Distributed Archive for the Cherenkov Telescope Array

Eva Sciacca[1], S. Gallozzi[2], A. Antonelli[2,3], A. Costa[1]

1 INAF, Astrophysical Observatory of Catania

2 INAF, Astronomical Observatory of Rome

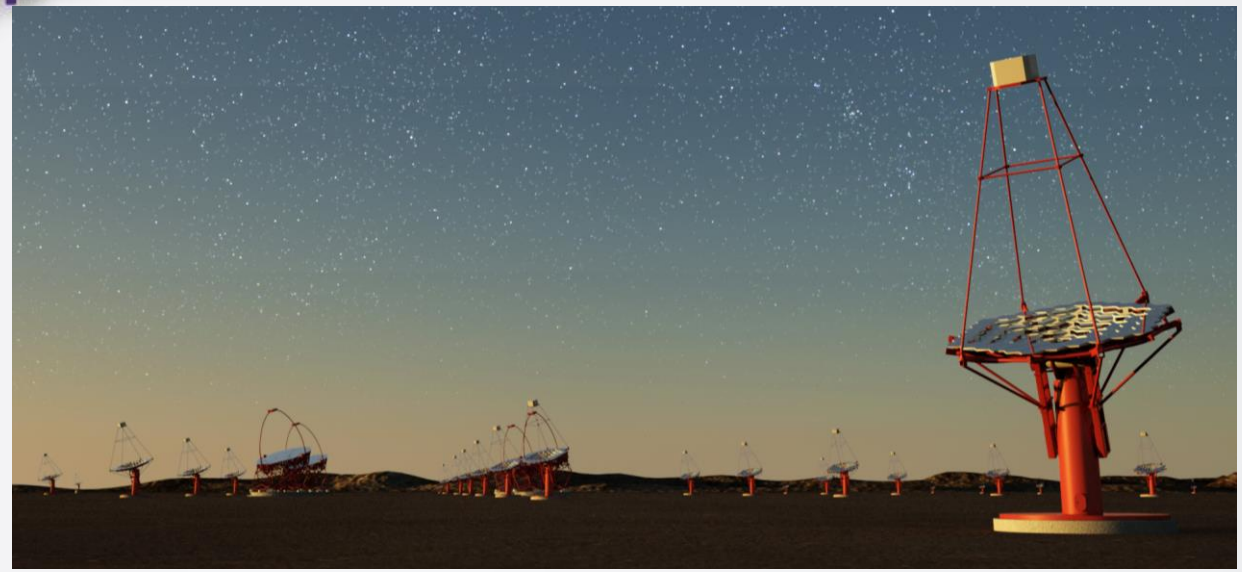
3 ASDC, ASI-Science Data Center

eva.sciacca@oact.inaf.it

RIA-653549

INDIGO-DataCloud is co-founded by the Horizon 2020Framework Programme

Cherenkov Telescope Array
https://cta-observatory.org



WHAT: CTA is the worldwide project for the future of Very High Energy gamma-ray astronomy.
~20 telescopes for the North-site (Canarie)
~100 telescopes for the South-site (Chile)

WHO: the CTA Consortium consists of scientists and engineers from 32 countries from 5 continents and has become a truly global (ESFRI) project.

OUR AIM: One of the major technological challenge is related to the data-handling and archiving of the huge amount of data (from 20 to 100 PB/year) coming from the observatory facilities.

CTA Data Model

Data Level	Short Name	Description
DL0	DAQ-RAW	Acquired raw data.
DL1	CALIBRATED	Calibrated camera data.
DL2	RECONSTRUCTED	Reconstructed shower parameters (such as energy, direction, particle ID).
DL3	REDUCED	Sets of selected events with associated instrumental response characterizations needed for science analysis.
DL4	SCIENCE	High Level binned data products (such as spectra, sky maps, or light curves).
DL5	OBSERVATORY	Legacy observatory data (such as survey sky maps or source catalog).

Data Requirements

Without data compression and assuming 165 operational nights/yr:

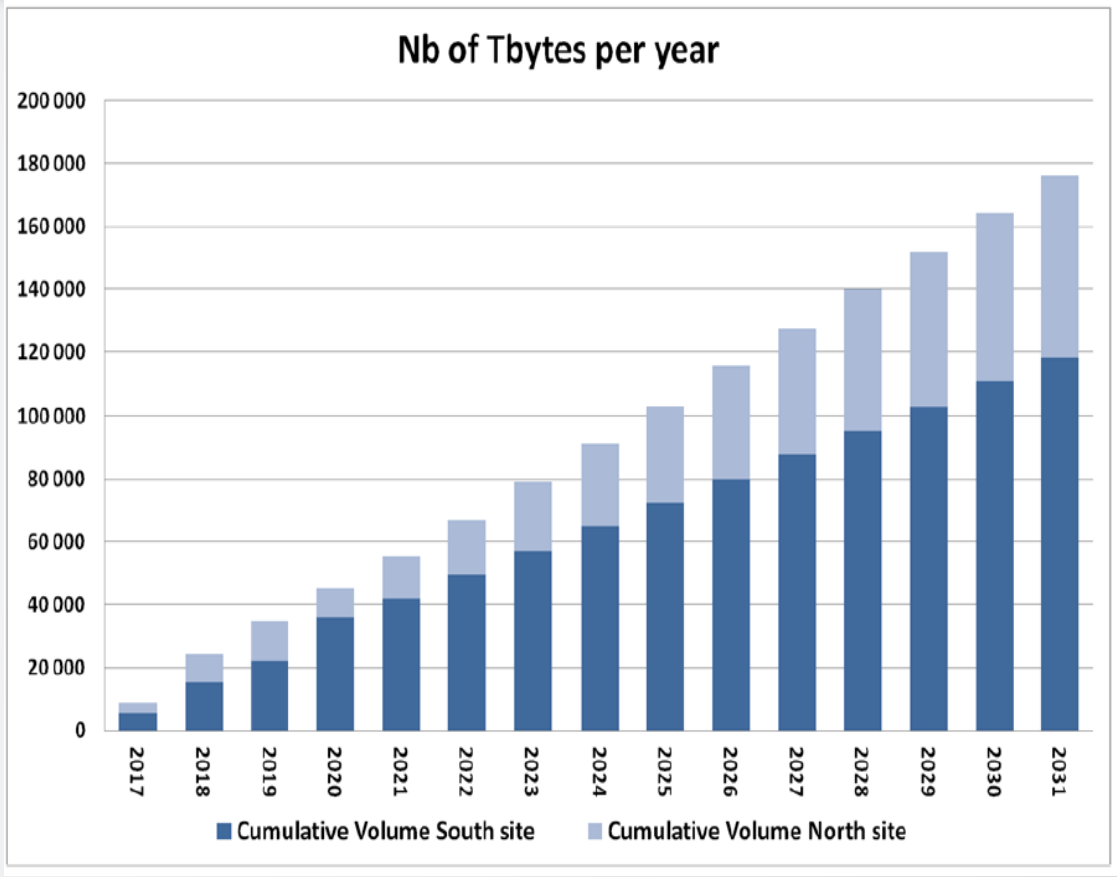
ASTRI/Prot. → ~0.8 TB/night
→ **~0.3 PB/year**

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

CTA → ~8.5 GB/s
→ ~40 TB/night
→ **~4 PB/year**
→ **~20 PB/year**

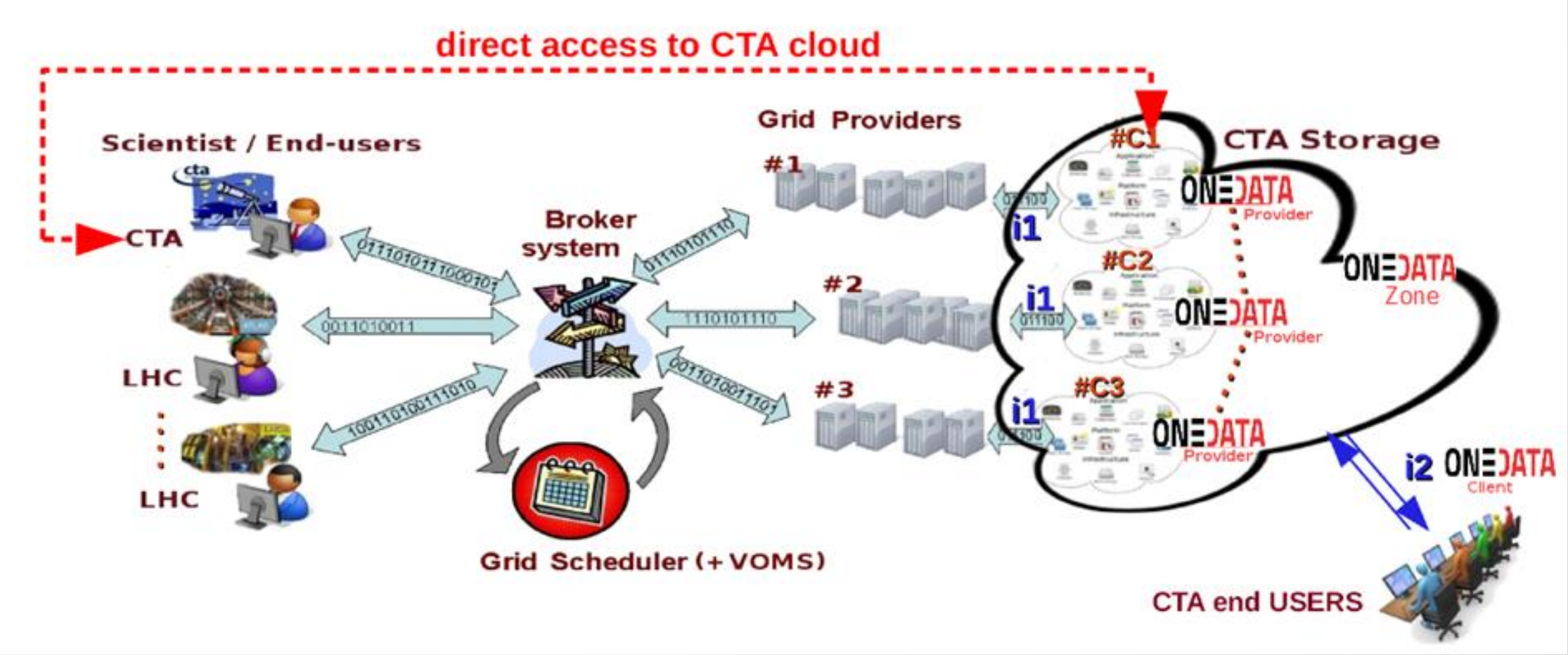
A.R.
(A.R. = After Reduction → input+processed data including calibs, intermediate reduction and MC simulation data)

this is the **OPTIMISTIC SCENARIO**
The pessimistic one can take ~>100PB/year !

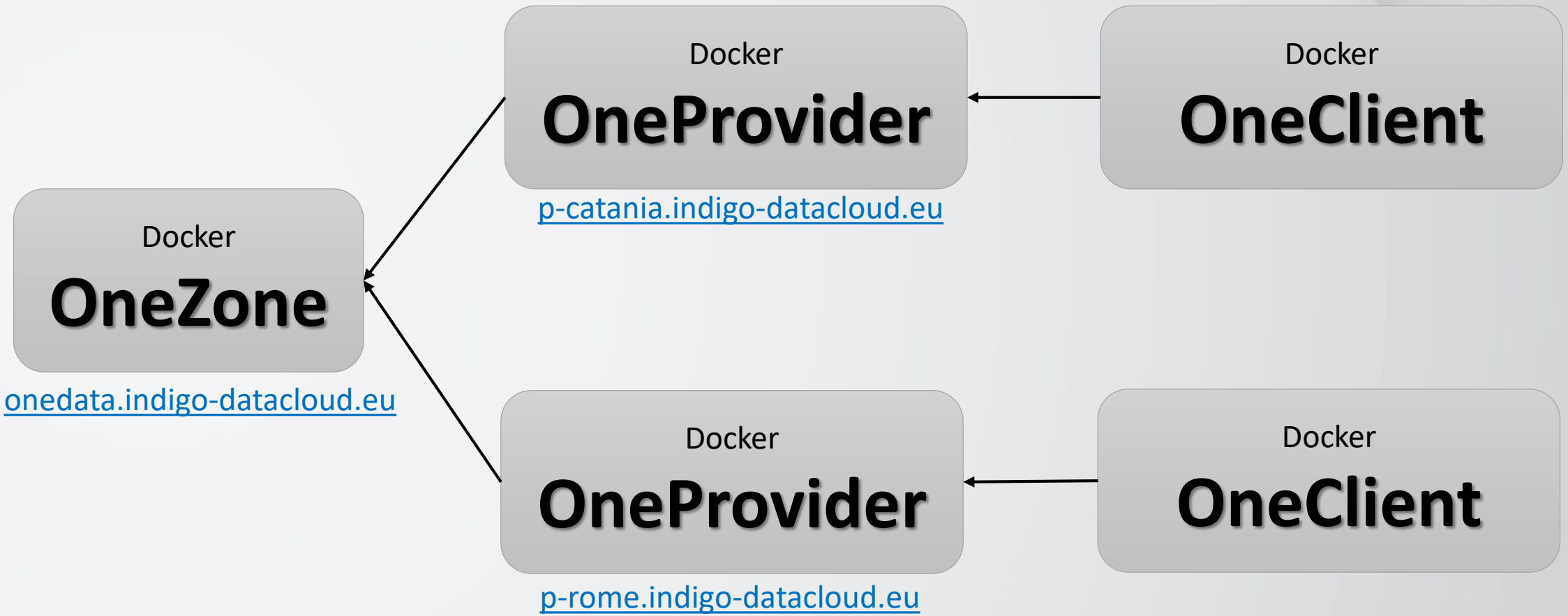


The CTA Archive system must store, manage, preserve and provide easy access to such huge amount of data for a long time.

Architecture



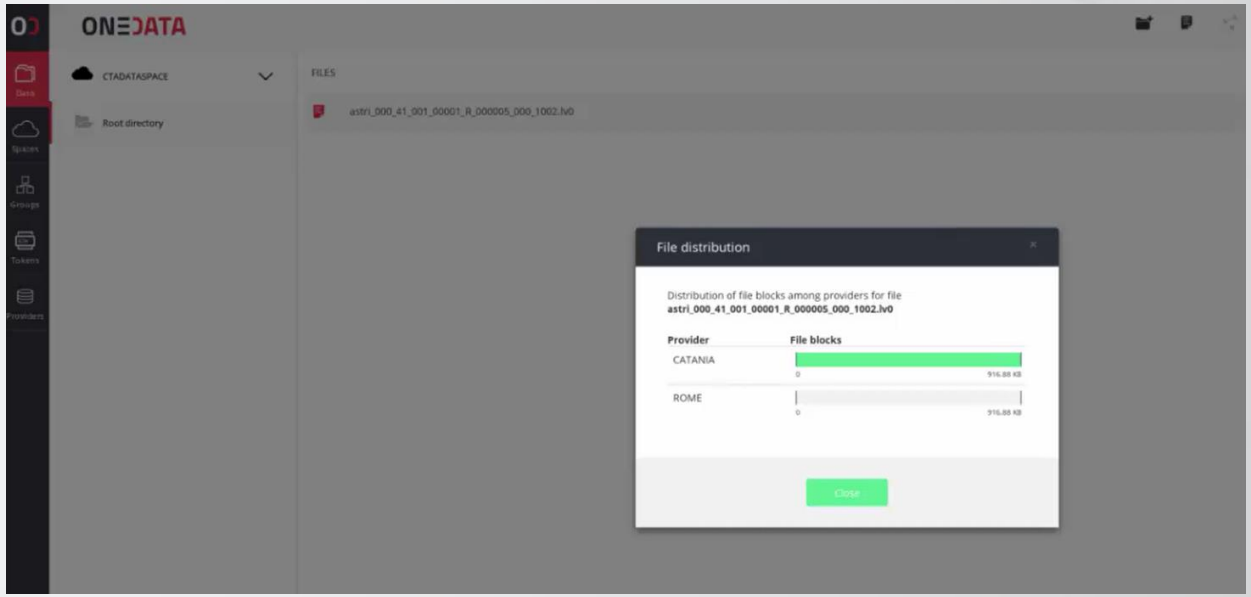
Demo Infrastructure



Thanks to the OneData team for the technical support (especially to Krzysztof Trzepla and Konrad Zemek)

Running the Demo

- Demo **datasets** coming from the **ASTRI project** (prototype of small size Cherenkov telescope) have been uploaded to the **CTA OneZone** within a space supported by the two providers.
- The ingested data are enriched with metadata thanks to the Cloud Data Management Interface (CDMI).



- The **CouchBase** database (embedded in OneData) is used to query and retrieve the metadata using **Elastic Search engines** (e.g. N1QL) or common **MapReduce functions** using the standard CouchBase console and the SDK from the client side. This will enable versatile access to the whole CTA dataset to higher level application frameworks and end-users analysis tools.

Distributed Archive Advantages

- **lower costs** respect to a single huge data center,
- **easy manageability & maintenance** by single-site human resources
- **distributed database** of meta-data within the architecture
- **easily scalability** by adding new nodes on the system
- **disaster recovery free** by different sites redundancy

INAF CTA @ INDIGO

eva.sciacca@oact.inaf.it

stefano.gallozzi@oa-roma.inaf.it

alessandro.costa@oact.inaf.it

angelo.antonelli@oa-roma.inaf.it