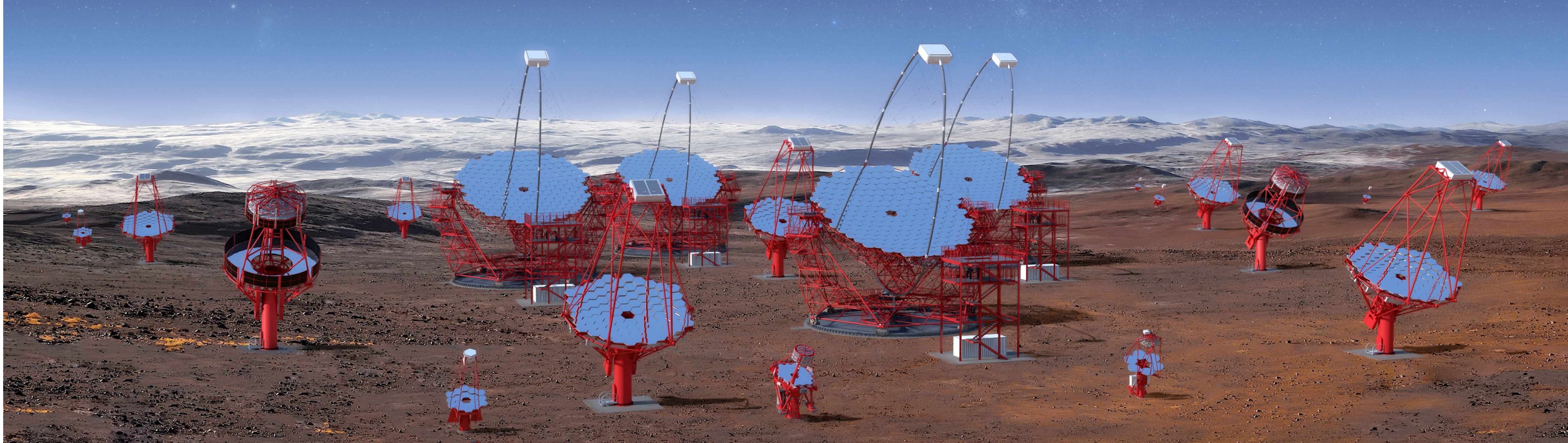
“INAF Contribution to Data Management in CTA”

S.Gallozzi, S.Lombardi, F.G.Saturni, C.Bigongiari, F.Lucarelli,
L.A.Antonelli, A.Costa, E.Sciacca, A.Bulgarelli, N.Parmiggiani
(INAF — OARm — SSDC — OACt — OAS-Bo)



2019 ICT Workshop

21st – 25th October 2019, Milano, Italy

INVOLVED PROJECTS

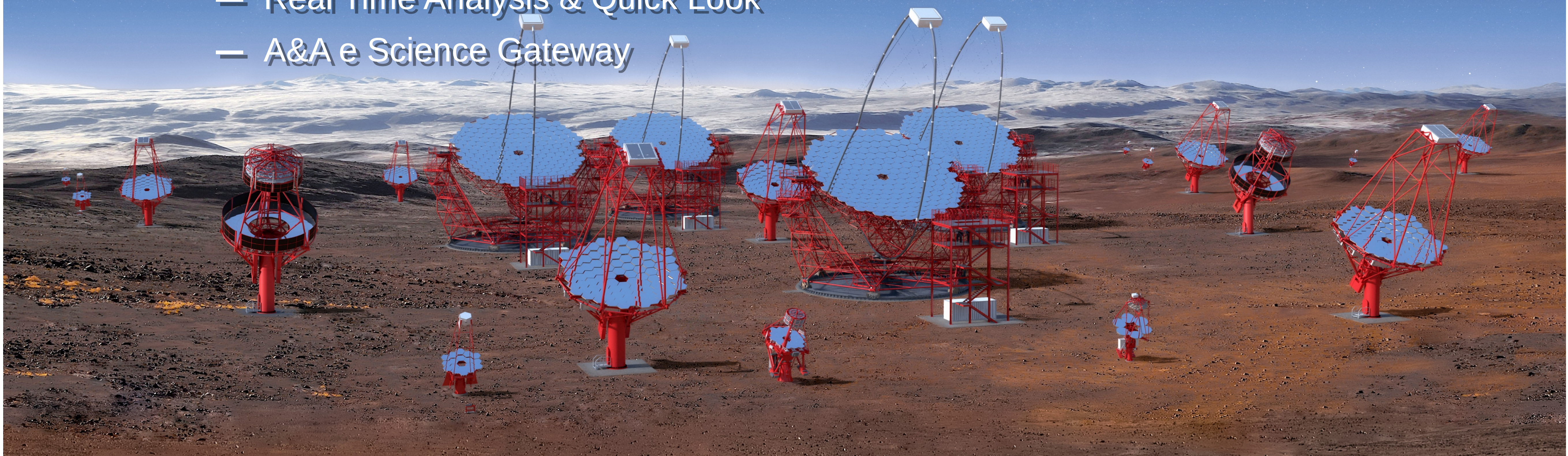
OUTLINE

Imaging Atmospheric Cherenkov Telescopes (IACTs)

- Next-generation IACTs: **the Cherenkov Telescope Array (CTA)** and the **ASTRI mini-array**

INAF Activities on CTA Data Management

- CTA Data Model and Data Management Organization
- ASTRI Prototype Data Model, Pipelines & Archive
- Miniarray / Pathfinder (pre-production for CTA?)
- Real Time Analysis & Quick Look
- A&A e Science Gateway

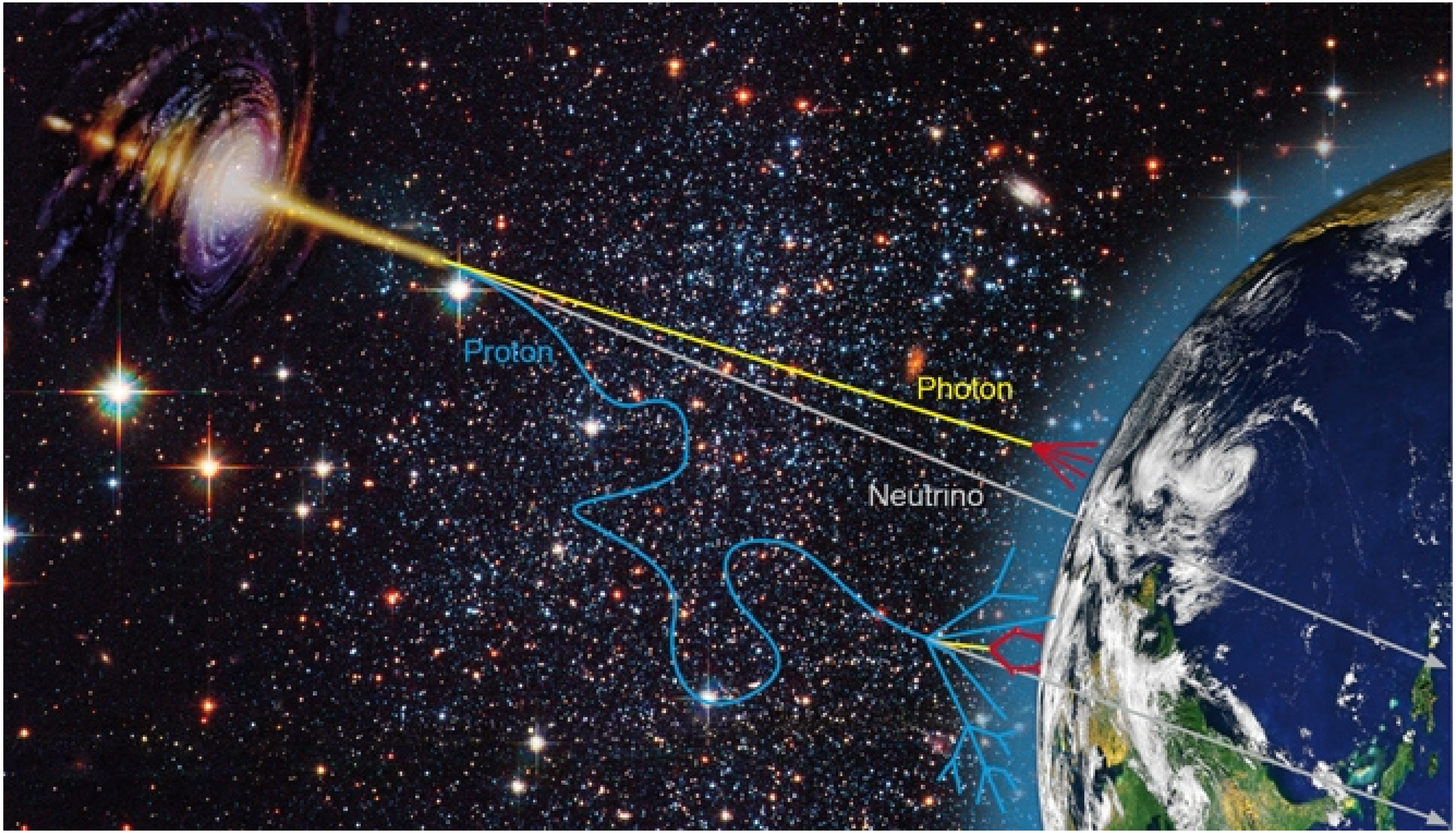


INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

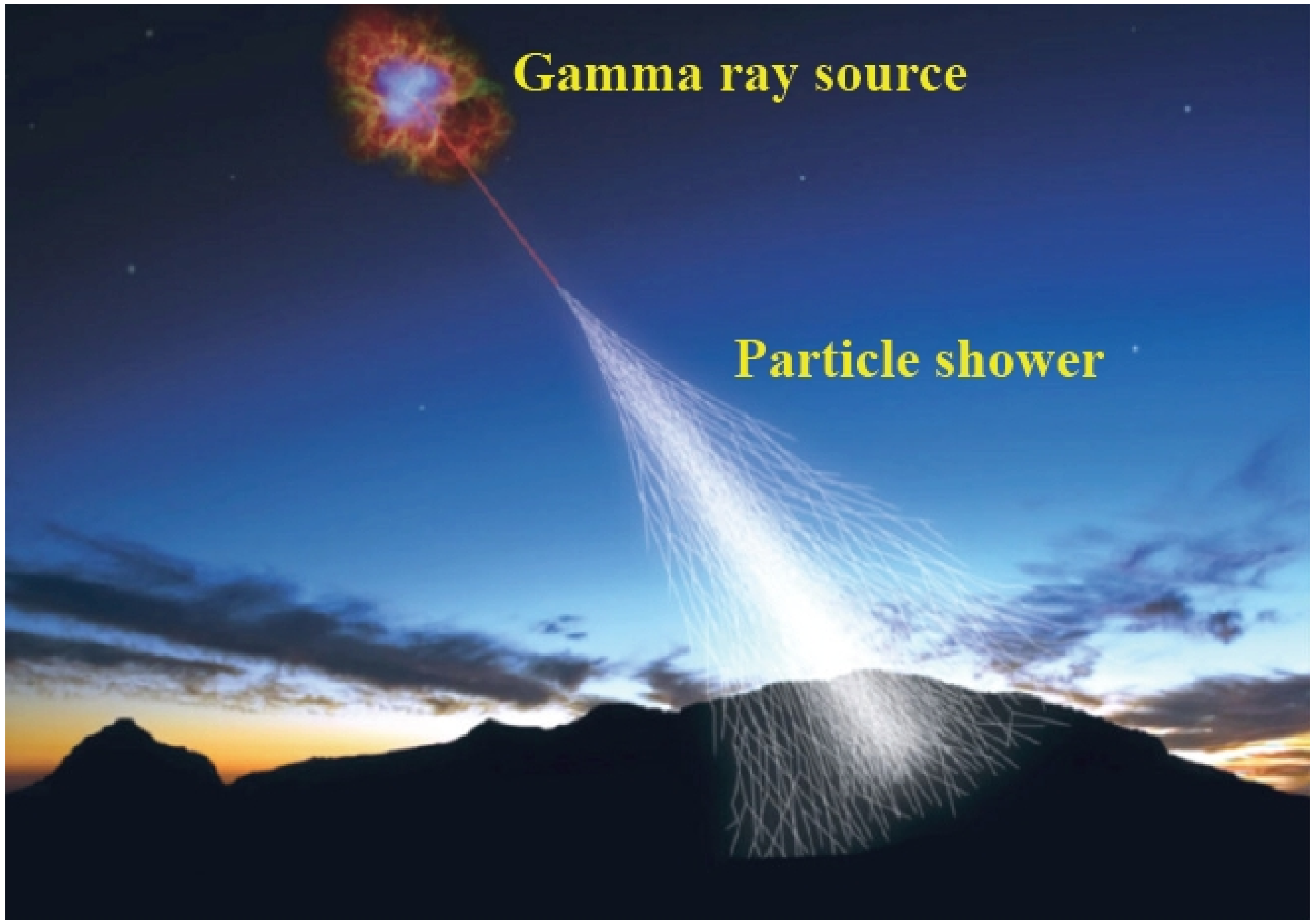
Gamma-ray instruments at HE/VHE (HE > ~50 MeV / VHE > ~50 GeV)

and others…

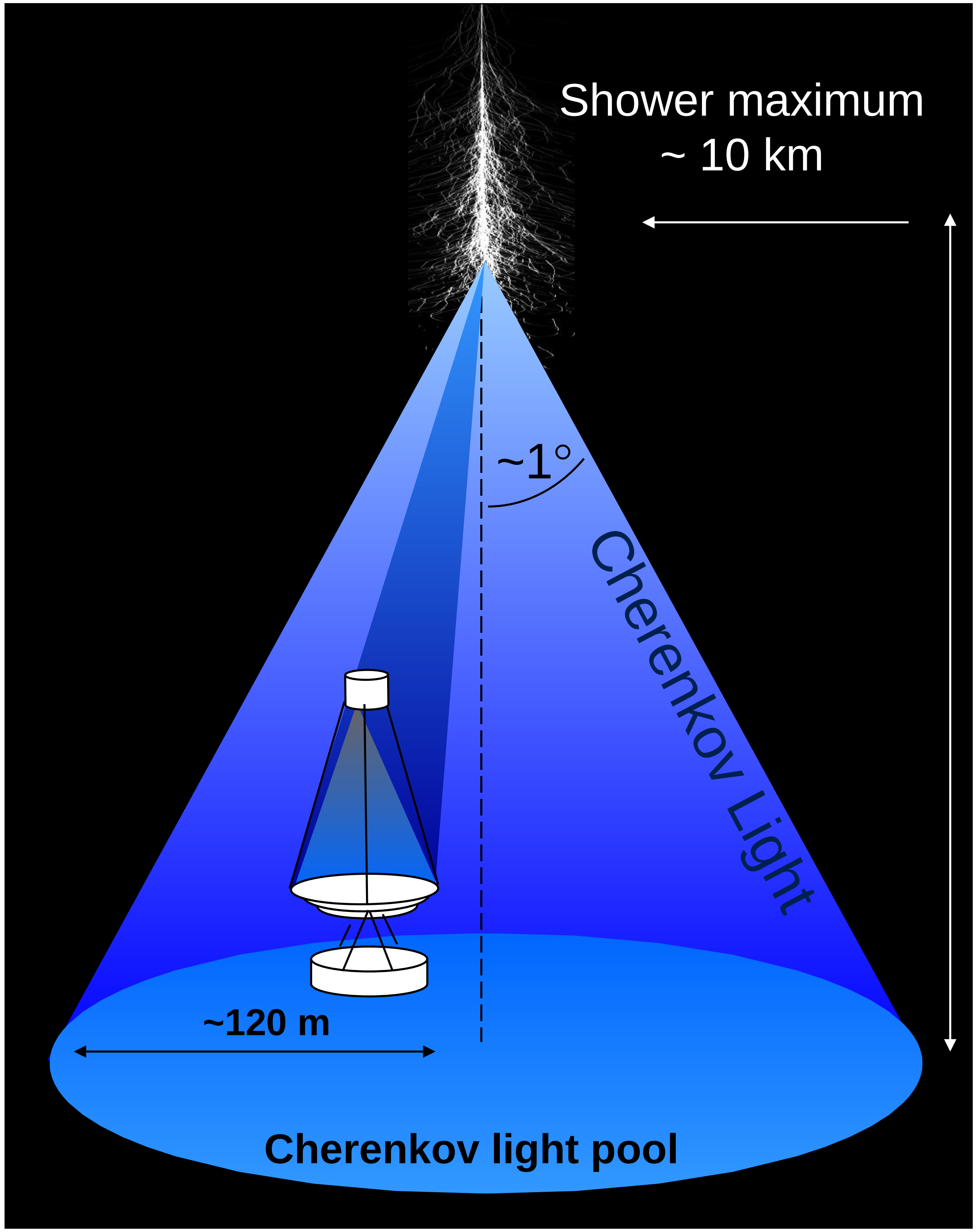
Imaging Atmospheric Cherenkov Technique



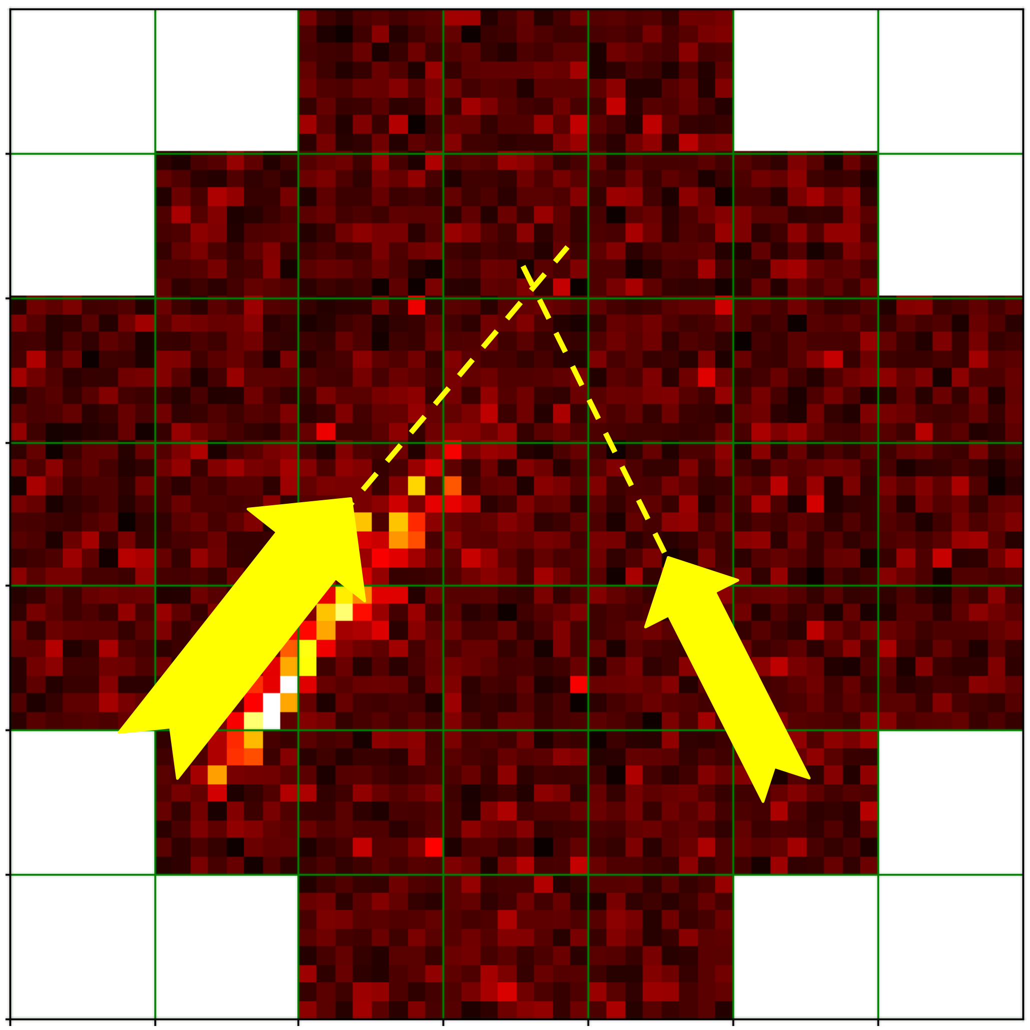
Imaging Atmospheric Cherenkov Technique

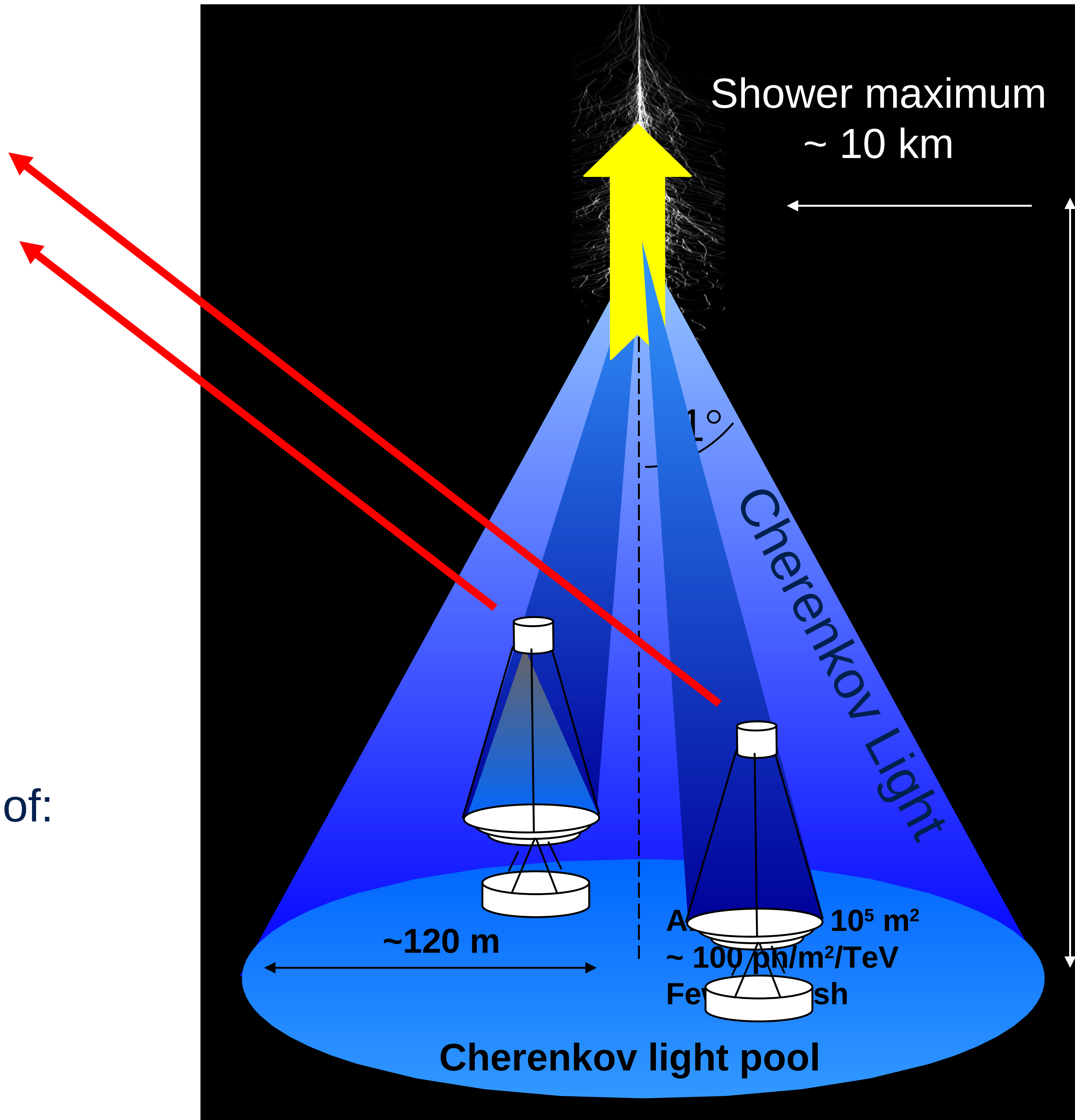


Gamma-rays (and charged cosmic rays) produce showers in the atmosphere. The charged particles in the shower emit Cherenkov light that can be detected by ground-based Cherenkov Telescopes.



Imaging Atmospheric Cherenkov Technique





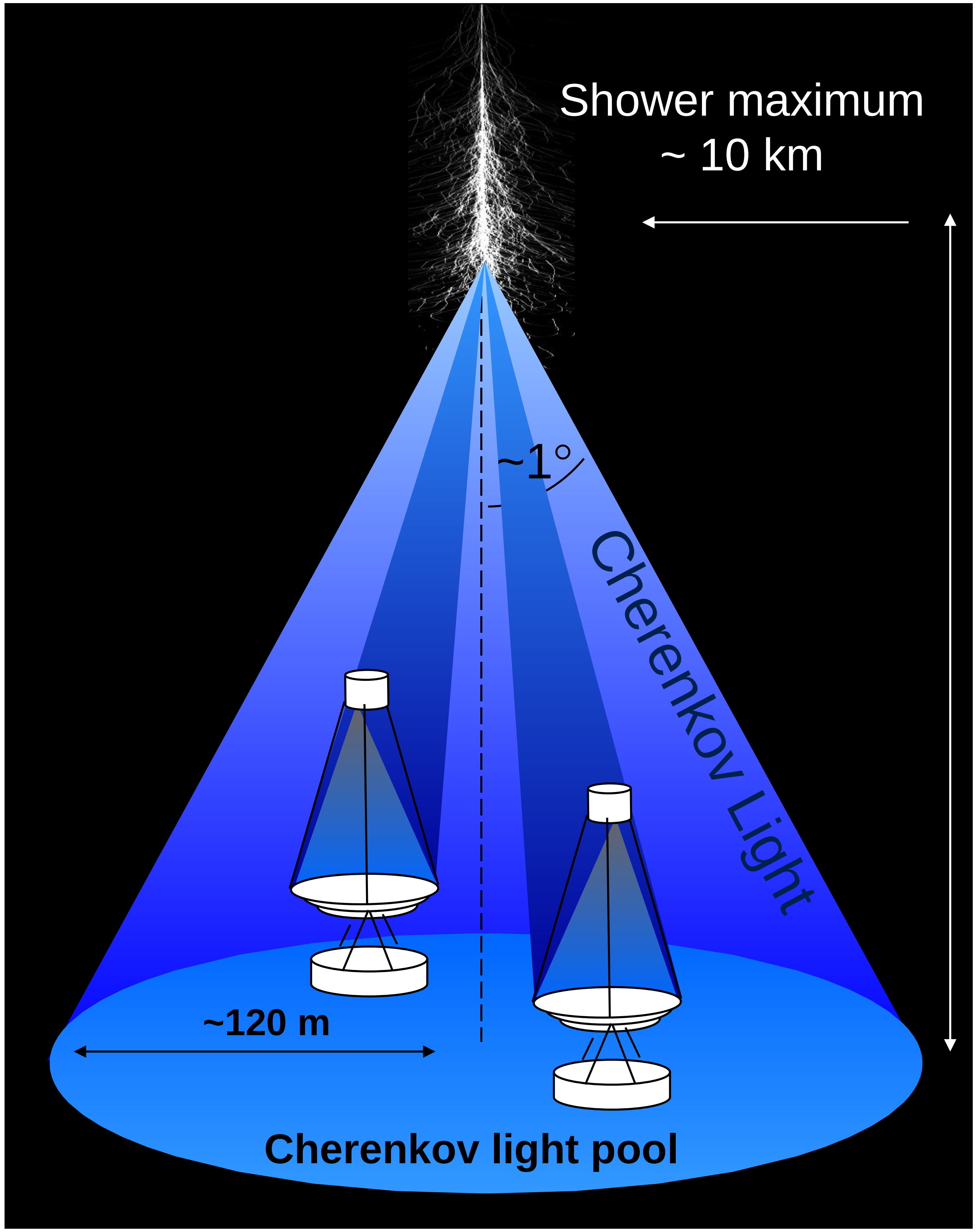
Stereoscopy ➔ better determination of:

- Energy of primary
- Direction of primary
- Kind of primary

➔ Higher Sensitivity ➔ Less Data Acquired

Imaging Atmospheric Cherenkov Technique

- Typical exposures on sources: from tens of minutes to hundreds of hours (depending on the flux of the sources)
- DAQ rate >~ 10^3 showers/second
- 1 shower ➔ multiple images
  (depending on the number of telescopes and the energy/impact-parameter of the shower)
- 1 image ➔ >~10^3 signal+time/pixels/telescopes
- $N_{\gamma s}/N_{hadrons}$ <~10^{-4}
  (even for bright sources like the Crab Nebula)
- Monte Carlo simulations
- Data reduction (for each recorded shower):
  - Pixels' signals calibration
  - Images' parametrization of the images
  - Merging of information of different images
  - Reconstruction of the shower properties
  - Generation of gamma-like showers' event-list
  - High-level scientific products: sky-maps, spectra, light-curves, …

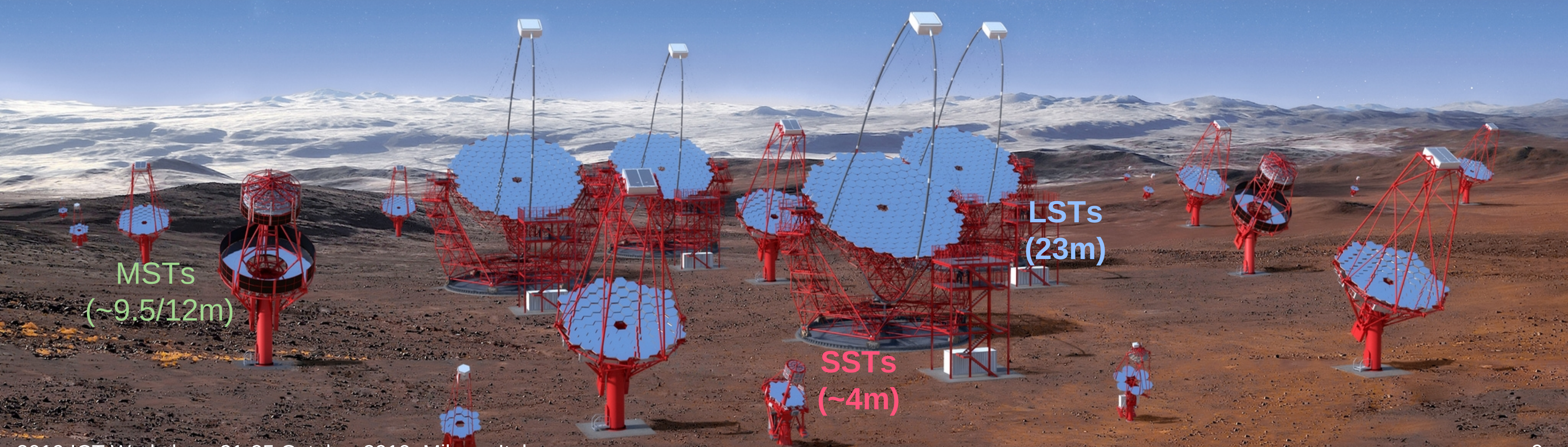


The Cherenkov Telescope Array

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Science optimization under budget constraints:
different telescope sizes for different energy ranges

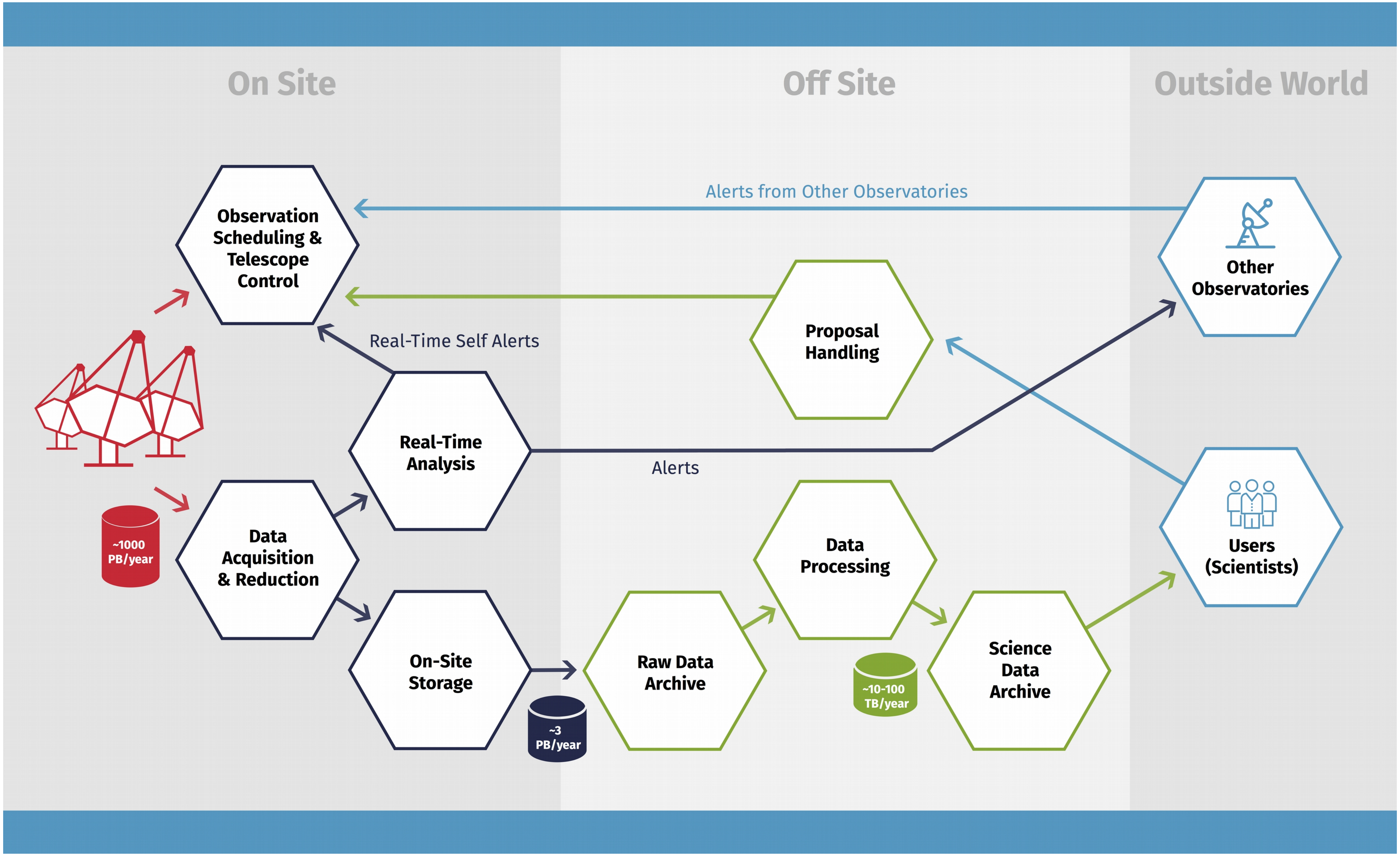
- **Southern Site (~4km²)**:
  - 4 Large-size Telescopes (LSTs)
  - 25 Medium-size Telescopes (MSTs)
  - 70 Small-size Telescopes (SSTs)
- **Northern Site (~0.4km²)**:
  - 4 Large-size Telescopes (LSTs)
  - 15 Medium-size Telescopes (MSTs)



The ASTRI Mini-Array

- **INAF flagship project selected as precursor/pathfinder of CTA**
- PI experiment designed to operate for 3 — 10 years
- Site @ Teide Observatory, Tenerife (Canary Islands, Spain) with total 9 — 12 SSTs
- 3 countries (Italy, Brazil, South Africa)

C.T.A. Data Flow



C.T.A. Data-Life Cycle

Three LOGICAL UNIT in C.T.. DM

- ACADA → Data Control & Acquisition
  (Obs Scheduling, Array/Tel Control and Data Acquisition, Real Time Analysis & QuickLook)
- DPPS → Data Processing & Preservation
  (Pipeline Analysis, Simulation and low-level Archive = **Bulk Archive**)
- SUSS → Science User Support
  (Obs Planning, Proposal Handling, Science support, Science Gateway + high-level Archive = **Science Archive**)

ARCHIVE for ASTRI & MiniArray

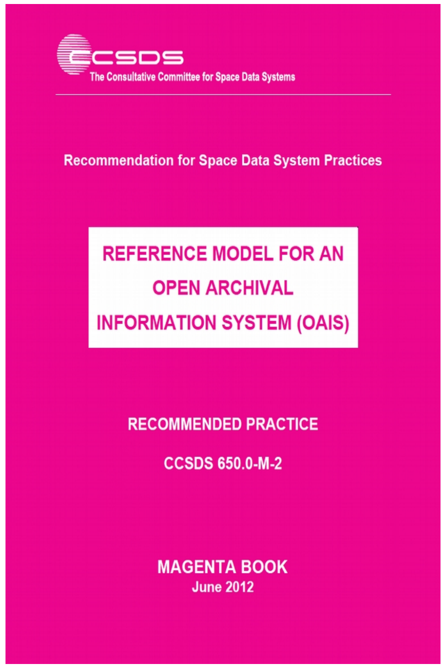
INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

In the scientific data lifecycle of any **OBSERVATORY** the role of the Archive is central.

The major aim of a Scientific Archive is to guarantee **data preservation and access** information for the **Long Term** and for **all data science products**.

The **archived information** must be also usable by **different user categories** (***data consumers***) who are separate in time, space and background from the ***data producers***.

Archive **MUST** be accessible well beyond the end of the operational life of the observatory and **MUST** follow the **O.A.I.S Standard**



CCSDS
The Consultative Committee for Space Data Systems

Recommendation for Space Data System Practices

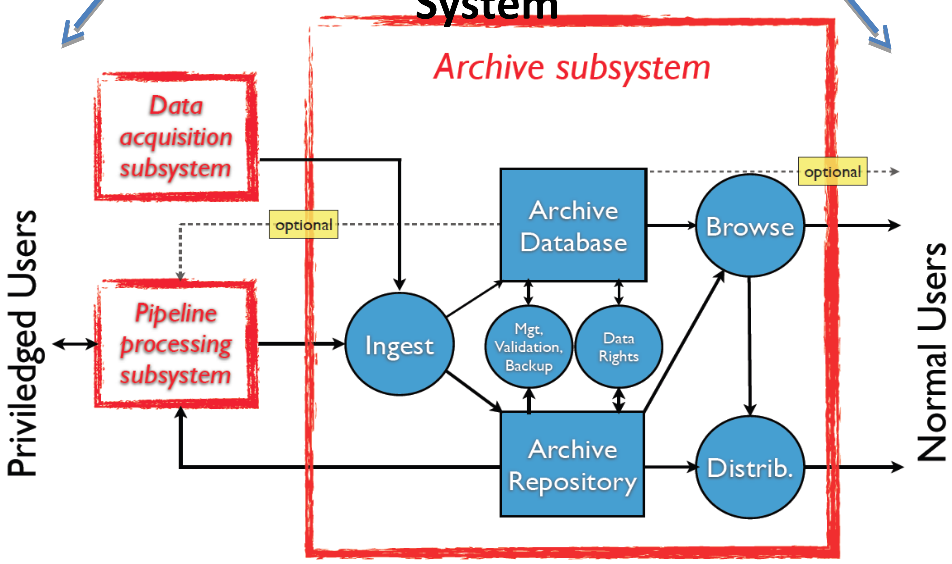
REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS)

RECOMMENDED PRACTICE

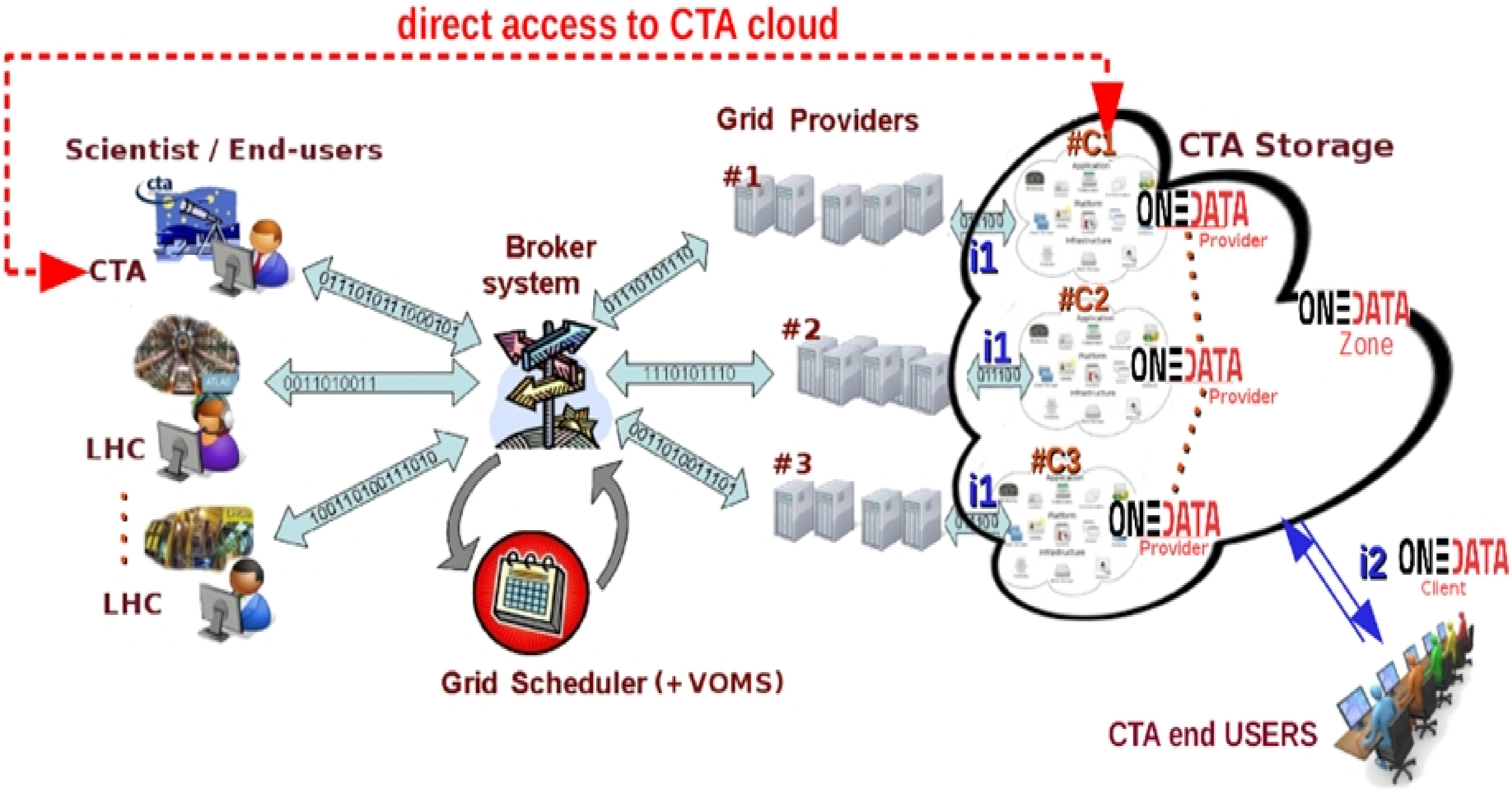
CCSDS 650.0-M-2

MAGENTA BOOK
June 2012

Data Producers and **Data Consumers** are rigidly separated by the **Archive System**



Prototype for CTA ARCHIVE

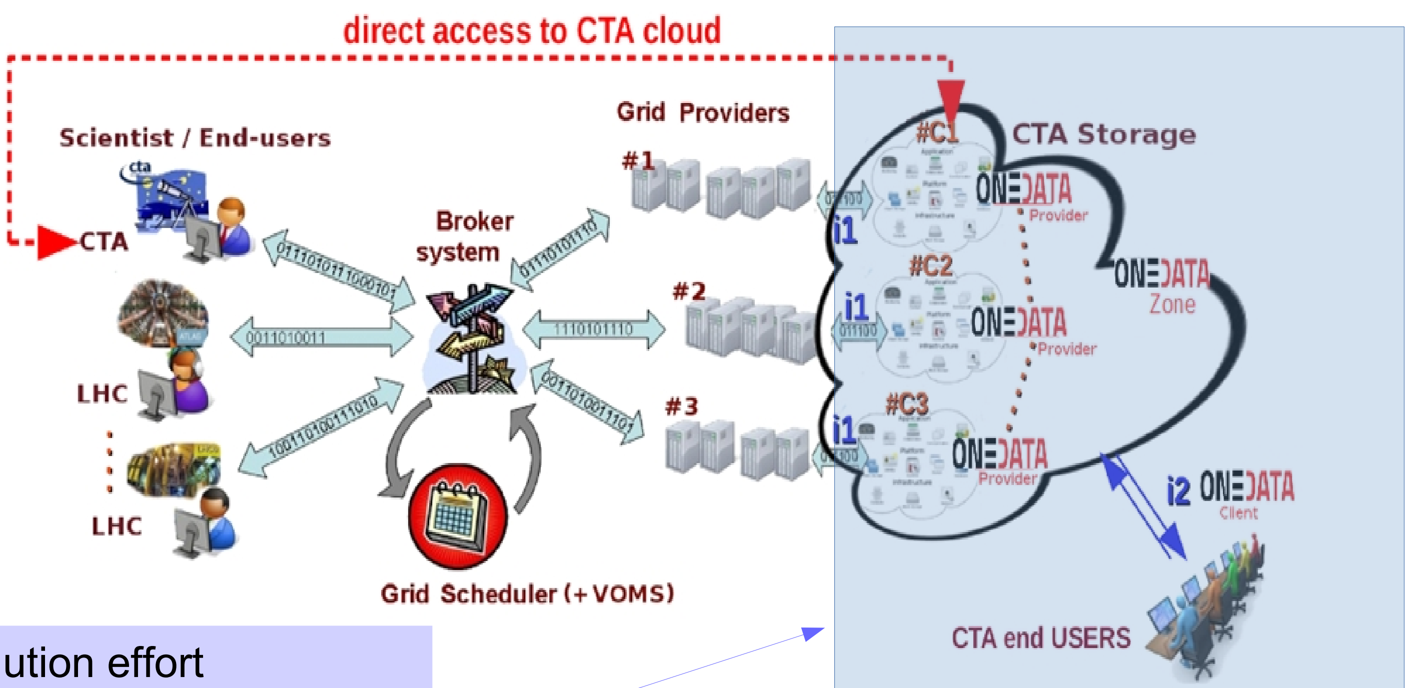


Prototype for CTA ARCHIVE

INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

CTA Collaboration & Community participated to the INDIGO-Data Cloud H2020 Project AS **"Use Case"** for the **INDIGO infrastructure**.

The aim of our commitment was the very fruitful multi-disciplinar collaboration with INDIGO Communities in order to include the **BigData challenges** coming from the CTA Archive as an INTERNAL INDIGO Use CASE / Case Study
→ to be investigated with a distributed approach ←



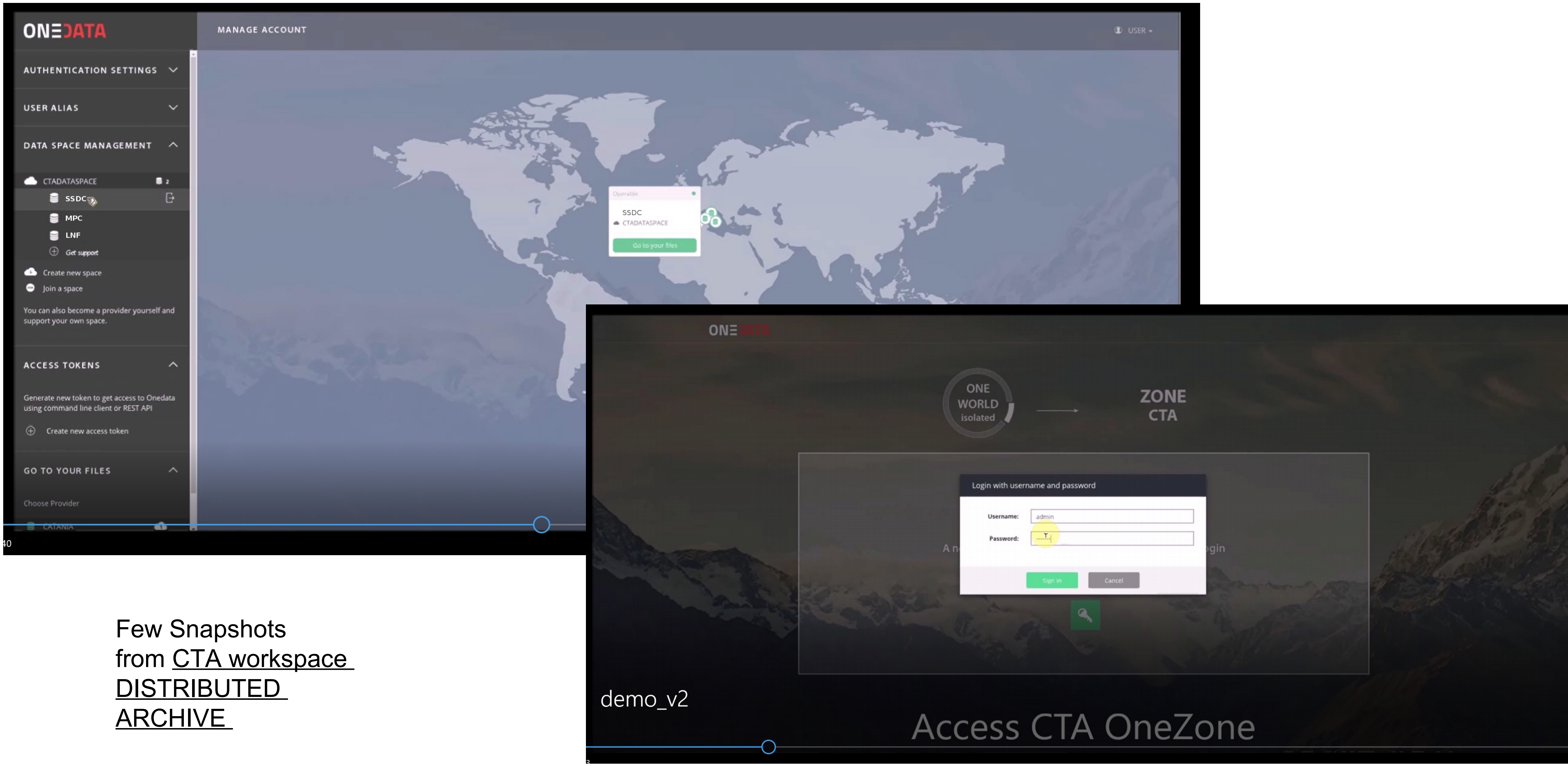
INDIGO solution effort
In the Distributed
Federation of Storage

INDIGO - DataCloud

OneDATA solutions are ready for *CTA A&A*

Prototype for CTA ARCHIVE

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS



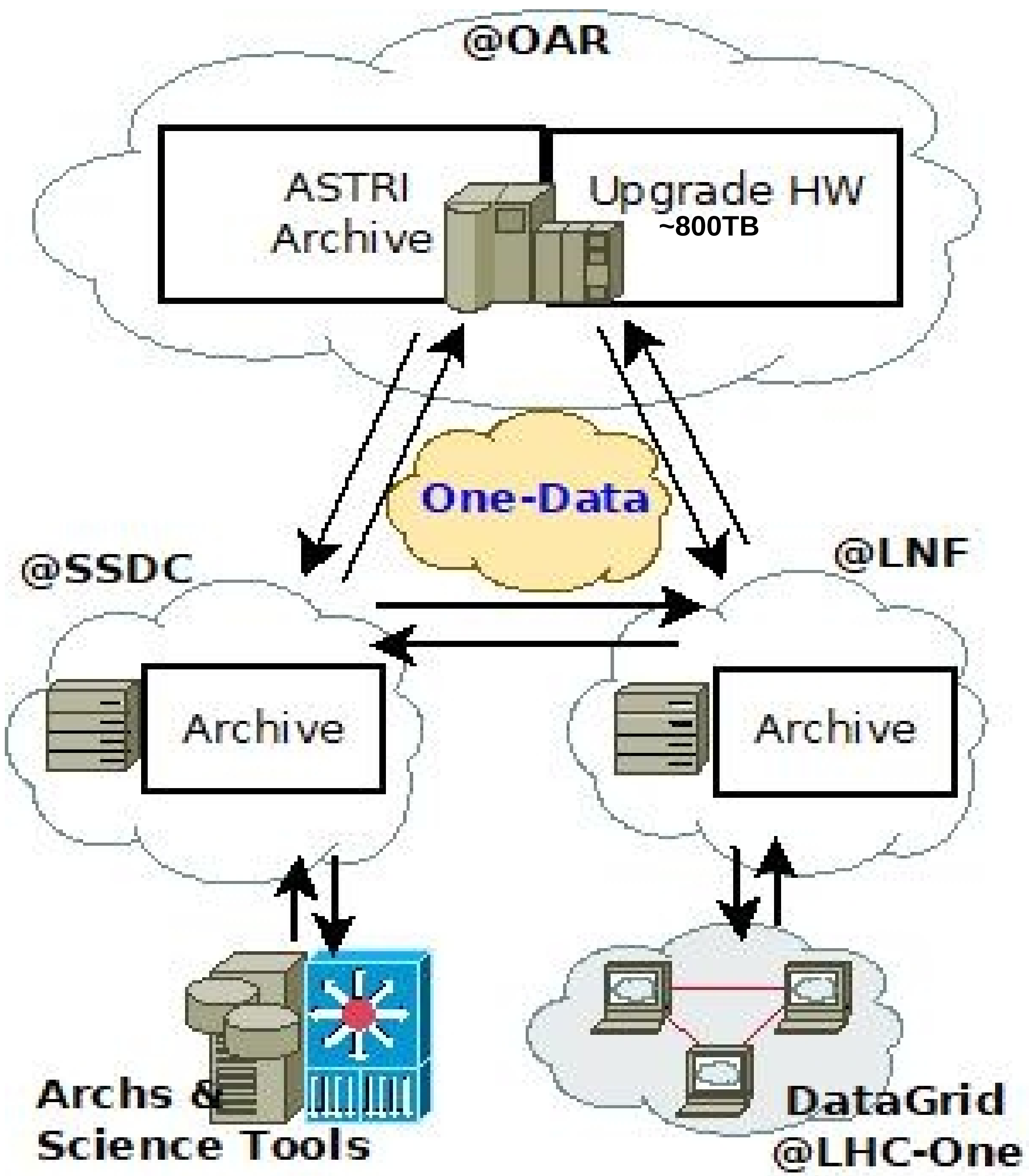
Few Snapshots
from CTA workspace
DISTRIBUTED
ARCHIVE

Prototype → Onedata's REST API's as well as oneclient command line tool for mounting virtual Onedata filesystem on the local machine

Prototype for Miniarray ARCHIVE

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

Archive Prototype testbed … (REAL+SIMUL ASTRI&CHEC DATA)



Currently working the **Archive Prototype Solution** using:

→ the **ASTRI camera real data**

→ the **INAF-PRIN ASTRI CTA Data Challenge (AC-DC)** for mini-array based simulation

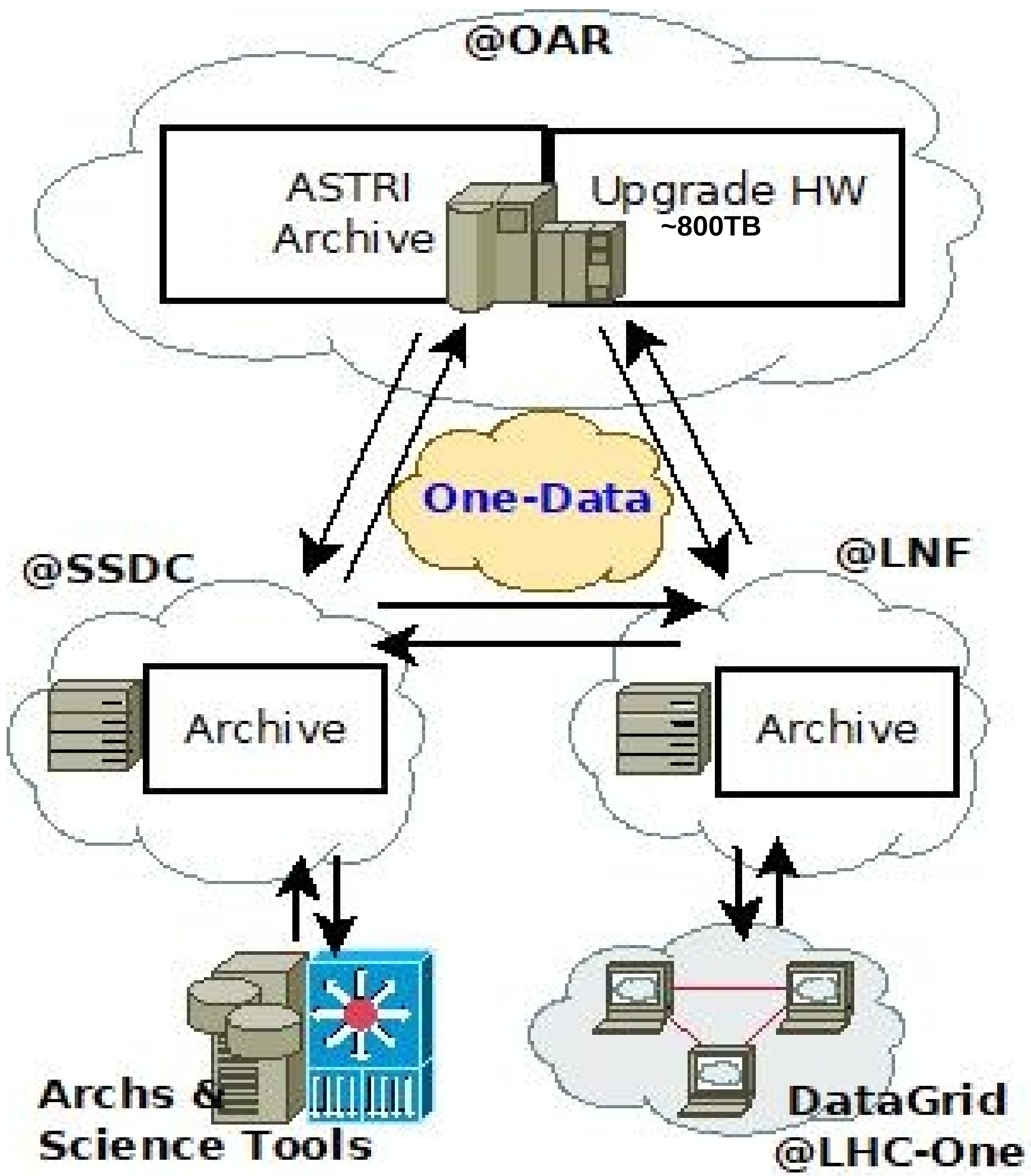
AC DC Astri/CTA Data Challenge

→ the **CHEC-M Camera Data**

since in last weeks CHEC-M Camera was hosted to the ASTRI prototype design and The ASTRI Archive is going to manage and store CHEC real data. Three CTA SST: GATE (left), ASTRI (centre), SST-1M (right).

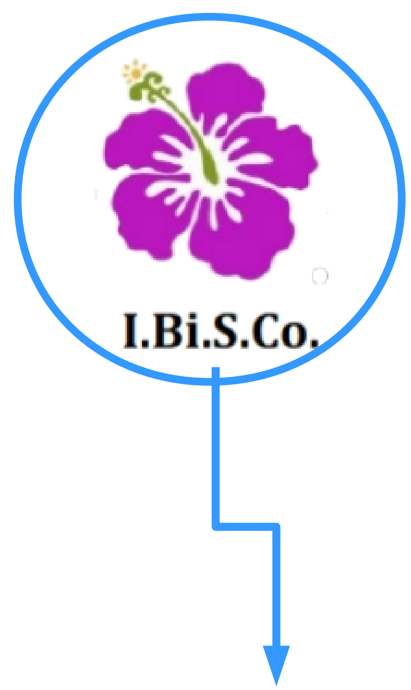
DATA SIZE TO BE Handled, Reduced & Provided by ent-to-end Miniarray

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS



Without data compression & assuming 165 operational nights/yr:

ASTRI/Prot.	→ **~0.8 TB/night**
	→ **~0.3 PB/year**
Mini-Array	→ **~3 TB/night**
	→ **~6.1 TB/night A.R.**
	→ **~1.0 PB/year A.R.**



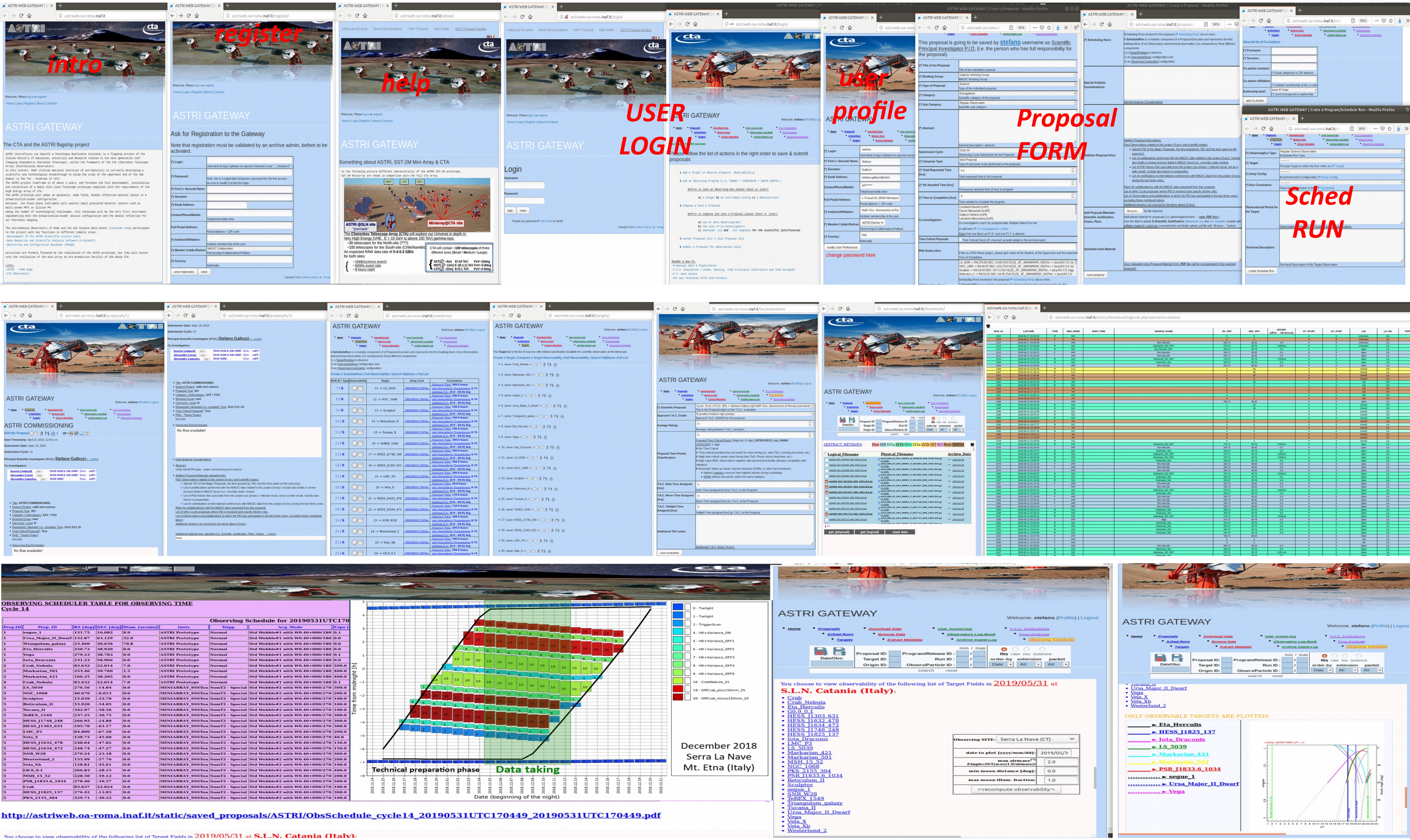
~520 kEuro
+1PB
~1500 cores

... and for CTA ?
A pessimistic scenario can involve more than 100PB/year !
→ triggering systems?
→ data reduction on site?
or...
→ challenging with big data

PON INFN
Memorandum Of Understanding
(INAF ↔ INFN)

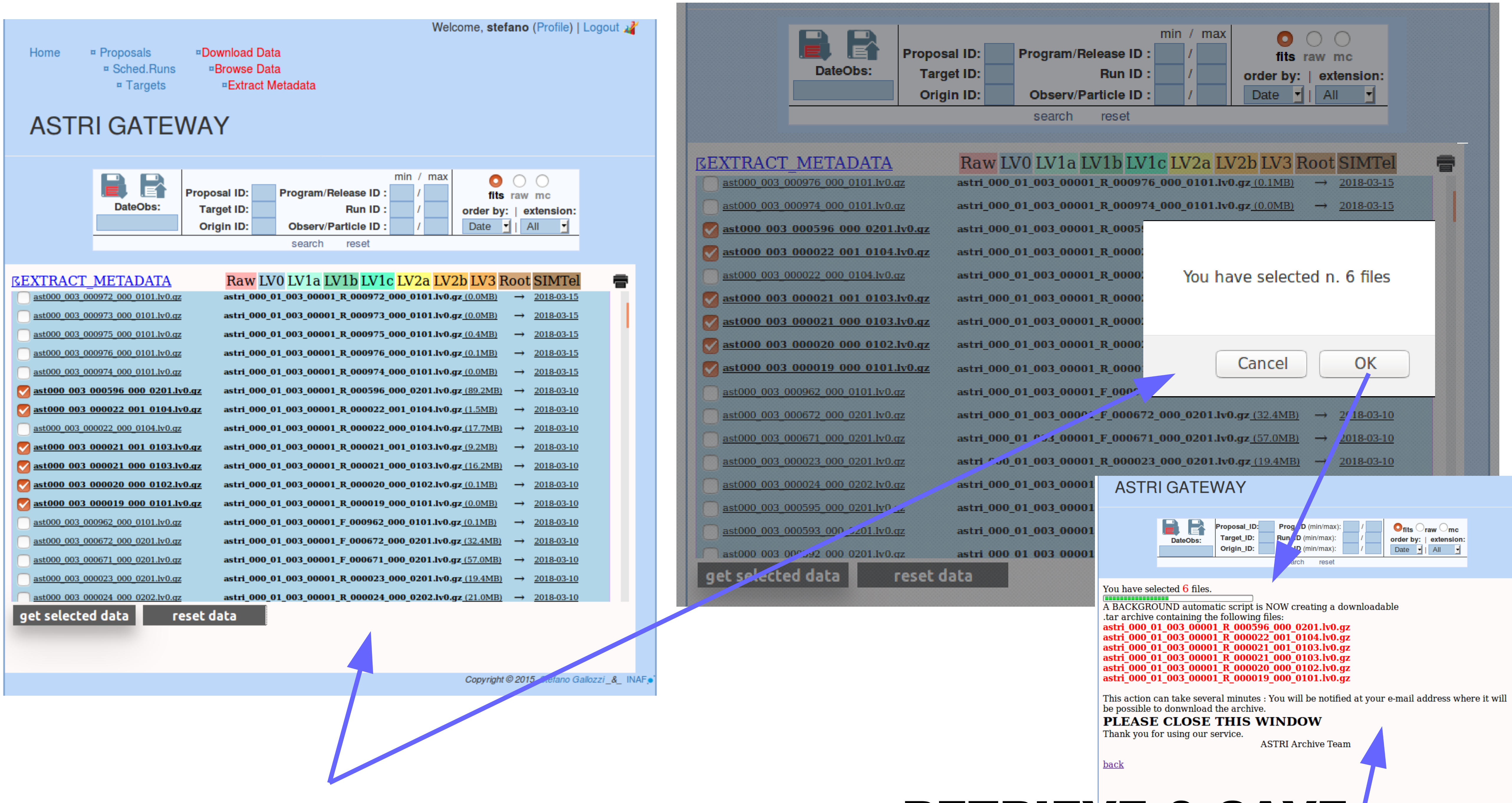
Gateway & Proposal Handling System

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS



USER Data ACCESS

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

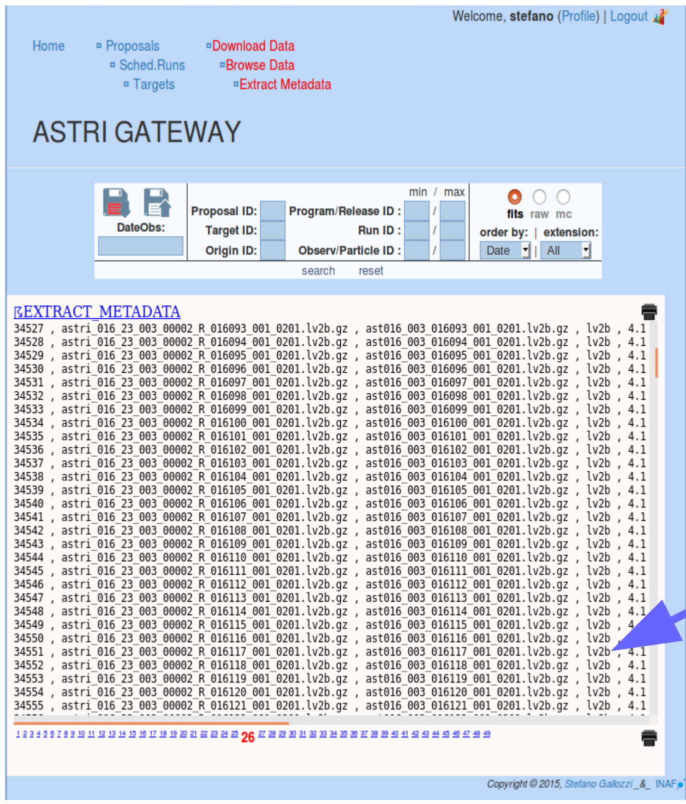


QUERY & SELECT DATA
(from lev0 to lev5 / photon list)

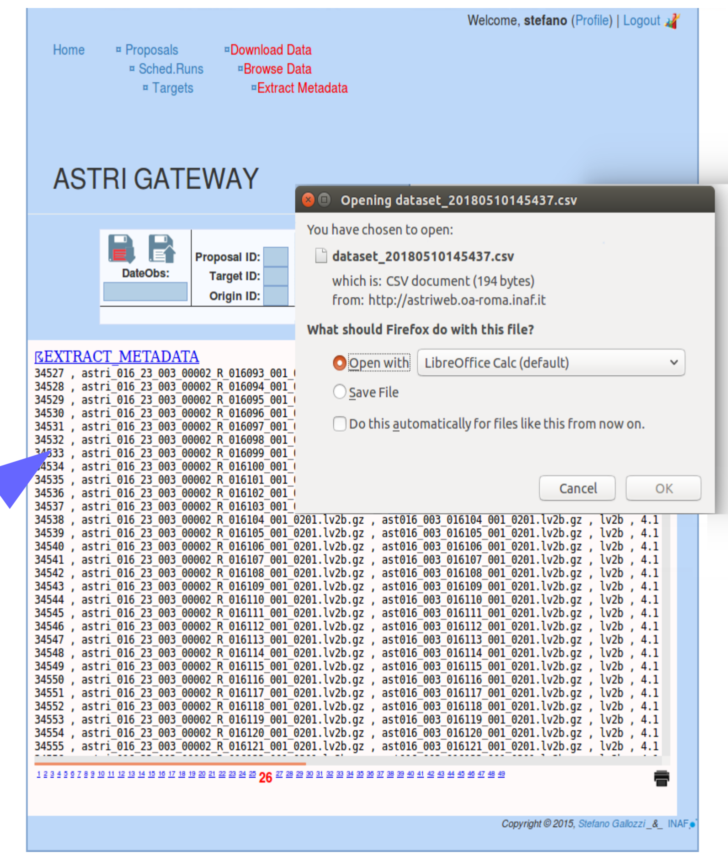
RETRIEVE & SAVE ALL LEVEL DATASETs

USER Data ACCESS

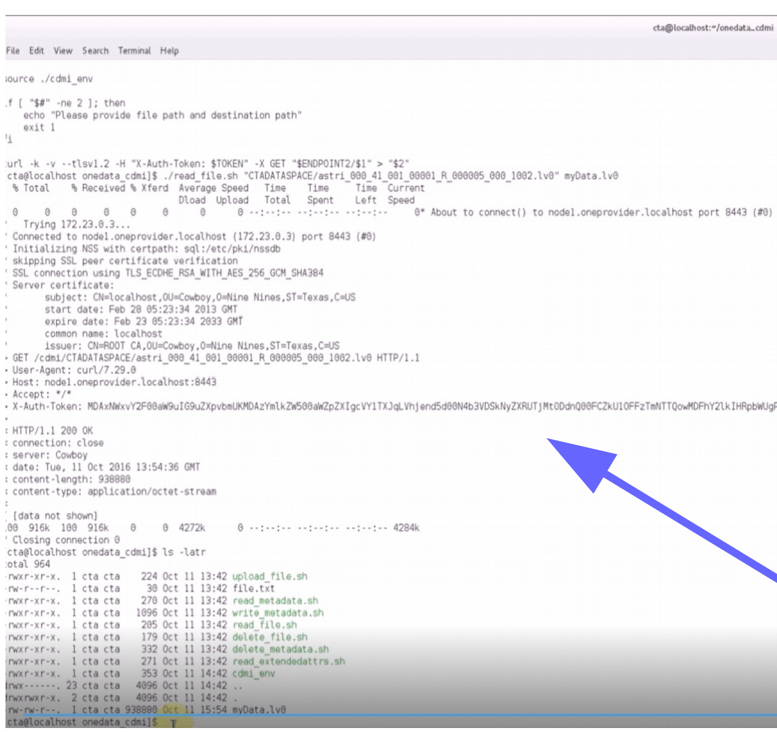
INAF ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

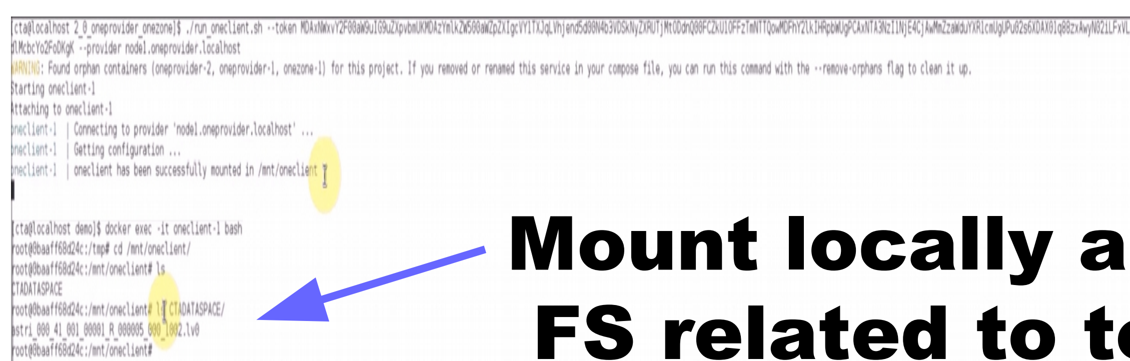


BROWSE, INSPECT DATA & SIMULATIONS (from lev0 to lev5)



EXTRACT META-DATA SETs

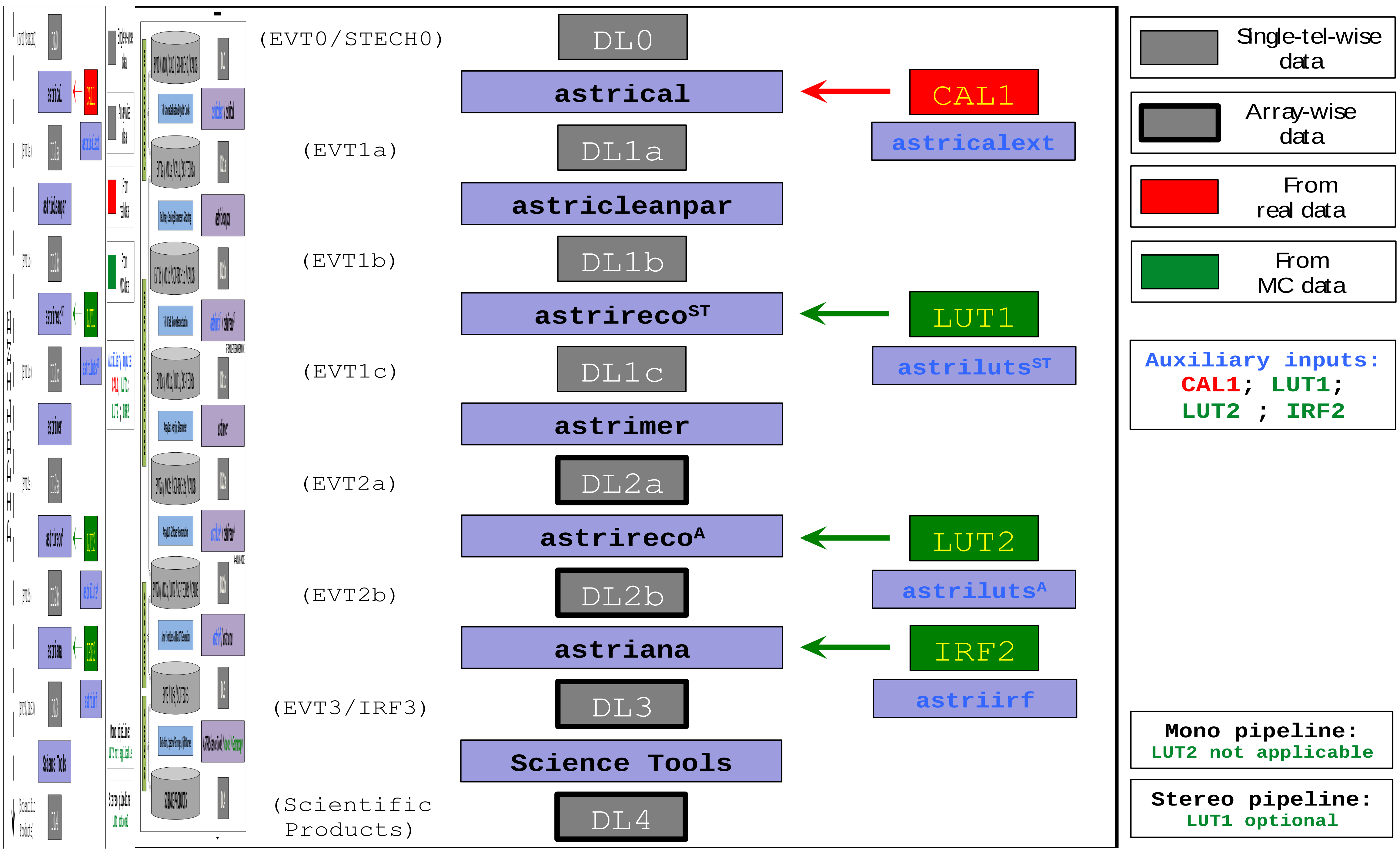




Mount locally a VIRTUAL FS related to token auth

OneData Client: get a token to access to DATA SETs related to a custom query

Data Processing in ASTRI-Mini Array



Breakdown stages; Basic functionalities; **Pipeline modules;** I/O data levels.

SPIE 991315 (2016); SPIE 107070 (2018);

Data Processing in ASTRI-Mini Array

INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

→ R0 (raw data streaming from telescope cameras) ←

- **Level 0 (DL0)**: raw data from the hardware/software data acquisition components that are permanently archived; **+ stereo info**

- **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*)

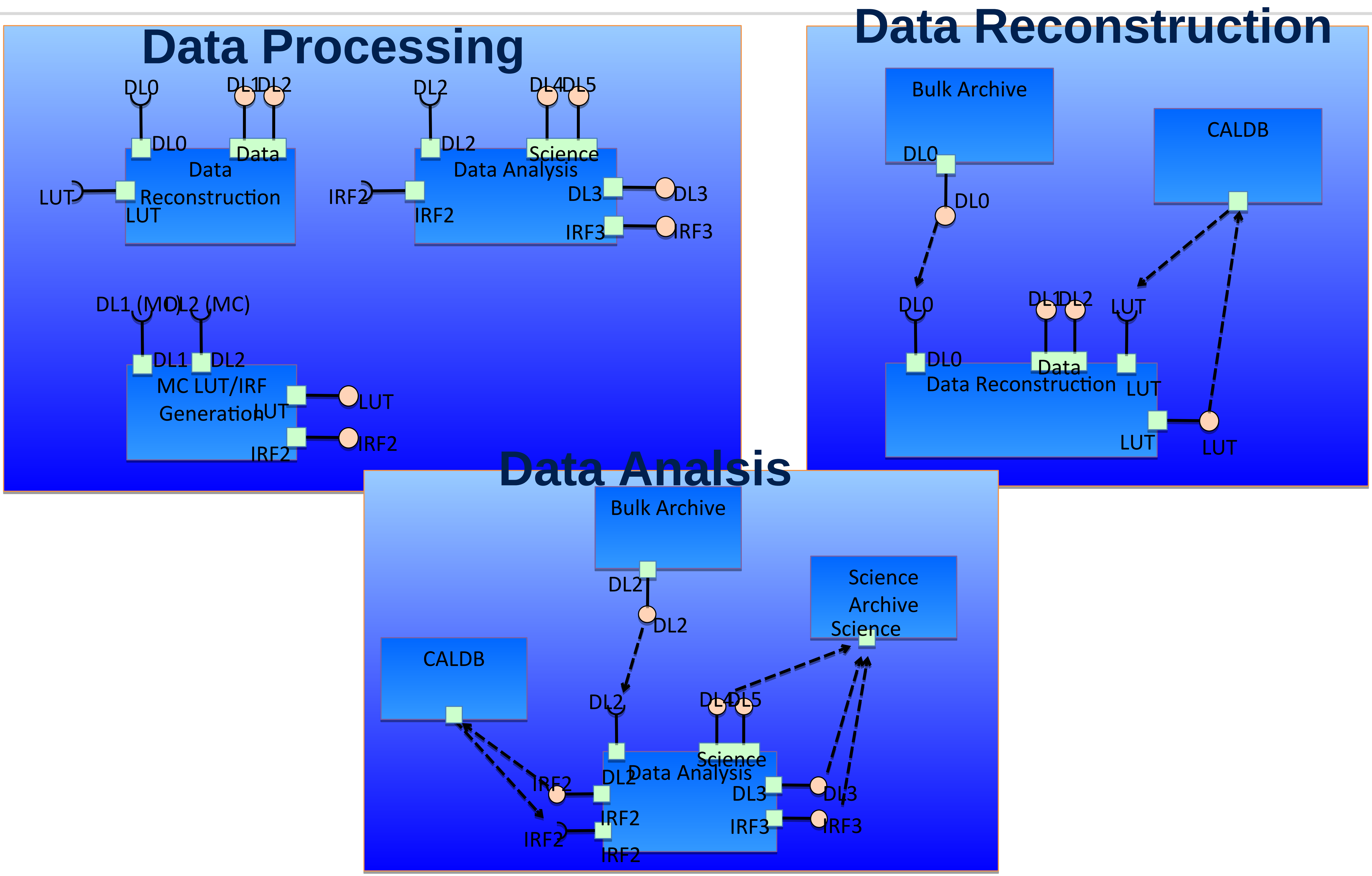
- **Level 3 (DL3)**: reduced data (*selected list of events plus corresponding instrument response functions*);

- **Level 4 (DL4)**: science data (*high-level scientific data products*);

- **Level 5 (DL5)**: observatory data (*legacy observatory data and catalogs*).

FITS data format at all levels from DL0

Functional Overview

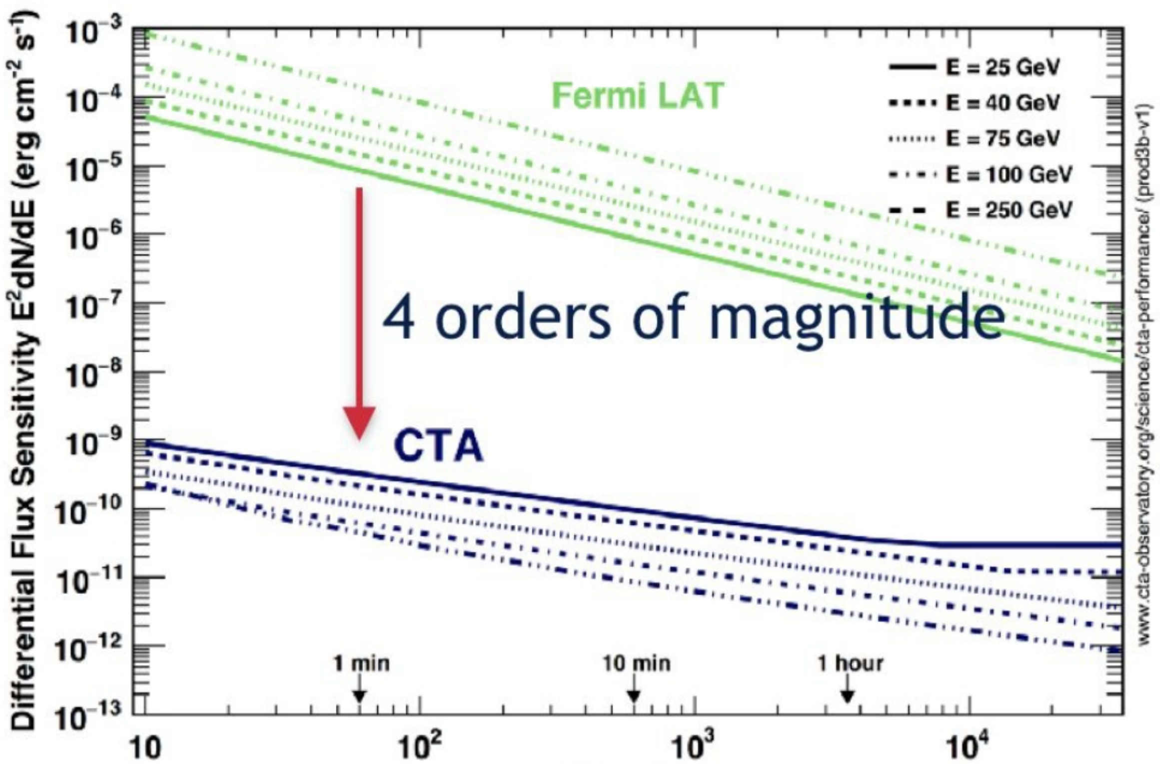


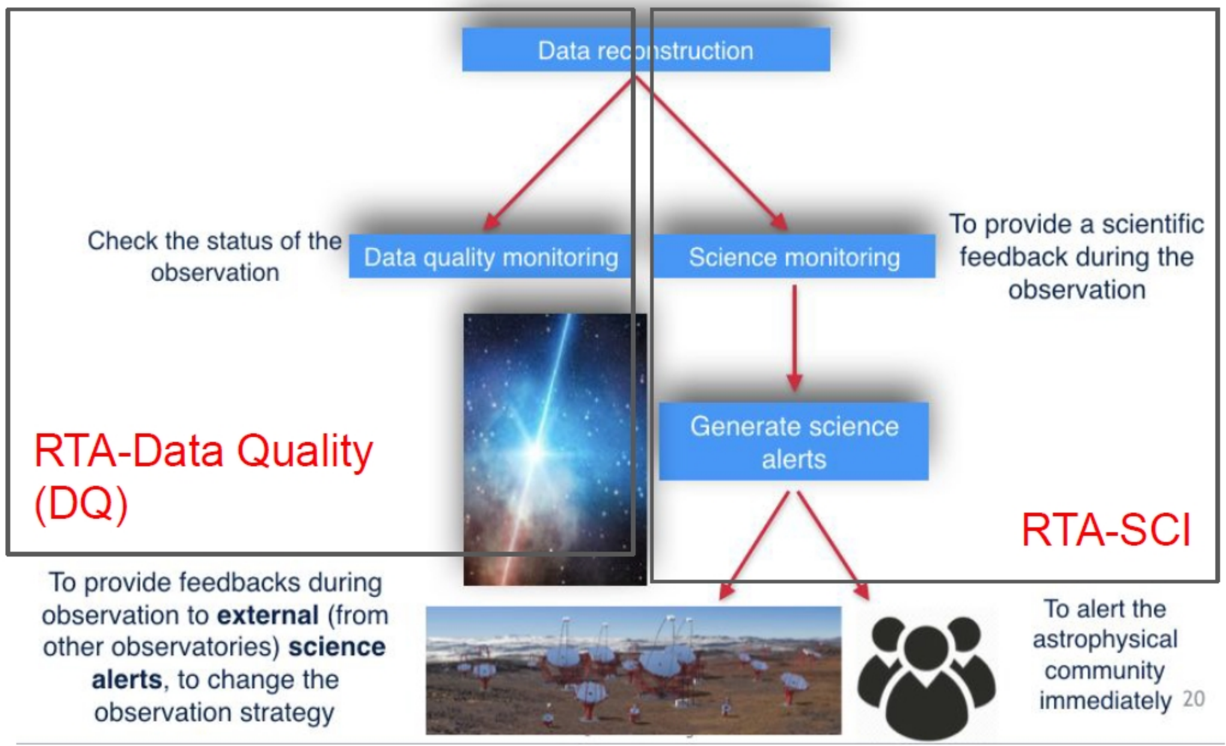
Data Processing in ASTRI-Mini Array

- **EVT*n*** (*scientific event-list data*): EVT(0,1a,1b,1c,2a,2b,3)
  - from raw data to high-level fully reduced event-list data
- **MC*n*** (*Monte Carlo event-list data*): similar definitions as in EVT*n*
- **CAL*n*** (*calibration data*): CAL(0,1,2)
  - used for cameras, optics, and array calibrations
- **MC-CAL*n*** (*Monte Carlo calibration data*): similar definitions as in CAL*n*
- **SCI-TECH*n*** (*set of technical data for scientific data reduction*)
- **LUT*n*** (*look-up-tables data*): LUT(1,2)
  - used for telescope-/array-wise event reconstruction
- **IRF*n*** (*instrument response functions data*): IRF(2,3)

The CTA Real-Time Analysis (@ OAS-Bo)

- The **Real-Time Analysis (RTA)** is a **software system** that analyse CTA data **during the observation.** The CTA data rate is 2-5 GB/s.
- **On-site** with the telescopes
- The RTA must be capable of **issuing science alerts when new transient phenomena are detected, with a latency of 30s**
- The RTA must **search for transient** phenomena on different timescales **from 10 seconds to 30 minutes,**
- The **sensitivity** of the analysis is required not to be worse than the one of the final analysis by more than a factor of 2
- INAF/OAS leads activities of the CTA Real-Time Analysis and CTA short-term sensitivity
- The CTA Real-Time Analysis is the key system in the context of **multi-messenger** and **multi-wavelength** astronomy.

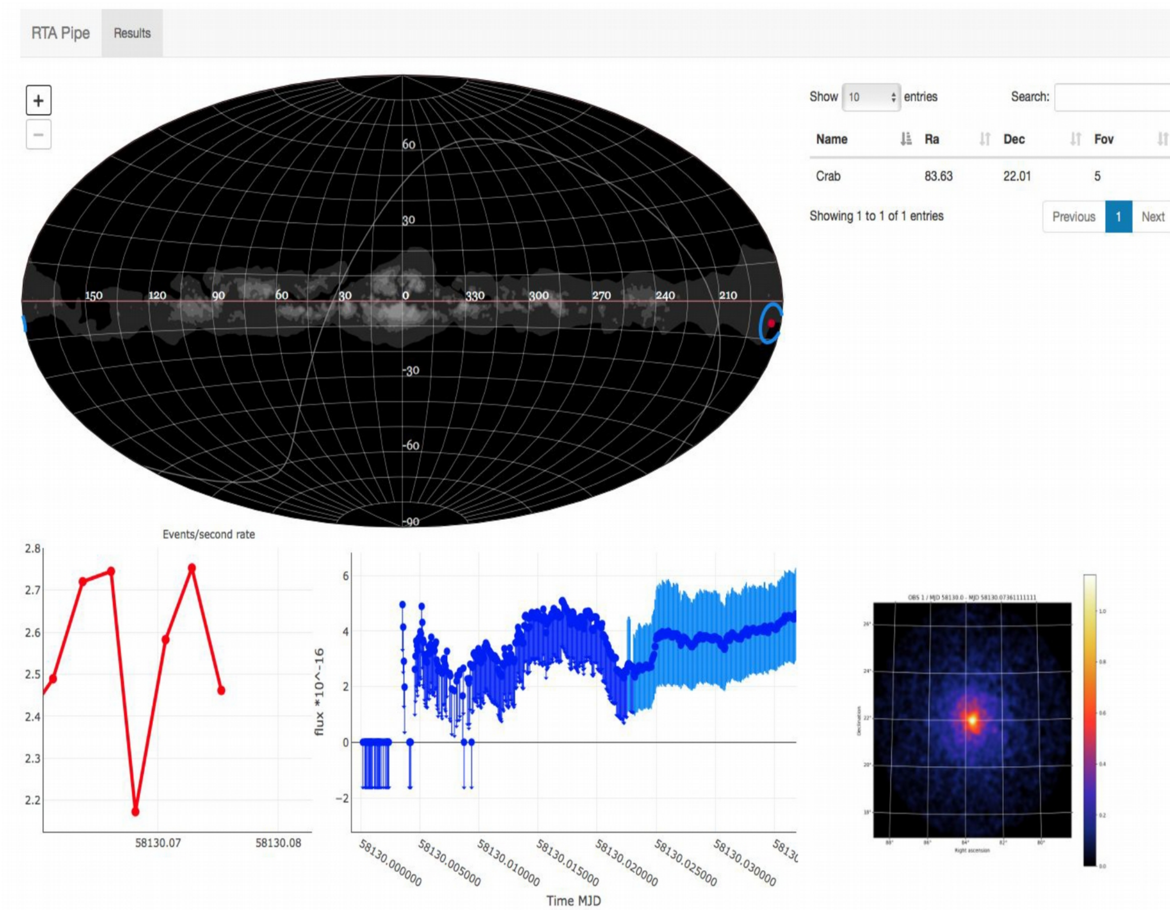




The CTA Real-Time Analysis (@ OAS-Bo)

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

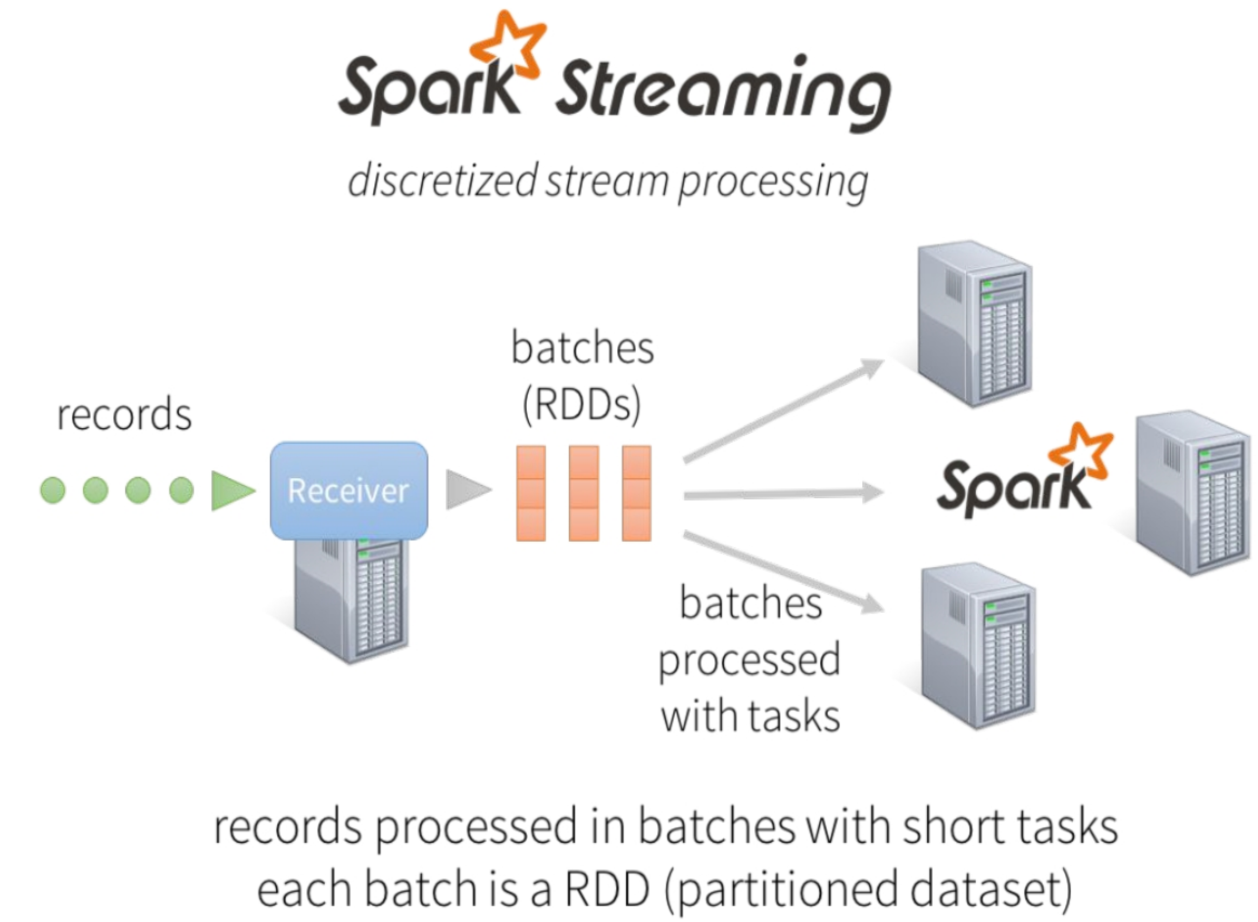
- The RTA-SCI is the software component that performs real-time scientific analyzes on CTA data, starting from the reconstructed event list
- The RTA-SCI produces different scientific output: Light Curves, Counts Maps, Alerts, Detections and TS Maps.
- The RTA-SCI runs multiple analysis in parallel: for this reason it requires a framework for scalability and flexibility
- The RTA-GUI queries the results database to show the results of the RTA-SCI using plots

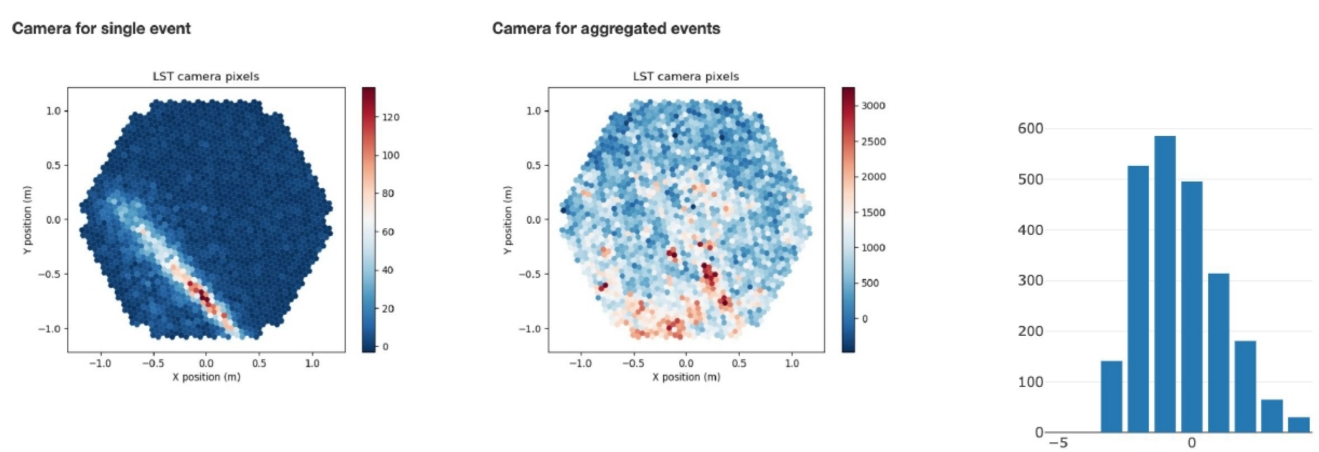


Real Time Data Quality & Quick Look

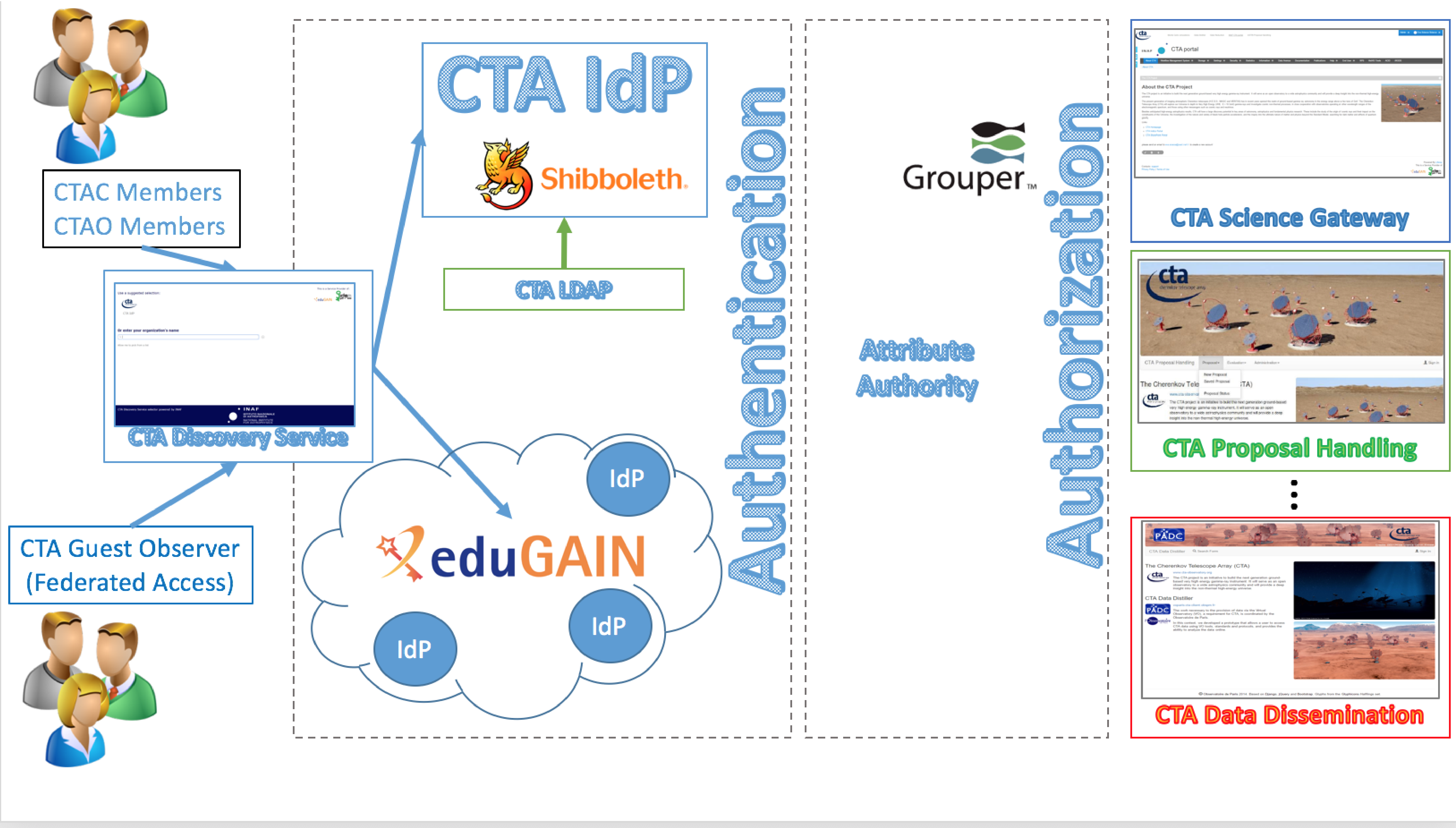
INAF ISTITUTO NAZIONALE DI ASTROFISICA NATIONAL INSTITUTE FOR ASTROPHYSICS

- During the reconstruction and the data quality analysis we have high data volume and data rate (kHz/s, GB/s).
- The data output from different reconstruction steps must be stored or buffered to run data quality analysis
- The data reduction streaming process can be done in parallel using framework like Spark in order to scale the application on a Cluster
- The RTA-DQ-GUI shows to the operator the results of the DQ checks: warning, alarms and plots.

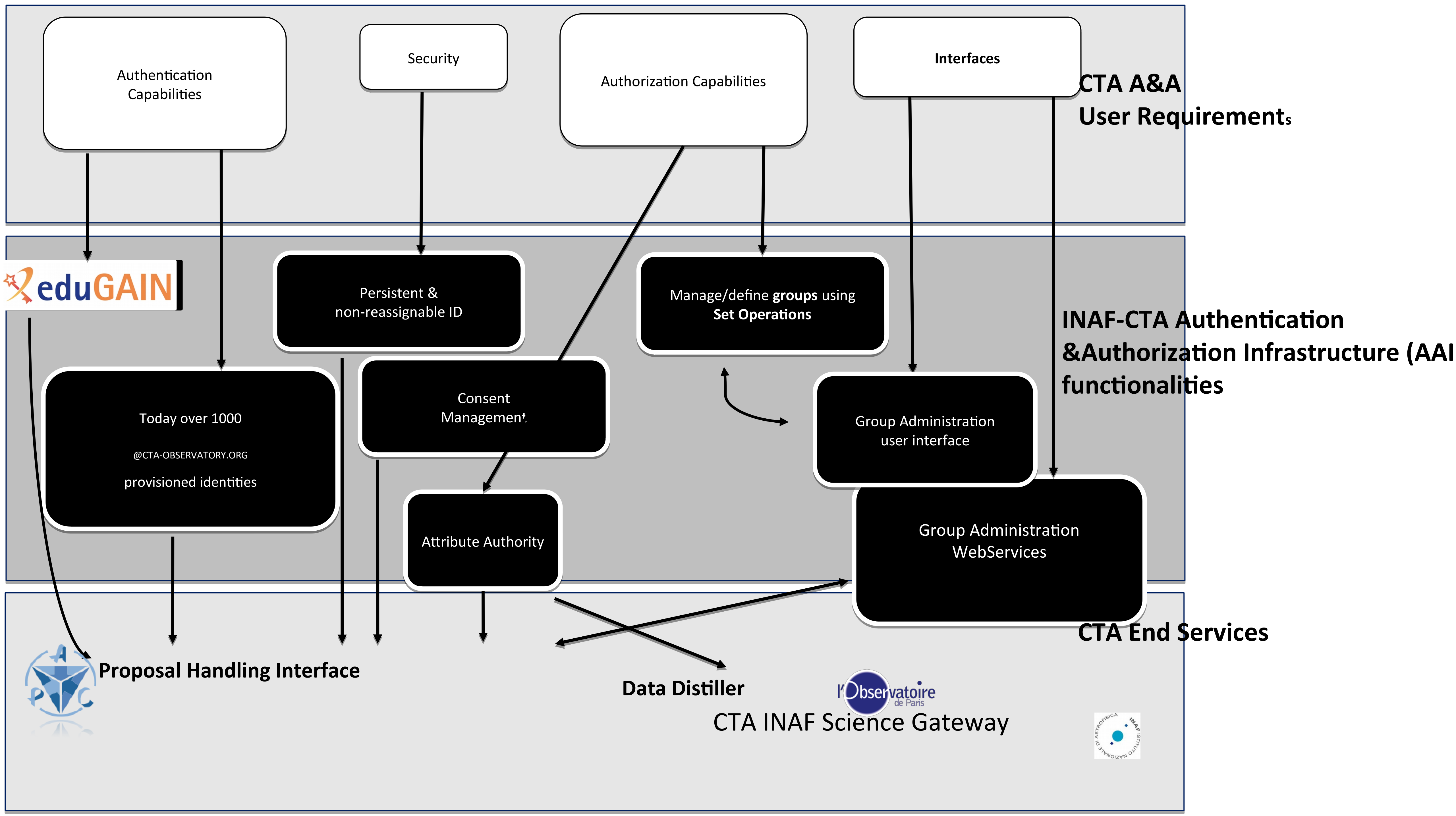




INAF AAI in CTA (@ OA-Ct)



User Req → Research Infra



INAF CTA Science Gateway

https://cta-sg.oact.inaf.it/

- Offers an environment to **orchestrate** scientific **applications** and **tools** with little concern for learning and managing the underlying **e-infrastructures** to execute them through a **Workflow Management System**.
- A **demonstrator** has been implemented following a typical CTA analysis performed with the **Fermi Science Tools**.
- Allows **interactive tasks** performed by a virtual desktop environment (ACID).
- **Connected** to the **INAF CTA AAI**
- Recently part of **EduGAIN Service Providers**
- See a **Demo** at: https://youtu.be/Qru6joO-Vw8

INAF CTA Science Gateway

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

https://cta-sg.oact.inaf.it/

cta cherenkov telescope array

Monte Carlo simulations | Data Distiller | Data Reduction | INAF CTA portal | ASTRI Proposal Handling

Sign In

INAF CTA portal

About CTA | Statistics | Documentation | Publications

About CTA

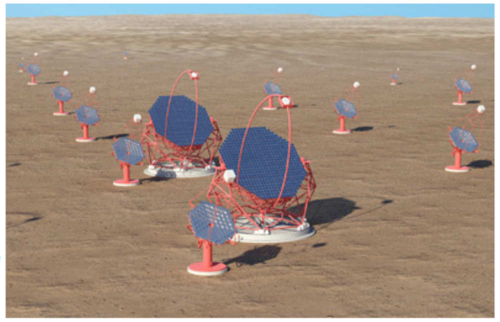
Web Content Display

About the CTA Project

The CTA project is an initiative to build the next generation ground-based very high energy gamma-ray instrument. It will serve as an open observatory to a wide astrophysics community and will provide a deep insight into the non-thermal high-energy universe.

The present generation of imaging atmospheric Cherenkov telescopes (H.E.S.S., MAGIC and VERITAS) has in recent years opened the realm of ground-based gamma ray astronomy in the energy range above a few tens of GeV. The Cherenkov Telescope Array (CTA) will explore our Universe in depth in Very High Energy (VHE, E > 10 GeV) gamma-rays and investigate cosmic non-thermal processes, in close cooperation with observatories operating at other wavelength ranges of the electromagnetic spectrum, and those using other messengers such as cosmic rays and neutrinos.

Besides anticipated high-energy astrophysics results, CTA will have a large discovery potential in key areas of astronomy, astrophysics and fundamental physics research. These include the study of the origin of cosmic rays and their impact on the constituents of the Universe, the investigation of the nature and variety of black hole particle accelerators, and the inquiry into the ultimate nature of matter and physics beyond the Standard Model, searching for dark matter and effects of quantum gravity.



Links:

- CTA Homepage
- CTA Indico Portal
- CTA SharePoint Portal

please send an email to eva.sciacca@oact.inaf.it to create a new account

Contacts: support
Privacy Policy | Terms of Use

LIFERAY

Powered By Liferay
This is a Service Provider of:
eduGAIN idem garr aai

- ✓ Based on LIFERAY
- ✓ Integrates a workflow management system (**gUSE/WS-PGRADE**) executing jobs on the major DCIs
- ✓ Integrated with the other CTA web applications

QUESTIONS?

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

BACKUP SLIDES

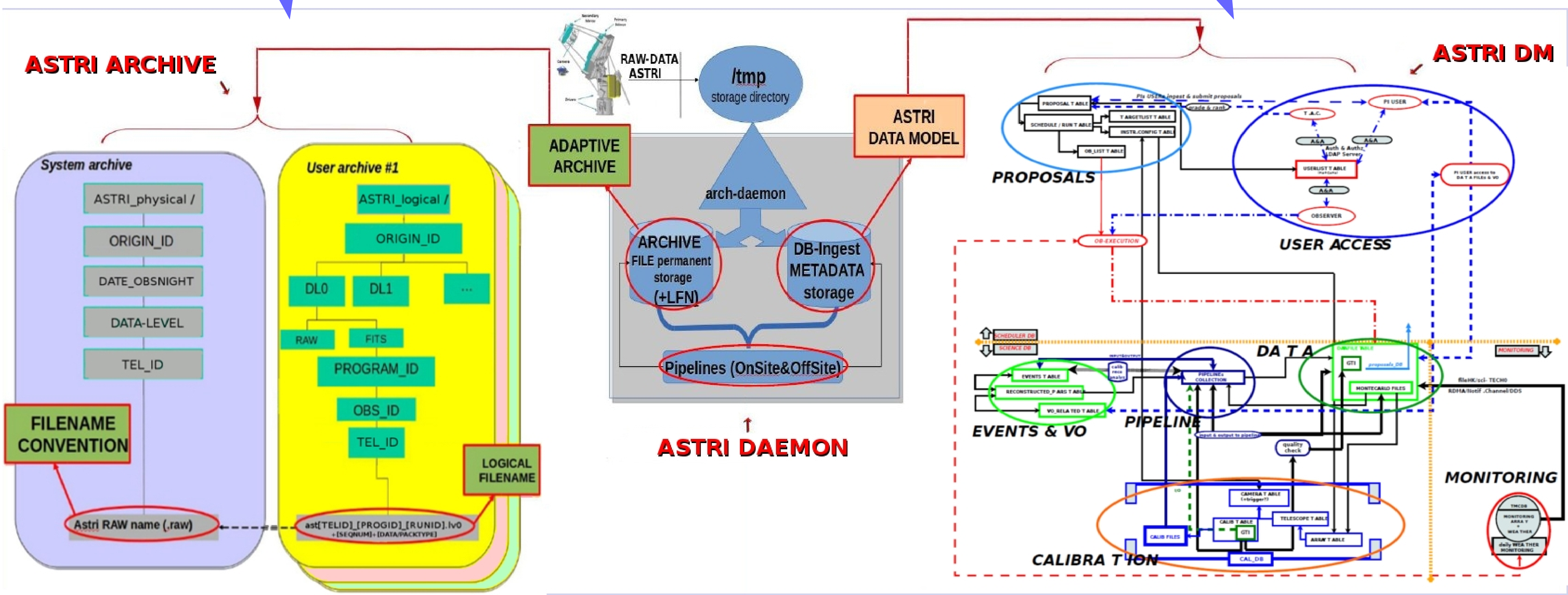
The ASTRI Archive System

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

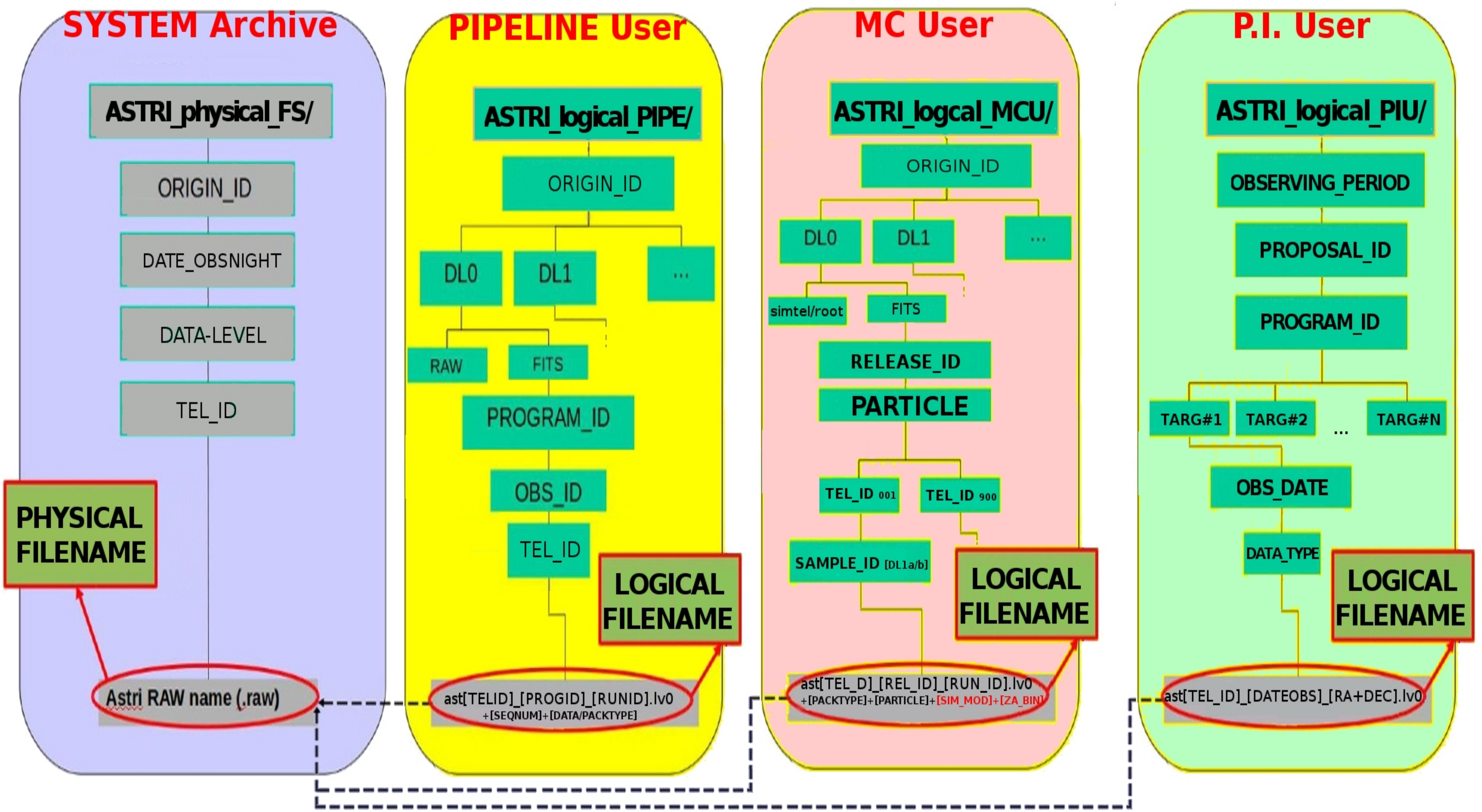
REPOSITORIES
(mySQL & mongo)

DATABASES (physical & logical)

+ couch-base



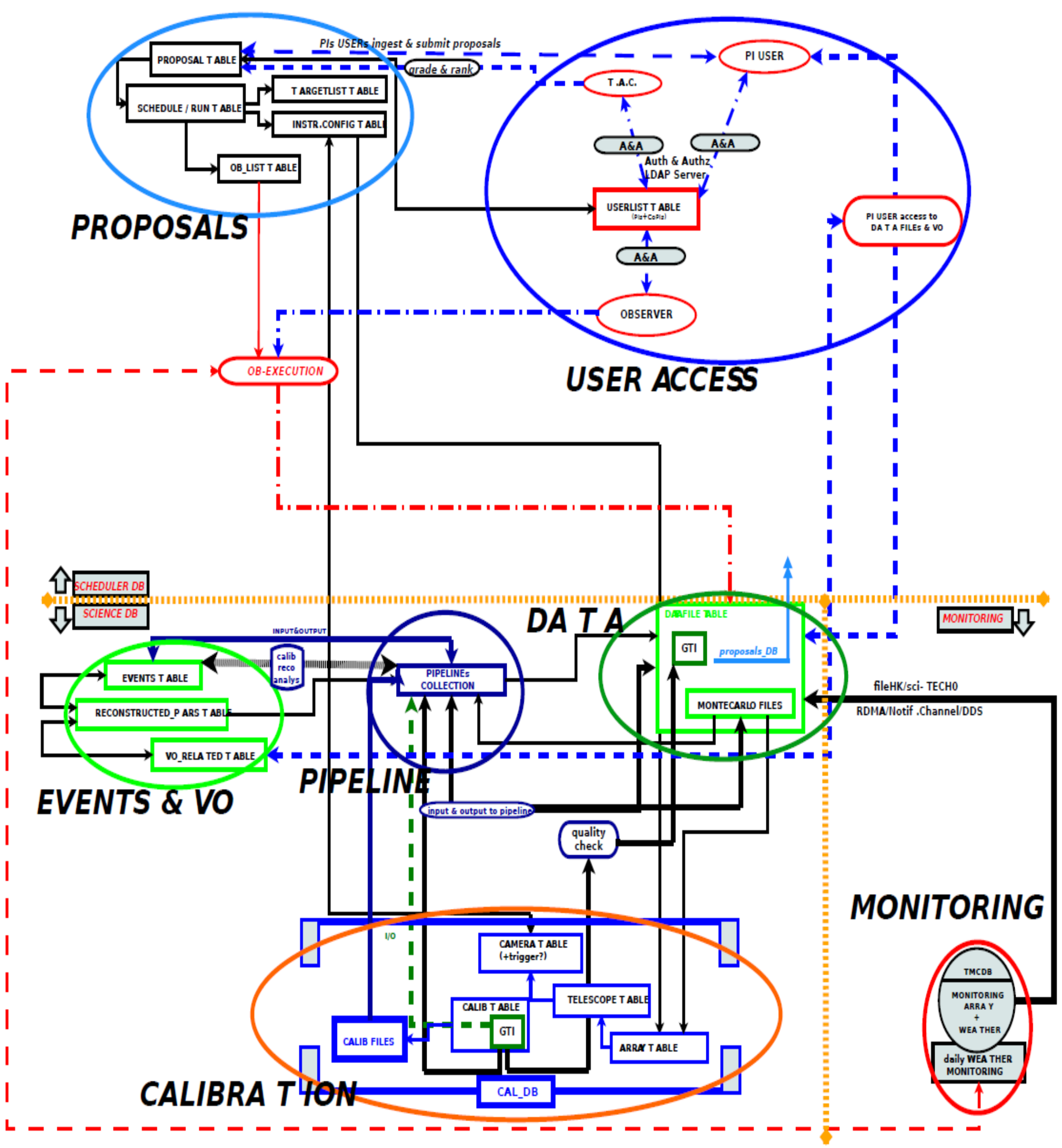
ASTRI Logical and Physical Archives

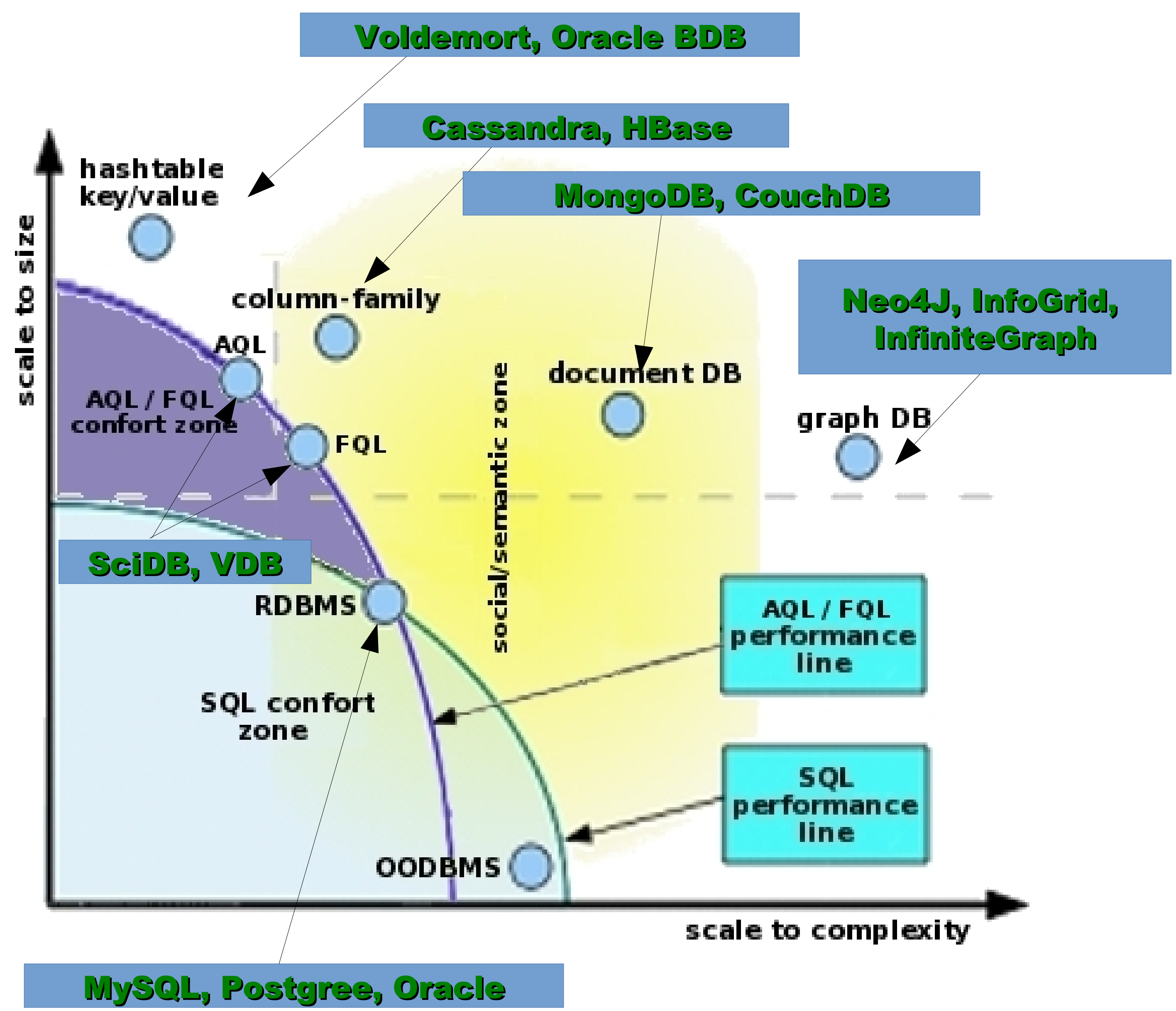


Knowledge Discovery DBs

INAF
ISTITUTO NAZIONALE DI ASTROFISICA
NATIONAL INSTITUTE FOR ASTROPHYSICS

...testing databases technology using ASTRI data-model...



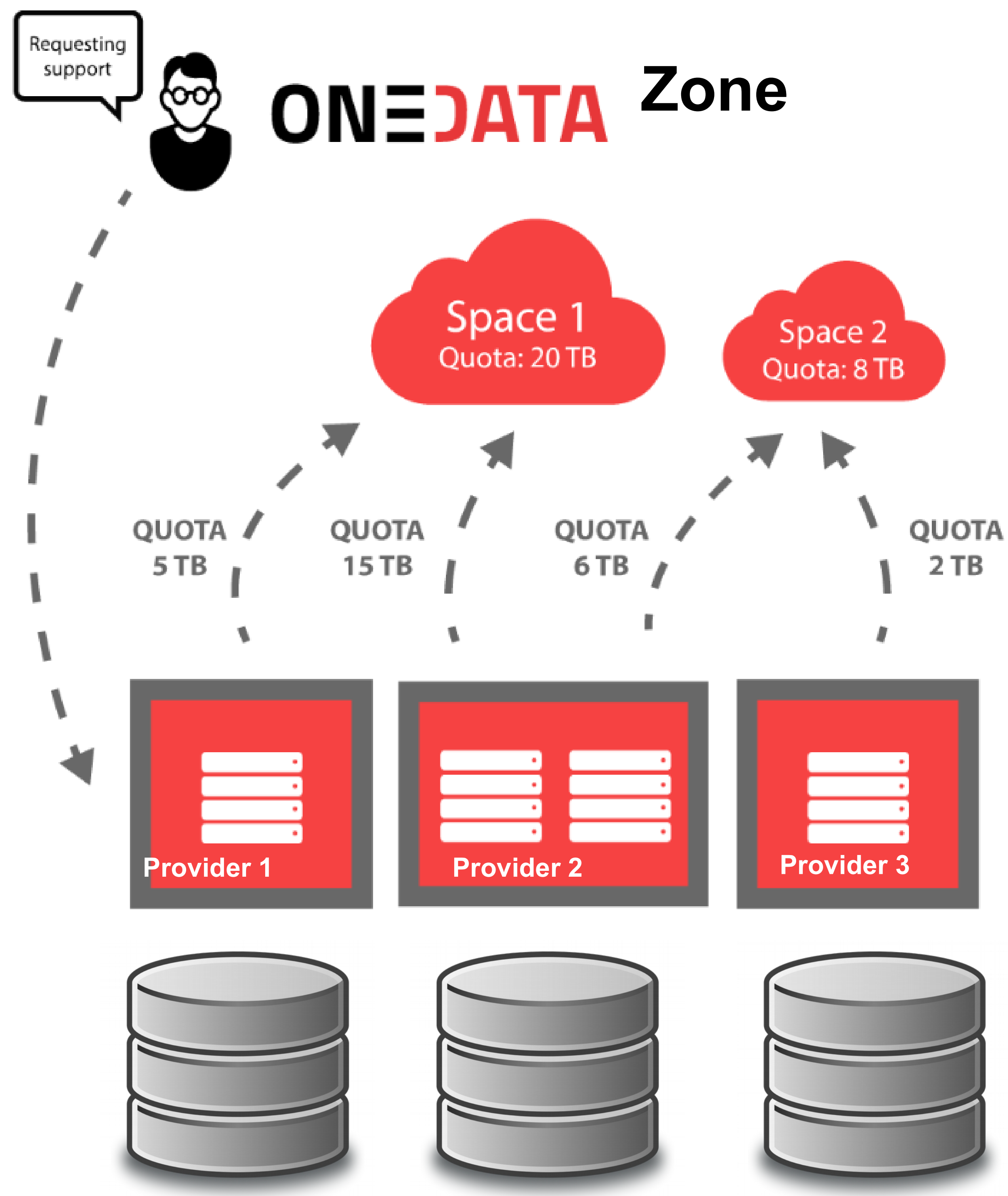


OneData Overview

OneData system **virtualizes** storage systems provided by storage resource providers **distributed** globally.

The most important concepts of the platform are:

- **Spaces** - distributed virtual volumes, where users can organize their data
- **Providers** - entities who support spaces with actual storage resources
- **Zones** - federations of providers which enable creation of closed or interconnected communities.

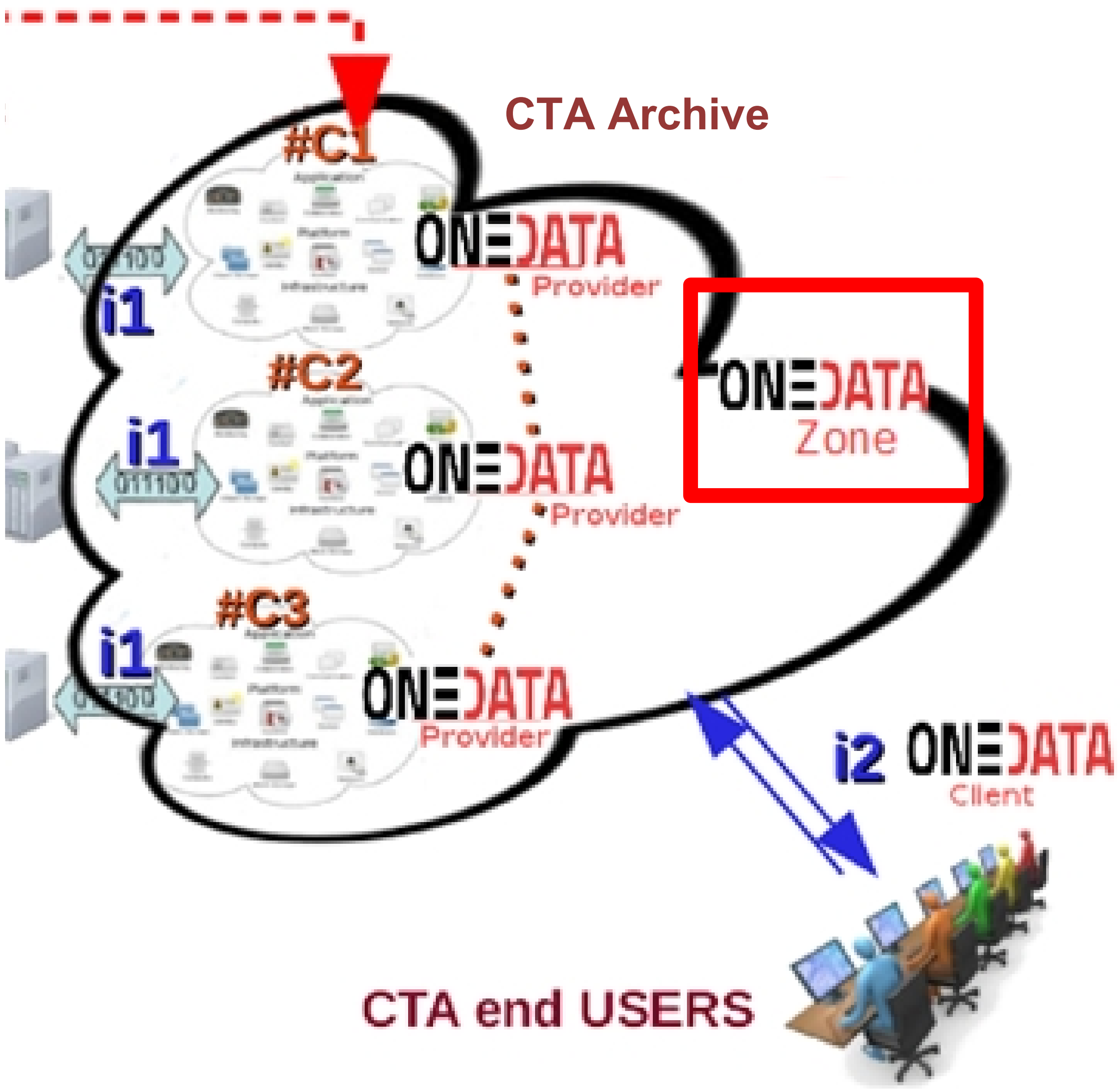


CTA OneZone

- **OneZone** is the gateway for users to the OneData system. It is responsible for connecting to the **authentication** and **authorization** infrastructure.

- It allows users to:

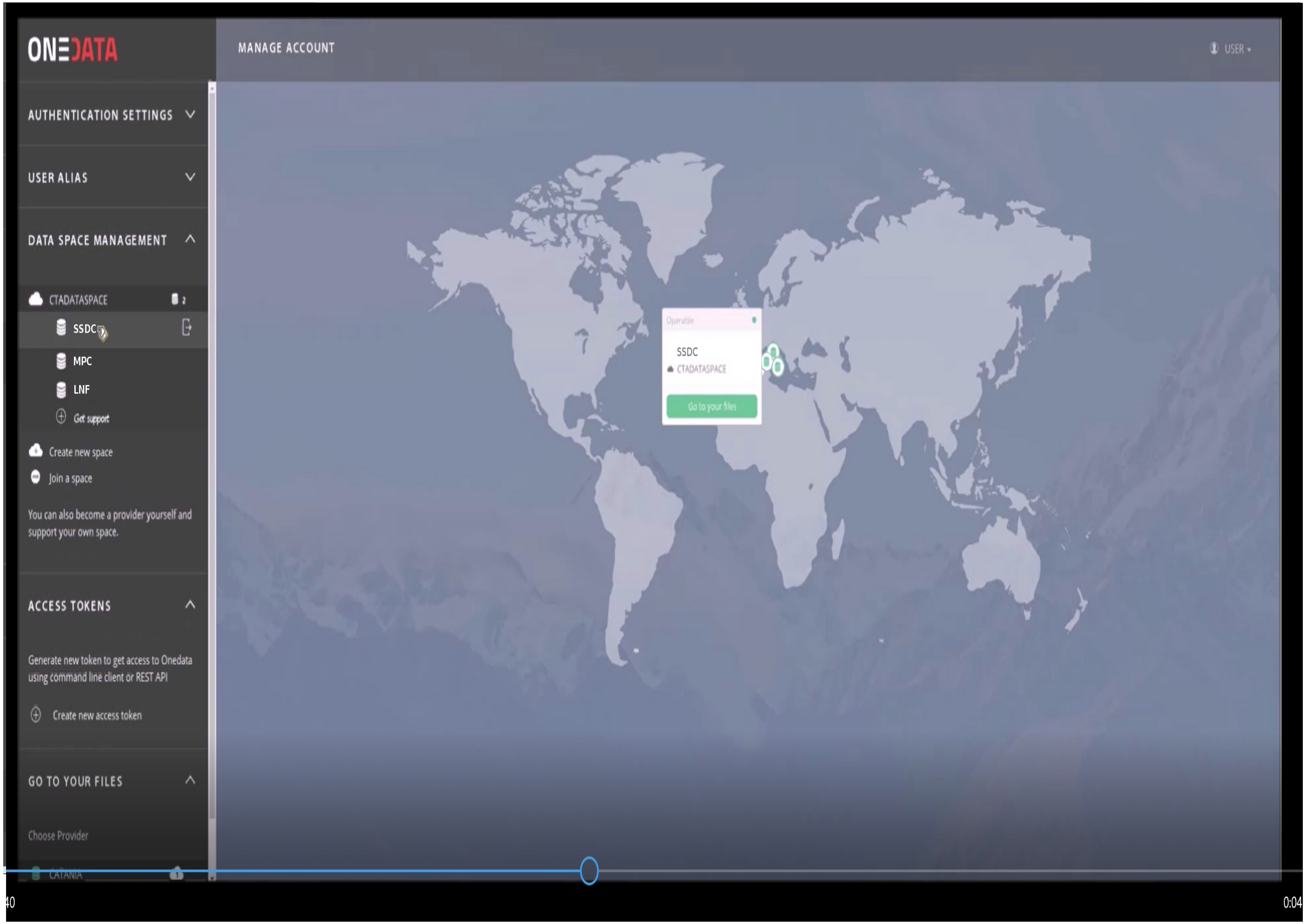
✓ create **user spaces**
✓ generate space **support tokens**, that can be used to support user spaces with storage from a dedicated storage provider
✓ **monitor availability** of storage providers that support user spaces
✓ see the **geographical distribution** of storage providers
✓ **choose storage provider** for spaces



CTA OneZone

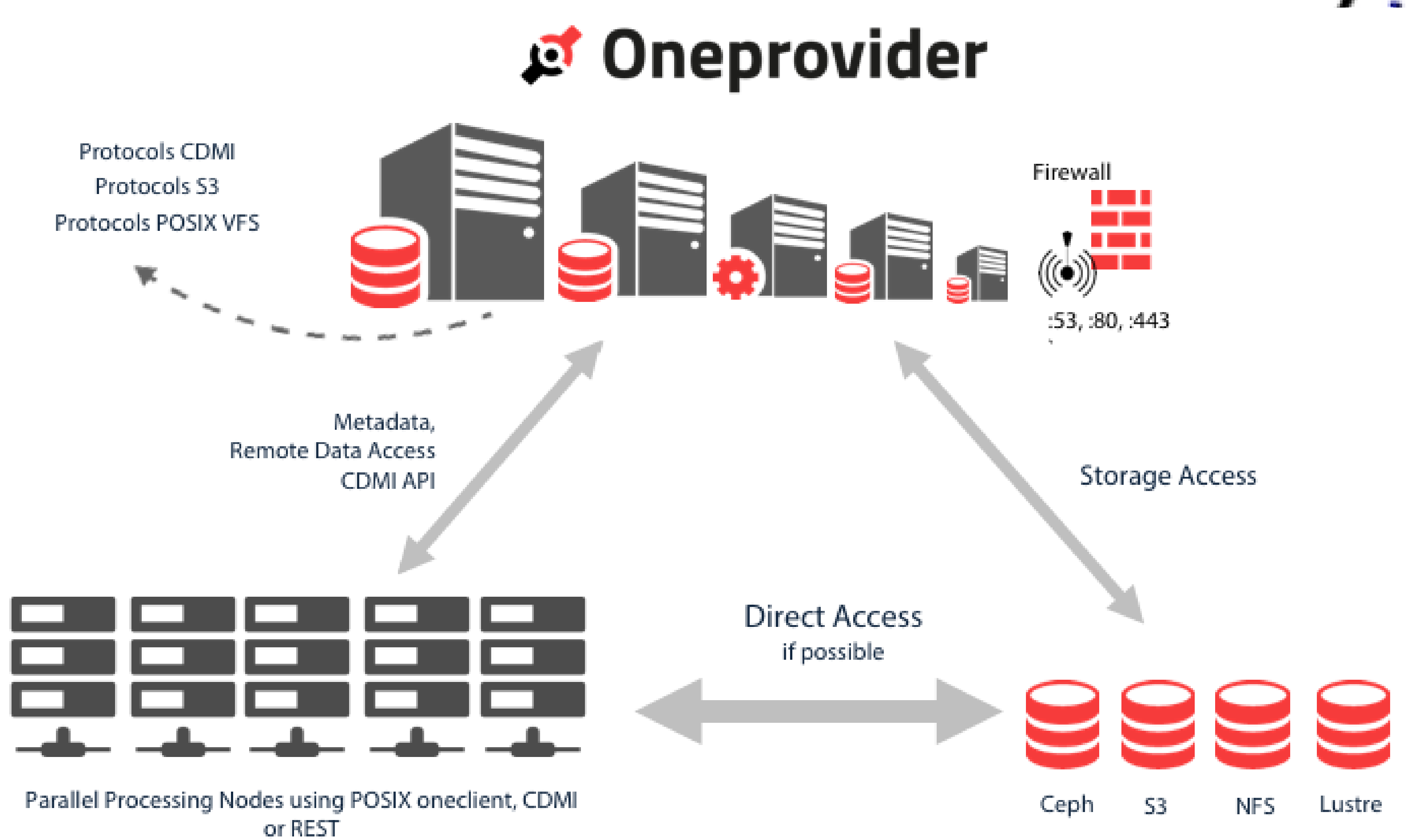
Spaces

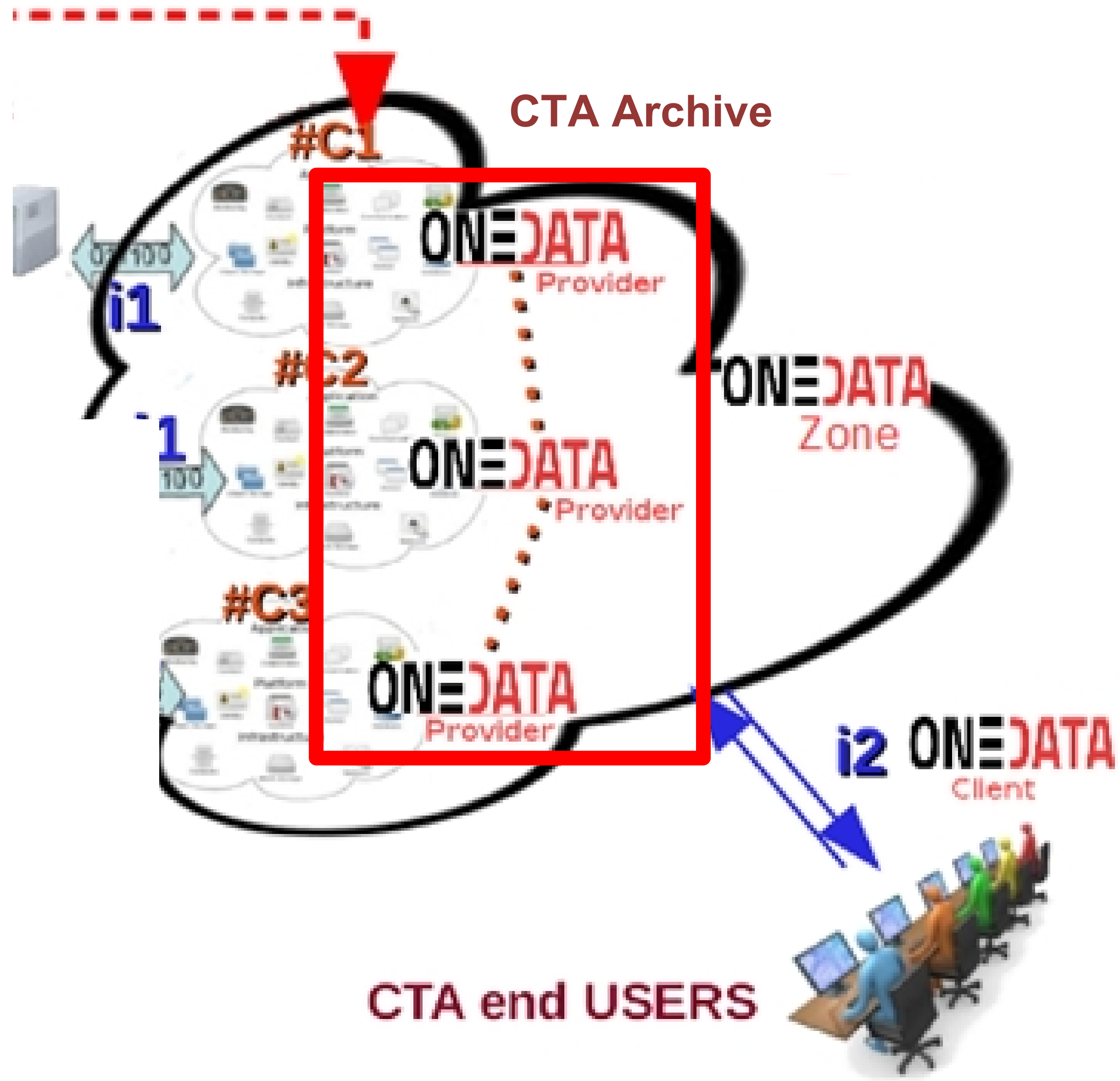
Providers



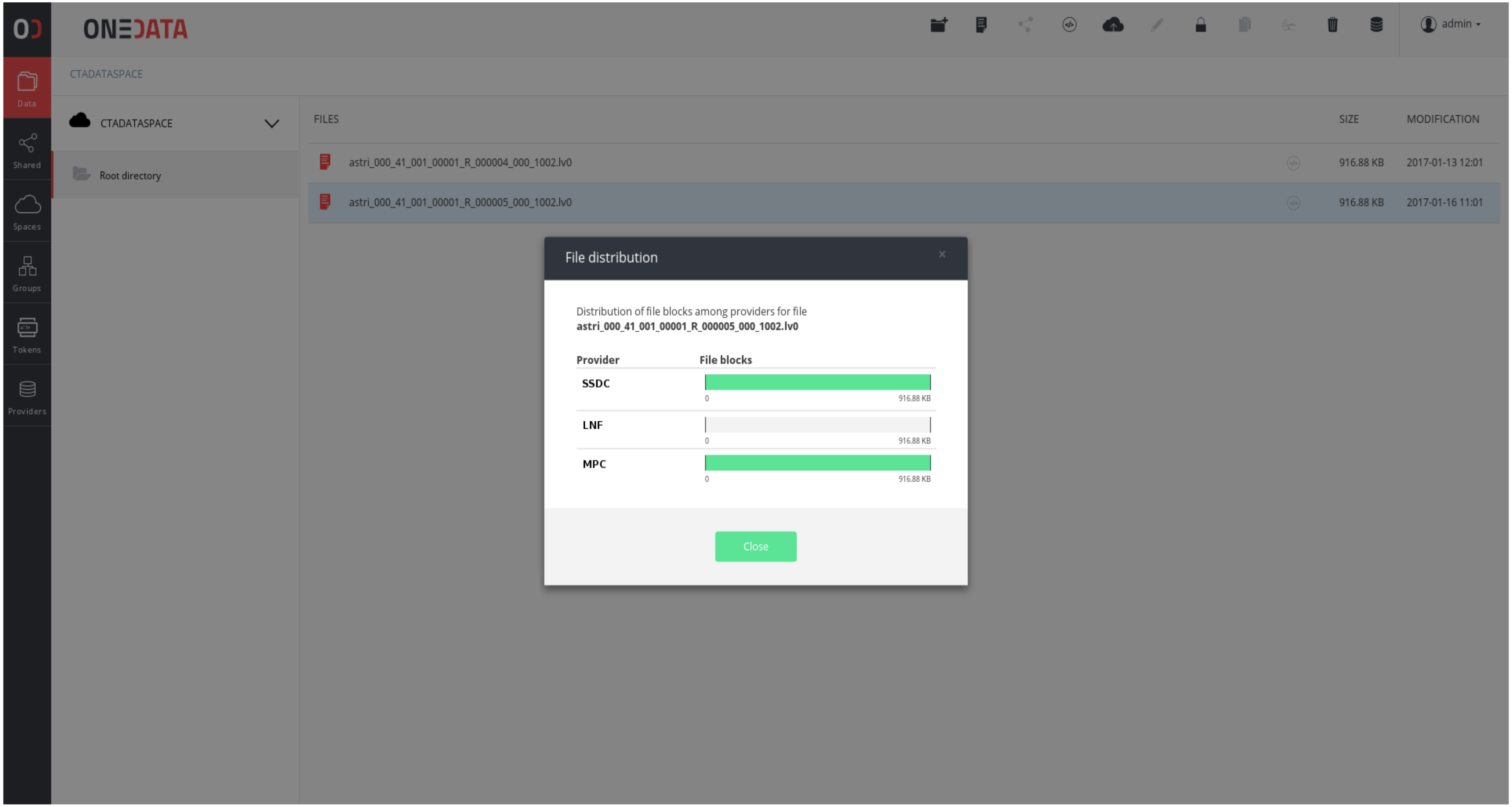
CTA OneProvider(s)

- **OneProvider exposes** storage resources. It is deployed in a data or computing center, on the nodes equipped with high speed **connections** to **storage resources**.





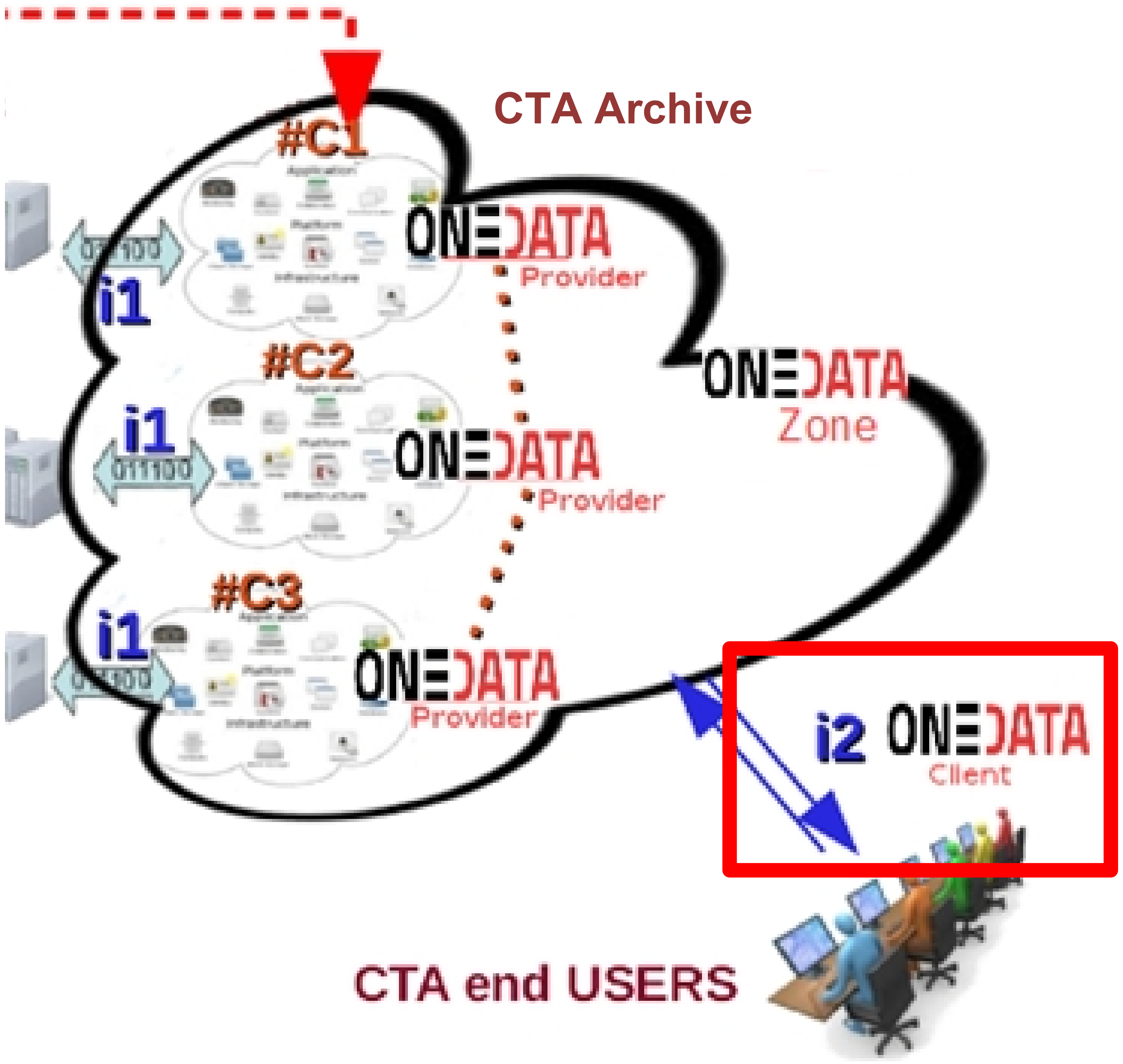
CTA OneProvider(s)



CTA OneClient

- **OneClient** is a command-line based application for **accessing** and **managing** user spaces via **virtual file system**.

- User spaces are **mounted** in the local file system tree (i.e. in a Grid Storage-Element FS as well).



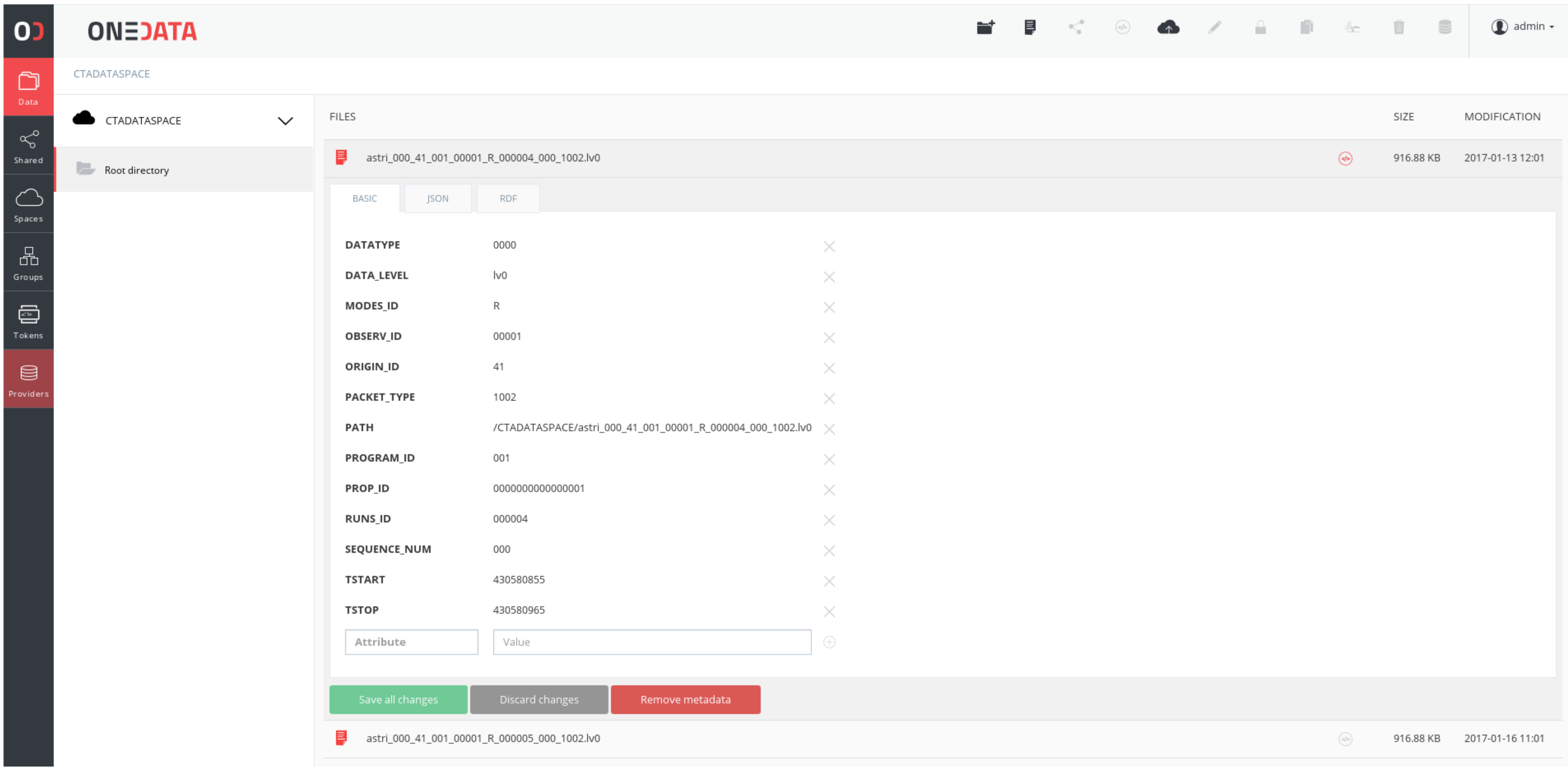
Metadata

Metadata in OneData are organized into 3 levels:

- ✓ **Filesystem attributes** - basic metadata related to file system operations such as file size, creation and modification timestamps, POSIX access rights, etc.,
- ✓ **Extended attributes** - these attributes enable assigning custom key-value pairs.
- ✓ **User metadata** - this level provides most flexibility and OneData itself does not assume any schema related with these metadata. For each resource, user can assign a separate document in one of supported metadata formats (currently JSON and RDF).

The filesystem and extended level attributes are accessible via **REST-API** and **CDMI** or directly through queries to the embedded database.

Metadata



ONEDATA
admin
Data
Shared
Spaces
Groups
Tokens
Providers
CTADATASPACE
CTADATASPACE
Root directory
FILES
SIZE
MODIFICATION
astri_000_41_001_00001_R_000004_000_1002.lv0
916.88 KB
2017-01-13 12:01
BASIC
JSON
RDF
DATATYPE 0000
DATA_LEVEL lv0
MODES_ID R
OBSERV_ID 00001
ORIGIN_ID 41
PACKET_TYPE 1002
PATH /CTADATASPACE/astri_000_41_001_00001_R_000004_000_1002.lv0
PROGRAM_ID 001
PROP_ID 0000000000000001
RUNS_ID 000004
SEQUENCE_NUM 000
TSTART 430580855
TSTOP 430580965
Attribute
Value
Save all changes
Discard changes
Remove metadata
astri_000_41_001_00001_R_000005_000_1002.lv0
916.88 KB
2017-01-16 11:01

Metadata

Sample Ingestion

```
curl -k -H $TOKEN_HEADER -H $CDMI_VSN_HEADER -H 'Content-Type:
application/cdmi-object'
-d '{"metadata" : {"PROGRAM_ID" : "001"}}' -X PUT "$ENDPOINTDATA"
```

Sample indexing function

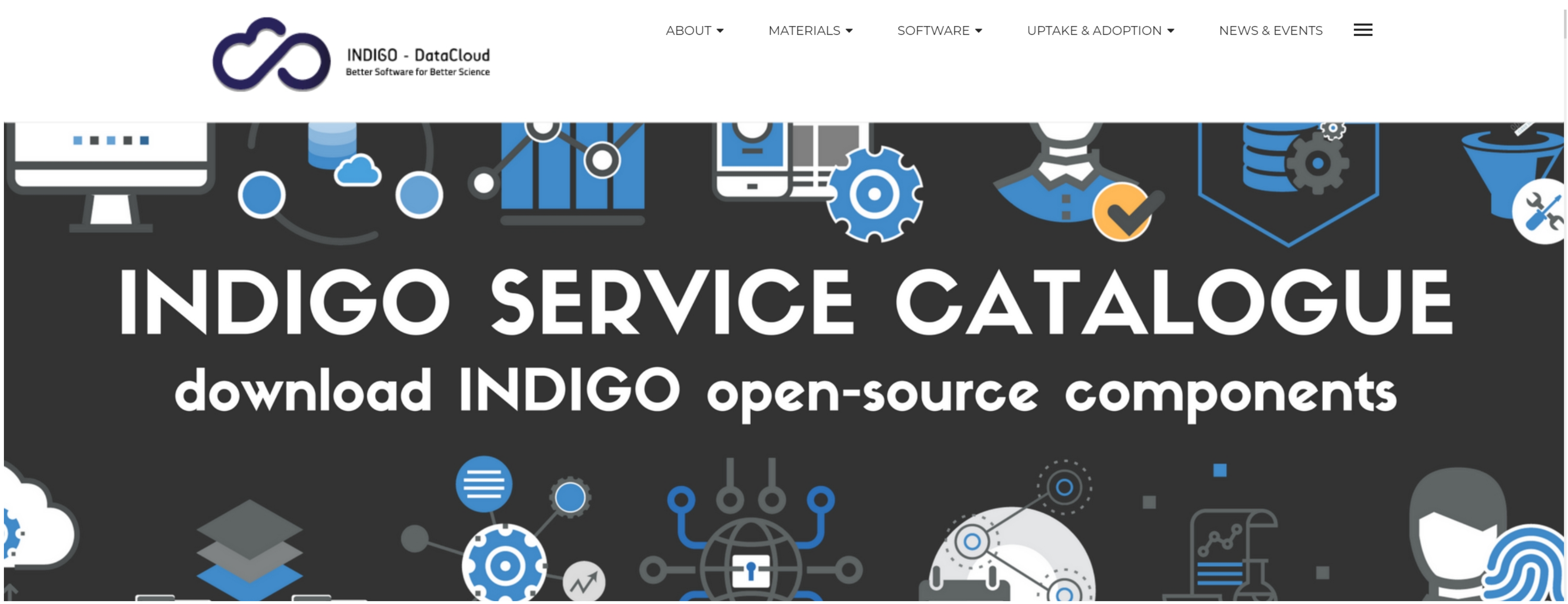
```
function(meta)
{
    if(meta['PROGRAM_ID'])
    {
     return meta['PROGRAM_ID'];
    }
    return null;
}
```

Query using a REST-API call

```
curl -v -k --tlsv1.2 -Ss -H "X-Auth-Token: $TOKEN" \
-X GET
"https://$HOST:8443/api/v3/oneprovider/query-index/$INDEX_ID?
key=\"0001\"&stale=false"
```

References

- CTA web page: http://www.cta-observatory.org/
- ASTRI web page: http://www.brera.inaf.it/astri/
- YouTube demo: https://youtu.be/UhOWnJluIgE
- INDIGO Data Cloud: https://www.indigo-datacloud.eu
- OneData documentation: https://onedata.org/docs/index.html
- OneData @ docker hub: https://hub.docker.com/u/onedata/



SSDC as server of CTA data products

- 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 betweeen resident and external catalogs.
  - Powerful tools to extract SED of sources and modelization.
  - VHE catalog products from literature already integrated in the TeGeV Catalogue.