

## Distributed Archive Prototype for the Cherenkov Telescope Array

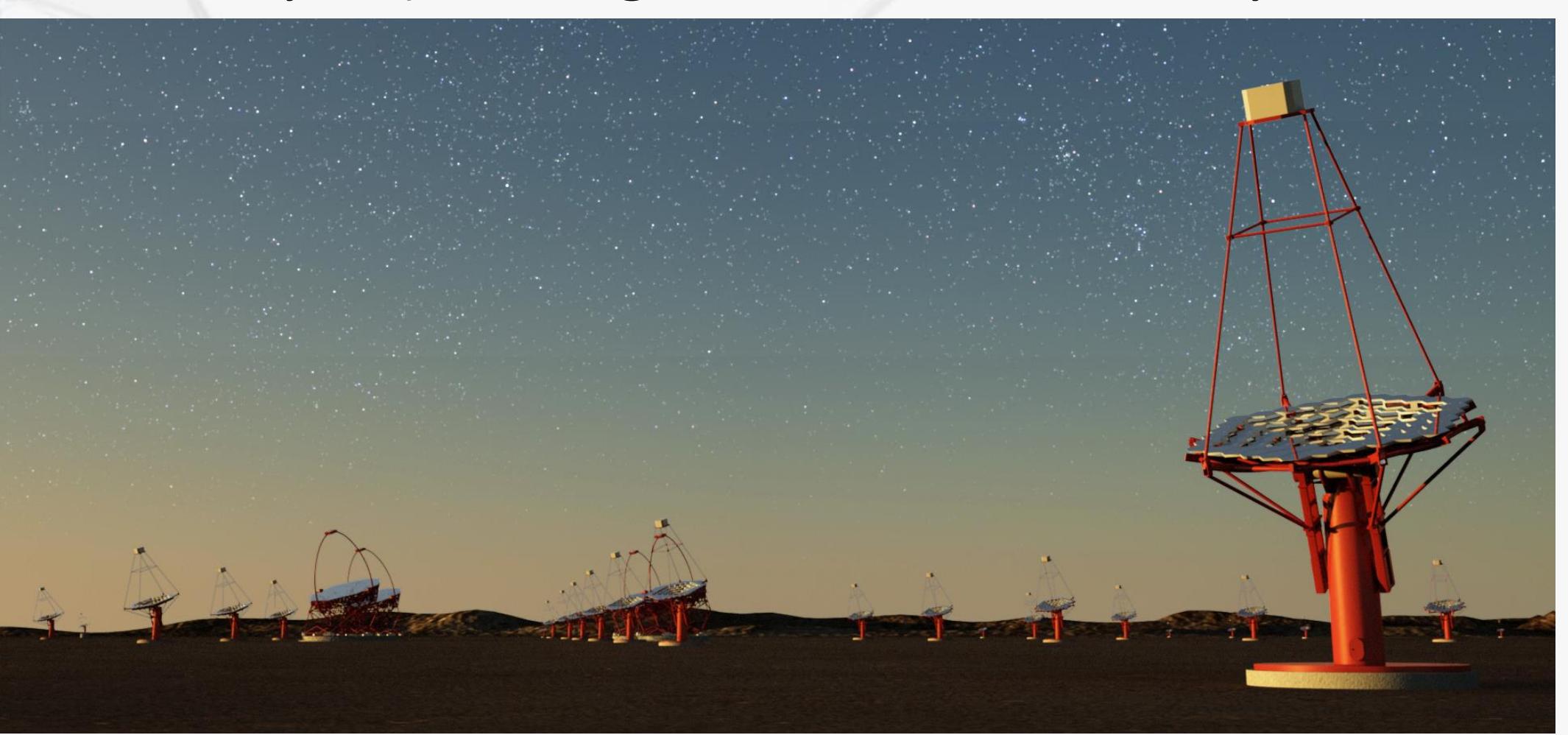
E. Sciacca<sup>1</sup>, S. Gallozzi<sup>2</sup>, A. Antonelli<sup>2,3</sup>, A. Costa<sup>1</sup>

### Cherenkov Telescope Array (CTA)

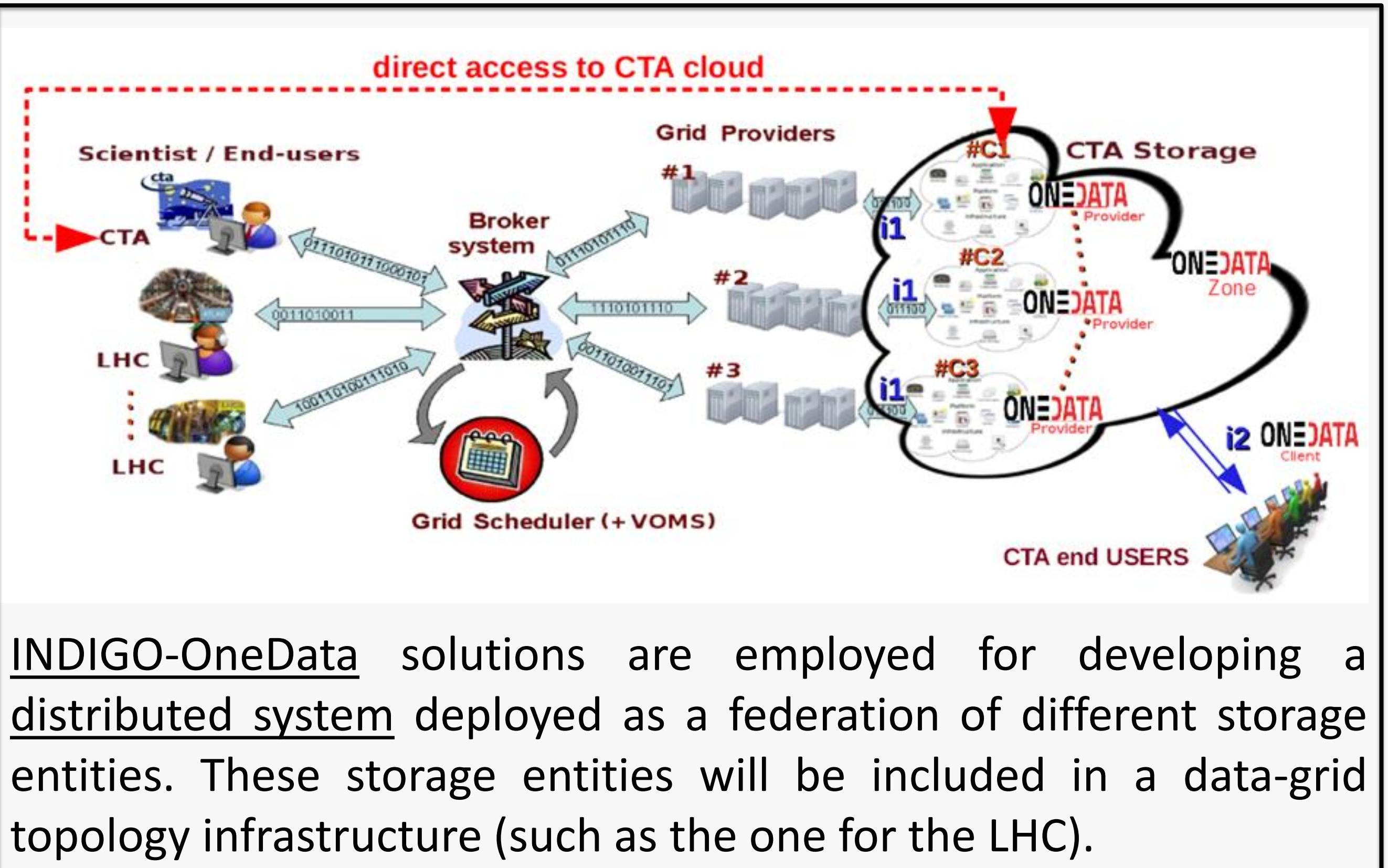
**WHAT:** CTA is the worldwide project for the future of Very High Energy gamma-ray astronomy.

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

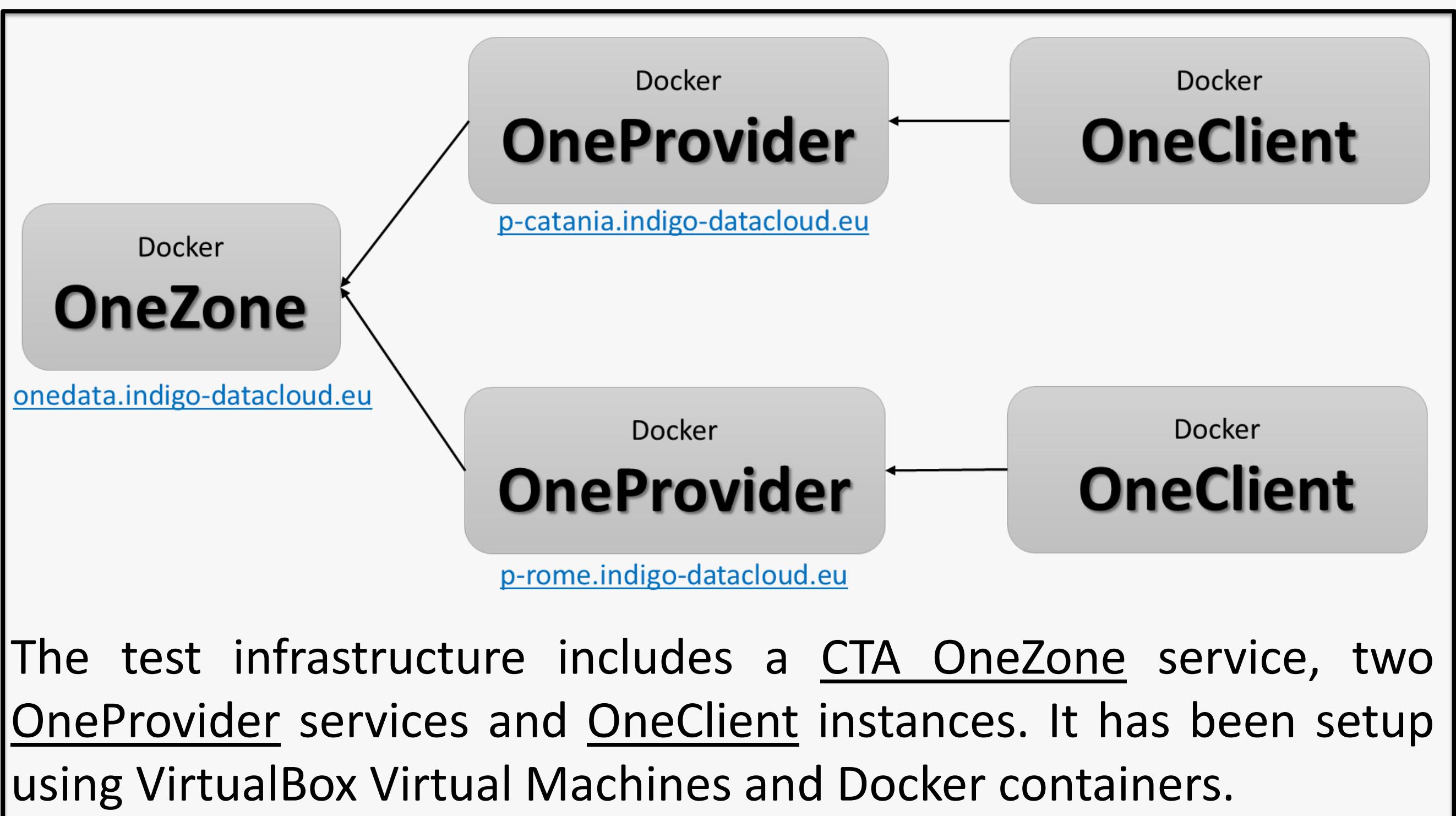
**OUR AIM:** One of the major technological challenge is related to the data-handling and archiving of the huge amount of data (from 20 to 100 PB/year) coming from the observatory facilities.



### CTA Archive & INDIGO-DataCloud



### DEMO Architecture



<sup>1</sup> INAF, Astrophysical Observatory of Catania

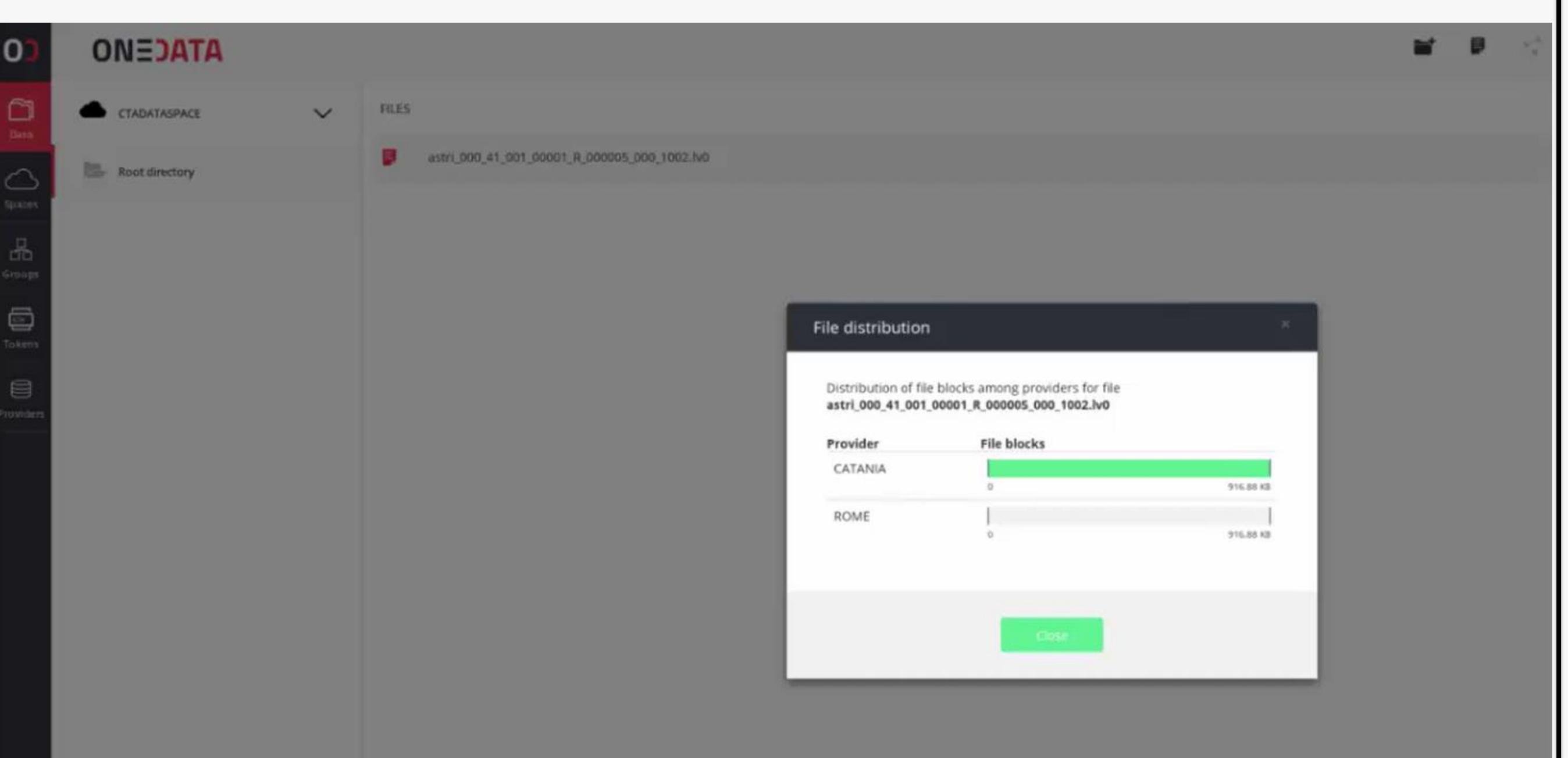
<sup>2</sup> INAF, Astronomical Observatory of Rome

<sup>3</sup> ASDC, ASI-Science Data Center



### Running the DEMO

The CTA OneZone allows users to create Spaces supported by one or more providers. In this demo we upload datasets coming from the ASTRI project. The data are distributed in the providers.



The ingested data are enriched with metadata (such as the PROGRAM\_ID or the OBSERVATION\_ID) thanks to the Cloud Data Management Interface (CDMI) or, alternatively, the REST API can be used.

Metadata queries are performed using REST-API and indexing functions (associated to the Space) on pre-defined extended attributes.

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

*Sample indexing function*

```
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"
```

*Sample REST-API call*

### Expected Results

- ✓ The distributed architecture allows to lower costs with respect to a single huge data center including easy manageability and maintenance.
- ✓ The solution takes care of redundancy policy: involved databases of metadata are distributed together with the storage sites allowing a very high throughput and availability of inter-communications with the best data-model scheduling organization.
- ✓ It is fault tolerant and risk-management free: it has no single point of failure and can easily solve any disaster recovery event thanks to the redundancy of the distributed approach with a robust database management system.

### Useful Links

CTA web page: <http://www.cta-observatory.org/>

ASTRI web page: <http://www.brera.inaf.it/astri/>

YouTube demo: <https://youtu.be/UhOWnJlulgE>

OneData documentation: <https://onedata.org/docs/index.html>

OneData @ docker hub: <https://hub.docker.com/u/onedata/>