
- Open Observatory
  - Follow FAIR principles
  - Support open data and open science
  - Adhere to astronomical standards
  - Support VO

CTA OBSERVATORY OPERATIONS

- "Data" is the final product of CTAO
- Users will receive their data fully calibrated in FITS format, and be provided analysis tools
- After a one-year proprietary period, data are open
- During 1st decade, available observation time split roughly evenly between Key Science Projects (in particular surveys) and open time

- Operational Challenges
  - Sub-array operation, wide field of view, instrument response generation, background modelling, rapid alert generation and response, data volume, science operations during construction
- CTA is a Software instrument
  - Software plays a critical role in all steps of the Observatory

CTAO SCIENCE OPERATIONS RESPONSIBILITIES

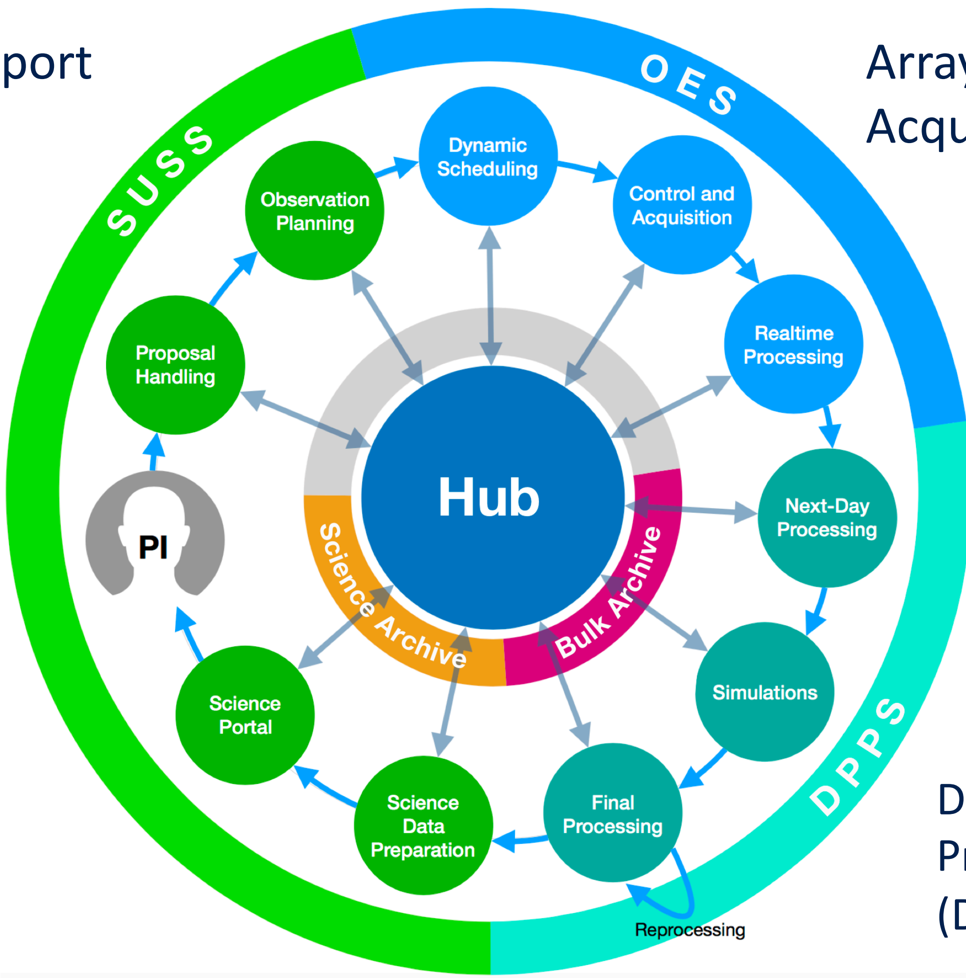
- CTAO responsible to operate the Observatory and deliver data
  - CTAO ERIC to be created to build and run the Observatory
  - Responsibilities of the CTA Observatory during operations is to organize, coordinate and perform all tasks to operate the Observatory
- Science Operations = Data processing + User Support Services + Computing Facilities Support
- Main elements for efficient and sustainable science operations include
  - Processing software
  - Computing facilities and storage
  - Expert people to run
    - Processing software
    - MC simulations
- List of science operations responsibilities (SDMC, HQ, Sites)
  - Prepare and deliver science data to users
  - User Support Services

- Call for Proposals
- Proposal Handling
- User support services
- Science Portal, FAQ, Newsletter
- PI interactions, incl. science alerts
- Provide tools for proposal preparation, data analysis
- Science planning
- Data processing and data quality monitoring
- Data archiving
- MC simulations
- Provide instrument response functions
- Science Analysis tools
- …

CTA SCIENCE OPERATIONS

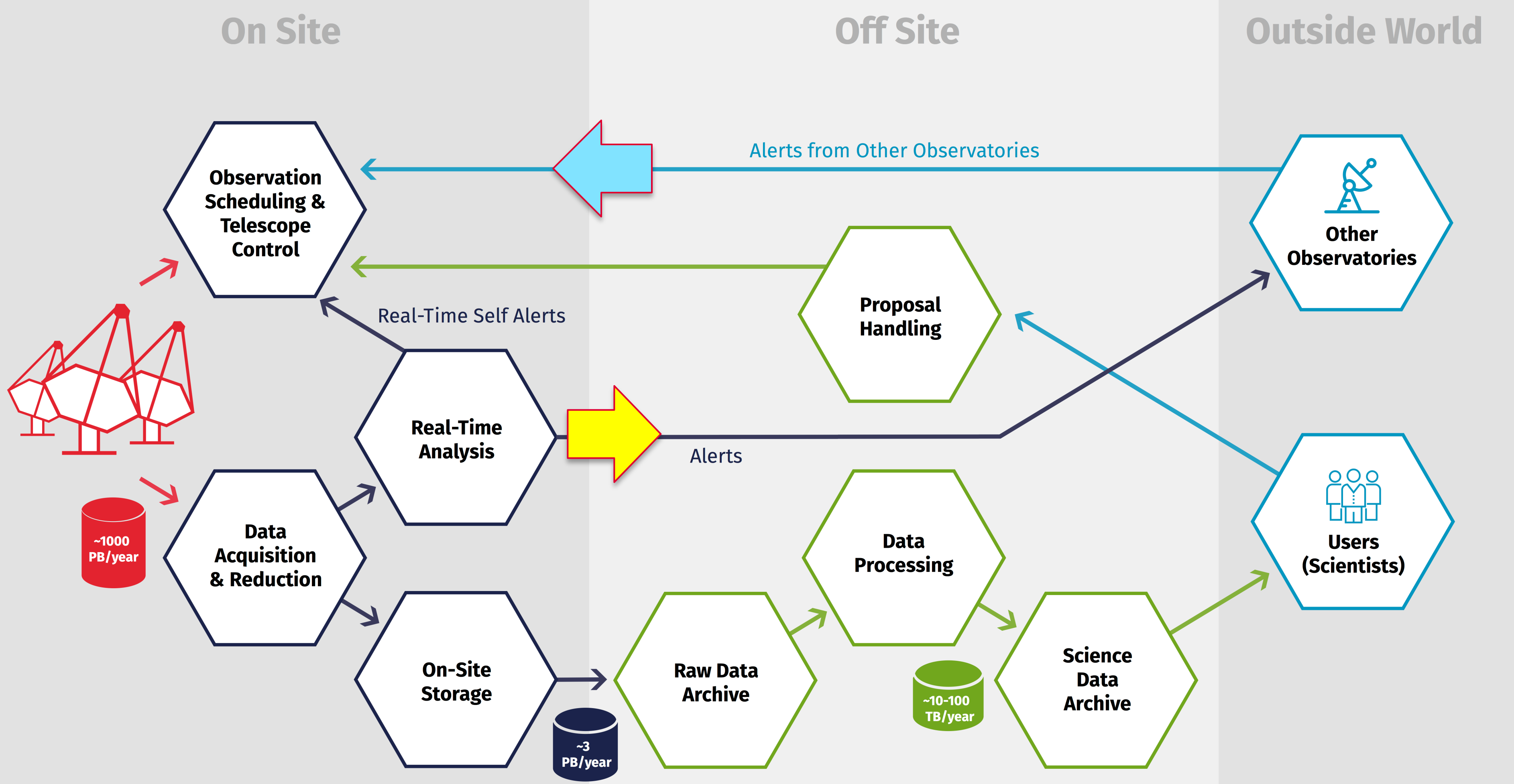
Science User Support System (SUSS)

Array Control and Data Acquisition (ACADA)



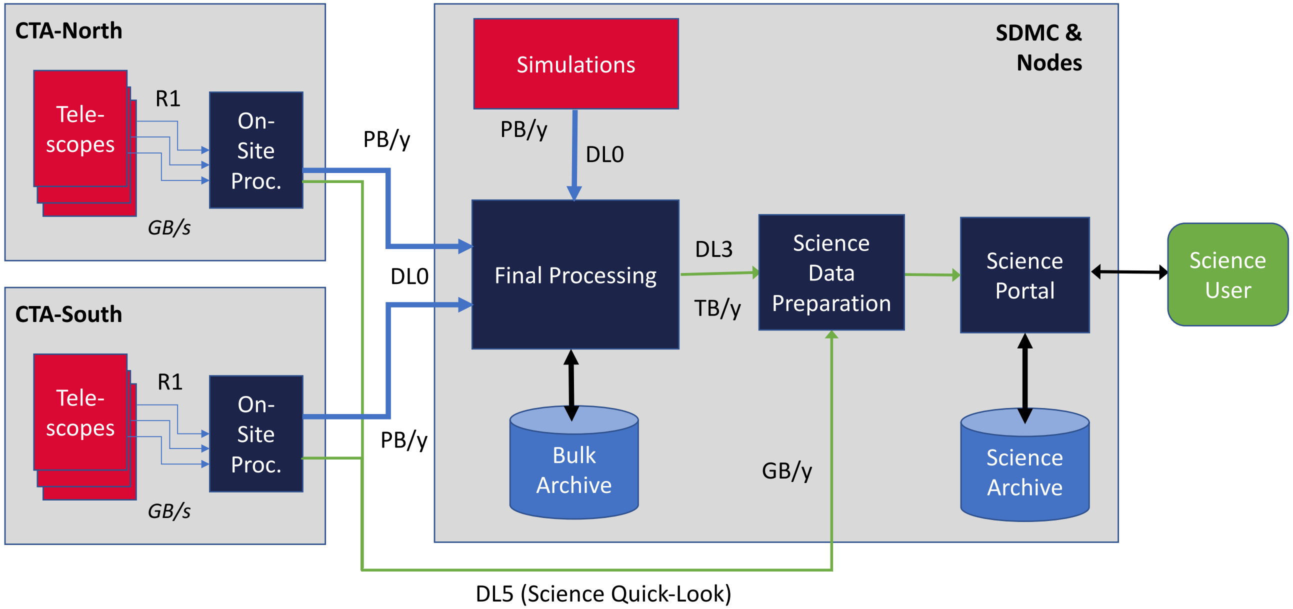
Data Processing and Preservation System (DPPS)

CTA SYSTEM ARCHITECTURE & INFORMATION FLOW



DATA FLOW & ANALYSIS CATEGORIES

- Data Processing at different steps
  - Timescales of near real-time and next-day for quicklook and science alerts (on-site processing), months for final products (final processing)
  - Involves strict data quality assurance and verification of data products
- Strong data reduction along the processing steps
  - From PB/y (at raw data level) to GB/y (high-level science data)
- Open access through Science Portal
  - access to science archive, to science analysis tools
  - Exploration of quick-look data products



DATA LEVELS

Data Level	Short Name	Definition
R0	Raw Internal	On-site streamed raw data, not normally preserved long-term in this form. R0 content and format is internal to each device / controllable system, such as raw data transmitted from the physical device / system to its respective server in the on-site Data Centre.
R1	Raw Common	On-site stream raw data meeting common standards, transmitted on-site from a Camera or other on-site system to the OES. This is the first level of data seen by the OES, that will typically need some pre-processing from the R0 data format. Exceptionally, some R1 data may be stored for engineering purposes.
DL0	Raw Archived	All archival data from the data acquisition hardware/software, transmitted from the OES to the DPPS. This is the lowest level of data that are intended for long-term storage in the bulk archive. This includes both camera event data and technical data from other sub-systems, such as non-camera devices or software.
DL1	Processed	Processed DL0 data that may include telescope-level (TEL) data and parameters derived from them. Typical contents include calibrated image charge, Hillas parameters, and a usable telescope pattern. DL1 data is not normally stored long-term.
DL2	Recon-structed	Reconstructed shower parameters such as energy, direction, particle ID, and related signal discrimination parameters. Does not include telescope-level (TEL) information. For each event this information may be repeated for multiple reconstruction and discrimination methods. DL2 data is not normally stored long-term.
DL3	Reduced	Sets of selected events with a single final set of reconstruction and discrimination parameters, along with associated instrumental response characterizations and any technical data needed for science analysis.
DL4	Binned	Data product produced by binning of DL3 data, including data cubes and maps which are suitable for combination/summation to produce DL5 products.
DL5	Science	Data product produced by combination of DL4 products an extraction target specific region(s) of interest. Includes for example light-curves and spectra, along with associated data such as source models and fit results.
DL6	High-Level	High-level or legacy observatory data, such as survey maps and source catalogues.

Tel

OES

DPPS

SUSS

DATA LEVELS

Data Level	Short Name	Definition
R0	Raw Internal	On-site to each respect … nternal n to its
R1	Raw Common	On-site system process … -site rposes.
DL0	Raw Archived	All arch the low camera … . This is h ware.
DL1	Processed	Process Typical data is … n. n. DL1
DL2	Recon-structed	Recons parame repeate … ination be ng-term.
DL3	Reduced	Sets of associa … g with lysis.
DL4	Binned	Data pr combin … for
DL5	Science	Data pr Include results. … terest. and fit
DL6	High-Level	High-le

Image/Data Cube Per Telescope

Mixed Tel/Event (Temporary)

Event List

Binned

Tel

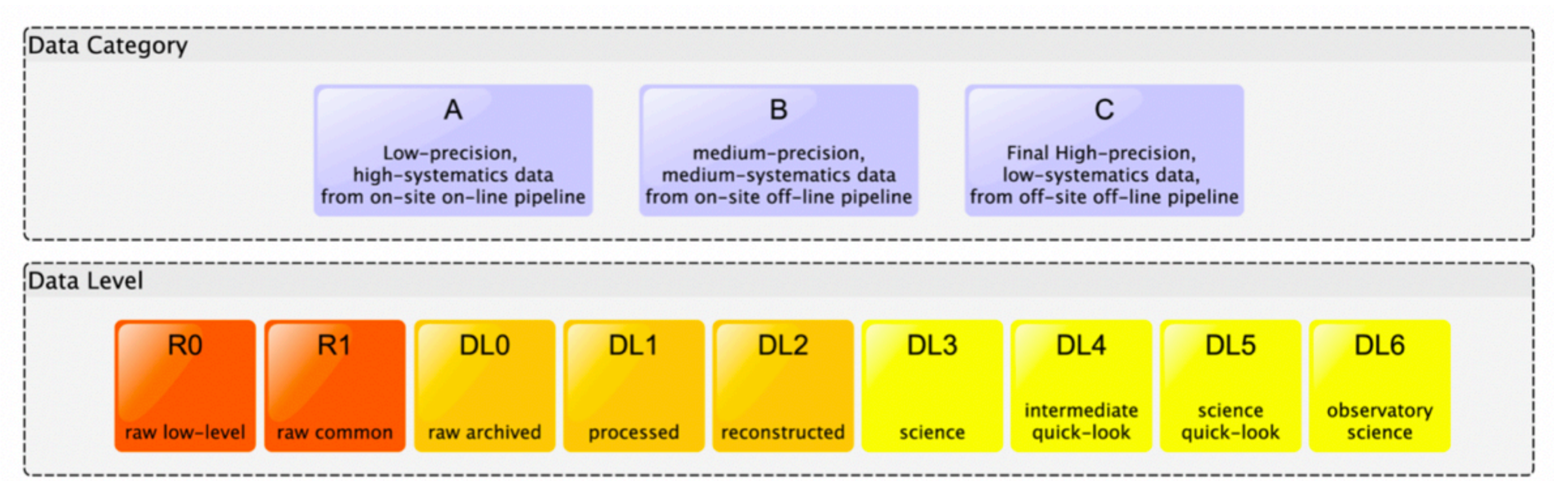
OES

DPPS

SUSS

DATA LEVELS AND CATEGORIES

- Data Processing performed for all data categories
  - Category A – real-time, rapid quick-look products for science alert generation, lowest precision and highest systematics
  - Category B – next-day, somewhat better precision and lower systematics with simplified analysis or calibration
  - Category C – final, full high-quality data processing, best calibration and algorithms
- Long-term Preservation of bulk data (DL0)
  - Large amounts of data, small number of users
- Long-term Preservation of science data (DL3)
  - Small amounts of data, large number of users
- Data access
  - Science Users get access to DL3-6 data, while raw data (DL0) is used internally by CTAO

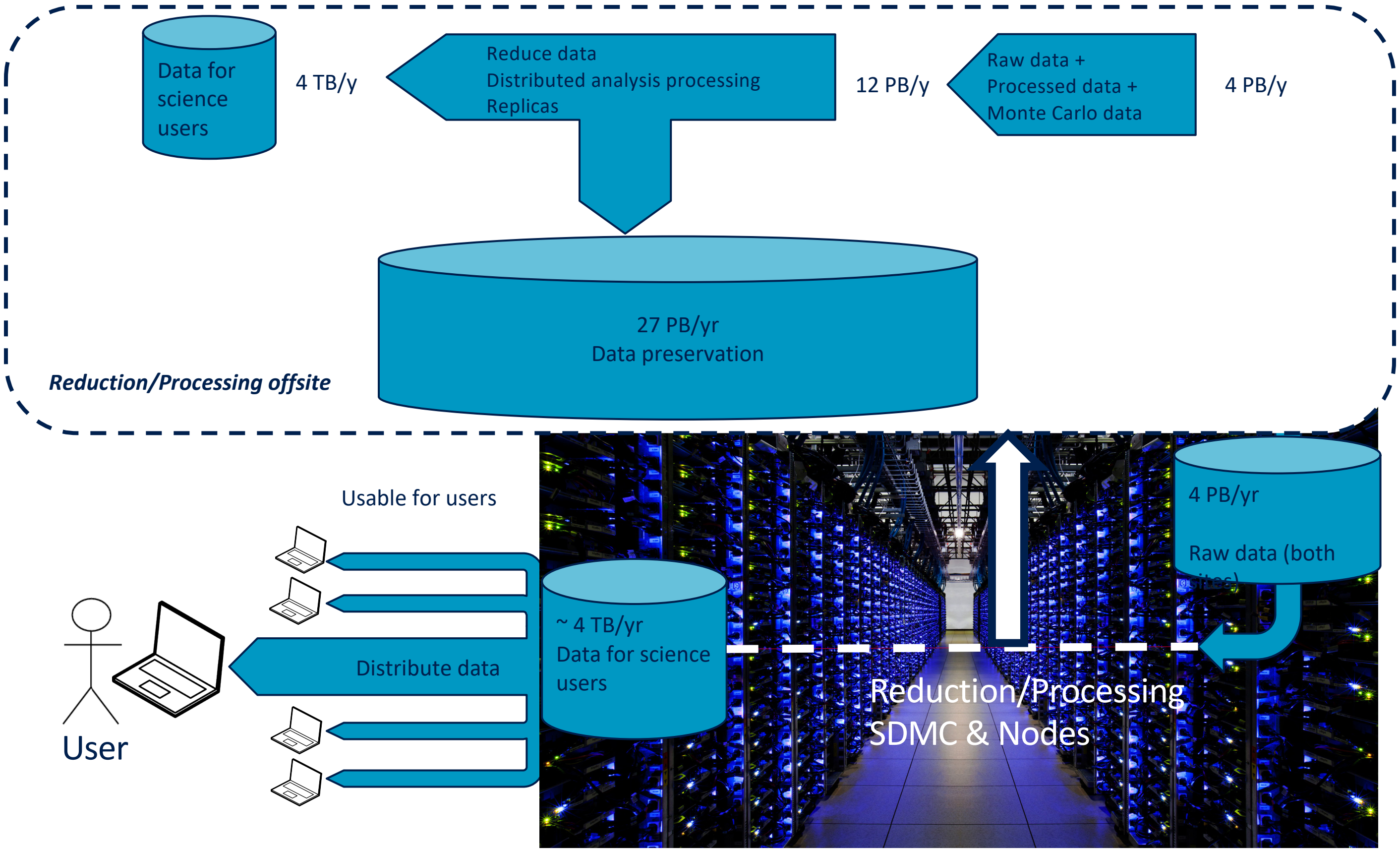


DATA MANAGEMENT CHALLENGES

- Reducing On-site Data Volume
  - 1000 PB/y directly coming out from cameras (mostly noise)
  - Impossible to transfer via internet link (see later)
- Storing and Processing Big Science Data:
  - Computing resource requirements imply distributed computing model
  - Data volume is too big to separate storage from computing
- Simulating CTA
  - Development of extensive air showers
  - Propagation of Cherenkov light
  - Cherenkov photon ray-tracing through the telescope optics to Camera
  - Camera electronics and photosensors simulation
- Developing a Data Processing Pipeline
  - PBs of data
  - Distributed team of developers.

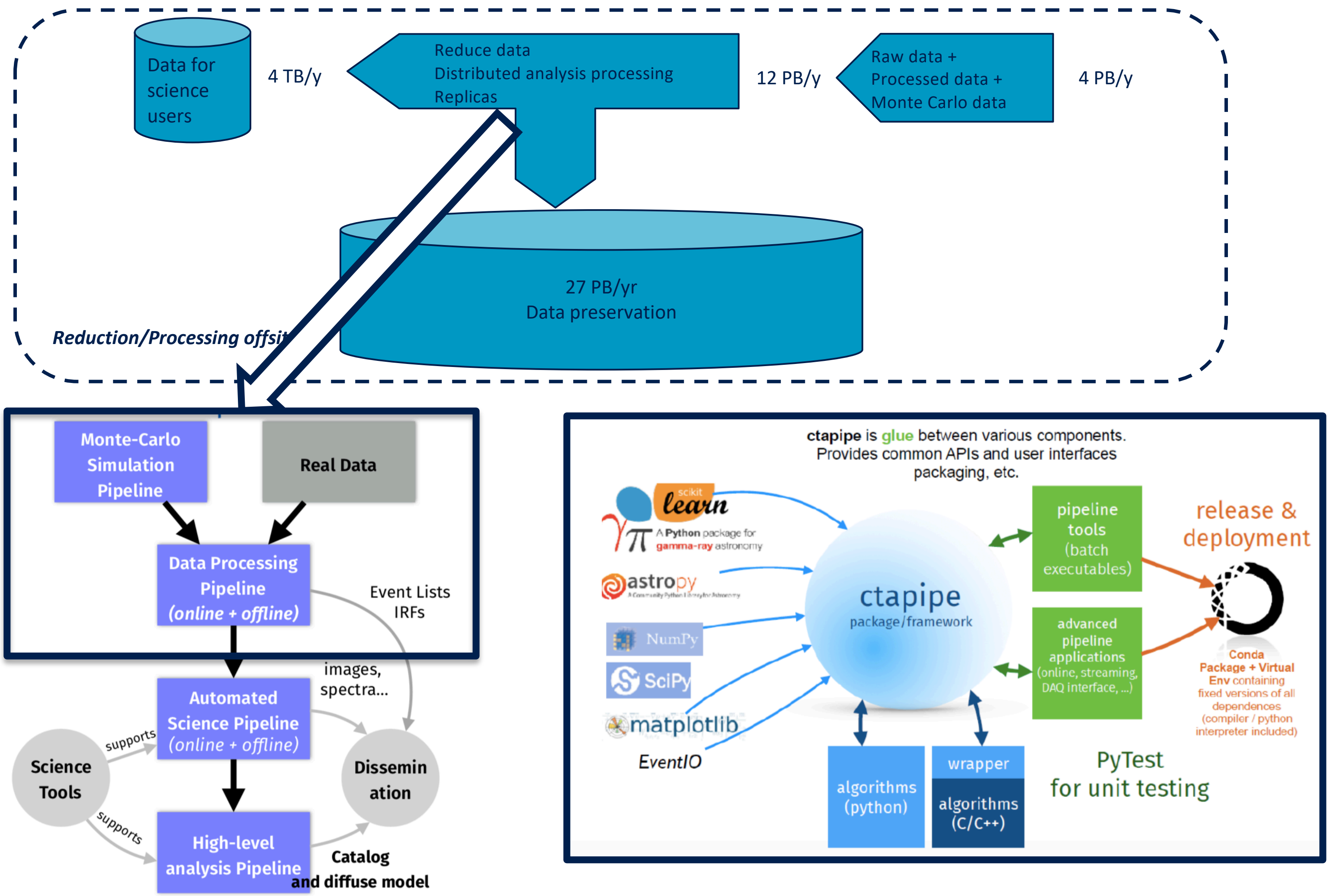
(+many more computing challenges for e.g. control and supervision of 100 telescopes, sub-array operation, scheduling)

SOLVING THE DATA VOLUME PROBLEM (II)

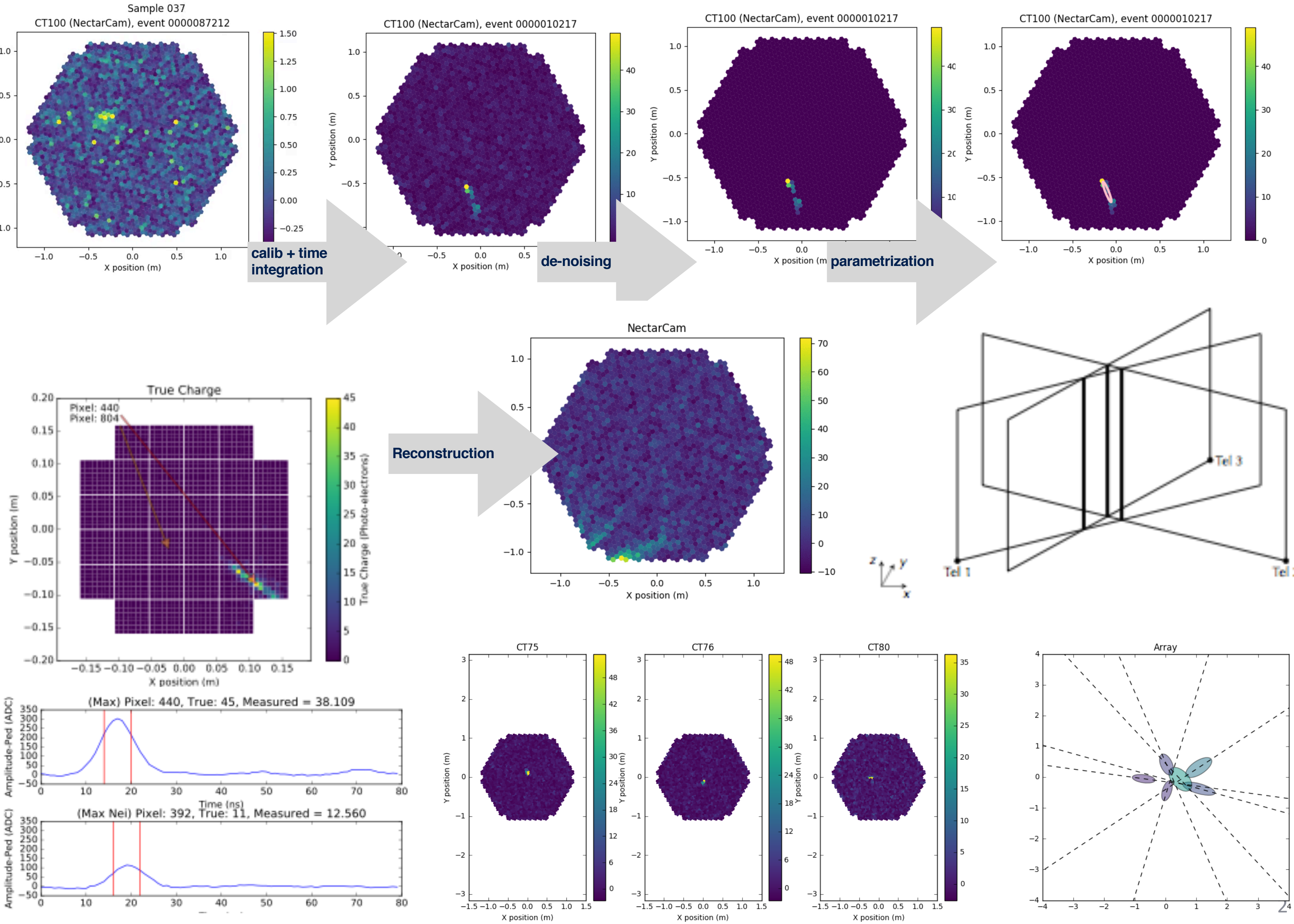


Computer clusters at mainland

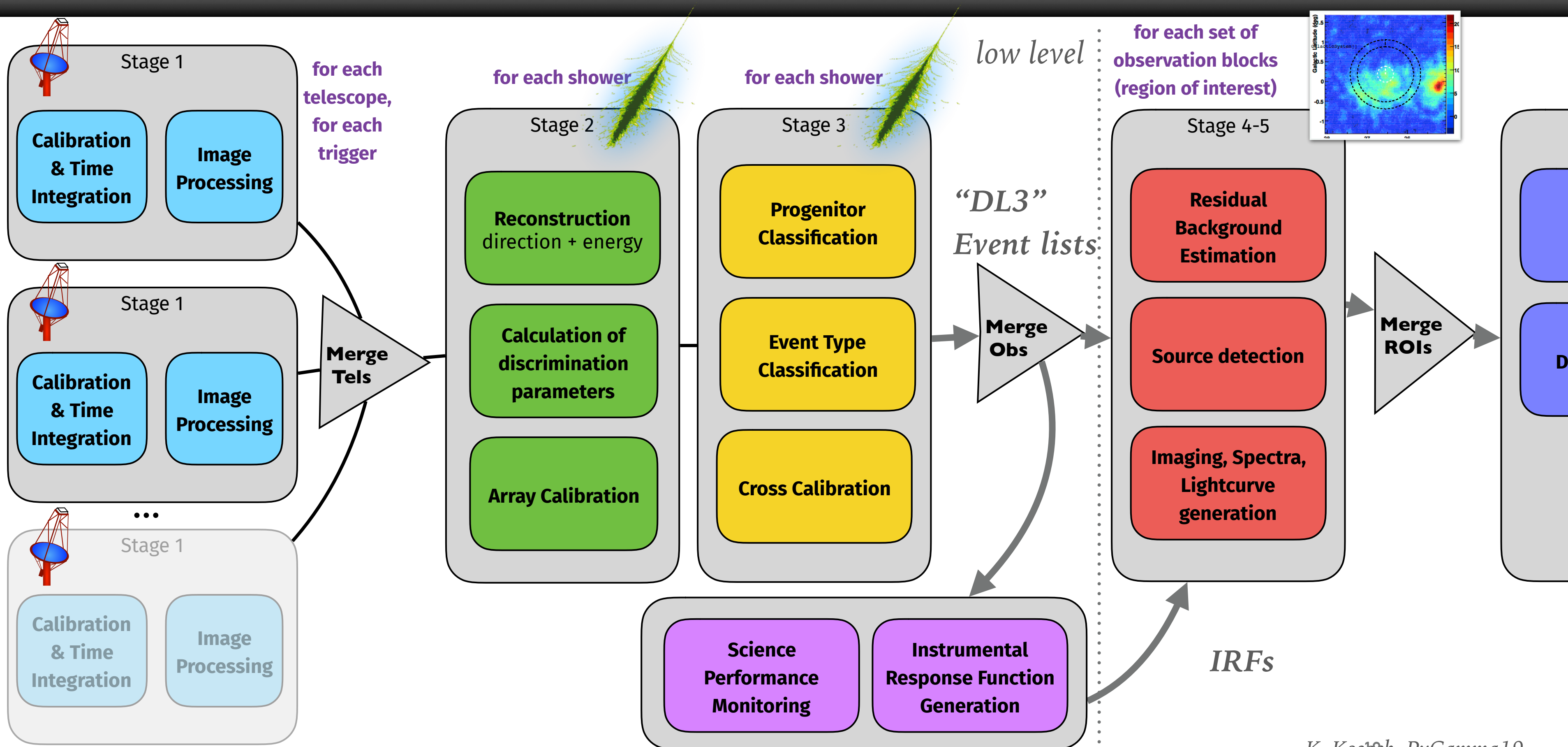
CTA PIPELINES: DATA PROCESSING



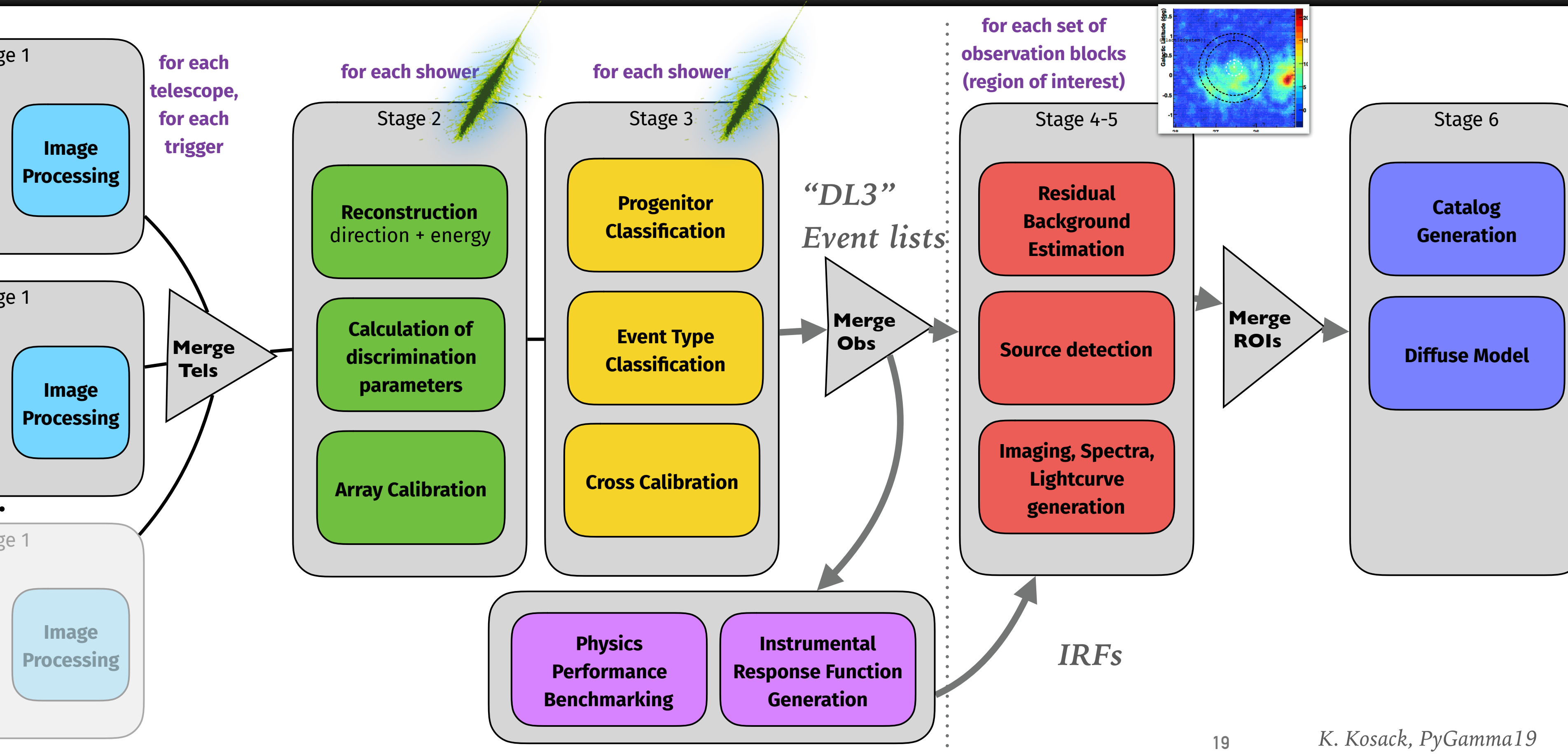
K. Kosack *ctapipe example*



Data Processing Pipeline (simplified)



Data Processing Pipeline (simplified)



K. Kosack, PyGamma19

Output: Science Data

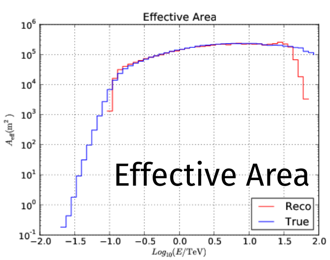
Event-List

event_id	RA	DEC	E	class	type	n_tels	...
1	23,3	-40,1	0,01			5	
2	24,6	-40,5	20,0			34	
3	23,5	-41,12	0,45			3	
4	21,3	-38,2	1,03			4	

Technical Tables (for sub-GTIs)

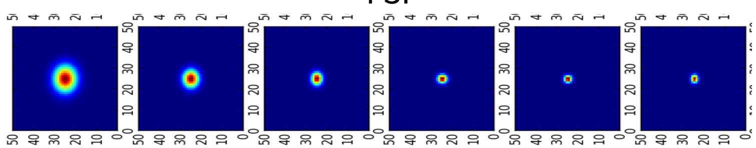
TIME	Transparency	Temperature	Trigger Rate
580234.34	0.8	32	12034
580234.35	0.94	32	13023
580234.36	0.70	33	12532

Instrumental Responses:

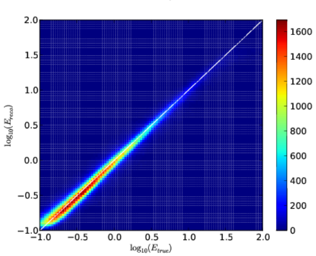


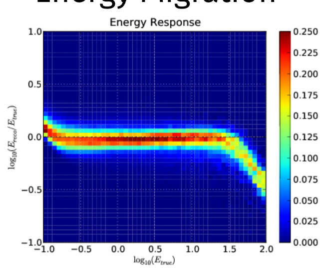
Effective Area

PSF



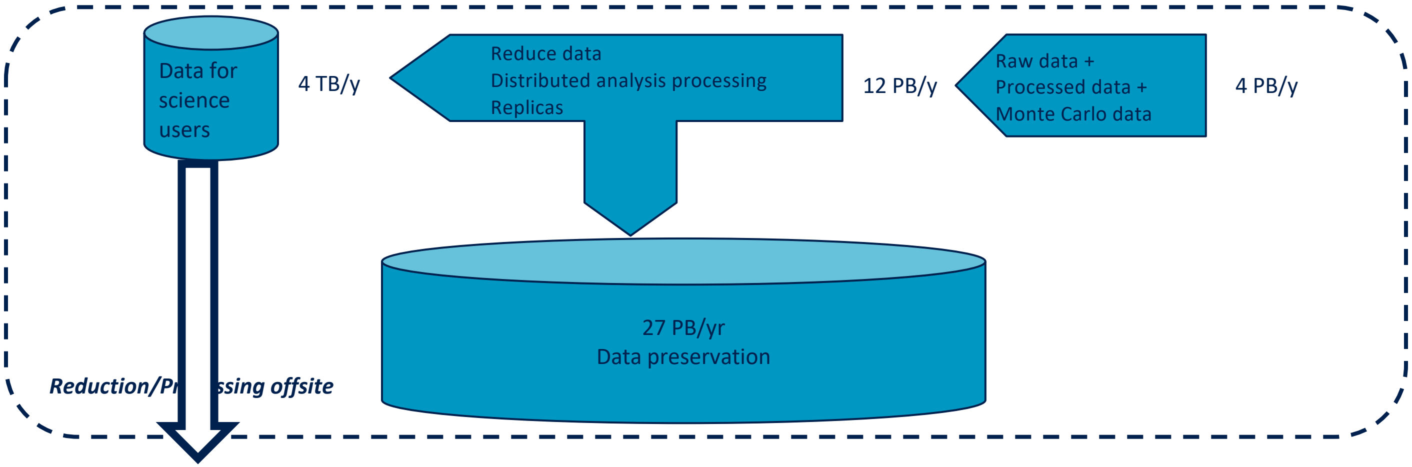
Energy Migration

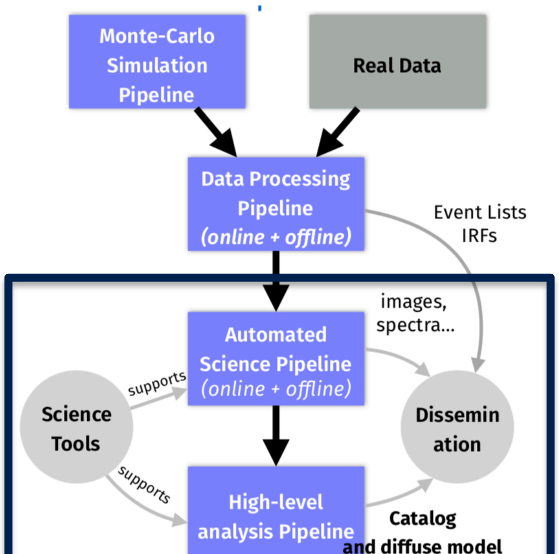


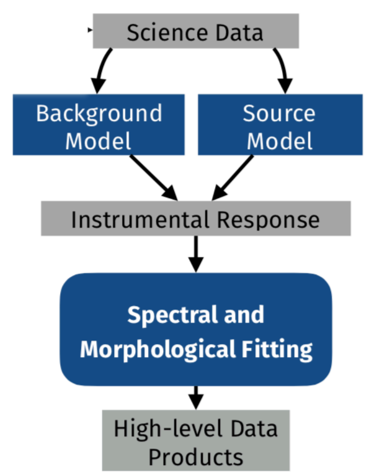


(note these are not CTA responses, just examples form HESS)

CTA PIPELINES: SCIENCE ANALYSIS







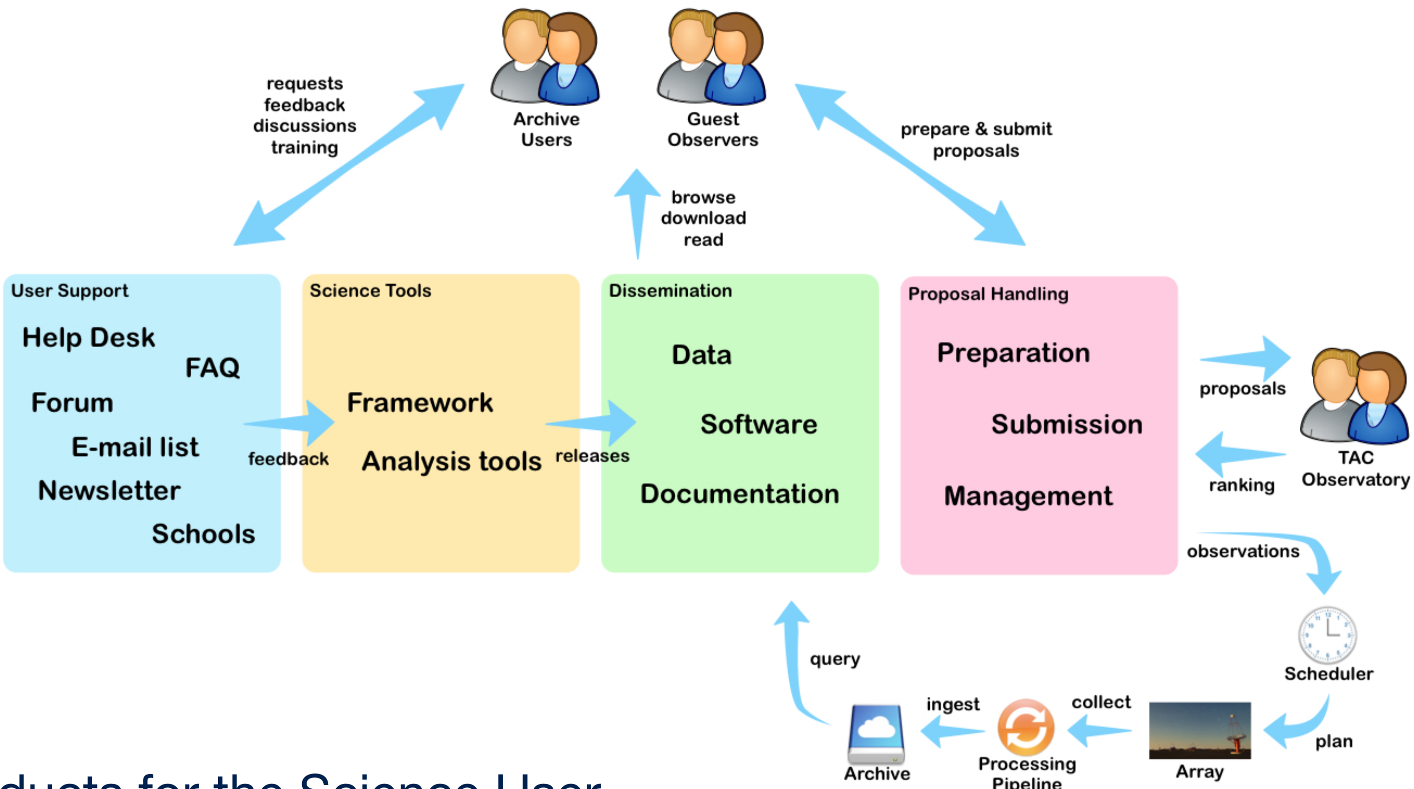
Prototypes for Science Tools:

Gammapy

Ctools

@ctoolssoftware

SCIENCE USER PERSPECTIVE



- Products for the Science User
  - Photon (candidate) event list data (FITS) (DL3)
  - Instrument response functions, background model
  - Science analysis tool suite, supporting documentation

DATA PROCESSING

- Large-scale data processing to bring the raw data (DL0) to reduced and calibrated science data products (DL3) is a service task by the Observatory
  - Includes all steps from calibrations, simulations, event reconstruction, data volume reduction, …
- However, there are clearly benefits from commonalities in the framework of the VO
  - Provenance (see talk by M. Servillat)
  - workflow management (e.g. CWL)
  - Machine-learning

SCIENCE ARCHIVE AND PORTAL

- CTA will serve the science users with science data products, tools and services via the science portal and archive
  - Access to DL3 data products and science tools
  - Higher-level data products available as quick-look products
  - User-contributed publication-ready science products
  - Access to public data and to proprietary data

- CTA aims to provide functions and services as expected by world-class observatories, making use of VO
  - Methods range from "making public data accessible via VO registry" to "building in VO application SW as part of the portal and archive"
  - Analysis to be started

SCIENCE DATA PRODUCTS AND SOFTWARE

- Science Data Products are accessible to users via the Science Portal and Archive
  - DL3 data products: event lists + Instrument Response Function
  - Higher-level data products: sky images, spectra, lightcurves
  - Source Catalogues
- Standardisation of data products within gamma-ray astronomy ongoing
- Exploration of commonalities with other wavelengths (e.g. KM3NeT)
  - Aim to support ESCAPE TSPs on MM/MWL
- Aim for making the data model VO-compliant and bring this to the IVOA as standards
  - What do the users need? What would support interoperability?
- Same for Science Tools
  - Provenance, metadata, …

SCIENCE ALERTS

- Support for MWL / MM physics is one of the key aspects of CTA
  - Follow-up on external Science Alerts or issuing of CTA-detected Science Alerts possible on O (1min) timescale
- Discussions started on enhancements of VO events (or similar format / supporting infrastructure)
  - Include provenance information (esp. important for upcoming era of brokers)
  - Allow for feedback loops (user-enhanced alerts)
    - Real-time negotiations to support "I will observe if facility XYZ observes too"
    - Issue notification that observations will start
    - Re-inserting the events with additional follow-up information?

SCHEDULES AND VISIBILITY

- CTA aims to be able to
  - exchange observation schedules to support MM / MWL
  - Inform on visibility
  - Support multi-observatory scheduling (TBC)

- Discussions not yet started in detail with the relevant groups in VO
  - Resource limitations

SUMMARY

- CTA will work as an open Observatory
  - User services and support in the core
- Data Challenges
  - Several PB/y raw data to be handled
  - High-quality science data products, software tools and services
- CTA software is entering the construction phase

CTA has interest in various topics related to VO and in particular within WP4:

  - Provenance (all topics)
  - Data models (DL3 and higher)
  - Support for multi-messenger approach (VO events, schedules)
  - Revision of all aspects of FAIR and VO principles for science operations, data, software and services (science portal, science archive)