Astrofisica con Specchi a Tecnologia Replicante Italiana

cta cherenkov telescope array

INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

Universidade de São Paulo
Instituto de Astronomia, Geofisica e Ciencias Atmosferica

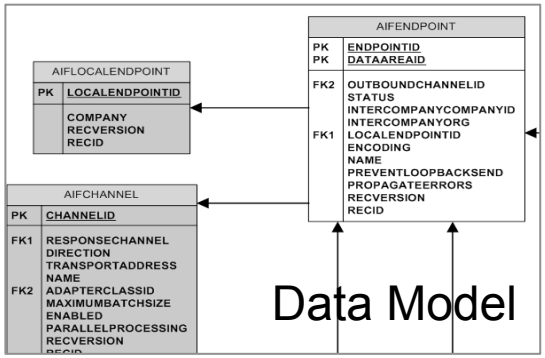
MIUR

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

ASTRI DATA MANAGEMENT.

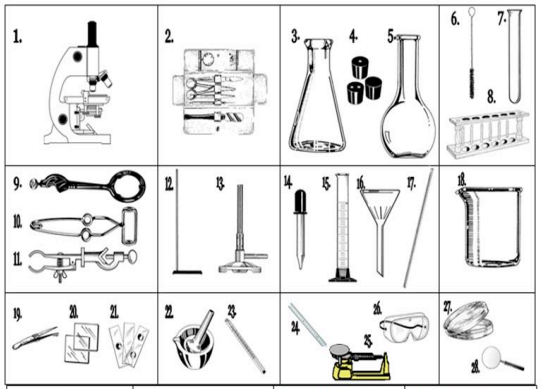
From ASTRI Prototype and Mini-Array to CTA.

L. Angelo Antonelli – INAF-OAR & ASDC
for the ASTRI Collaboration & the CTA Consortium



Data Model

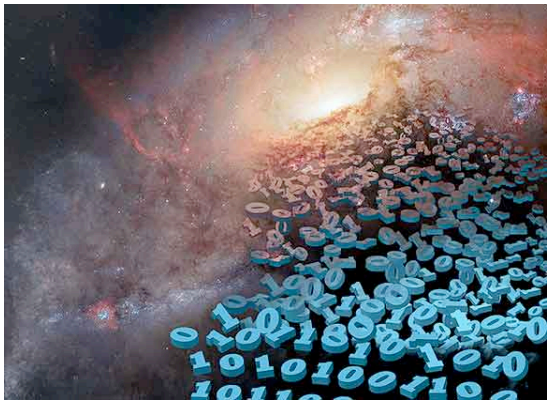
Reconstruction Pipelines



Science Tools

Archive(s)

Simulations



Data Access

ASTRI DATA MANAGEMENT TEAM

DATA RECO & ANALYSIS


Fabrizio Lucarelli


Angelo Antonelli


Denis Bastieri


Saverio Lombardi


Matteo Perri


Andrea Giuliani


Alberto Madonna


Michele Mastropietro


Sara Buson


Sabina Sabatini


Imma Donnarumma


Giovanni Piano

ARCHIVES


Andrea Di Paola


Vincenzo Testa


Stefano Gallozzi


Alessandro Carosi


Federica Moscato

MC SIMUL.

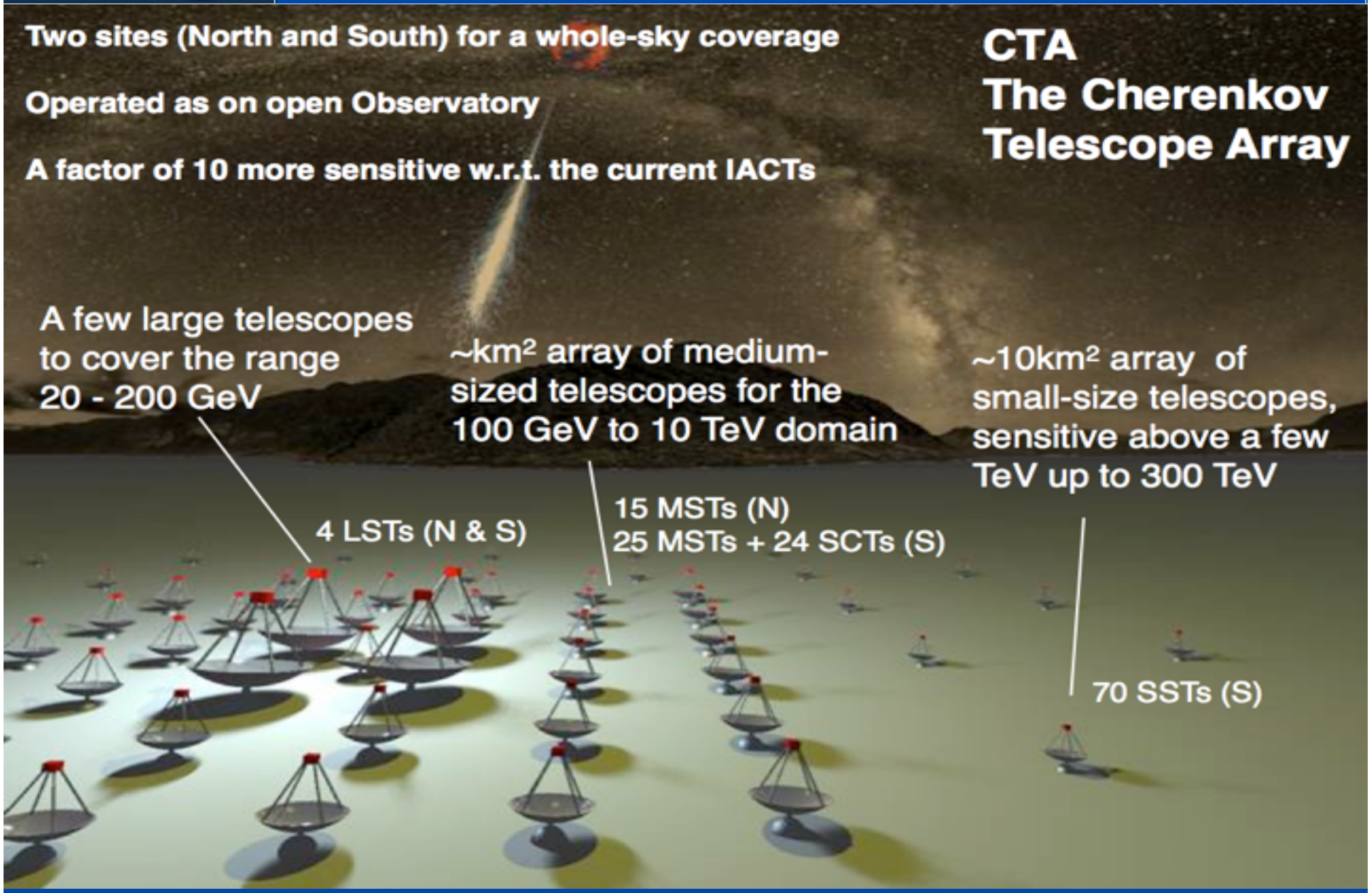

Antonio Stamerra


Ciro Bigongiari


Federico Di Pierro

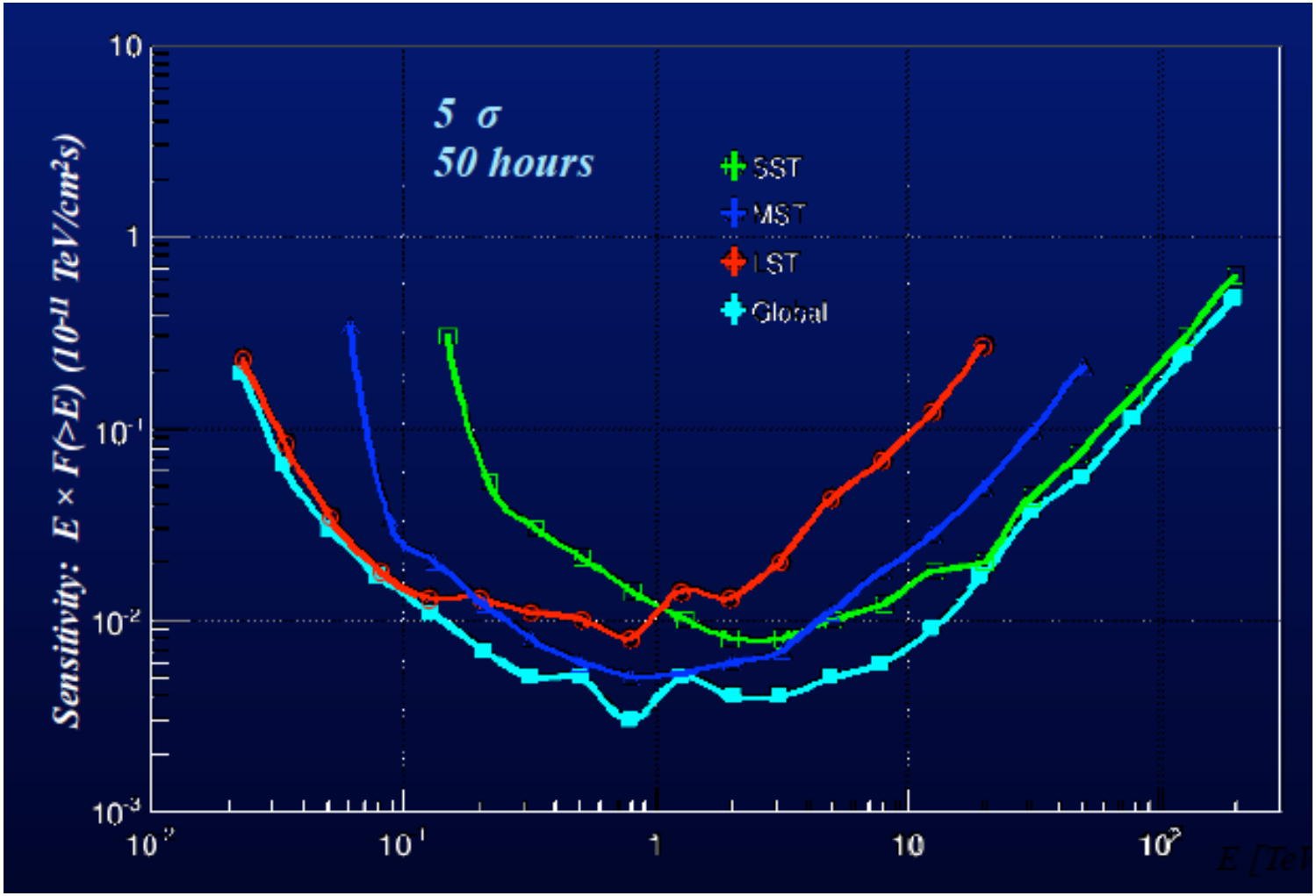
- Introduction: CTA & ASTRI
- Data Analysis Software.
- Archive.

- Introduction: CTA & ASTRI
- Data Analysis Software.
- Archive.


Two sites (North and South) for a whole-sky coverage
Operated as on open Observatory
A factor of 10 more sensitive w.r.t. the current IACTs
CTA
The Cherenkov
Telescope Array
A few large telescopes
to cover the range
20 - 200 GeV
~km² array of medium-
sized telescopes for the
100 GeV to 10 TeV domain
~10km² array of
small-size telescopes,
sensitive above a few
TeV up to 300 TeV
4 LSTs (N & S)
15 MSTs (N)
25 MSTs + 24 SCTs (S)
70 SSTs (S)

The CTA Project in a nutshell.

WHY CTA?



A deeper sensitivity (about a factor 10) on a wider band (from 20-30 GeV to 100 TeV) more then 1000 new gamma ray sources are foreseen to be discovered.

The CTA Project in a nutshell.

WHO IS CTA?

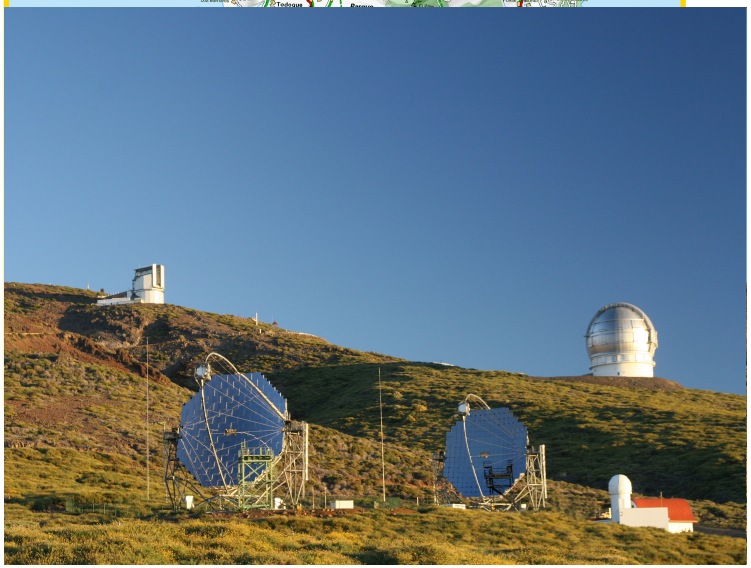
CTA is an ESFRI Project

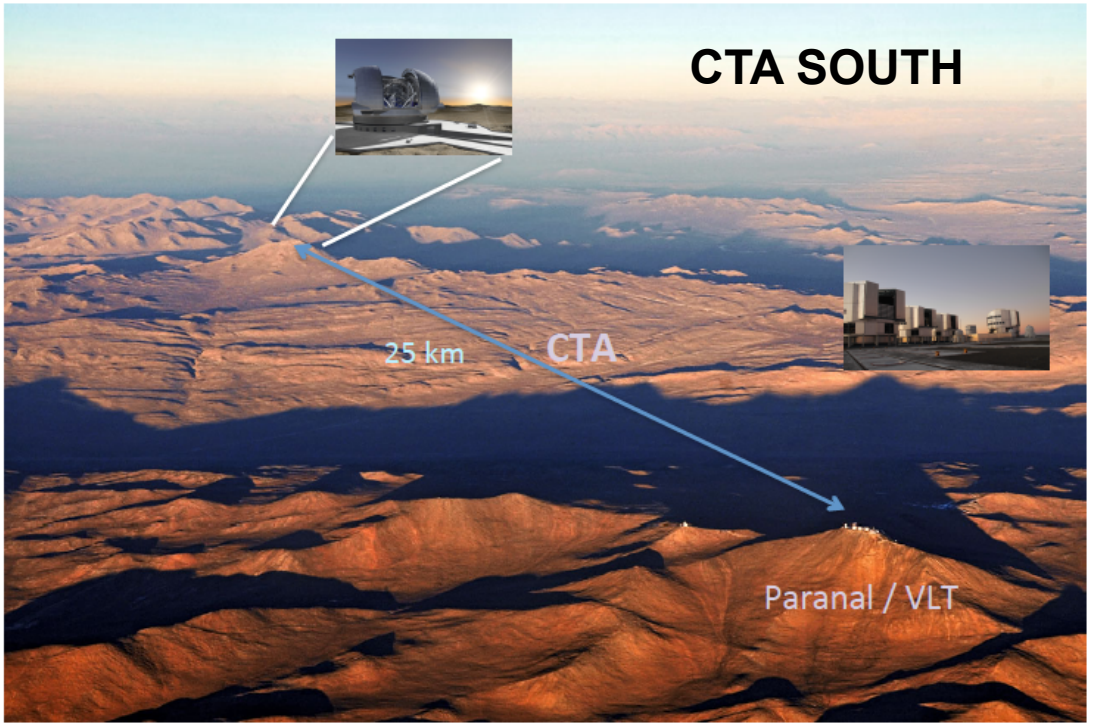
31 countries, ~1200 participants, ~180 institutes, ~400 FTE

The CTA Project in a nutshell.

WHERE CTA?

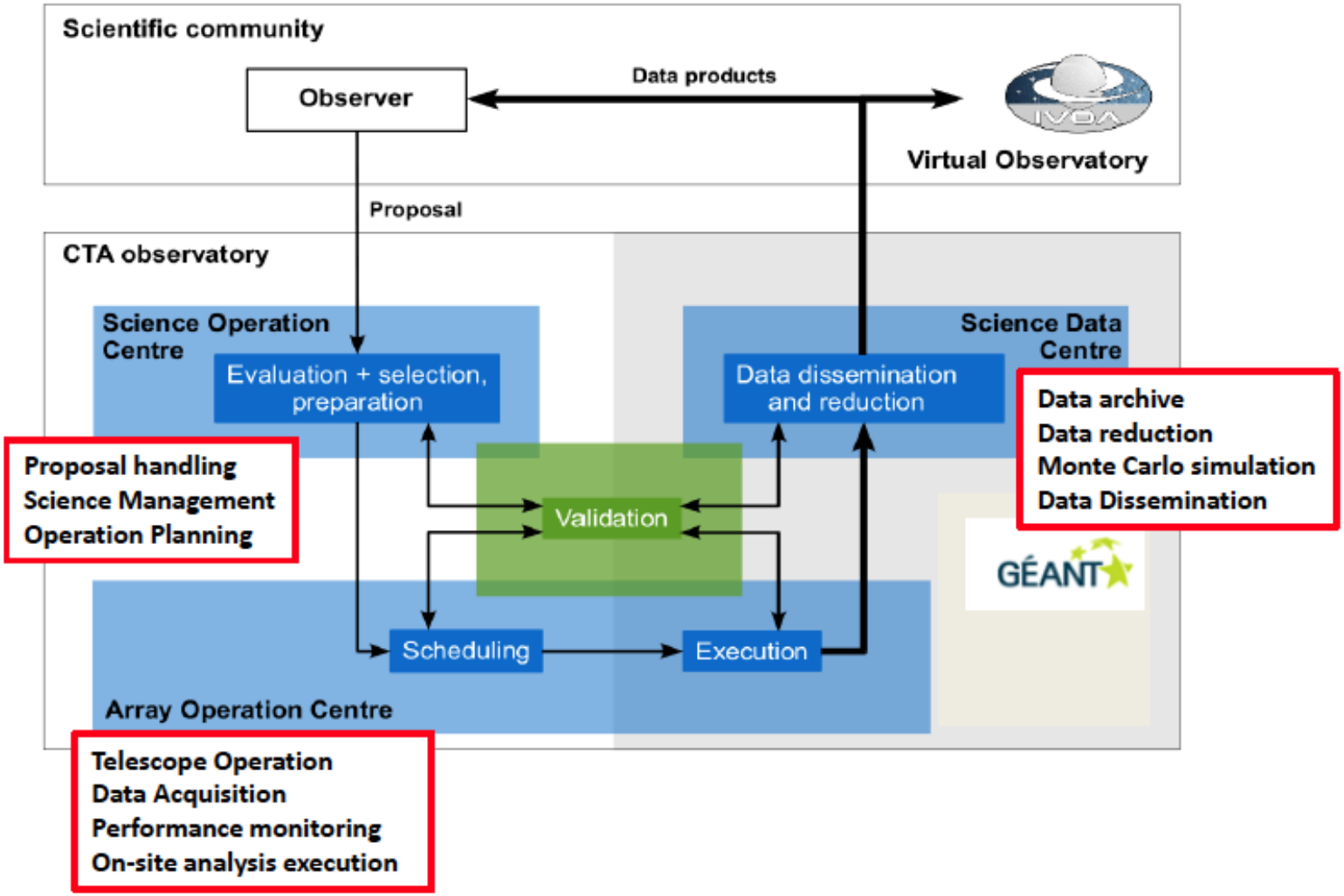
CTA NORTH



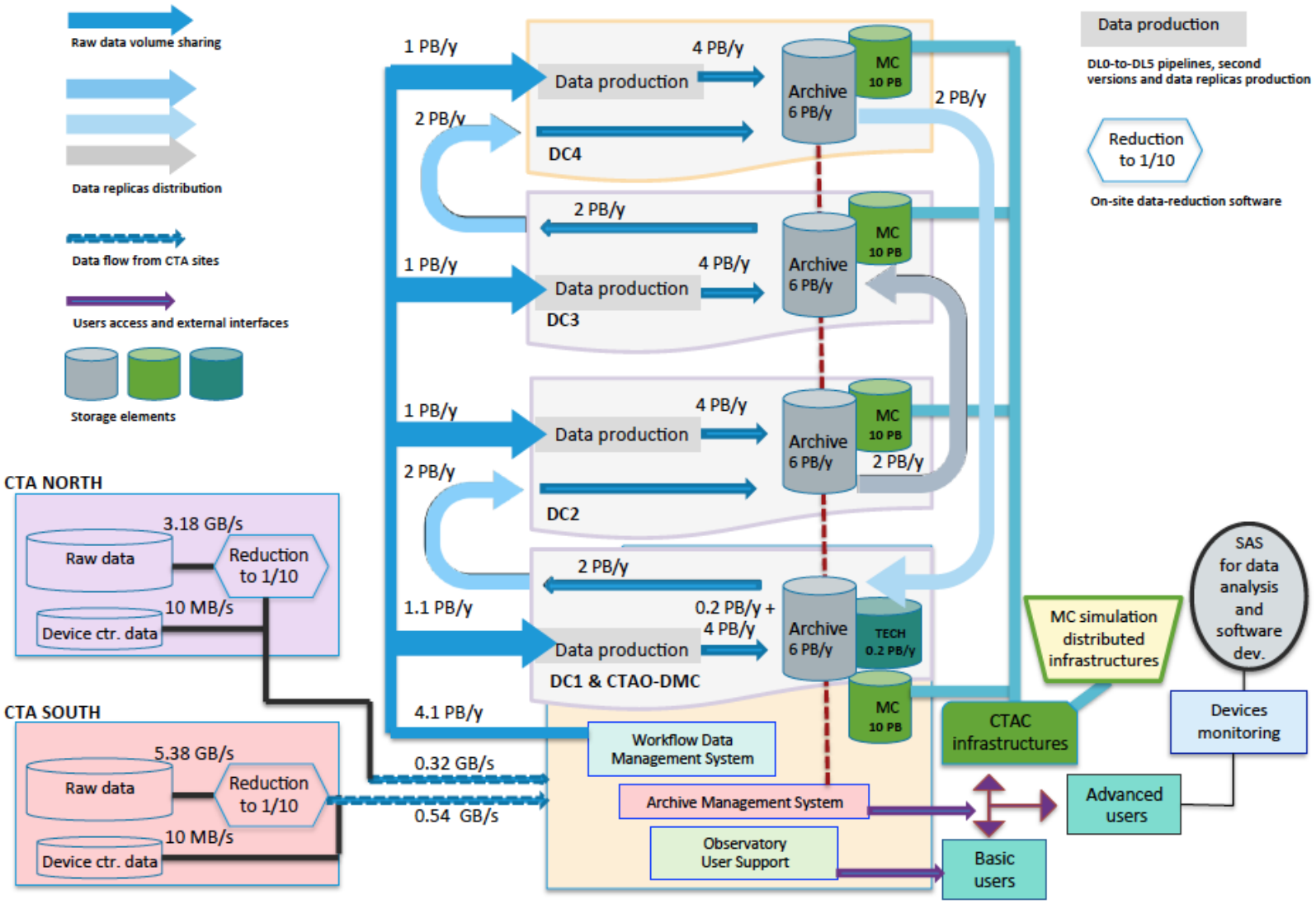


CTA as an Open Observatory

CTA will be a PB project operated as an observatory open to a large scientific community.



CTA Data Flow



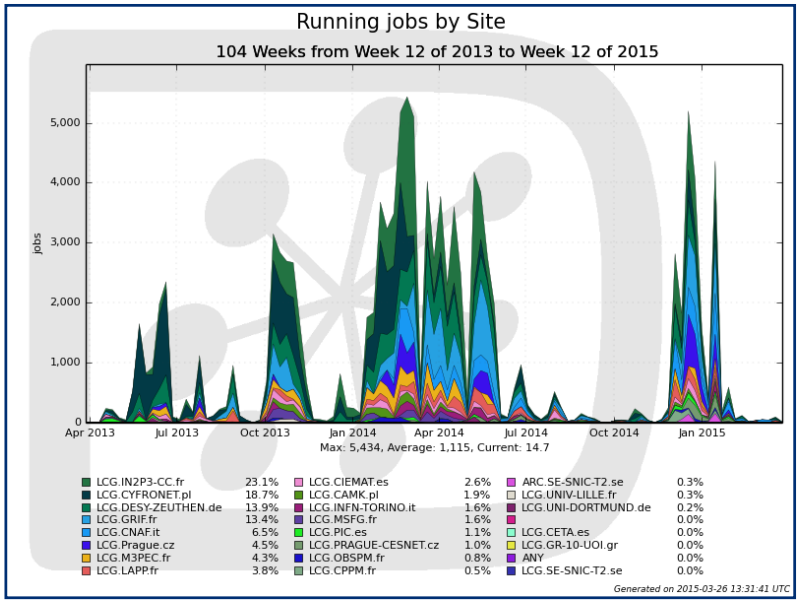
MC Production 2014-2015

- **About 20 centres** currently support the CTA VO. They are classified as:
  - ***Production centres*** if they provide CPU only **-> *used for MC production***
  - ***Analysis centres*** if they provide both CPU and storage resources **–> *used for MC production and Data Analysis***

CPU usage by Centre

- **Total used 2014: 94 M HS06**
- **87% CPU for MC production**
- **13% CPU for Users Analysis (7 active users)**
- **(2.6 M executed jobs in 2014)**
- **Total pledged 2015: 143 M HS06**

50% FRANCE; 20% POLAND; 14% GERMANY
8% ITALY; 4% SPAIN; etc.



Storage resources at 6 Analysis Centres

- **Total pledged 2015: 1.4 PB**
- **About 400 TB more disk than in 2014**

Centre	Allocated Disk 2014 (TB)	Pledged Disk 2015 (TB)	Used Disk (TB)
CYFRONET-LCG2 (**PL**)	448	600	206
DESY-ZN (**DE**)	336	336	204
IN2P3-CC (**FR**)	190	270	98
GRIF (**FR**)	50	120	61
IN2P3-LAPP (**FR**)	60	100	51
INFN-T1 (**IT**)	30	110	15
Total	**1004**	**1426**	**635**

Without data compression: prototypes and the final array.

ASTRI => ~0.8 TB/Night
=>**~0.3PB** per year

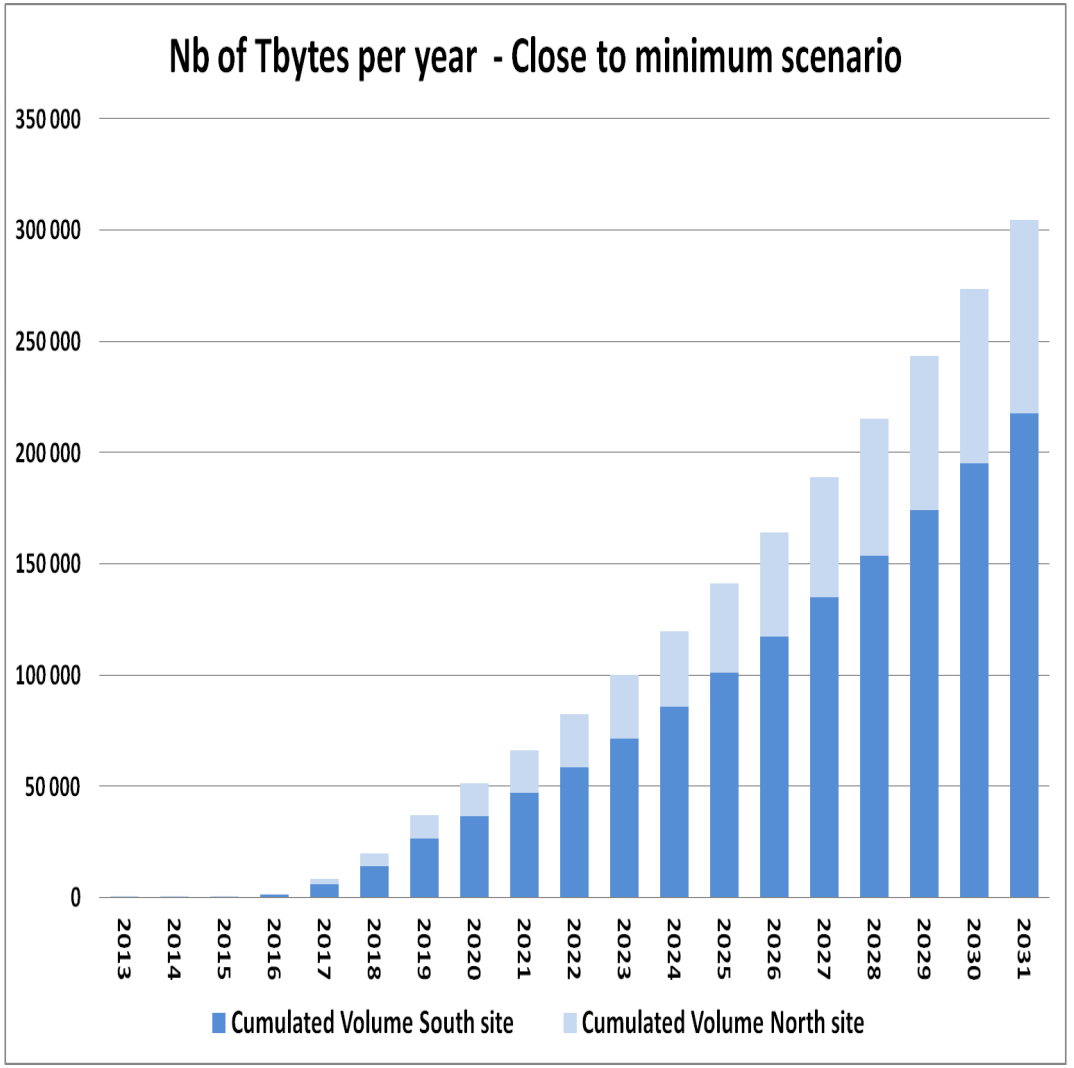
MA/SST => ~3TB/night
=>**~1PB** per year

CTA => ~24TB/Night
=>**~10PB** per year

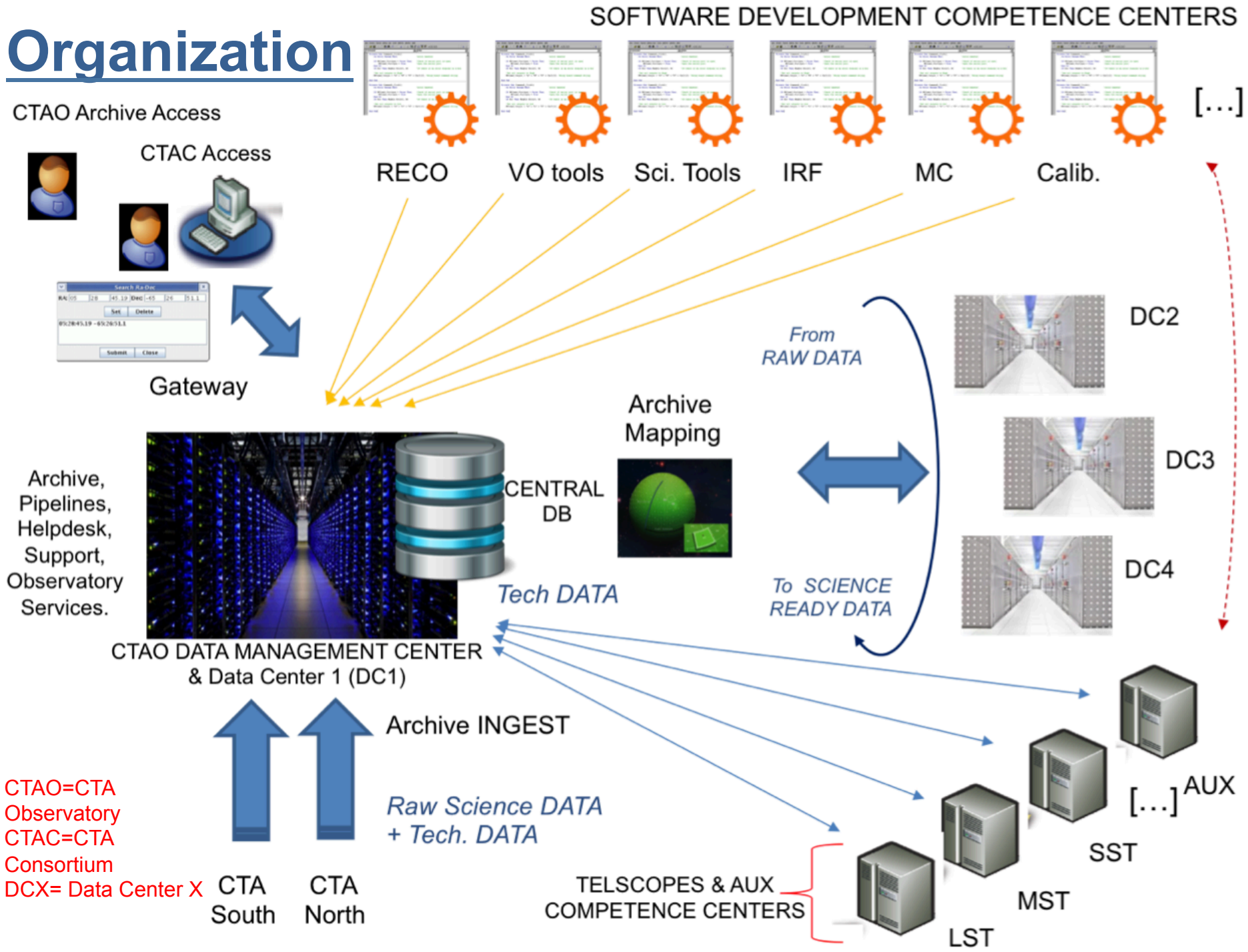
Thus **10 PB per year** including calibs, reduction and MC simulation data.

this is the OPTIMISTIC SCENARIO
The pessimistic one >100PB/year !

The CTA Archive system must store, manage, preserve and provide easy access (IVOA Access) to a such huge amount of data for a long time.

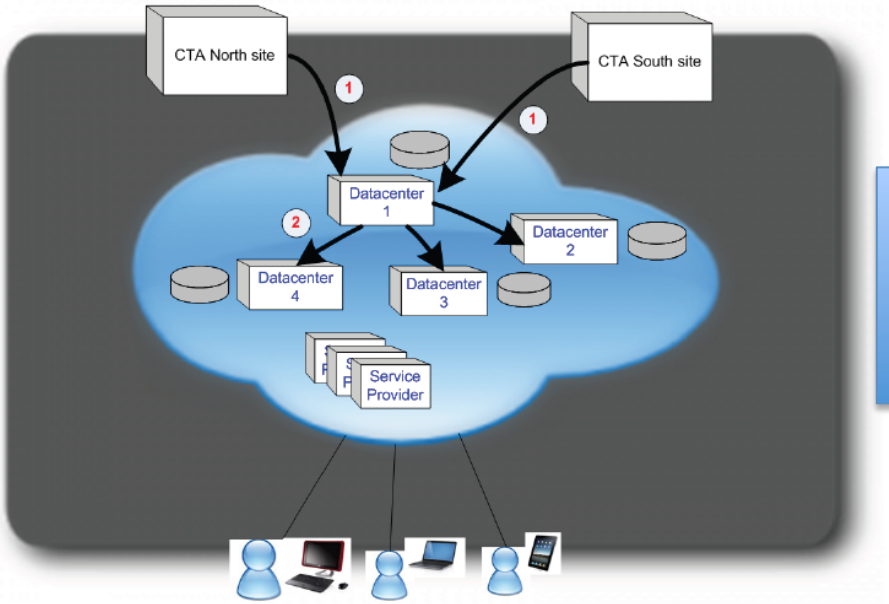


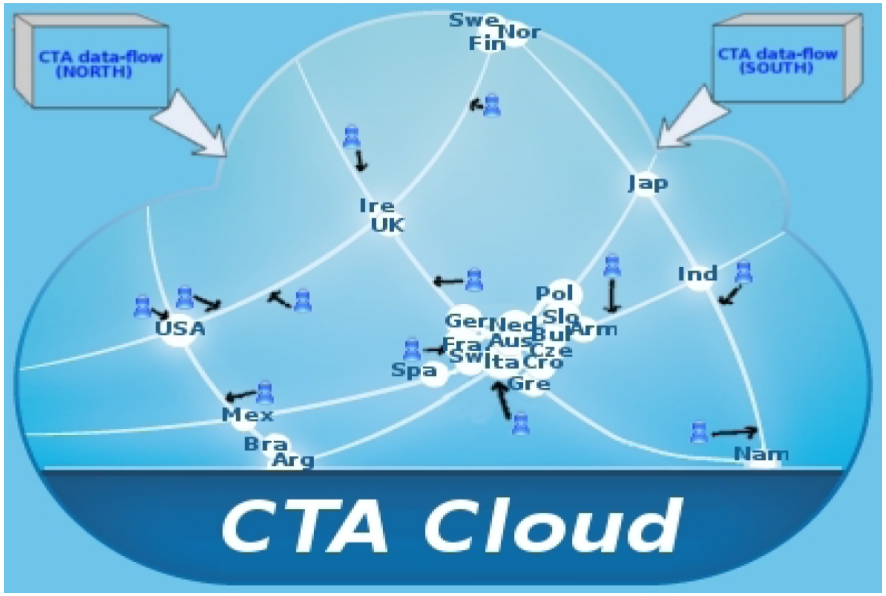
Organization



Toward CTA Preproduction: Distributed Archive System

(from CTA-DATA-TDR, v1.0, January 19, 2015)





- High Performance Storage System (massive I/O)
- High Throughput Storage System (optimized transfer _vs_ minimize latency)
- Long-term durable, persistent data storage
- High level of availability, low access latency

- Secure access to data (different level of access)
- Linear scalability (dimension increase with time)
- Hardware No-Single Point of Failure (no-SPOF)
- High Performance and fault tolerant DB (handle BigData)
- Minimize costs (costs _vs_ benefits)

To manage all this data-flow we identified a **Distributed Archive System** as a cloud resource composed by several storage entities (NODES).

An hardware/software entity

Storage Node (SN)
+
Distributed File System (DFS)
+
Distributed Database (DD)
+
Distributed Computing (DC)

INDIGO - DataCloud

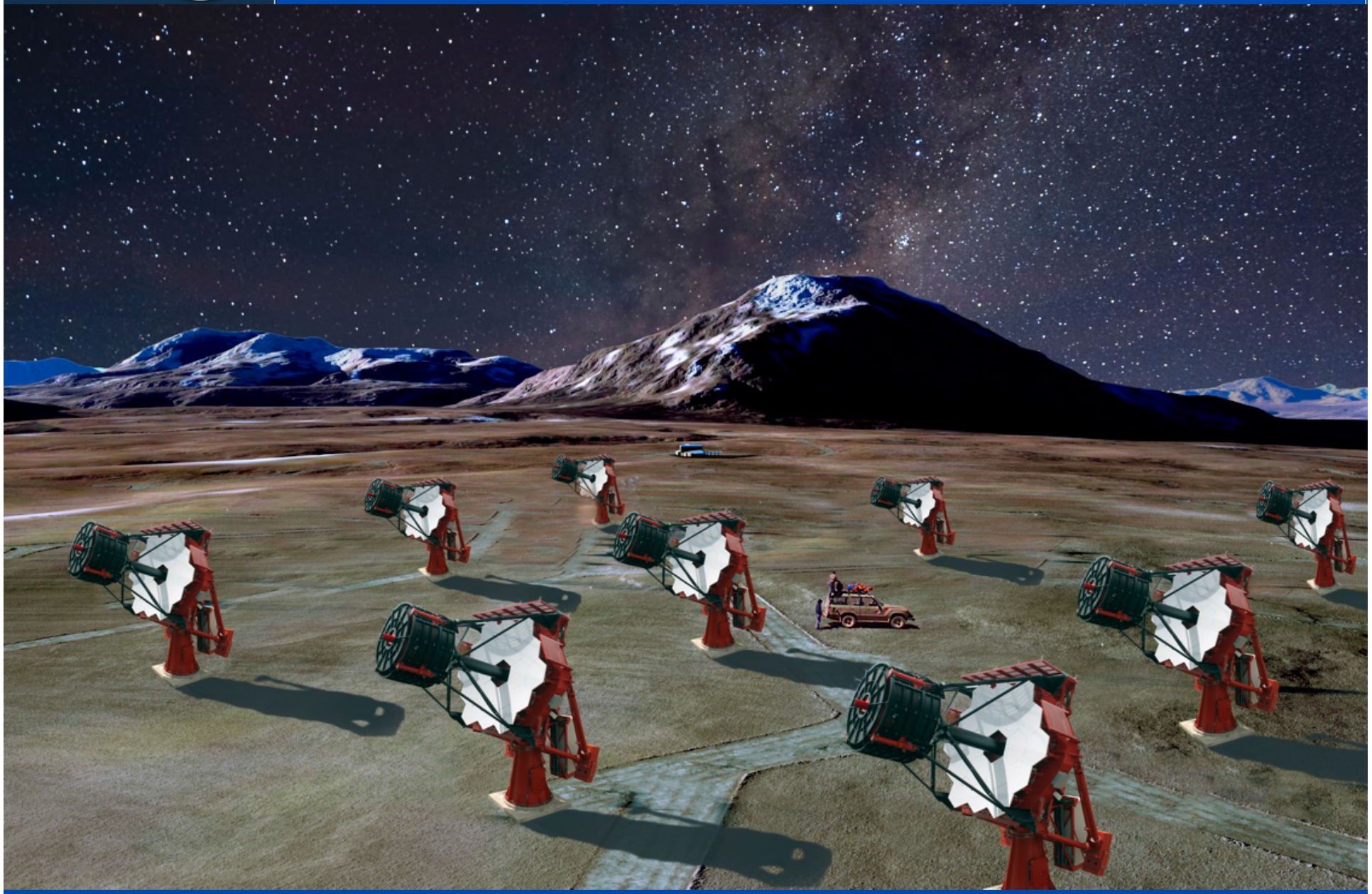
This is the "building-block" for a scalable distributed cloud storage.
We are developing the ASTRI prototype Archive as single SB.
→ several SBs → network of distributed resources (**CTA Cloud**) →
fault toleracy, geographical distribution (and commitment) for CTA partners.

Project ASTRI in a nutshell

- Project funded in 2010-2014 by the Italian MIUR “Progetto Bandiera” (> 40 FTE)
- Now the project continues with the support of MIUR (“Progetto Bandiera extension”) and MISE (“Industrial Astronomy” program) with the participation of Universities from South Africa and Brazil
- End-to-end SST-2M prototype:
  - Validation and commissioning of the telescope via Cherenkov astronomical observation
- End-to-end implementation of a mini-array (# 9) of SST-2M (pre-production) at the CTA southern site:
  - Validation and commissioning of the array (including trigger and SW) via Cherenkov astronomical observations, first CTA scientific data
  - Aiming at the construction of 35 out of the 70 SST units of the CTA southern array

ASTRI SST-2M Prototype

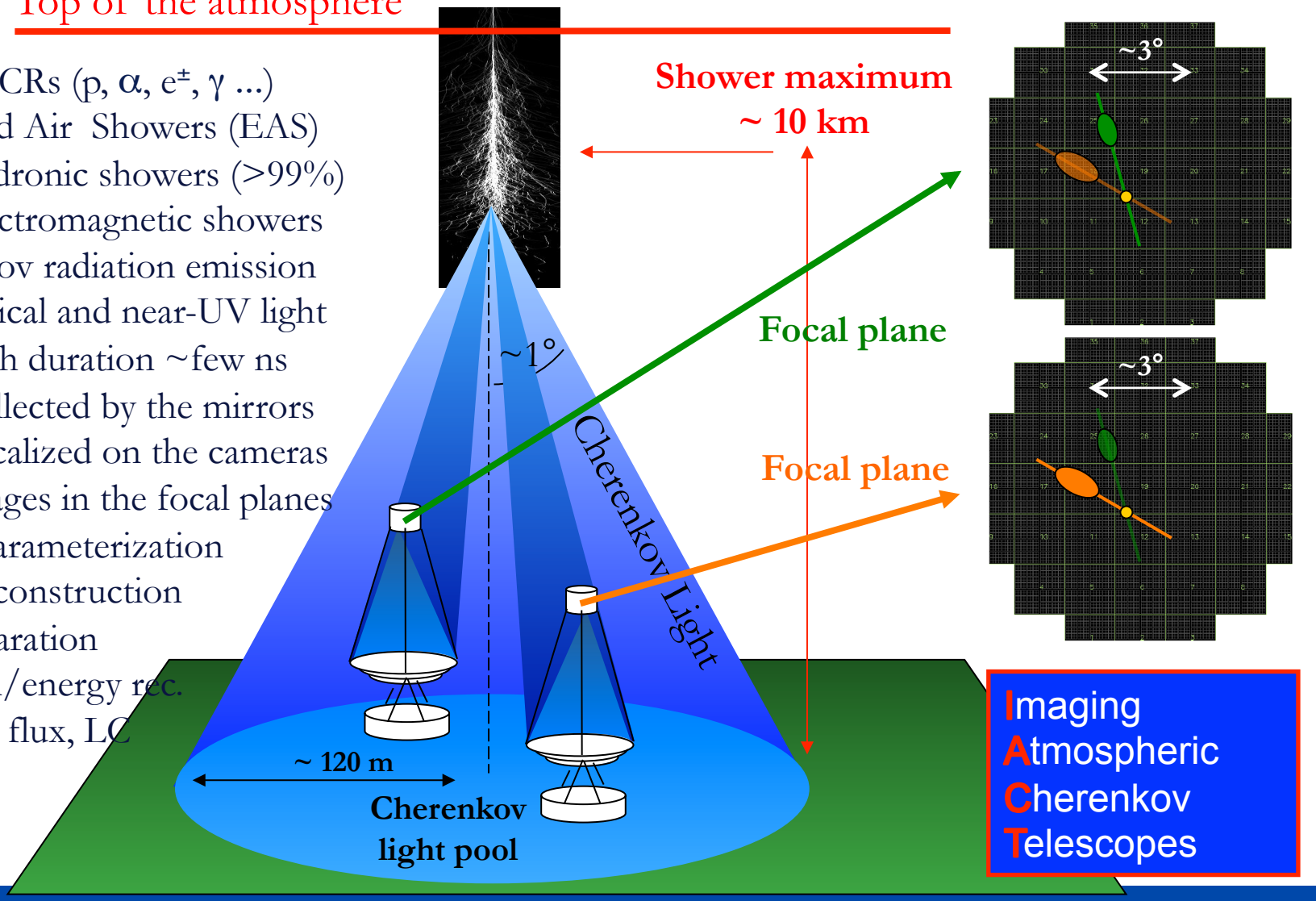
ASTRI/CTA Mini Array



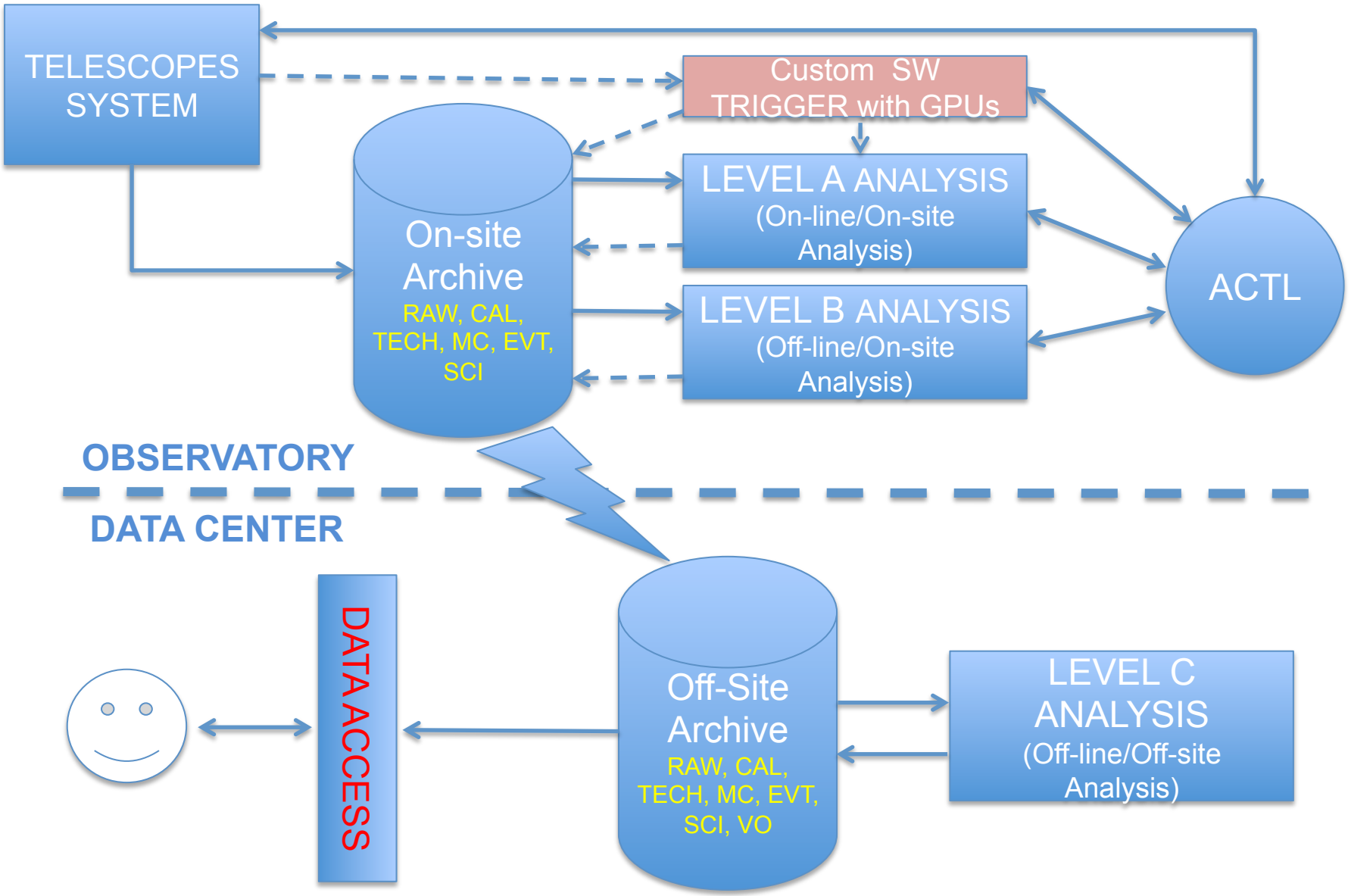
- Introduction: CTA & ASTRI
- Data Analysis Software.
- Arc

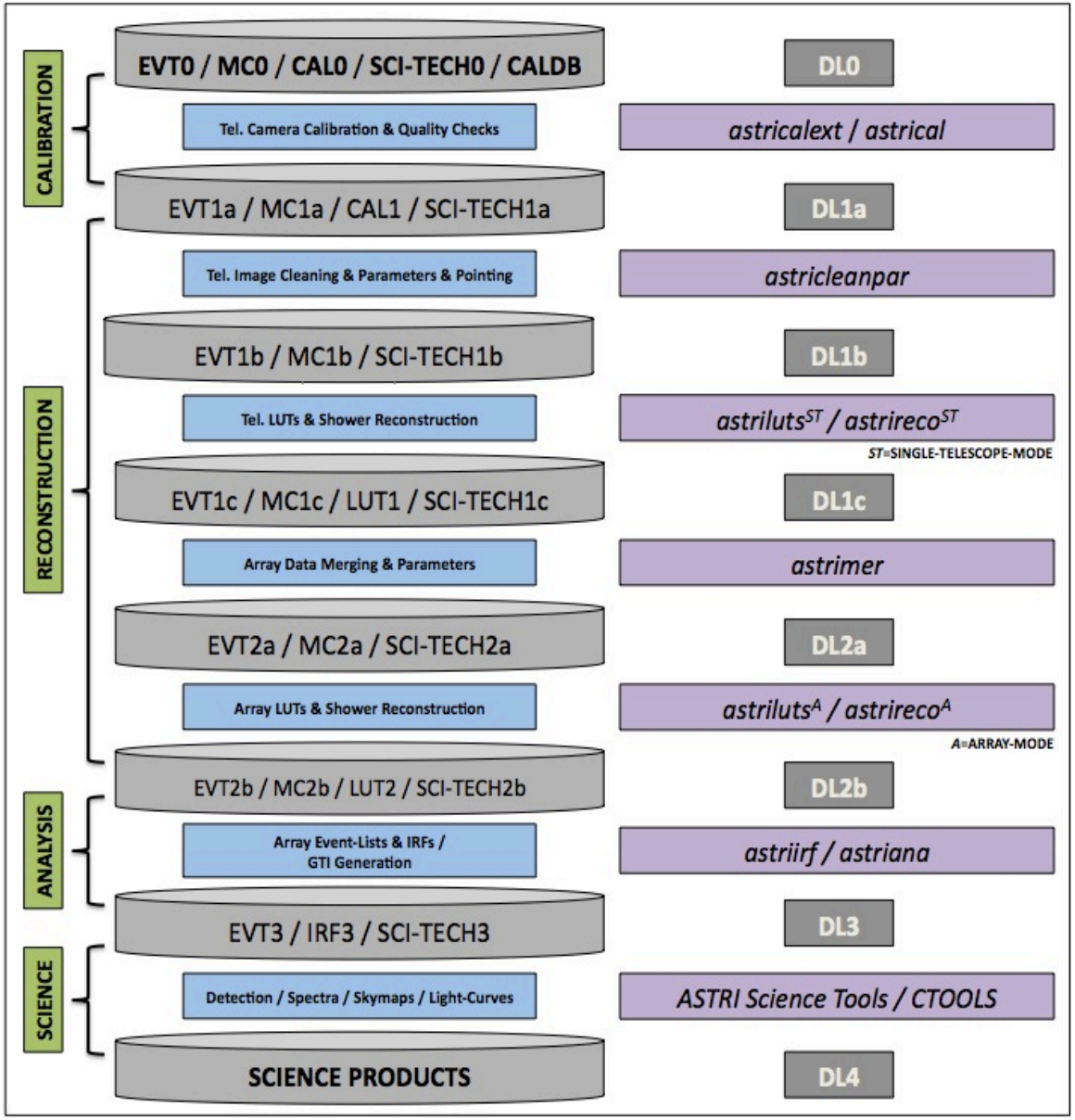
Imaging Air Cherenkov Technique

- Primary CRs (p, α, $e^{\pm}$, γ ...)
- Extended Air Showers (EAS)
  - Hadronic showers (>99%)
  - Electromagnetic showers
- Cherenkov radiation emission
  - optical and near-UV light
  - flash duration ~few ns
- Light collected by the mirrors
- Light focalized on the cameras
- EAS images in the focal planes
- Image parameterization
- Array reconstruction
- γ/h separation
- direction/energy rec.
- γ signal, flux, LC



ASTRI Prototype/Mini-Array Computing Model



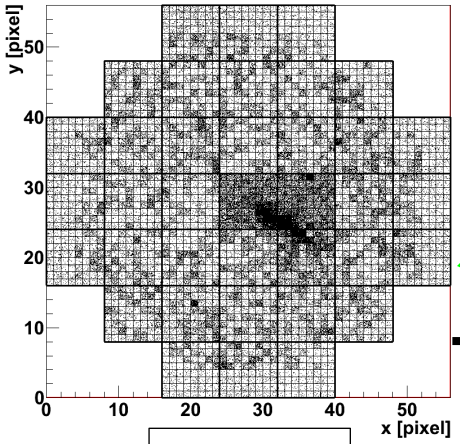


The ASTRI SST-2M Prototype and Mini Array Data Pipeline

- follows the general design and data model scheme defined in CTA Data Management
- manages **FITS data** (from DL0 to DL3) adopting CFITSIO/ CCFITS libraries;
- is written in **C++** (Unix environment) / **CUDA7** (for GPU/ARM coding);
- is developed in independent software modules linked by pipelines written in **Python**;
- will make use of *ad hoc* and official CTA Science Tools for final scientific results production (DL4).

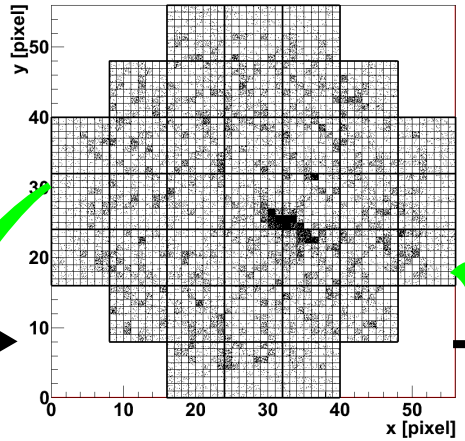
ASTRI SST-2M Prototype Data Reco & Analysis SW Status

DL0

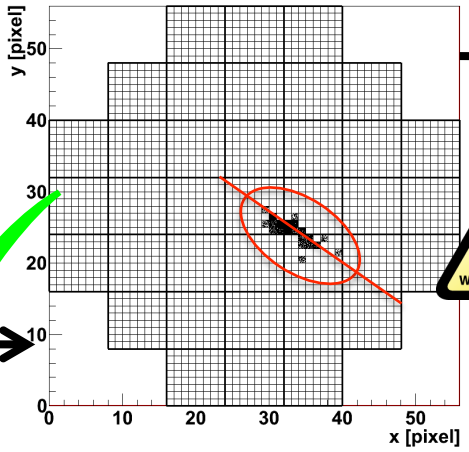


ADC Units

DL1a



DL1b/c



DL2a/b

(Event Lists + GTI and corresponding IRFs

WORK IN PROGRESS

DL3

Event Lists and corresponding IRFs

WORK IN PROGRESS

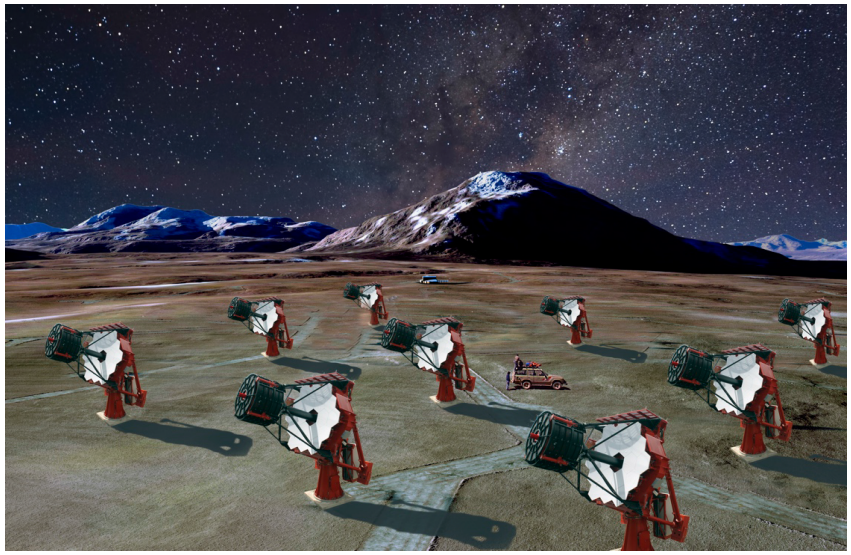
DL4

(SCIENCE PRODUCTS)

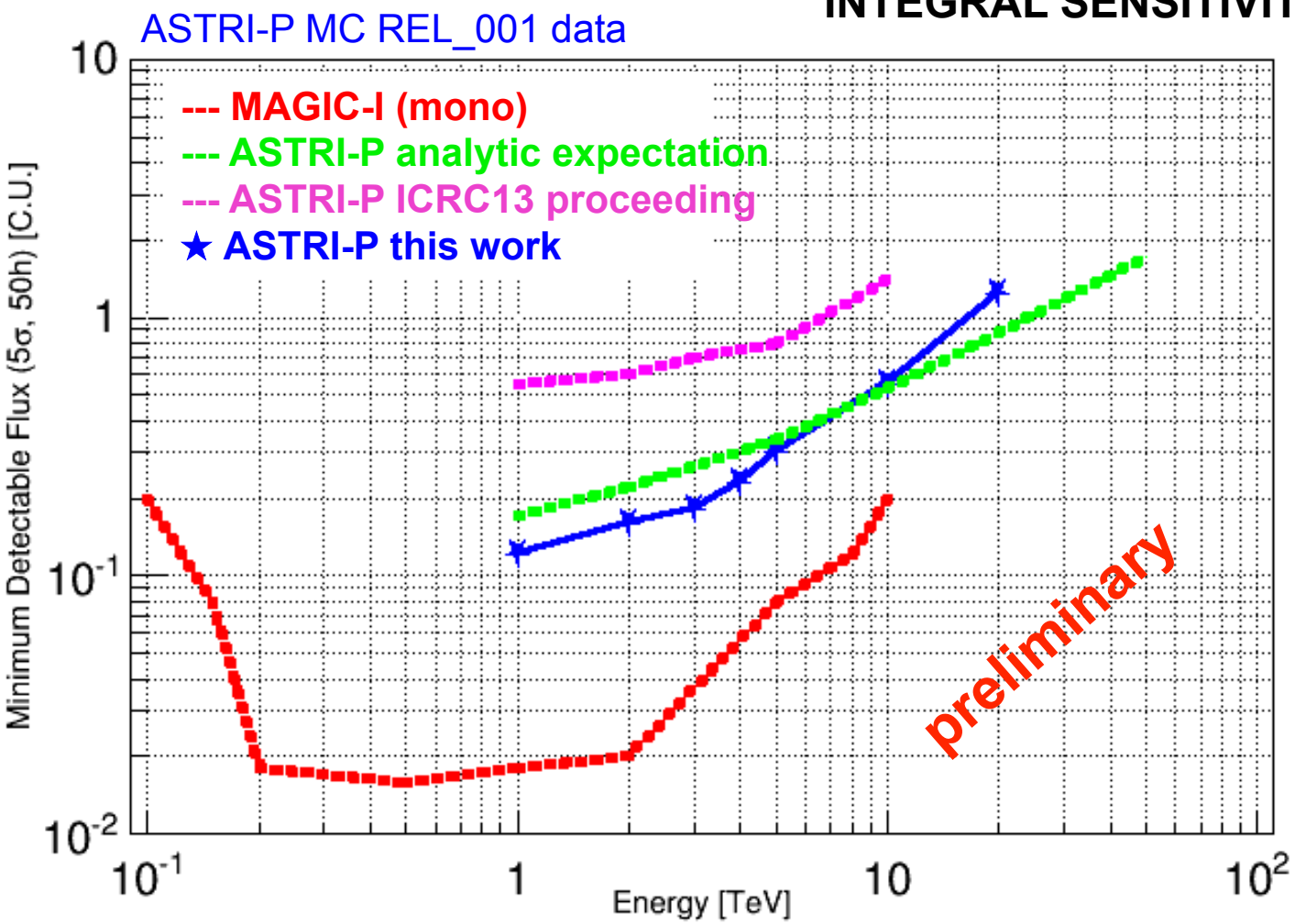
WORK IN PROGRESS

DL5

(OBSERVATORY PRODUTS)



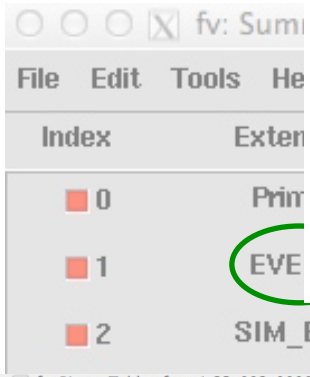
INTEGRAL SENSITIVITY

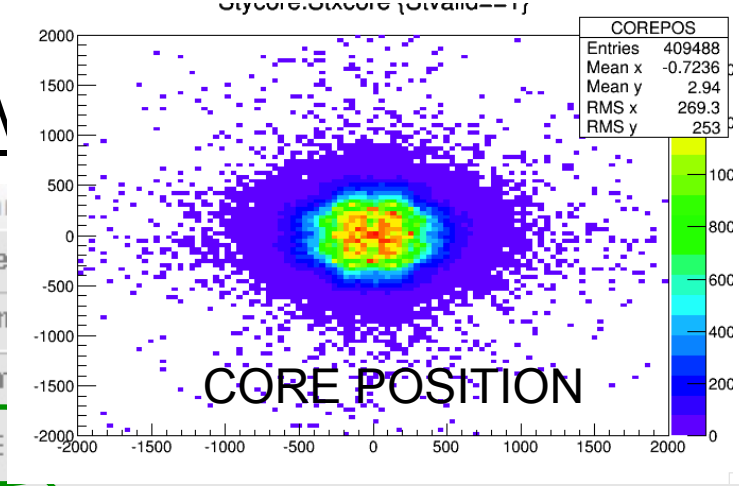


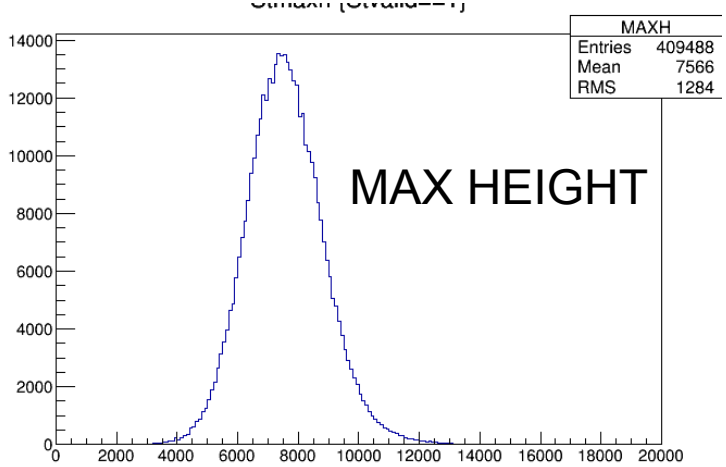
A-W RECONSTRUCTION (DL1c➔DL2a) Status

STEREO RECON.

- ASTRI







Binary 11 cols X 50000 rows

Header Hist Plot All Select

fv: Binary Table of astri_23_002_00001_R_009999_001_0201.lv2a[1] in /astri03/data1/lombardi/ASTRI/ASTRI_ANALYSES/ANALYSIS_REL002_ARRAY/SIM/DL2a/MC2a/gamma_ptsrc/arrayconf_1/teltype_1/rf1/test/

File Edit Tools Help

Select / All / Invert	STVALID 1L	STXDIR 1E [deg]	STYDIR 1E [deg]	STXCORE 1E [m]	STYCORE 1E [m]	STMAXH 1E [m]	STMINDIR 1E [deg^2]	STMINCORE 1E [m^2]	STMINMAXH 1E [m^2]	STIP 1E [m]	EVTNUM 1J	NTRIGTEL 1J	TRIGTEL 1J	TELTYPE 1J	NUSEDTEL 1J
Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify
1	T	1.116531E-01	5.010708E-02	6.028513E+02	1.422729E+02	0.000000E+00	2.414397E-17	1.697185E-09	0.000000E+00	6.194120E+02	1106	7	13	1	2
2	T	1.116531E-01	5.010708E-02	6.028513E+02	1.422729E+02	0.000000E+00	2.414397E-17	1.697185E-09	0.000000E+00	6.194120E+02	1106	7	1	1	2
3	T	1.180399E-01	-2.722386E-01	1.090776E+03	7.164079E+01	0.000000E+00	6.684230E-17	2.523765E-09	0.000000E+00	1.093126E+03	3507	3	13	1	2
4	T	1.180399E-01	-2.722386E-01	1.090776E+03	7.164079E+01	0.000000E+00	6.684230E-17	2.523765E-09	0.000000E+00	1.093126E+03	3507	3	15	1	2
5	T	-7.600708E-02	-4.709019E-02	4.399545E+02	-3.393379E+02	0.000000E+00	5.765386E-20	0.000000E+00	0.000000E+00	5.556169E+02	5303	10	16	1	2
6	T	-7.600708E-02	-4.709019E-02	4.399545E+02	-3.393379E+02	0.000000E+00	5.765386E-20	0.000000E+00	0.000000E+00	5.556169E+02	5303	10	15	1	2
7	T	-1.253837E-01	1.279560E-01	3.071549E+02	-4.125042E+02	0.000000E+00	8.248764E-02	2.842357E+05	0.000000E+00	5.142994E+02	5308	11	16	1	4
8	T	-1.253837E-01	1.279560E-01	3.071549E+02	-4.125042E+02	0.000000E+00	8.248764E-02	2.842357E+05	0.000000E+00	5.142994E+02	5308	11	15	1	4
9	T	-1.253837E-01	1.279560E-01	3.071549E+02	-4.125042E+02	0.000000E+00	8.248764E-02	2.842357E+05	0.000000E+00	5.142994E+02	5308	11	1	1	4
10	T	-1.253837E-01	1.279560E-01	3.071549E+02	-4.125042E+02	0.000000E+00	8.248764E-02	2.842357E+05	0.000000E+00	5.142994E+02	5308	11	13	1	4

Go to: Edit cell:

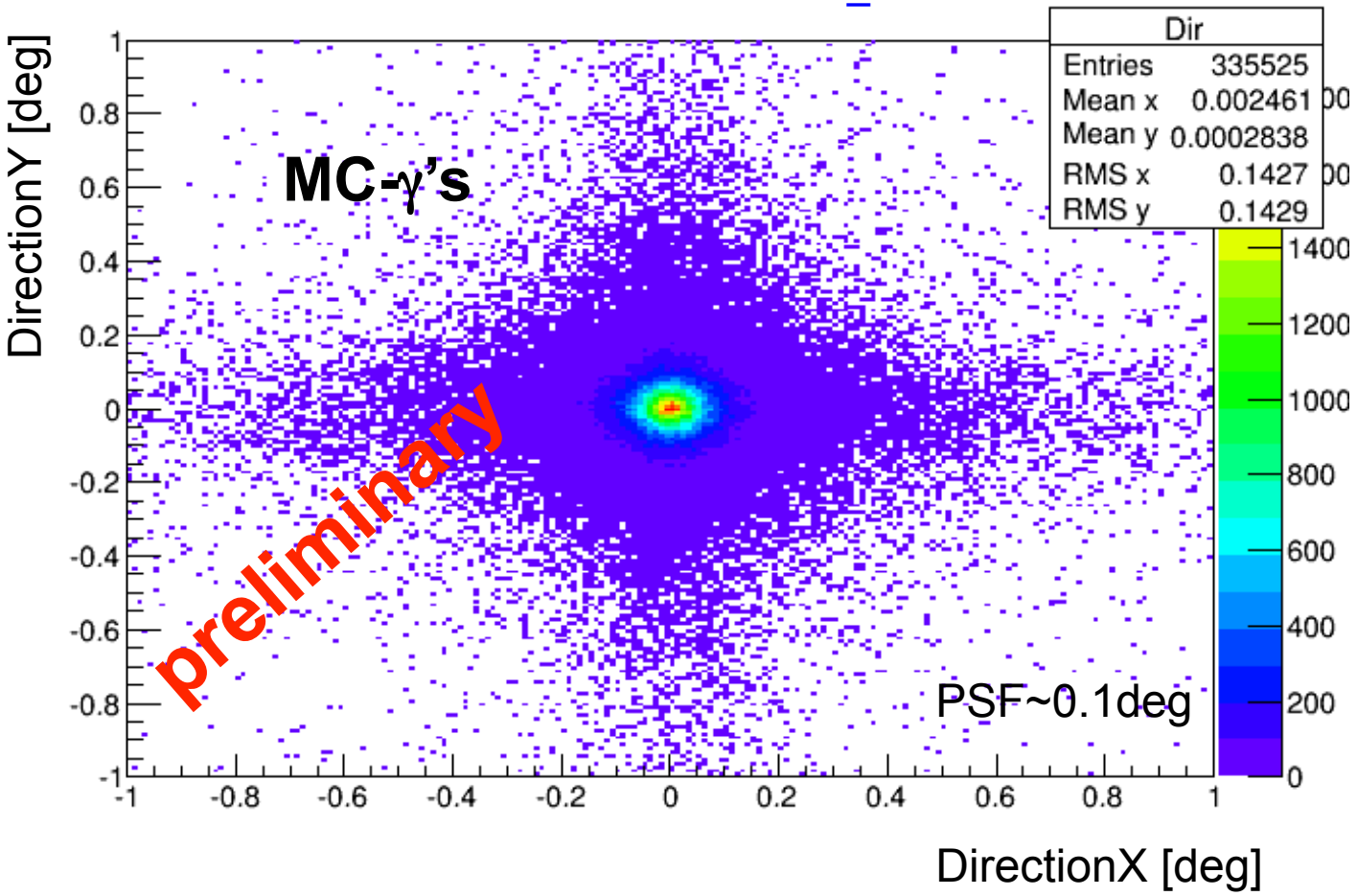
STEREO PARAMETERS

ARRAY INFORMATION

STEREO RECON.

DIRECTION ESTIMATION

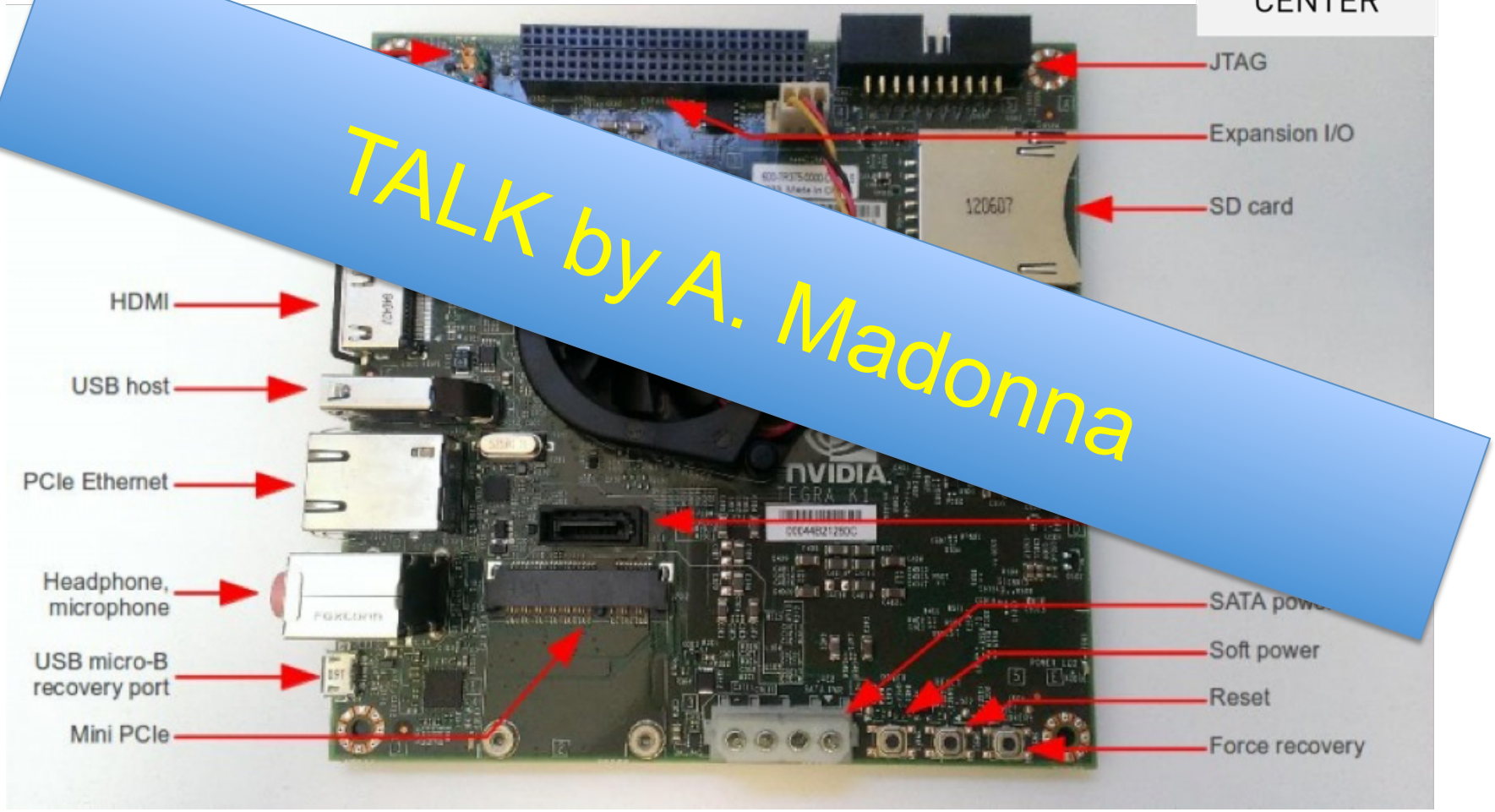
ASTRI-MA MC REL_002 data



Jetson TK-1

192 cores @ 192$

NVIDIA GPU RESEARCH CENTER



TALK by A. Madonna

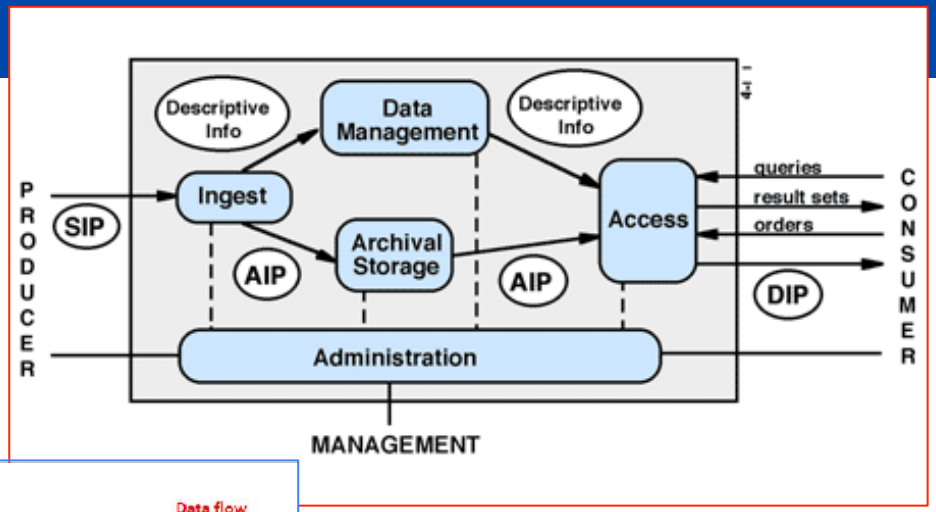
- Data reduction in <6 W (real-time) on Jetson TK1:
  - calibration (ADC → phe)
  - cleaning (single core)
  - image parametrization
- Photon list (img. parms → *E*, *hadronness*, *direction*)
  - via Breiman RF: running on Shark
  - distributed on GPU (K40)
  - test on DNN on-going

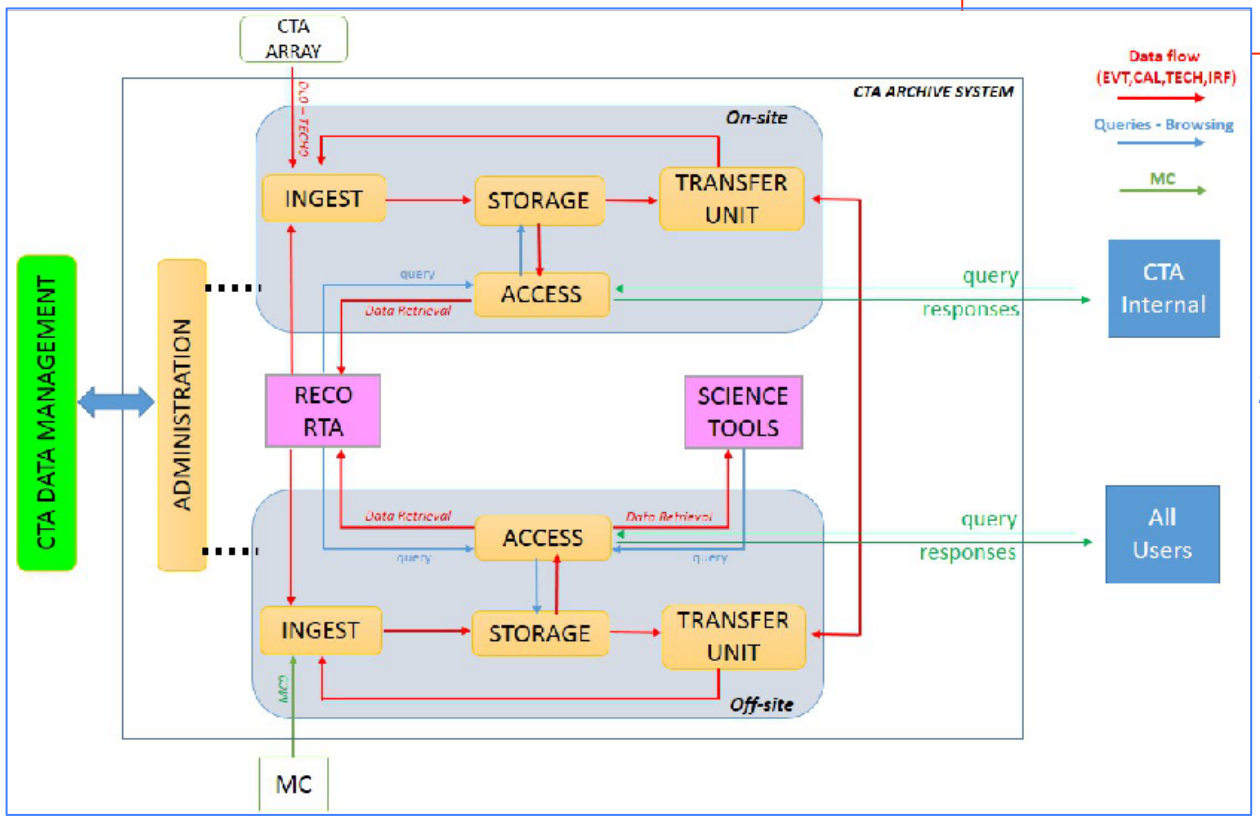
- 2kPxl × 4 B/Pxl × 1kHz = 8 MB/s
- 30 s batches = 240 MB
- 5 s I/O (FITS → input data; output data → FITS)
- 10 momenta ~ 20 kFLOP/Pxl ~ 40 GFLOP/1kEvt
  ~ 3J/1kEvt
- cleaning/clustering ~ 10 s on 1 core
- Average current over 30 s (2 batches): 0.43 A
- Average power over 30 s (2 batches): 5.2 W
- 2 kEvt/s on current version (twice the reqs)
- Expected 4 kEvt/s on TX1 (same power, but some tweaks)

- Introduction: CTA & ASTRI
- Data Analysis Software.
- Archive.

CTA Archive

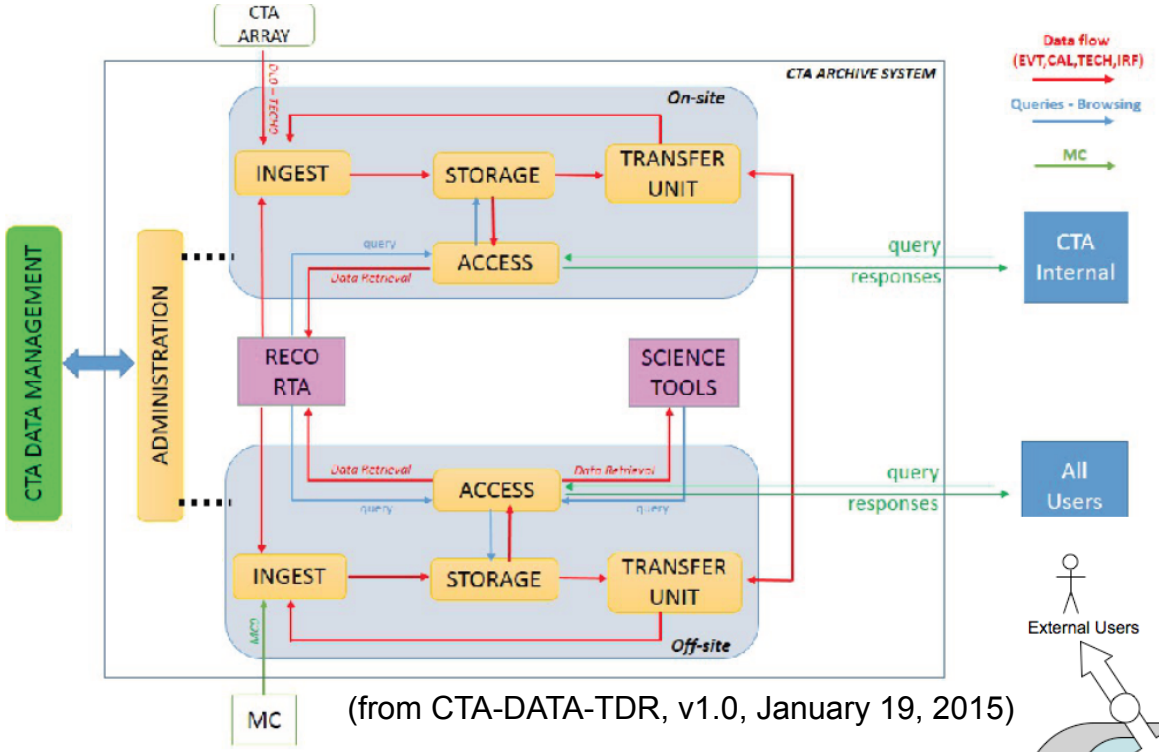
1 – Reference Model: OAIS





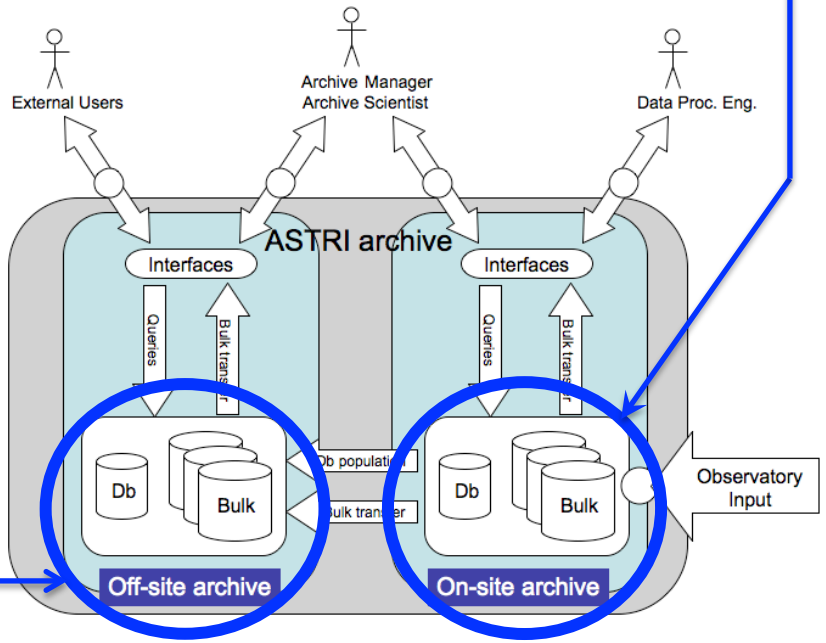
2 – CTA Archive Organization

CTA Archive System Concept Design & ASTRI

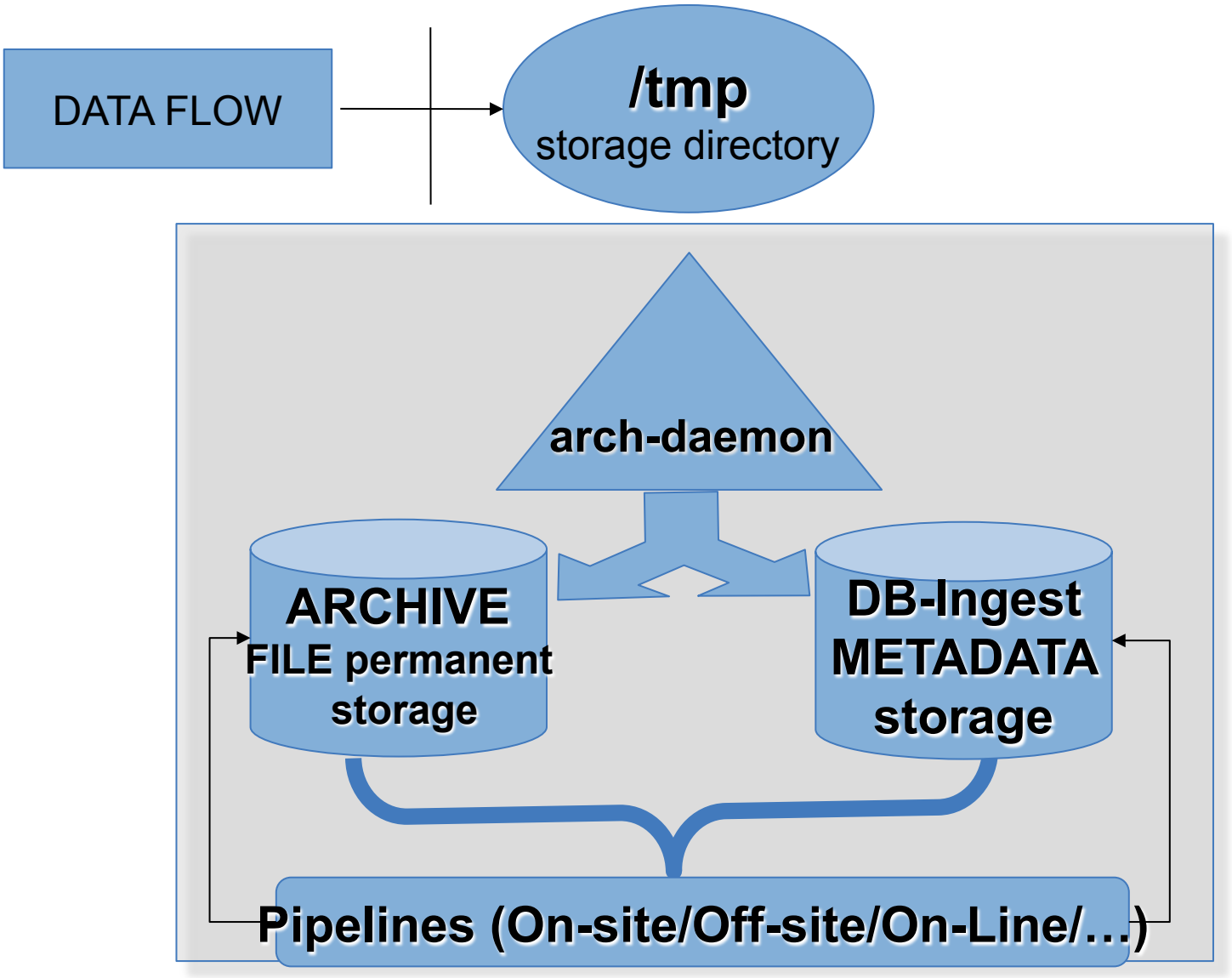


(from CTA-DATA-TDR, v1.0, January 19, 2015)

- ***ASTRI On-site archive*: *temporarily* contains DL0 data files (EVT0/CAL0/ TECH0) and products of the *in situ* OPA and DOPA.**



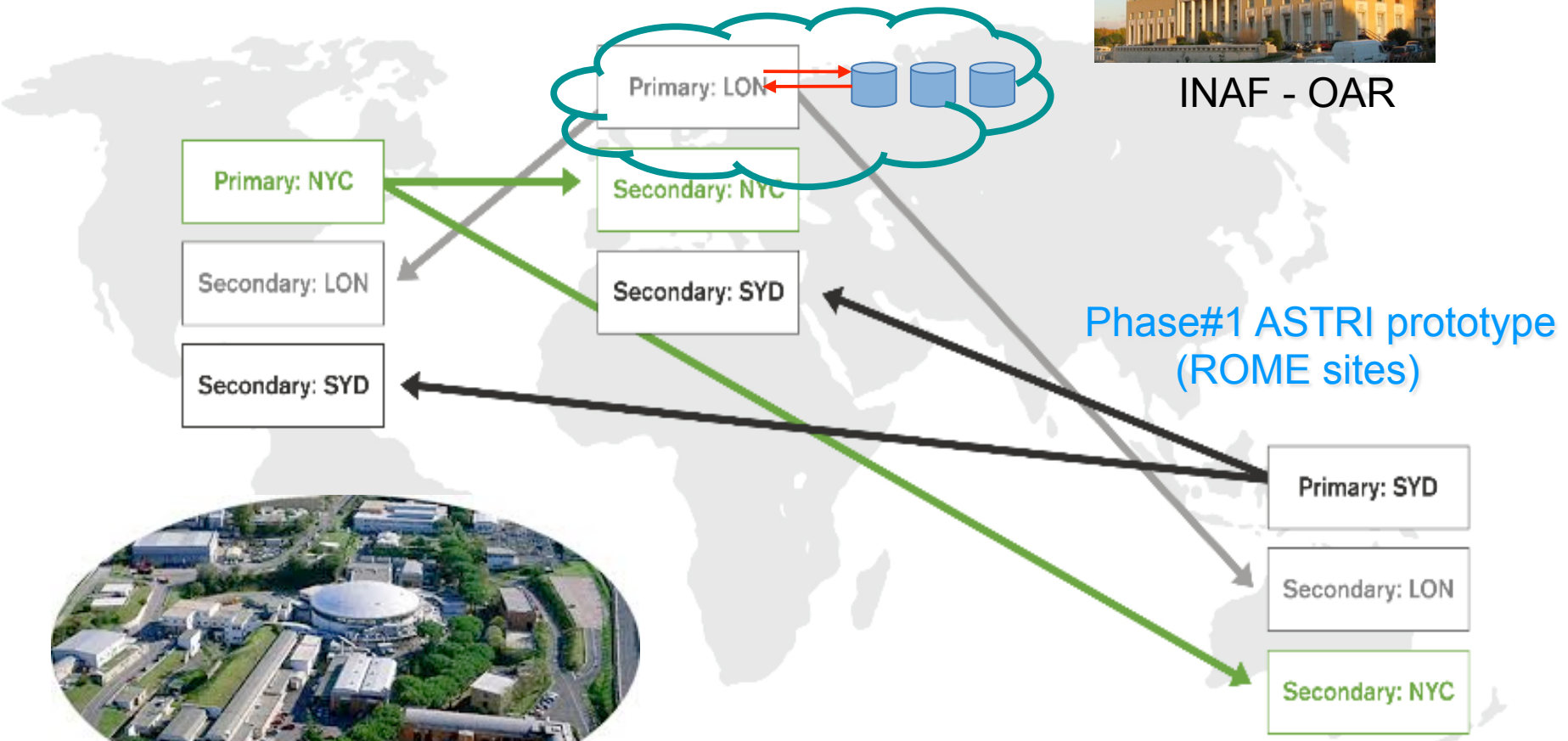
- ***ASTRI Off-site archive*: contains all RAW data, data reco products and high-level science products**


DATA FLOW
/tmp
storage directory
arch-daemon
ARCHIVE
FILE permanent
storage
DB-Ingest
METADATA
storage
Pipelines (On-site/Off-site/On-Line/…)

ASTRI ARCHIVE SYSTEM Deployment Phases

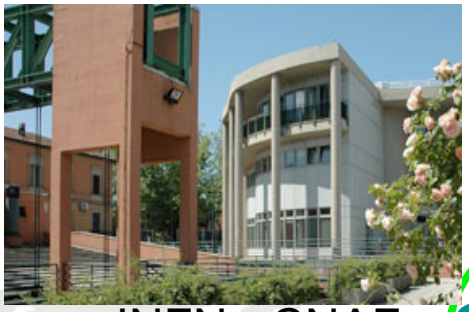
DATA CENTERS

INAF - OAR



INFN - LNF

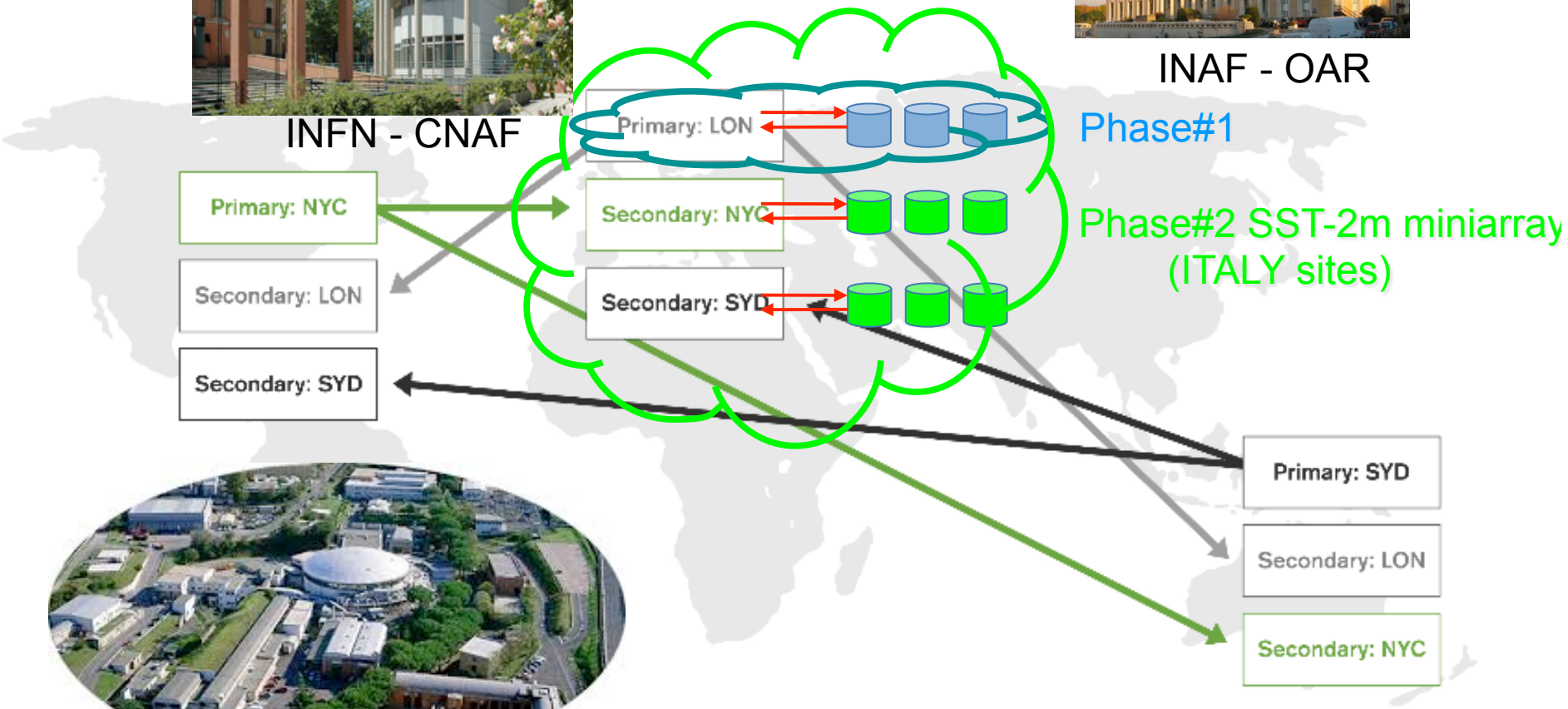
ASTRI ARCHIVE SYSTEM Deployment Phases



INFN - CNAF

DATA CENTERS

INAF - OAR

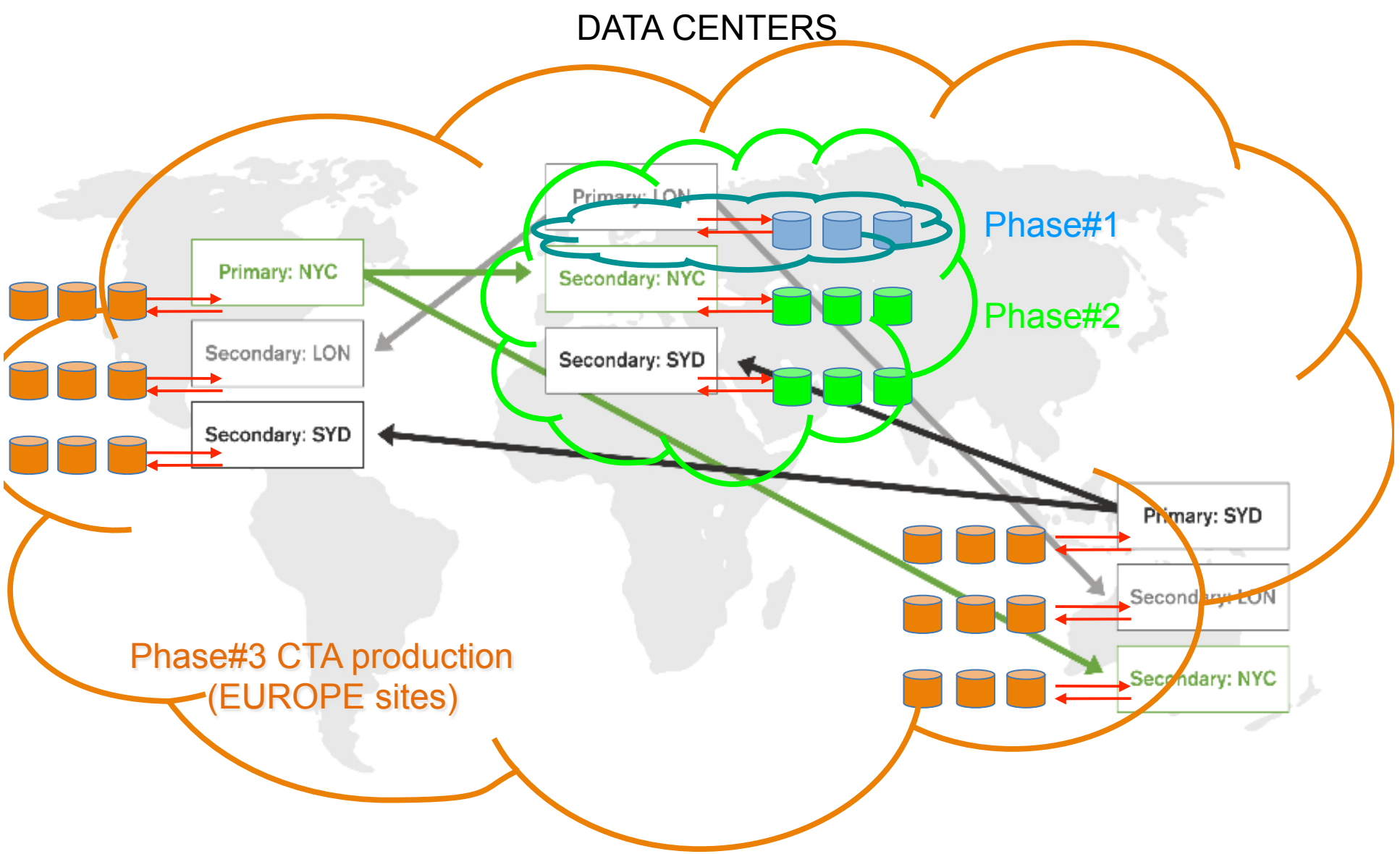


Phase#1

Phase#2 SST-2m miniarray (ITALY sites)

INFN - LNF

ASTRI ARCHIVE SYSTEM Deployment Phases



LEV-A: first prototype of the CTA archive is going to be realized in order to archive and manage all data products coming from the CTA/ASTRI SST-2M prototype.The CTA/ASTRI SST-2M archive system will provide full access to several archive users to the whole data set in order to archive, reduce, analyse and publish scientific data in a PI-oriented platform.

LEV-B: the second level is an archive system prototype capable of efficiently manage (for archiving and data-processing activities) data coming from several telescopes (mini-array) with several and (possibly) different kind of cameras technologies.

LEV-C: CTA Archive core. It is the final prototype version totally compliant with CTA requirements.

- The ASTRI Project is now moving from the prototype construction phase to the CTA pre-production phase.
- A complete data analysis and archiving system is under construction.
- Several solutions are tested in order to be suitable and finally selected for CTA.