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

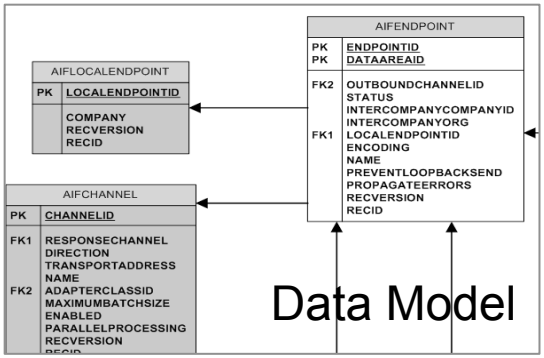
CTA & ASTRI DATA

From ASTRI Prototype and Mini-Array to CTA.

L. Angelo Antonelli – INAF-OAR & ASDC

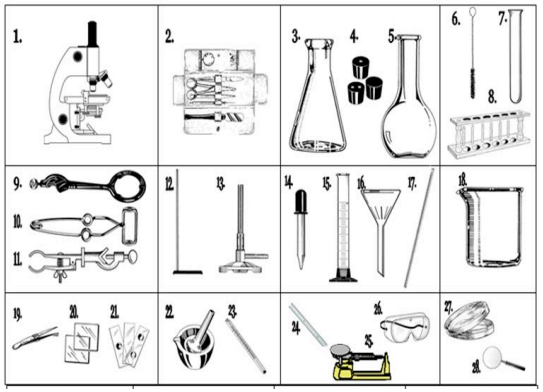
for the ASTRI Collaboration & the CTA Consortium

DATA MANAGEMENT



Data Model

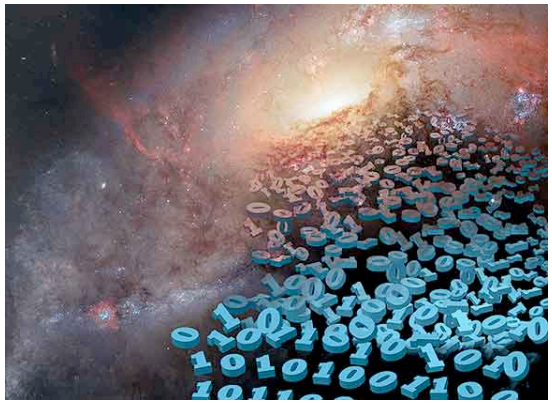
Reconstruction Pipelines



Science Tools

Archive(s)

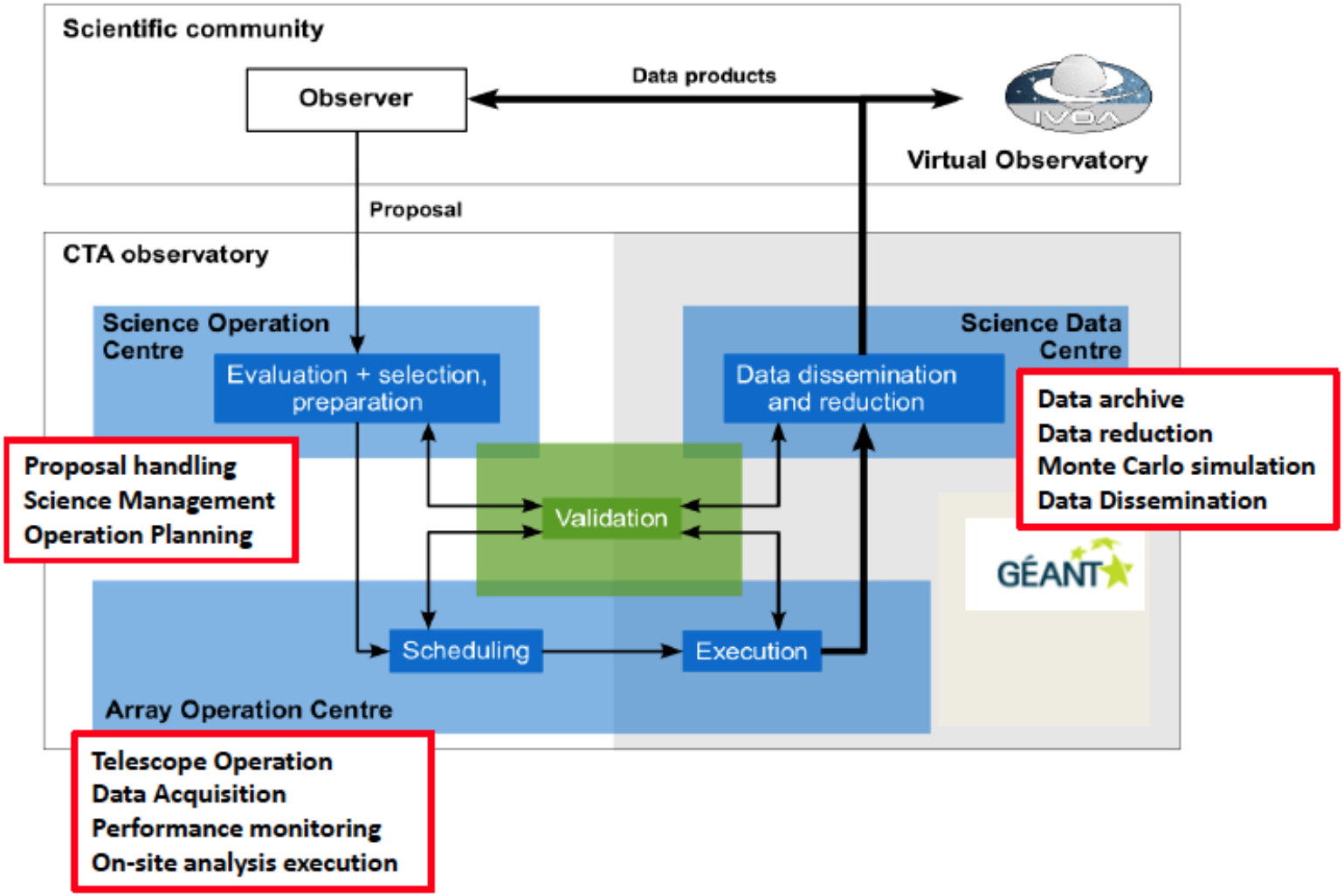
Simulations



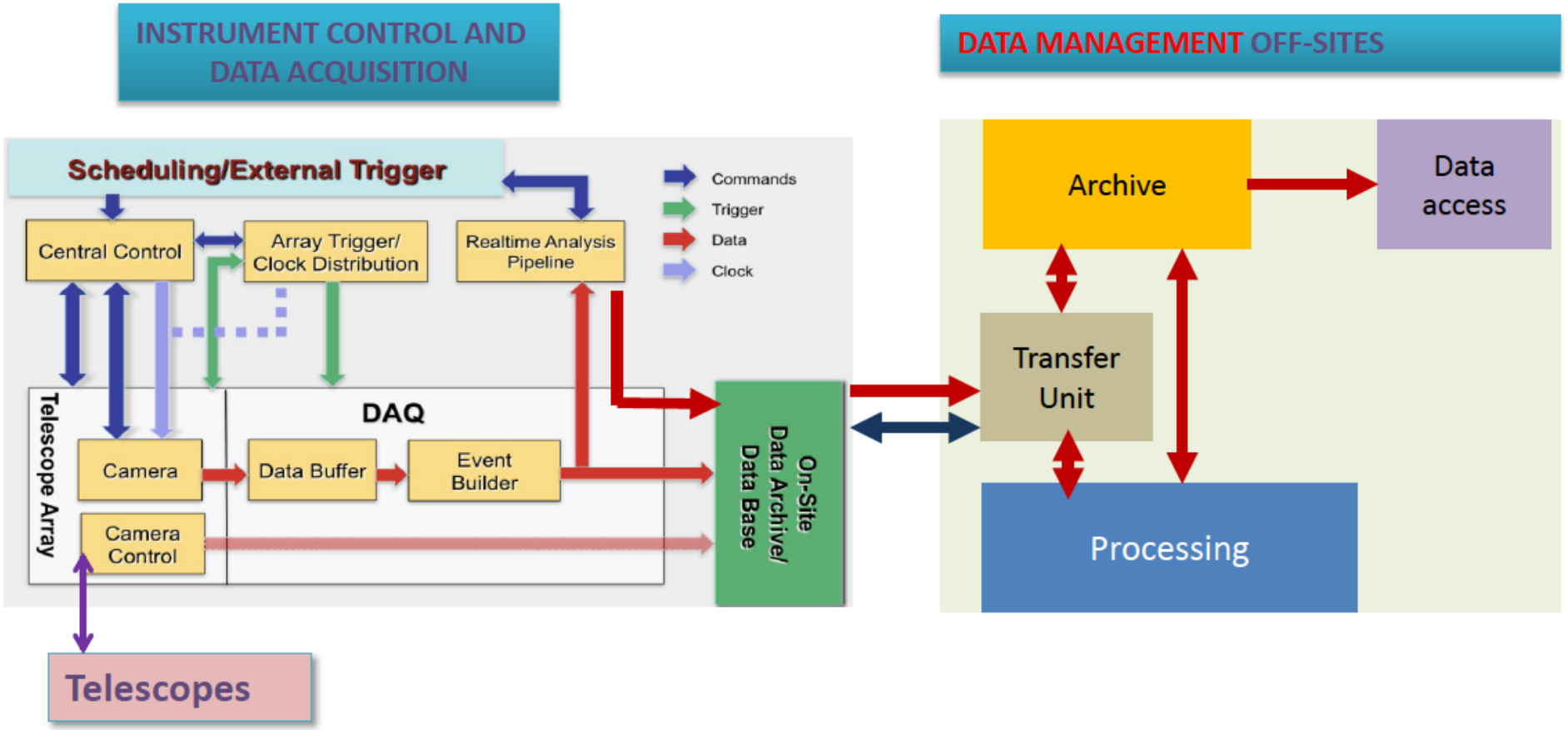
Data Access

CTA as an Open Observatory

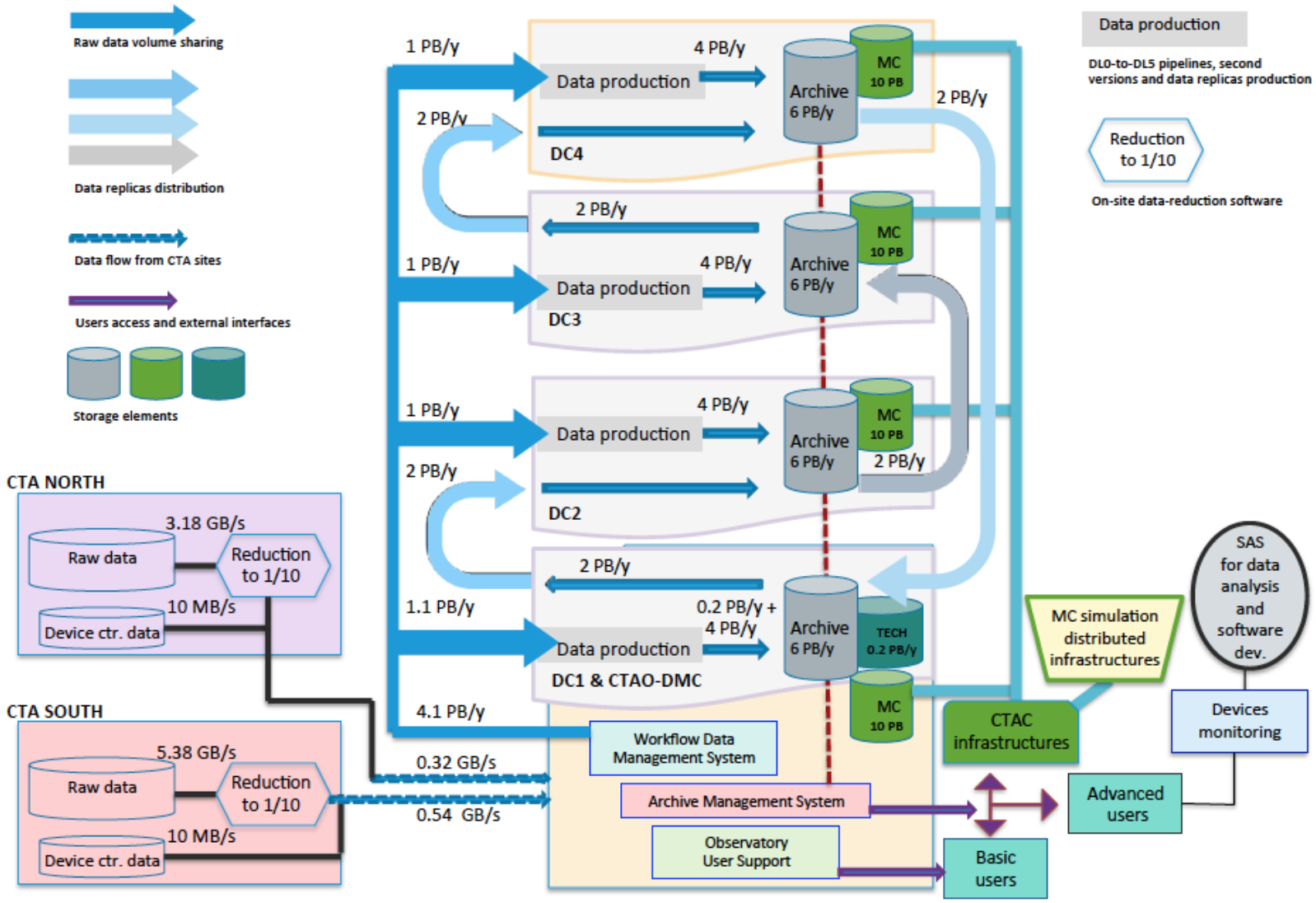
CTA will be a PB project operated as an observatory open to a large scientific community.



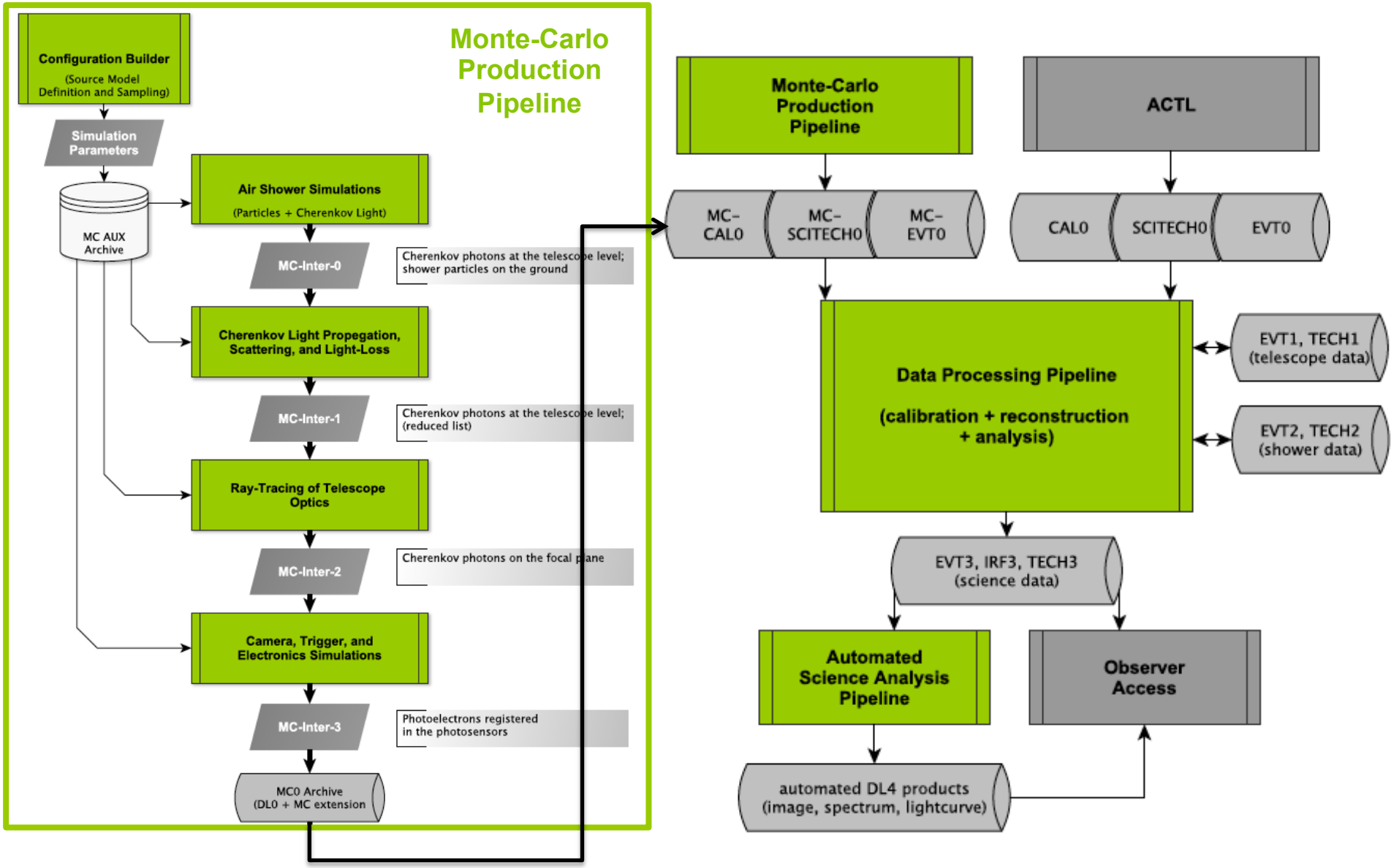
CTA DATA MANAGEMENT DESIGN OVERVIEW



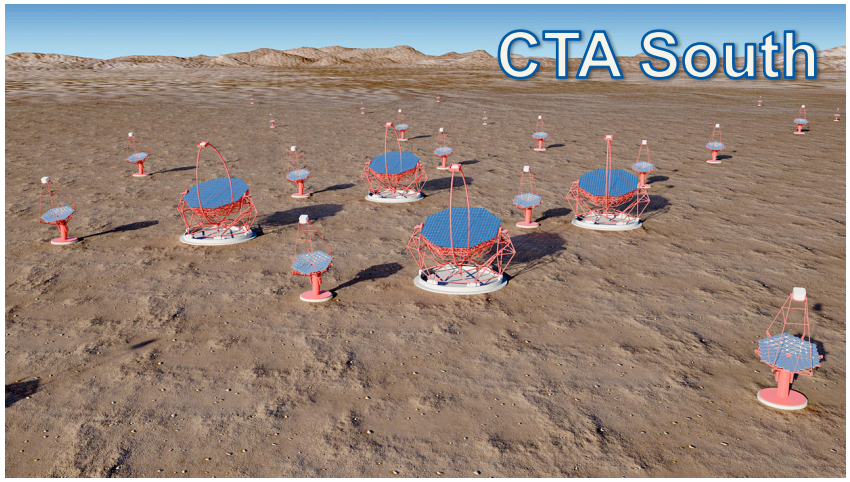
CTA Data Flow



CTA Pipelines



Cumulative Amount of Storage and Cores for CTA



Year	2017	2018	2019	2020	2021
Storage capacity (PB)	9	24	35	45	55
Computing (Milliard HS06.sec)	380	439	537	675	853
CPU cores (Peak number)	1204	1392	4083	4880	5931

Year	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031
Storage capacity (PB)	67	79	92	104	116	128	141	153	165	177
Computing (Milliard HS06.sec)	1050	1247	1444	1641	1838	2035	2232	2430	2627	2824
CPU cores (Peak number)	6836	7406	7786	8058	8262	8420	8547	8651	8737	8954

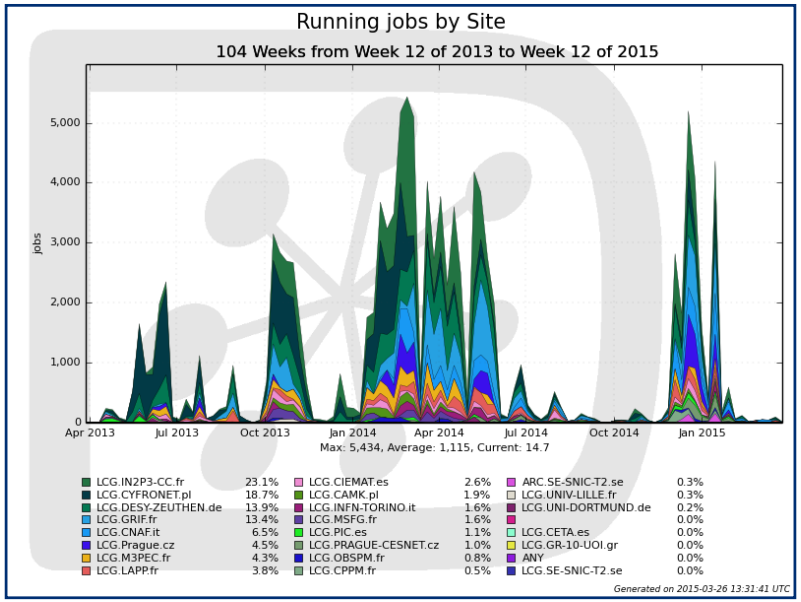
MC Production 2014-2015

- **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**

Prototypes & pre-construction: Project ASTRI

- 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 observations.
- 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

The amount of data

Without data compression: prototypes and the final array.

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

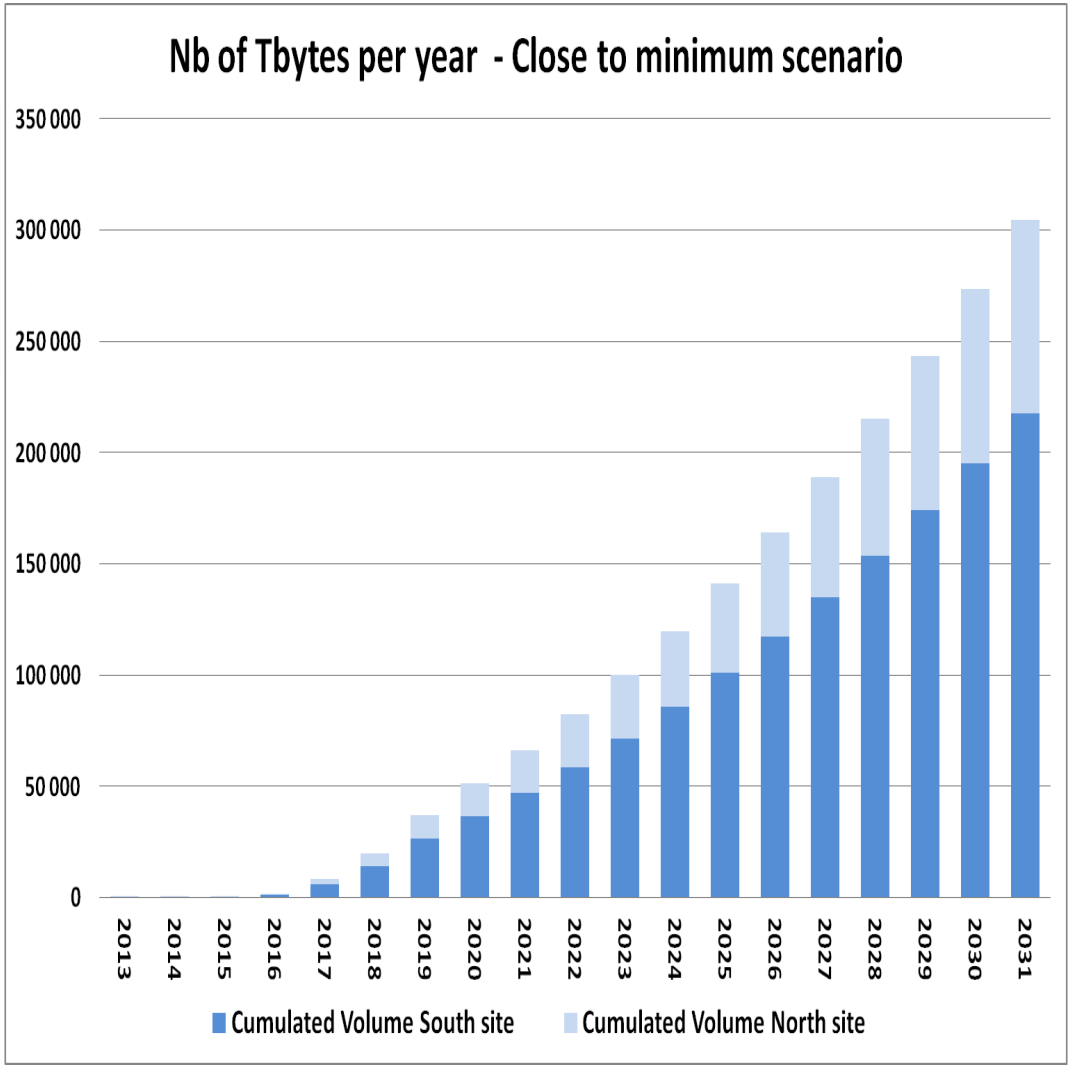
MA/SST => ~3TB/night
=>**~1PB** per year

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.



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.

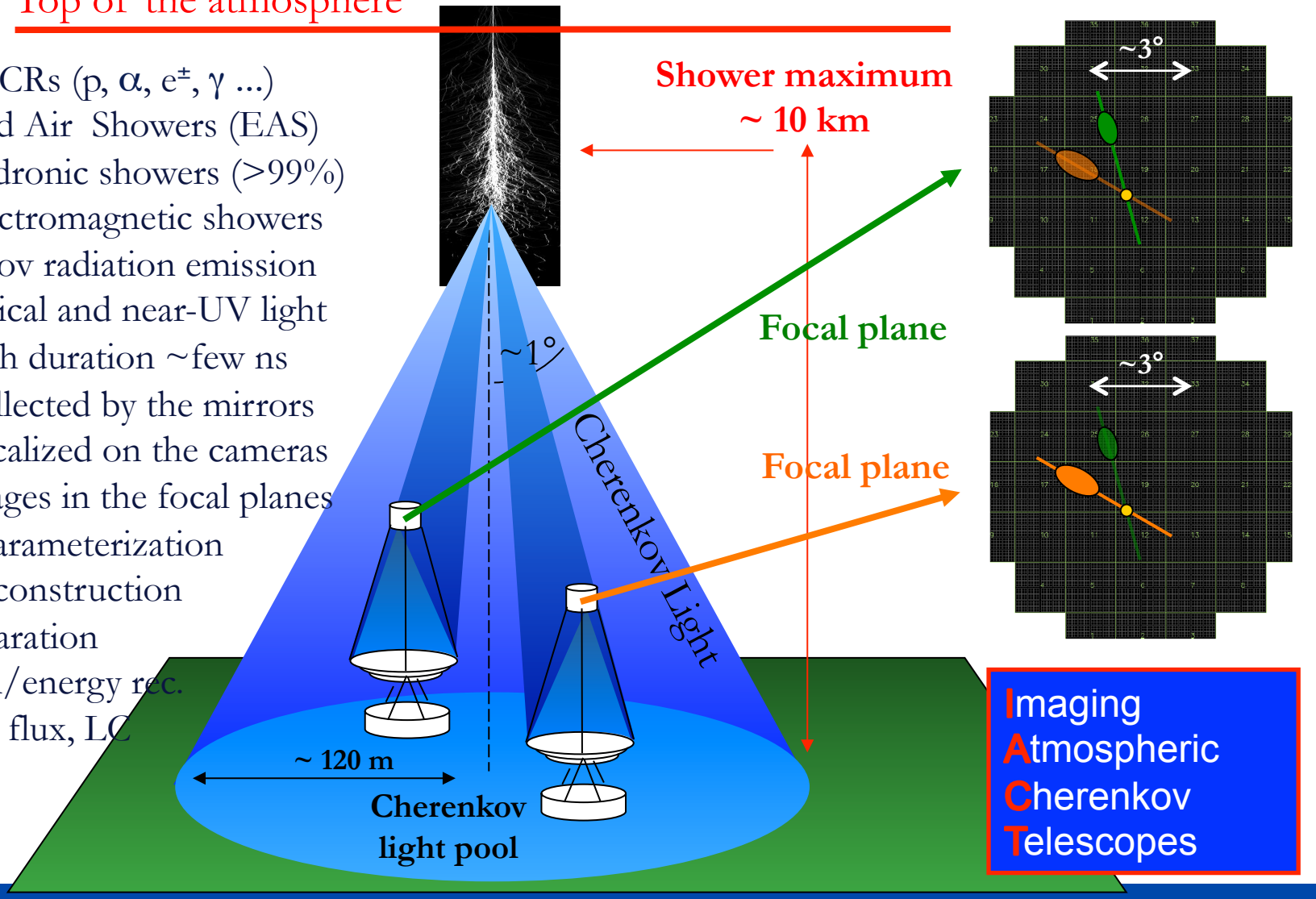

Antonio Stamerra


Ciro Bigongiari

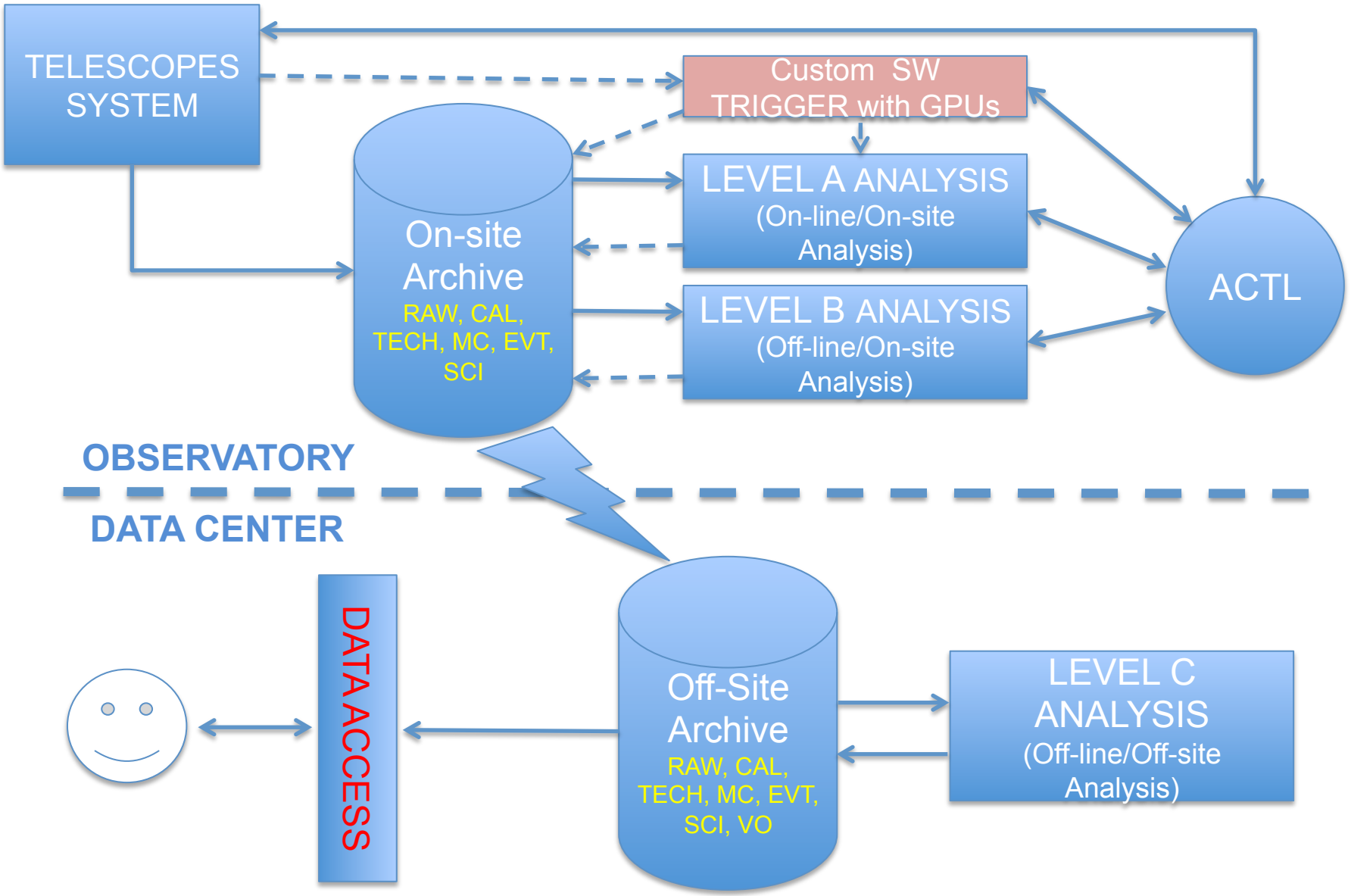

Federico Di Pierro

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



CUSTOM SW TRIGGER: Conceived to leverage GPU's parallel architecture for execution speed and low-power. Could be used to increase the multiplicity of events and reduce DAQ-rate in collaboration with INFN.

LEVEL A Analysis: (OPA - On-line Pipeline Analysis) It is an *almost real time* data analysis starting as soon as data are available in the archive (typically after 90-120 sec) and performs the complete data reduction and analysis. Streaming is an open option.
This is well suited for scientific alerts and monitoring purposes.

- Reduced calibrations and LUTs/IRFs.
- Faster analysis with low power consumption: the software is optimized for GPUs and for ARM processors.
- well suited for scientific alerts.

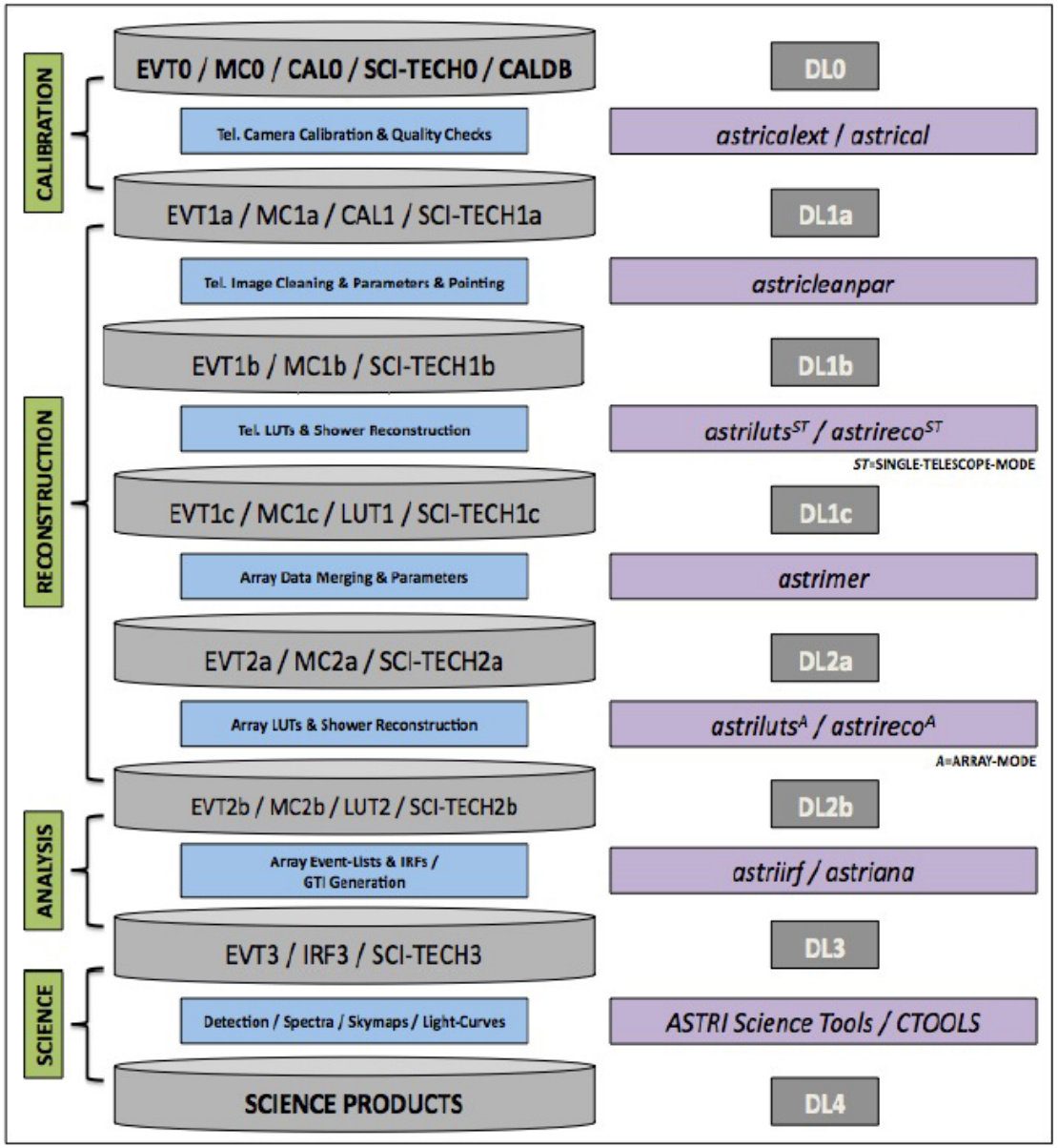
LEVEL B Analysis: (DOPA - Delayed On-site Pipeline Analysis). It performs the complete data reduction and analysis and starts at the end of each observation/night directly at the observatory. DOPA is aimed at monitoring sources providing scientific alerts also on stacked-analysis-bases (e.g. detection in >1 nights).

- Access the full dataset available in the local archive.
- Refined calibrations.

LEVEL C Analysis: (SCIOPA - SCIentific Off-line Pipeline Analysis). It performs the complete data reduction and analysis at the Data Center adopting the best calibrations available for the most refined data reduction and analysis from L0 to L4.

ASTRI Data Levels and Data Flow

Lombardi, Antonelli, Bastieri, Madonna, Mastropietro et al., in prep.

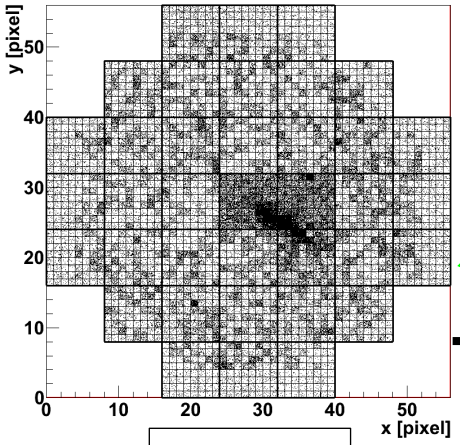


The ASTRI SST-2M prototype and mini-array Data Analysis:

- can handle *both prototype and mini-array data*
- *follows the general CTA design* and data model scheme defined in CTA Data Management
- is developed for *on-line/on-site/ off-site* scientific analysis
- manages *FITS data* from DL0 to DL4 (CFITSIO/CCFITS libraries)
- is written in *C++/Python/CUDA*
- can run on *x86 / ARM CPUs* & *NVIDIA GPUs*
- is developed in *independent software modules* linked by pipelines written in *Python*
- makes use of *ad hoc and official CTA Science Tools*

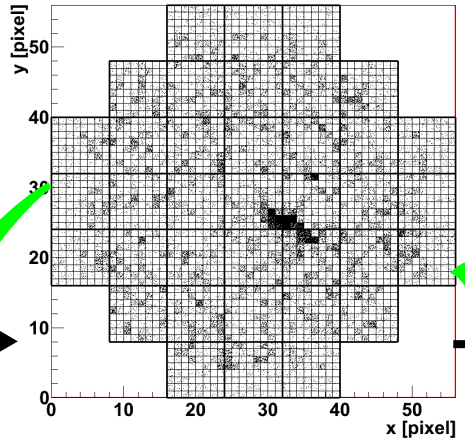
ASTRI SST-2M Prototype Data Reco & Analysis SW Status

DL0

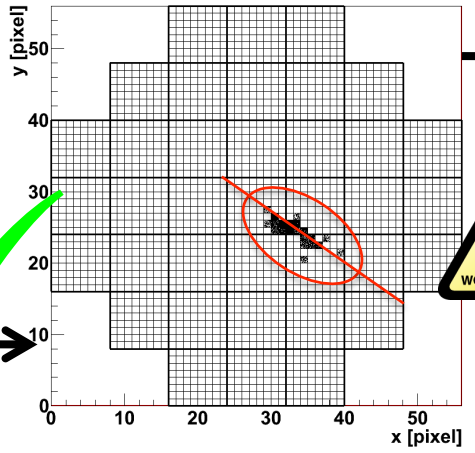


ADC Units

DL1a



DL1b/c



DL2a/b

(Event Lists + GTI and corresponding IRFs

DL3

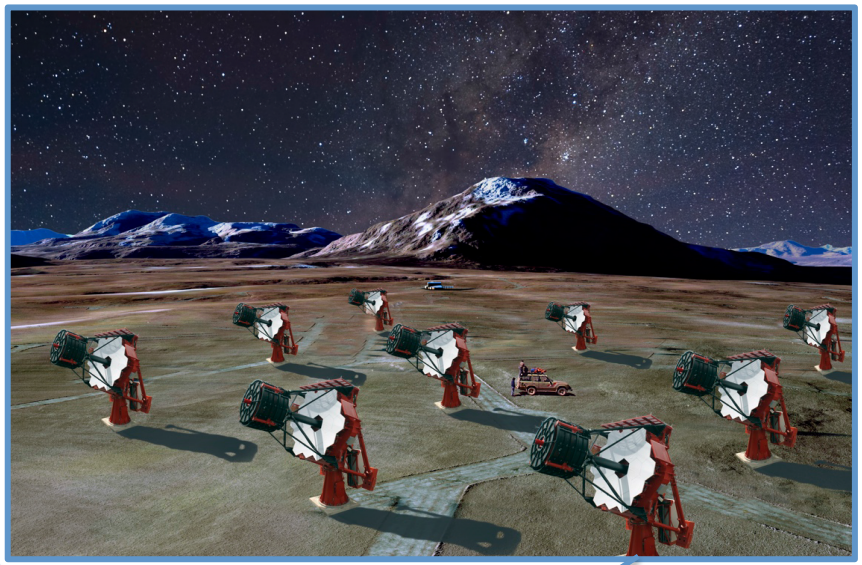
Event Lists and corresponding IRFs

DL4

(SCIENCE PRODUCTS)

DL5

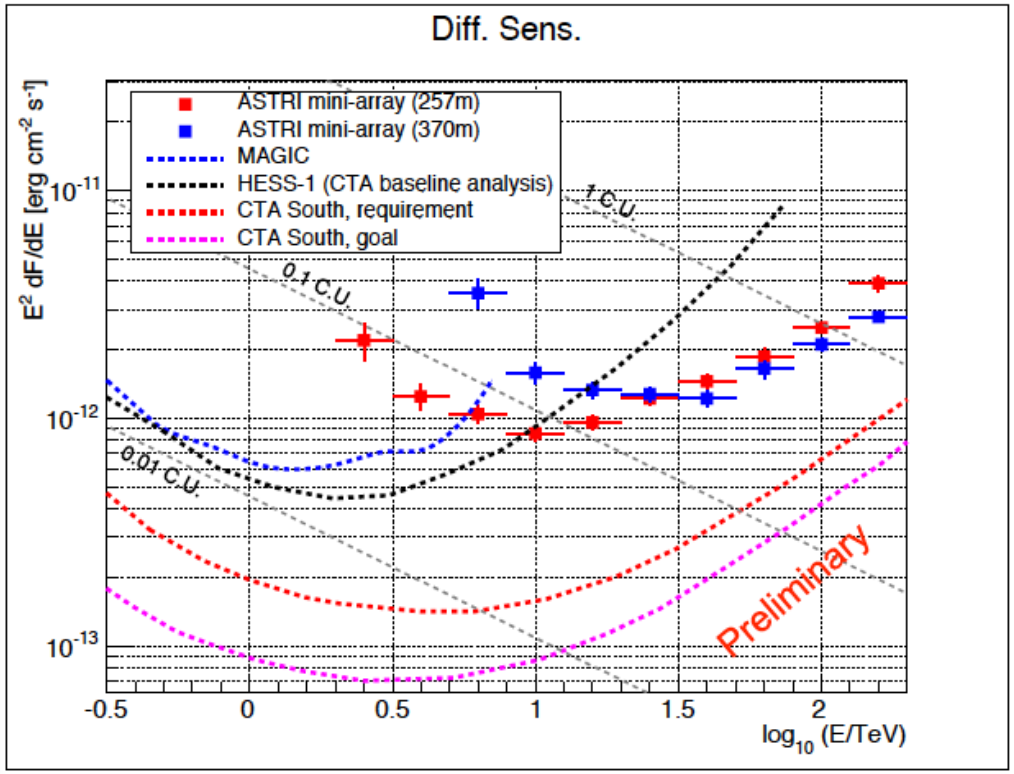
(OBSERVATORY PRODUTS)



ASTRI SST-2M Mini Array Performances

- **Preliminary performance**
  - based on MC-CTA Prod2
  - and official CTA-MC pipelines
- **Sensitivity**
  - slightly better than H.E.S.S.
  - above ~10 TeV for an array
  - composed of 9 telescopes
- **Angular resolution**
  - a few (4–5) arcmin
- **Energy resolution**
  - of the order of 10-15%
- **Wide field of view**
  - ~10°

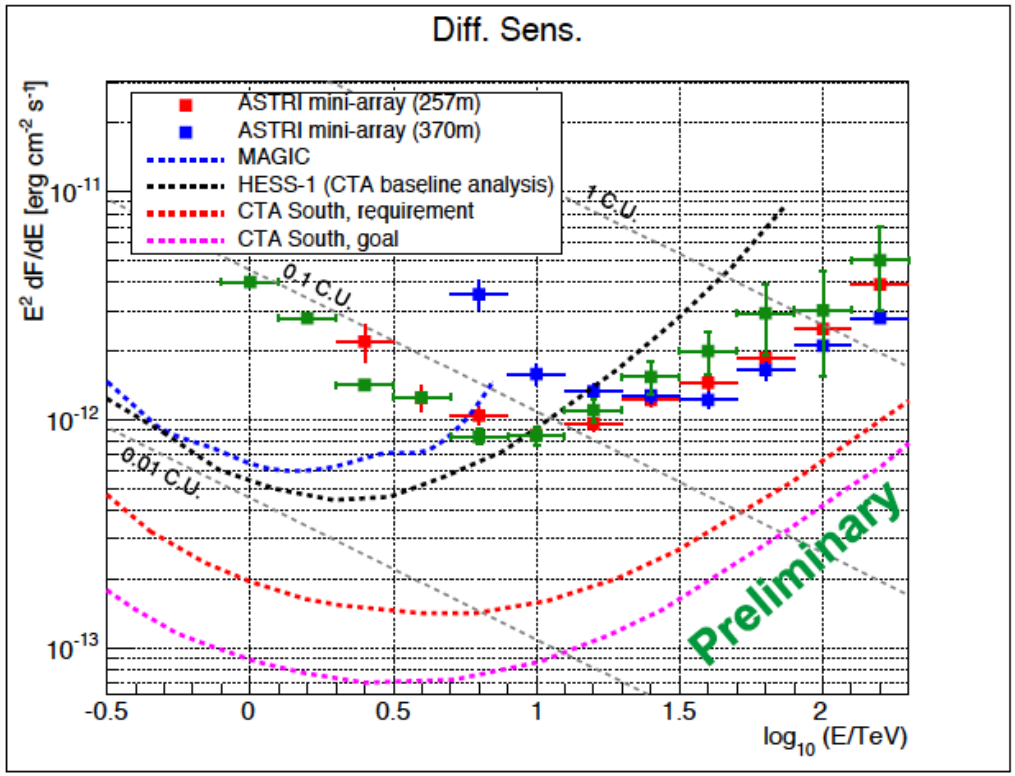
Di Pierro et al., in prep.



ASTRI SST-2M Mini Array Performances

- **Preliminary performance**
  based on MC-CTA Prod2
  and official CTA-MC pipelines
- **Sensitivity**
  slightly better than H.E.S.S.
  above ~10 TeV for an array
  composed of 9 telescopes
- **Angular resolution**
  a few (4–5) arcmin
- **Energy resolution**
  of the order of 10-15%
- **Wide field of view**
  ~10°

Lombardi et al., in prep.



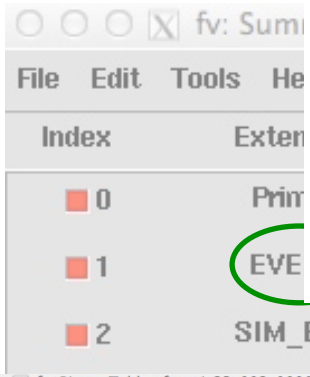
■ ASTRI mini-array (~250m) – ASTRI DATA PIPELINE

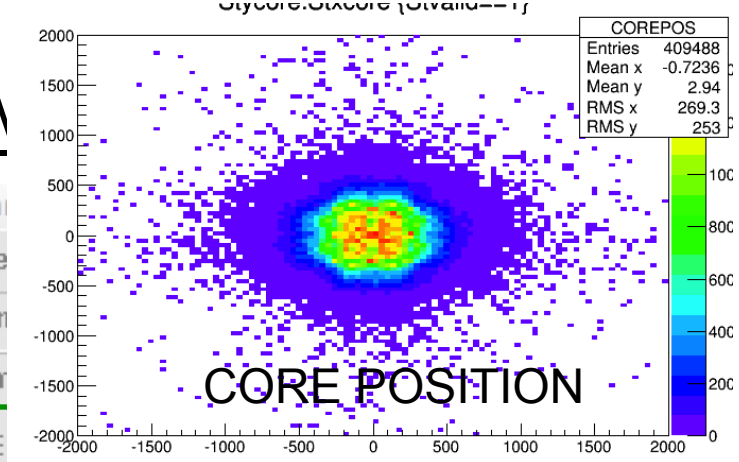
PRELIMINARY ESTIMATES ACHIEVED WITH ASTRI PIPELINE IN REASONABLE AGREEMENT WITH PERFORMANCE ACHIEVED WITH CTA-MC PIPELINE

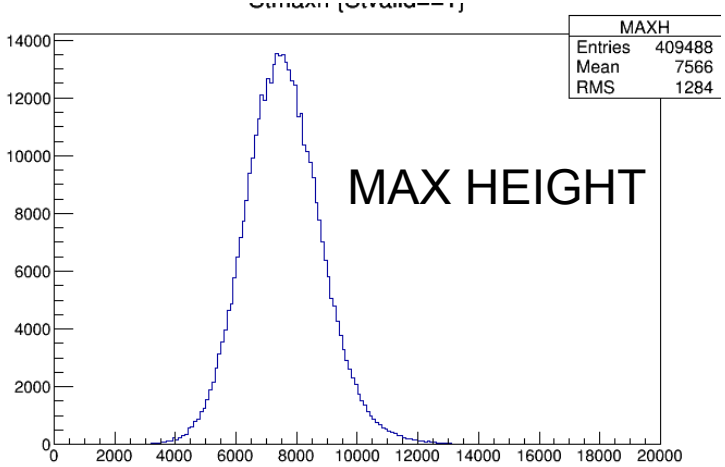
RECONSTRUCTION (DL1c➔DL2a) Status

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:

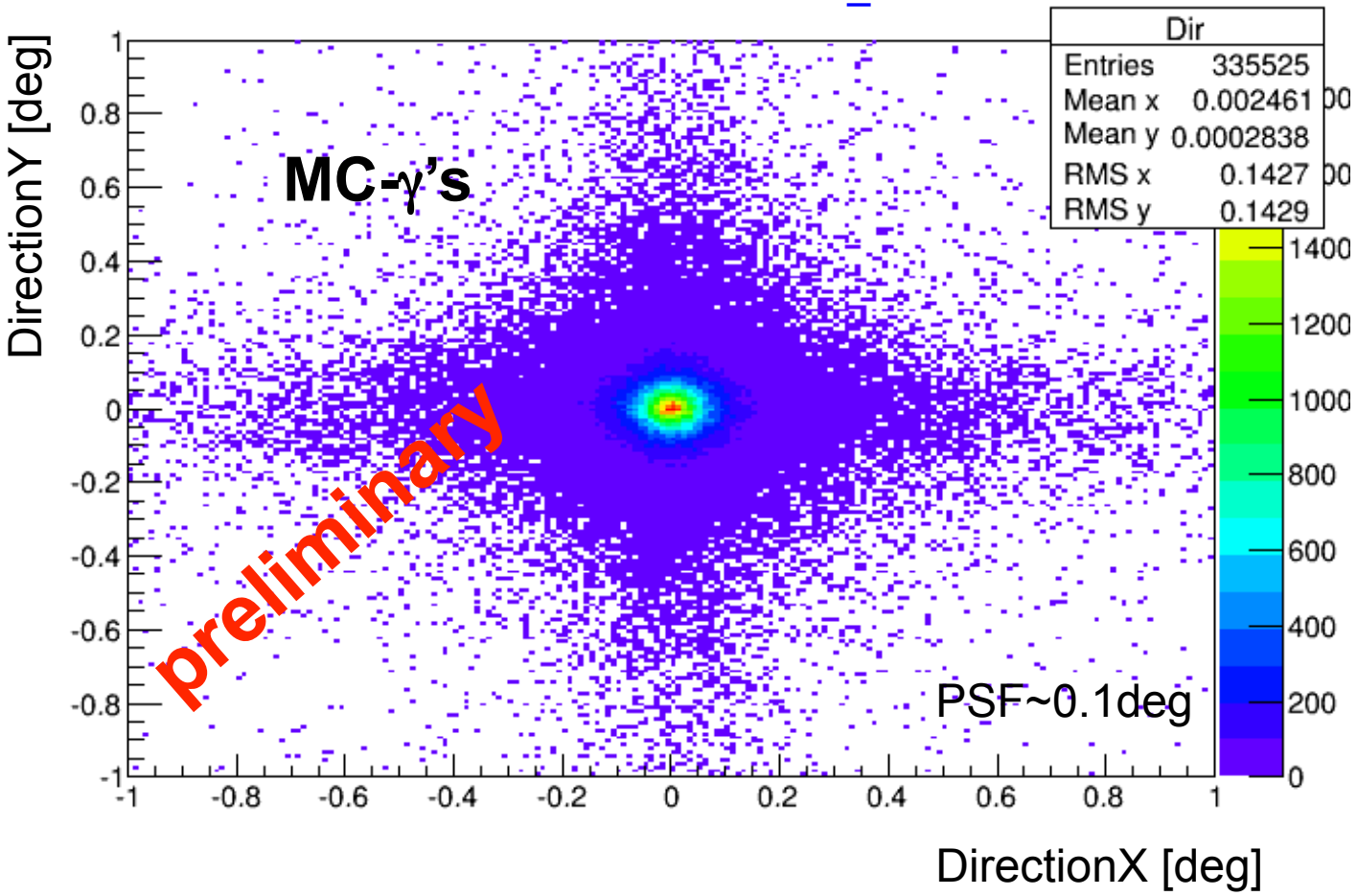
STEREO PARAMETERS

ARRAY INFORMATION

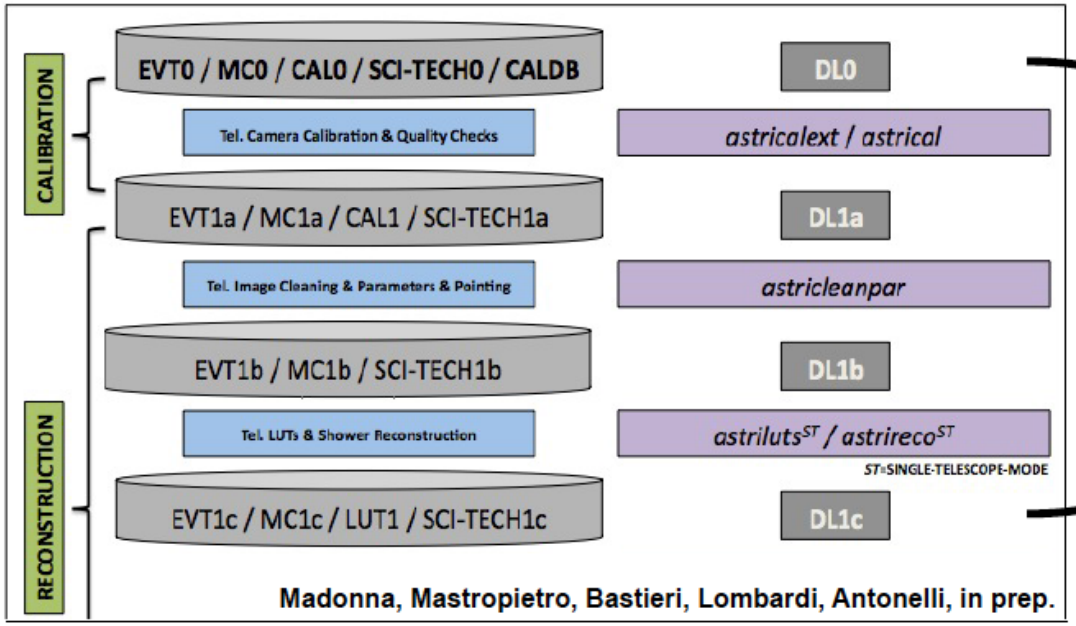
STEREO RECON.

DIRECTION ESTIMATION

ASTRI-MA MC REL_002 data



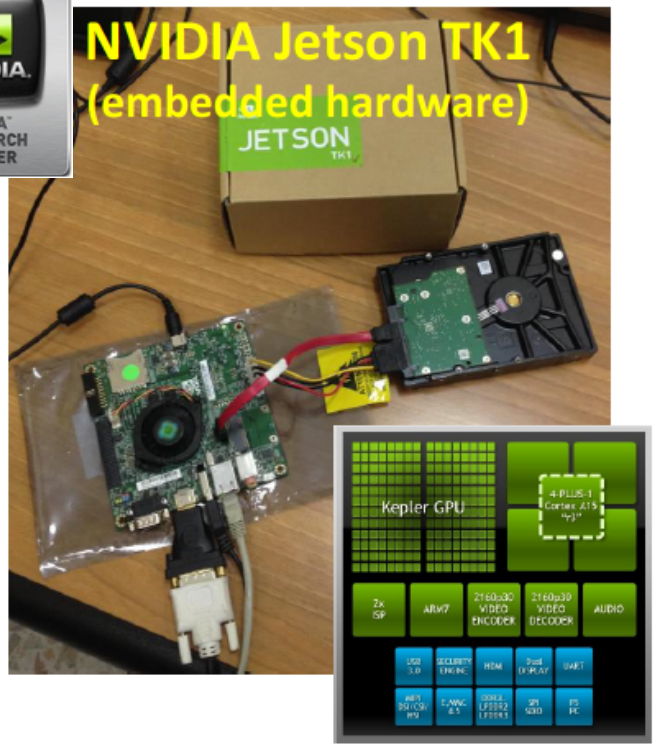
Testing new hardware solutions for On-site Analysis



KEY ADVANTAGE FOR *ON-LINE/ON-SITE* ANALYSIS IN REMOTE OBS. SITES

SINGLE TELESCOPE RECONSTRUCTION ANALYIS PORTED ON ARM/GPUs (*NVIDIA Jetson TK1, C++/CUDA*)

NVIDIA CUDA RESEARCH CENTER



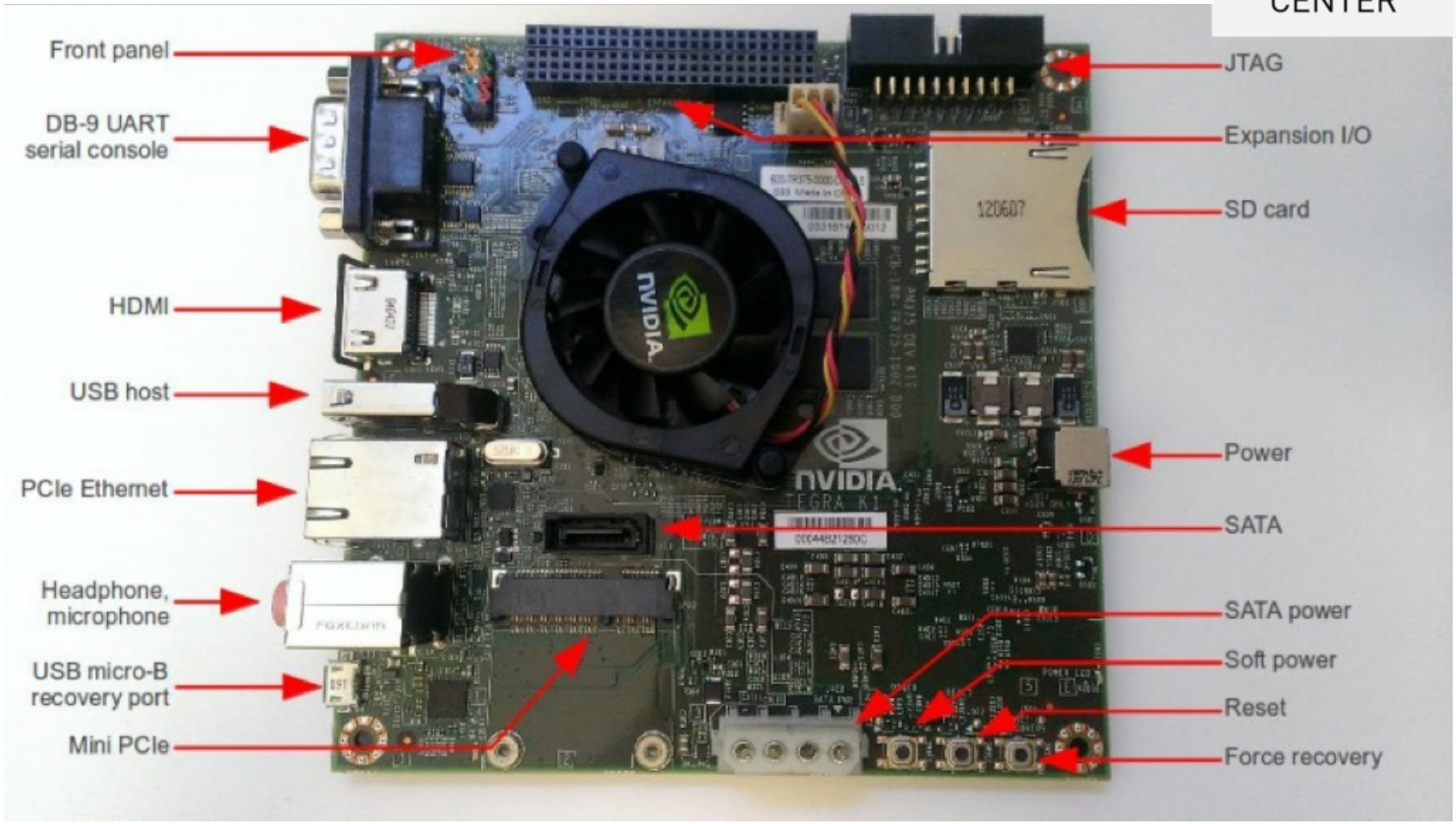
- Single codebase, 3 hardware platforms (x86, ARM CPUs, GPUs)
- Software modules automatically detect GPUs in the system (CPUs/GPUs execution switchable on user request)
- **MC-based *preliminary* performance on *Jetson TK1*:**
  - **DL0 → DL1(c) @ >2000 evt/s ≈ 2 × max DAQ rate!**
  - **Less than 10W!**
- Soon tests on real ASTRI SST-2M prototype data @ Mt. Etna (early 2016)

Testing new hardware solutions: ARM+GPUs

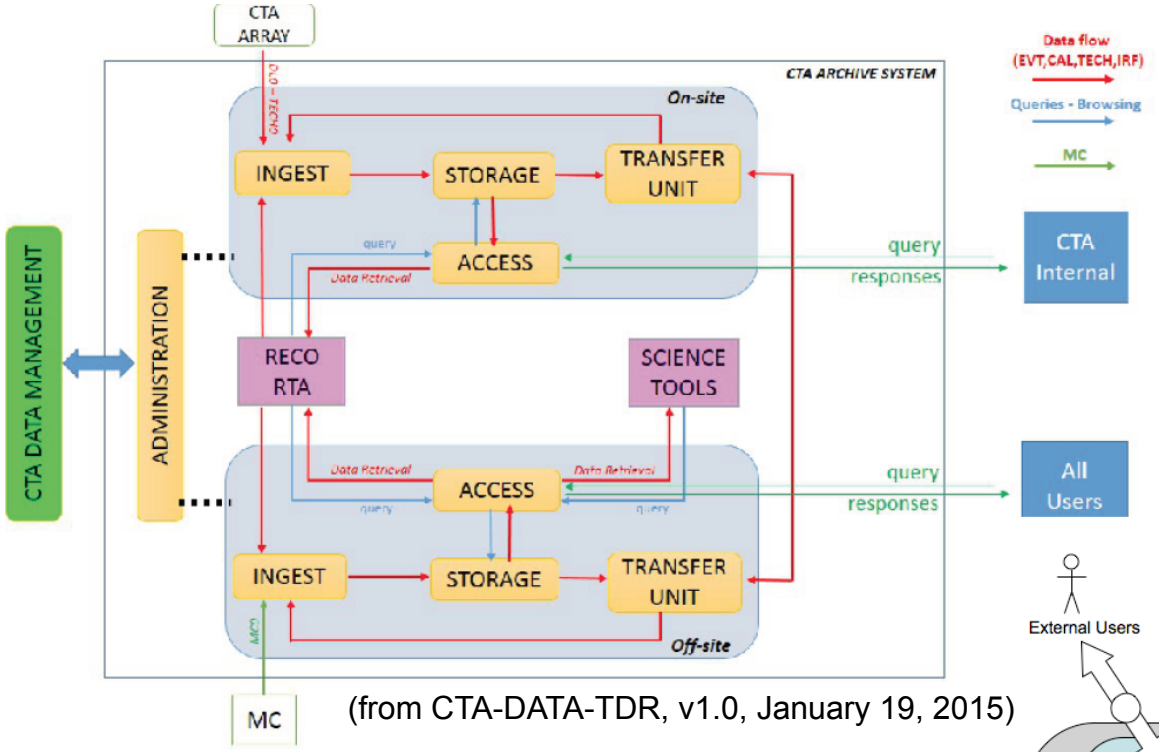
Jetson TK-1

192 cores @ 192$

NVIDIA GPU RESEARCH CENTER



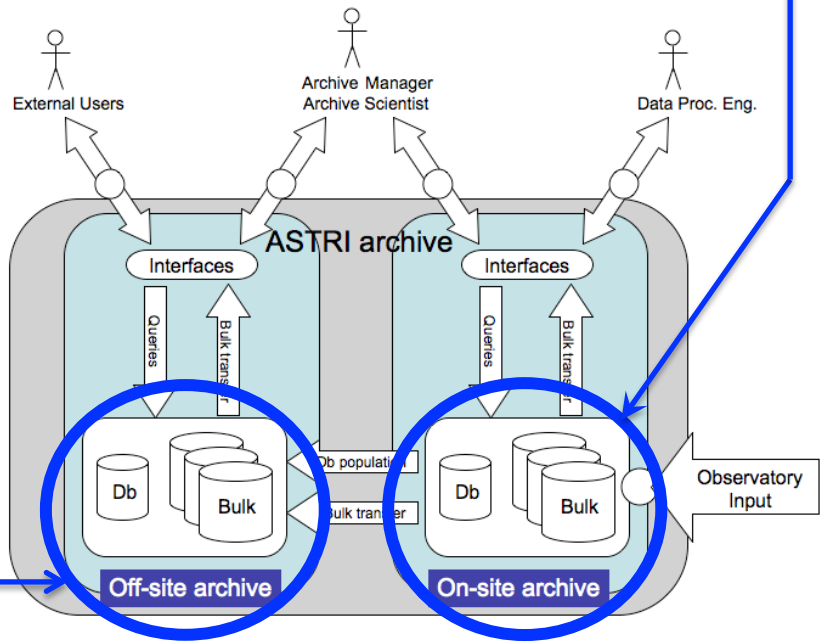
CTA Archive System Concept Design & ASTRI

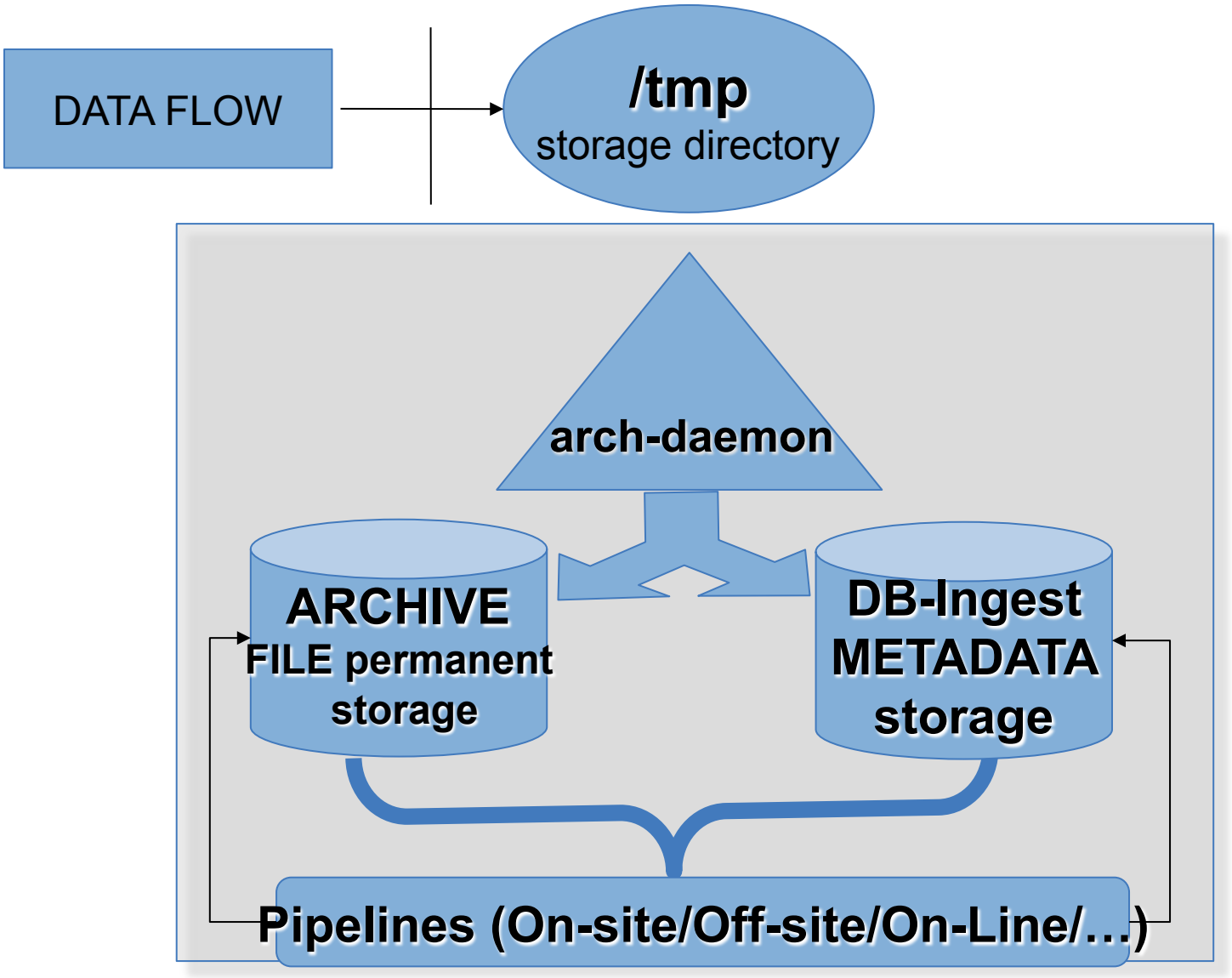


(from CTA-DATA-TDR, v1.0, January 19, 2015)

- ***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**

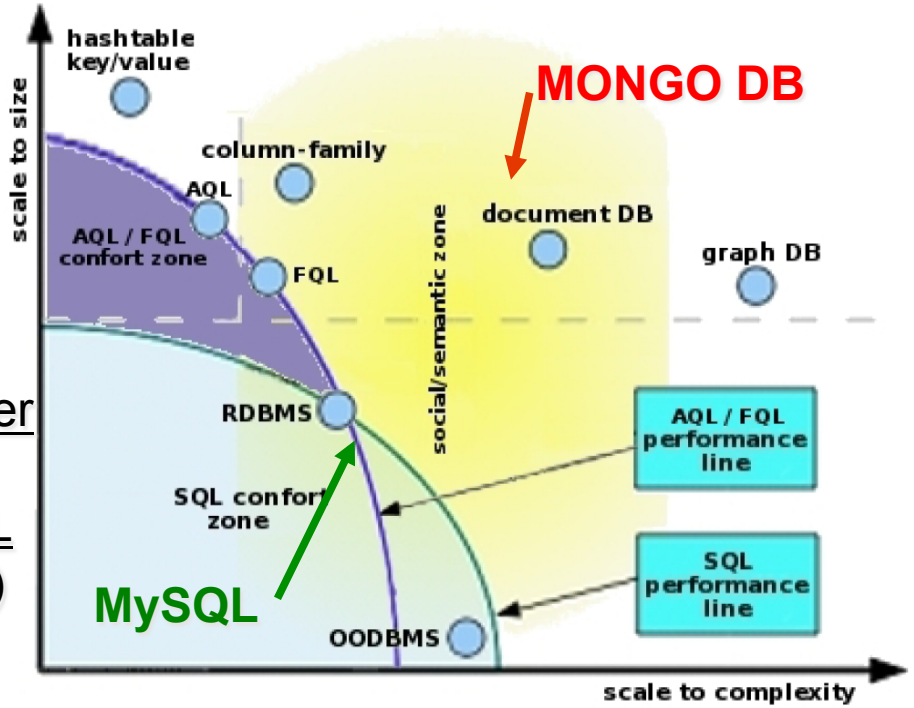



DATA FLOW
/tmp
storage directory
arch-daemon
ARCHIVE
FILE permanent
storage
DB-Ingest
METADATA
storage
Pipelines (On-site/Off-site/On-Line/…)

Database tests and concerns

-Explored several DB families characteristics
-Identified Archive structures & DB engines suitable for the CTA archive

- prototype phase → RDMS (mySQL)
- preproduction phase → NoSQL cluster (MONGO DB)
- CTA production → Distributed NoSQL (MONGO DB shard replica sets)



MONGO DB SHARED REPLICA SETS can be configured to work on a Distributed File Systems Environment and thus reproducing the desired distributed environment for the CTA Archive. A possible implementation is the HADOOP Distributed File system, which works well with Mongo DB, supports Analytics and permits to off load the operational database.

main requirement: **10GB network connections** between Archive/DB nodes

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.

ASTRI ARCHIVE SYSTEM Deployment Phases

DATA CENTERS

INAF - OAR

Primary: LON

Primary: NYC

Secondary: NYC

Secondary: LON

Secondary: SYD

Secondary: SYD

Phase#1 ASTRI prototype
(ROME sites)

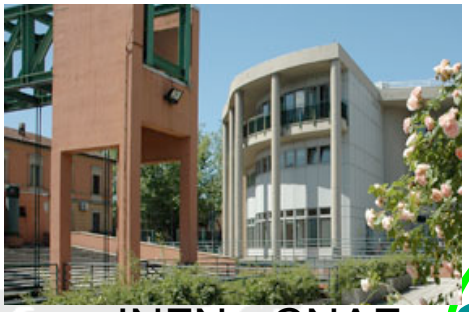
Primary: SYD

Secondary: LON

Secondary: NYC

INFN - LNF

ASTRI ARCHIVE SYSTEM Deployment Phases



INFN - CNAF

DATA CENTERS

INAF - OAR

INFN - LNF

Primary: LON

Secondary: NYC

Secondary: SYD

Primary: NYC

Secondary: LON

Secondary: SYD

Primary: SYD

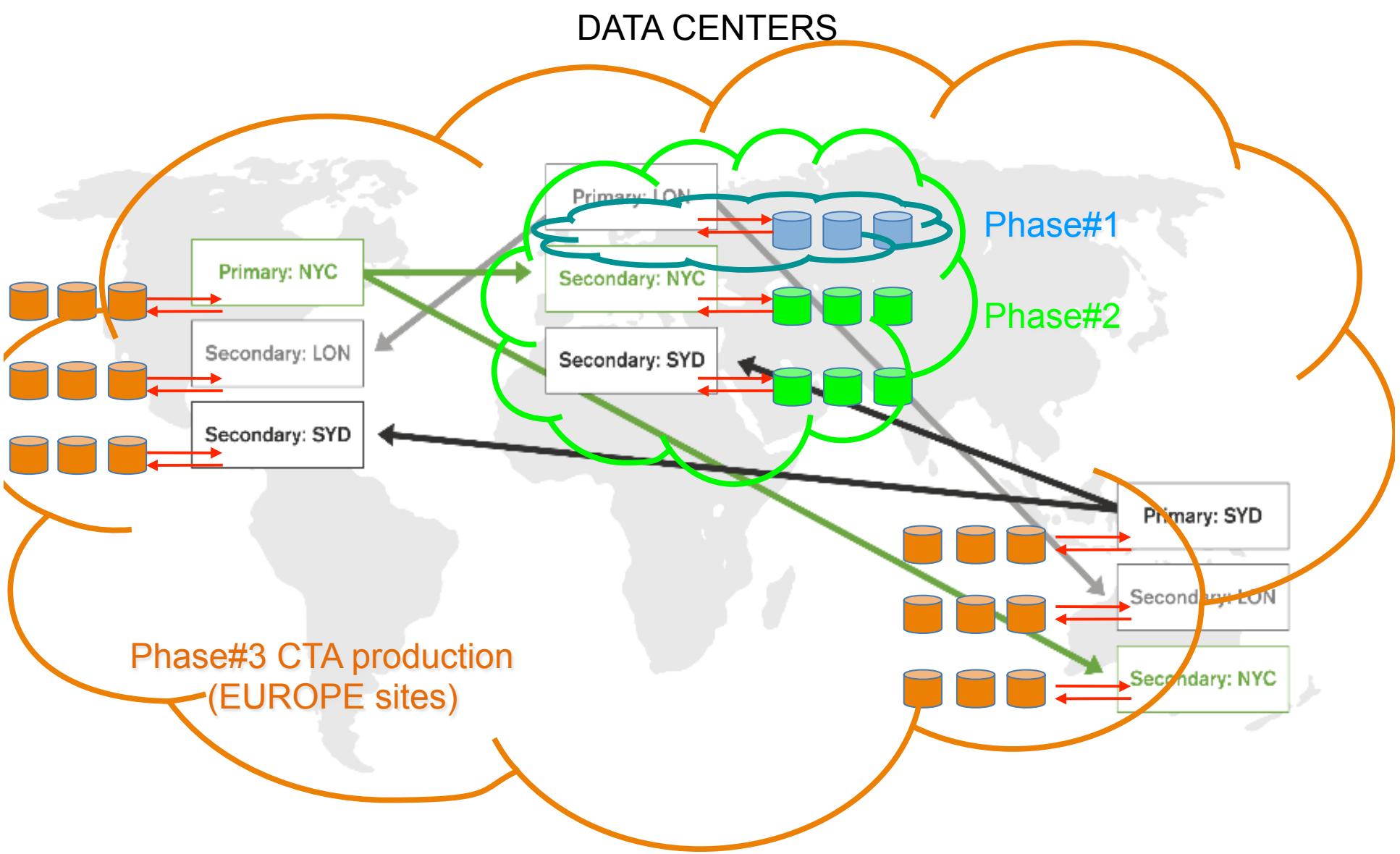
Secondary: LON

Secondary: NYC

Phase#1

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

ASTRI ARCHIVE SYSTEM Deployment Phases



Thank you!