Services and a distributed infrastructure pilot for the Cherenkov Telescope Array

Alessandro Costa, Eva Sciacca, Fabio Vitello
Ugo Becciani and Piero Massimino

INAF, Astrophysical Observatory of Catania, Italy

Overview

- The CTA Project
- Requirements
- Overall Architecture
  - The CTA Science Gateway
  - The Interactive Virtual Desktop
  - The Authentication and Authorization Infrastructure
- Conclusions

The CTA Project

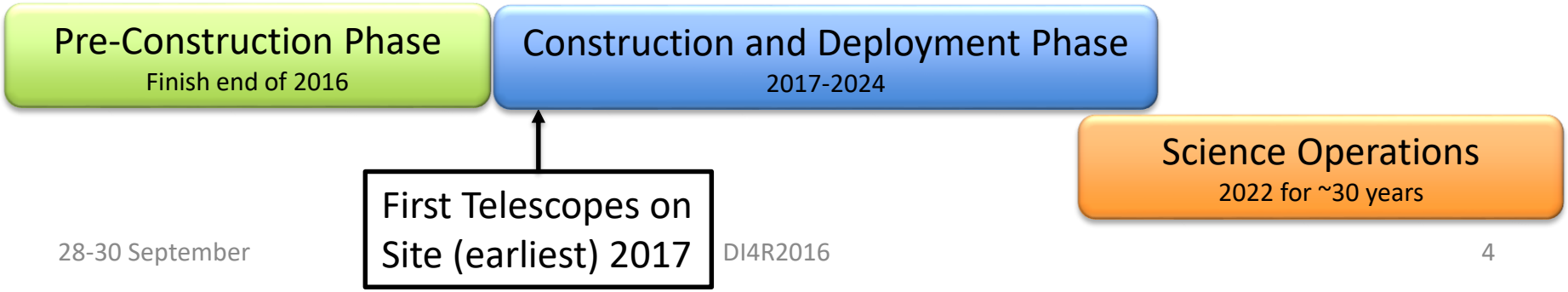
https://www.cta-observatory.org/

The CTA project is an initiative to build the world's largest and most sensitive ground-based high-energy gamma-ray observatory.

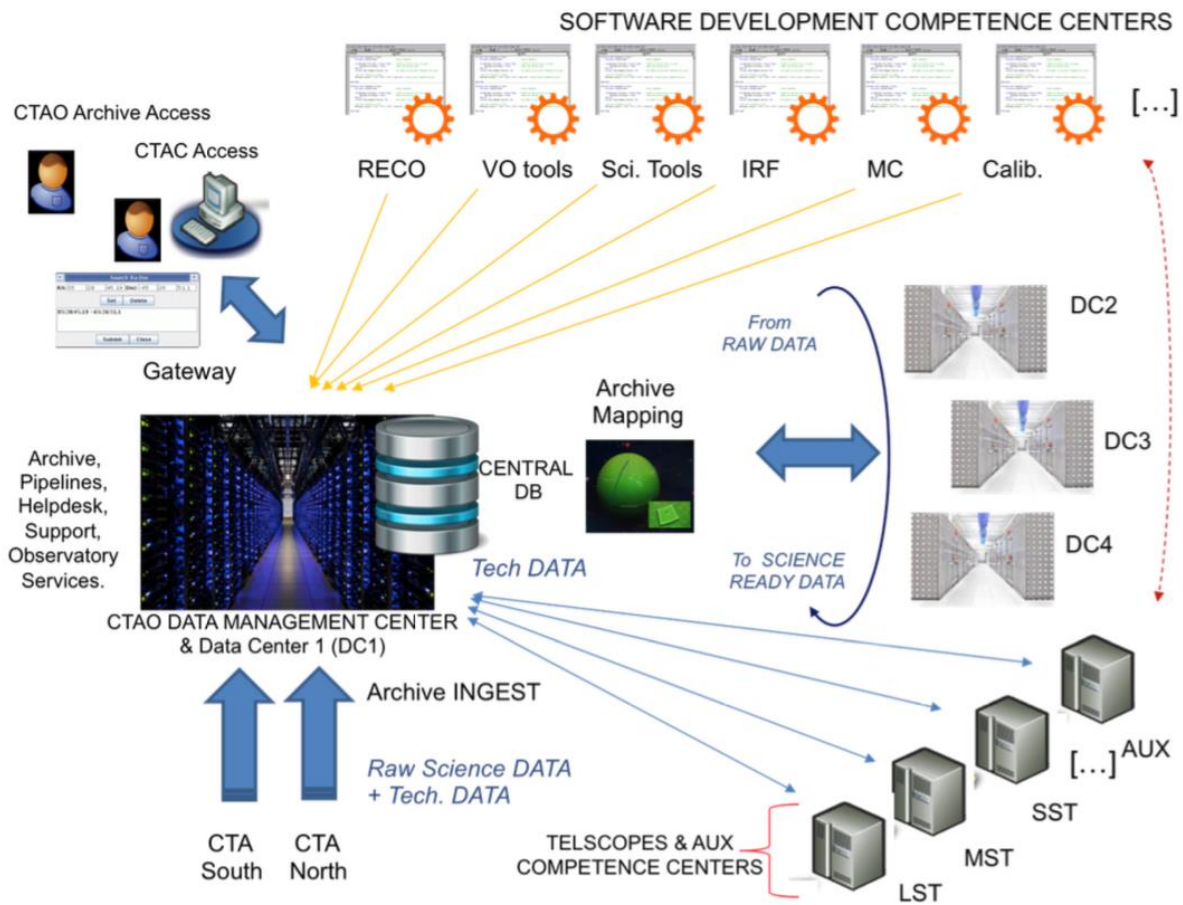
- Aims:
  - ✓ Understanding the **origin** of **cosmic rays** and their role in the Universe
  - ✓ Understanding the nature and variety of **high energy phenomena** such as particle acceleration around black holes
  - ✓ Searching for the ultimate **nature of matter and physics** beyond the Standard Model

The CTA Project

- Since 2008, CTA has been included in the roadmap of the European Strategy Forum on Research Infrastructures (ESFRI).
- The CTA Consortium started in 2007 to design the installation and to work towards its implementation. A Design Study phase has been ended in 2010.
- CTA is now in the Pre-Construction Phase, aimed to deliver a Technical Design Report and being ‘‘construction ready’’ towards the 2017.
- In order to optimize the coverage of the night sky, the CTA Observatory will consist of about 100 telescopes on the southern site (Chile) and about 20 telescopes on the northern site (Canarias).
- Project schedule:



The CTA Project

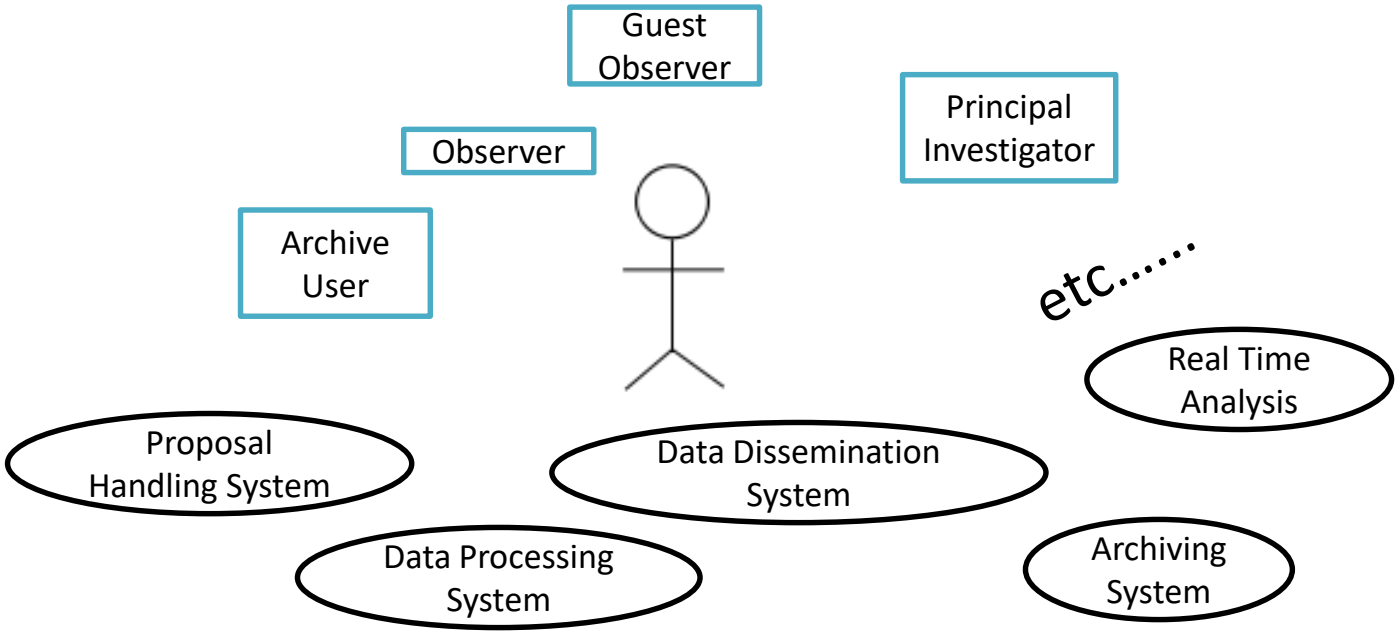


- 4 PB/yr RAW DATA => 25 PB/yr in total
- Transferred and/or generated by four off-site Data Centers
- Process data where they are stored
- Databases:
  - Proposal handling
  - Archives Management
  - Technical, engineering and monitoring

Use Case Actors

ACTOR Definition

- Human
  - NAME, ROLE, DESCRIPTION
- System
  - NAME, DESCRIPTION, Sub-System



Science Gateway Requirements

- The total data volume to be managed by the CTA Science Data Centre is of the order of 27 PB/year.
- The "CTA science gateway" will provide:
  - ✓ access to data and data center infrastructures across organizational and national boundaries,
  - ✓ support services,
  - ✓ analysis software,

  And will be:
  - ✓ developed in WEB2.0,
  - ✓ using open standards and sustainable licenses.
- Aims:
  - ✓ support workflow handling,
  - ✓ virtualization of hardware,
  - ✓ visualization as well as resource discovery,
  - ✓ job execution,
  - ✓ access to data collections, and applications and tools for high-level data analysis.

A&A Use Cases

Some of the defined Use Cases are:

- Authenticate an already registered CTA Consortium user
- Authenticate a user based on his institute/laboratory account
- Access to the developed services within the “CTA science gateway” and other CTA web resources will be based on each **user profile** and **category** (e.g. unsigned user, guest observer, advanced user, principal investigator, archive user, pipeline user, etc).
- Lost password management
  - User with a local A&A login/password must be able to ask for a new password in case of lost password
- Group creation
  - A&A Administrator or DATA Applications must be able to create a group, associate roles and define group owner(s) (e.g.: **ProposalId** group -> data access right)
- Group management
  - A group owner is able to manage his group from a central A&A management system or from specific DATA applications integrated or not in the Gateway: invite users, remove users, account expiration dates,...

A&A User Requirements

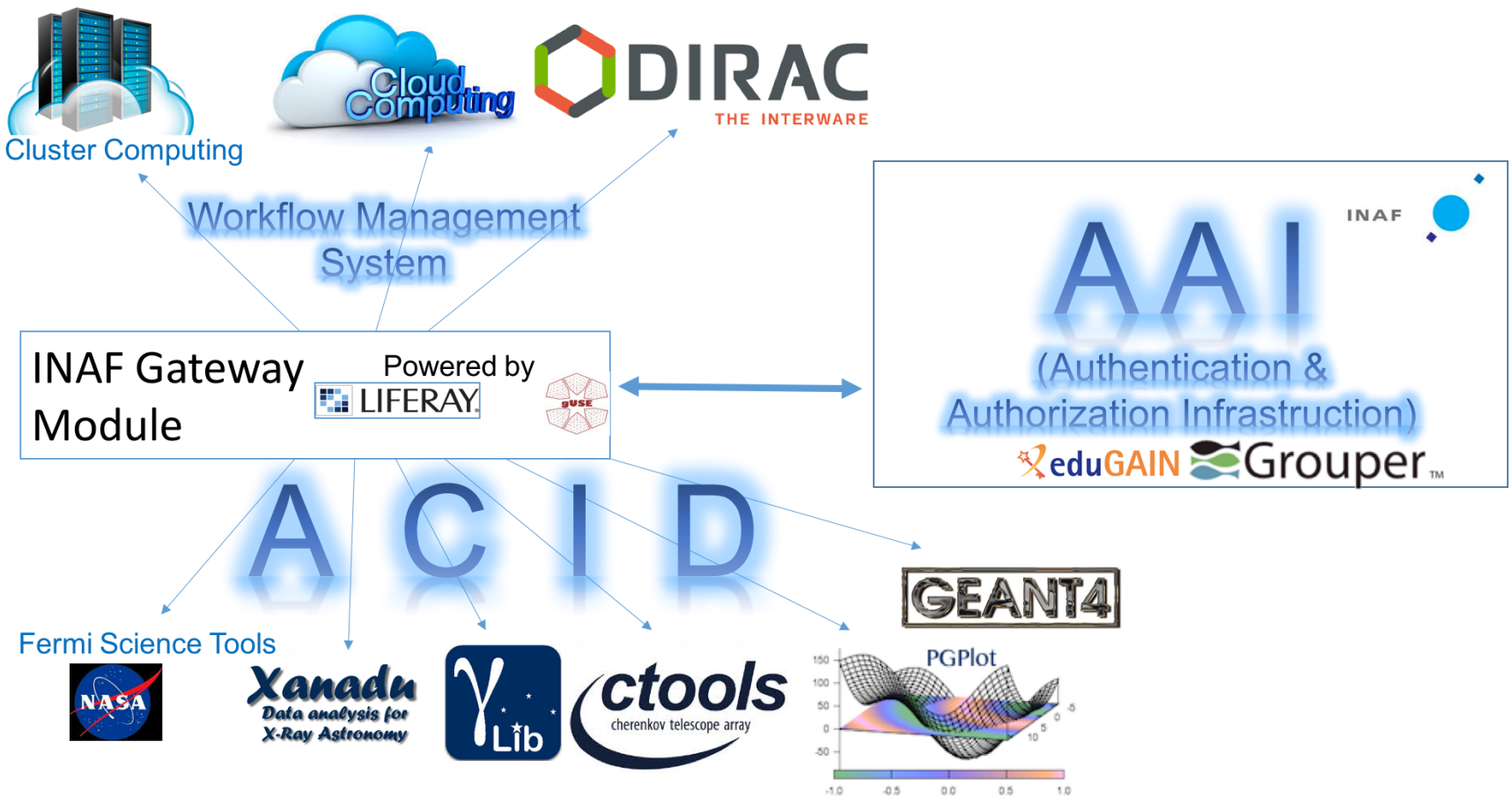
User Requirements are grouped in the following Classes:

- Authentication Capabilities
  - E.g. A Guest Observer that cannot be identified by a scientific community must be able to be identified by a local account protected by login/password.
- Authorization Capabilities
  - E.g. Authorization should be granted to users, groups of users, data and applications.
- Management Capabilities
  - E.g. A group is created per Proposal Id
- Availability and performances
  - E.g. The A&A system must be able to support 500 authentication/authorization requests per minutes
- Security
  - E.g. The user must be sure that the A&A system will garantee privacy of users information.
- Portability
  - E.g. The A&A system must be portable enough to be maintained over the period of operations and 10 years after CTA decommissioning.

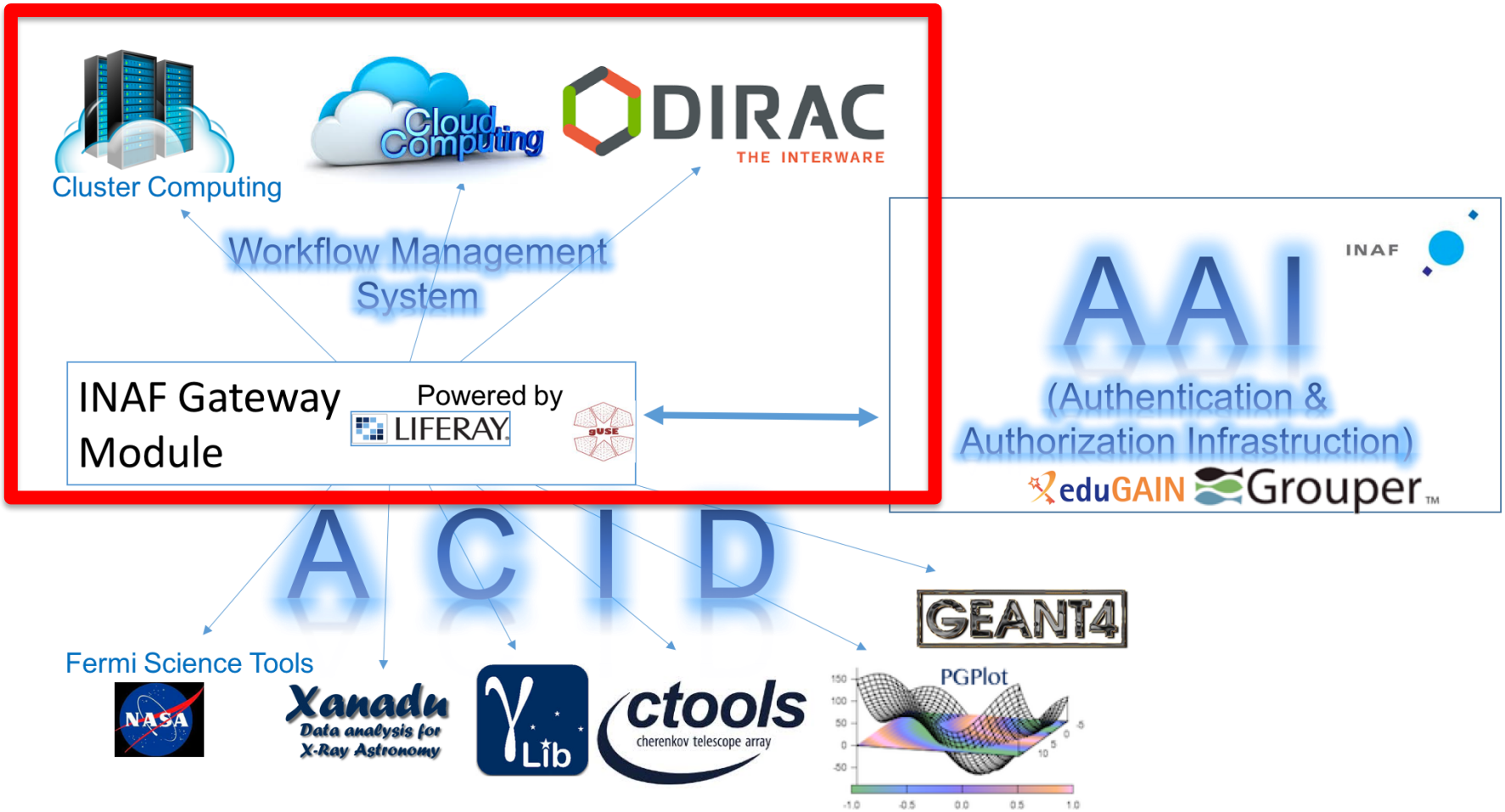
Prototyping Activities

- The “CTA science gateway” is implemented as a set of complementary modules.
- Three of them are being developed with different aims:
  1. Provides a workflow management system, it is powered by **WS-PGRADE/gUSE**, based on **Liferay** platform and an added value of this module is a **web desktop environment** (with a VNC-based User Interface) developed by INAF;
  2. Integrates existing CTA applications in a specific **InSilicoLab** platform developed by ACC Cyfronet AGH;
  3. Compliant with the Virtual Observatory and it is based on the **Django** platform, developed by the Observatoire de Paris

Overall Architecture

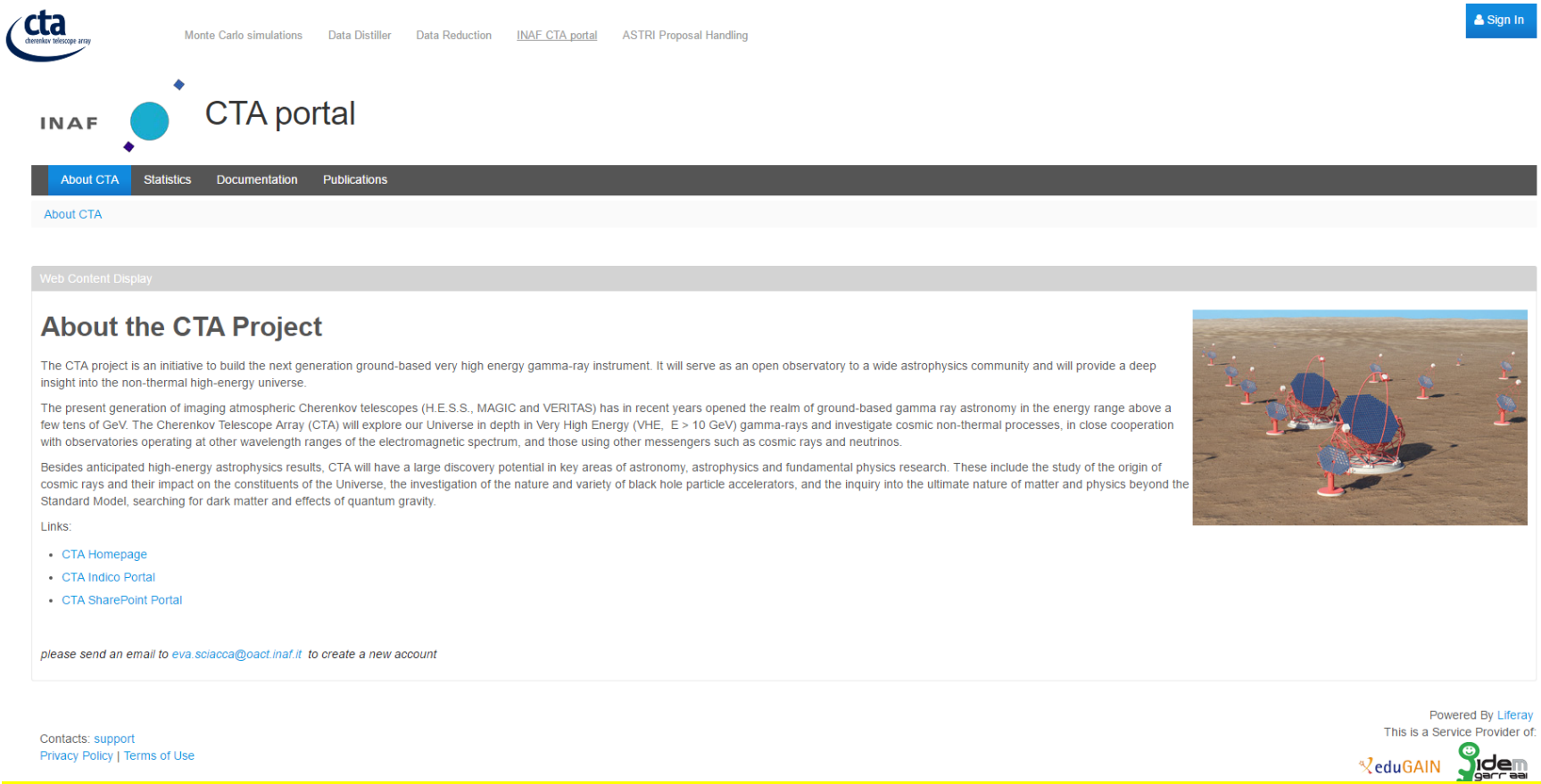


The Science Gateway



The Science Gateway

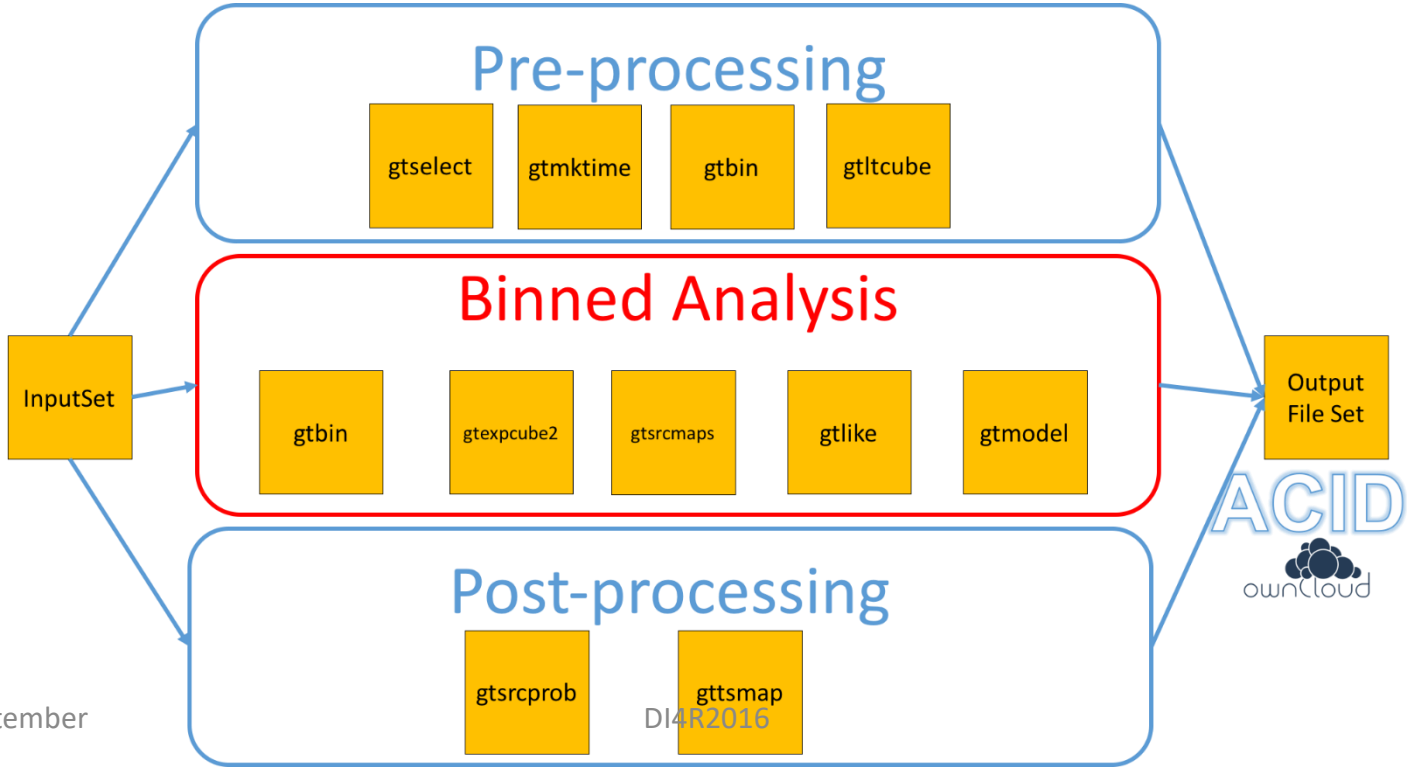
https://cta-sg.oact.inaf.it/



- ✓ Based on **Liferay**
- ✓ Integrates a workflow system (**gUSE/WS-PGRADE**) executing jobs on the major DCIs
- ✓ Connected to the AAI
- ✓ Integrated with the other “CTA science gateway” modules

The Science Gateway Fermi Workflow

A demonstrator has been implemented following a typical CTA analysis performed with the **Fermi Science Tools**.



The Science Gateway Fermi Workflow

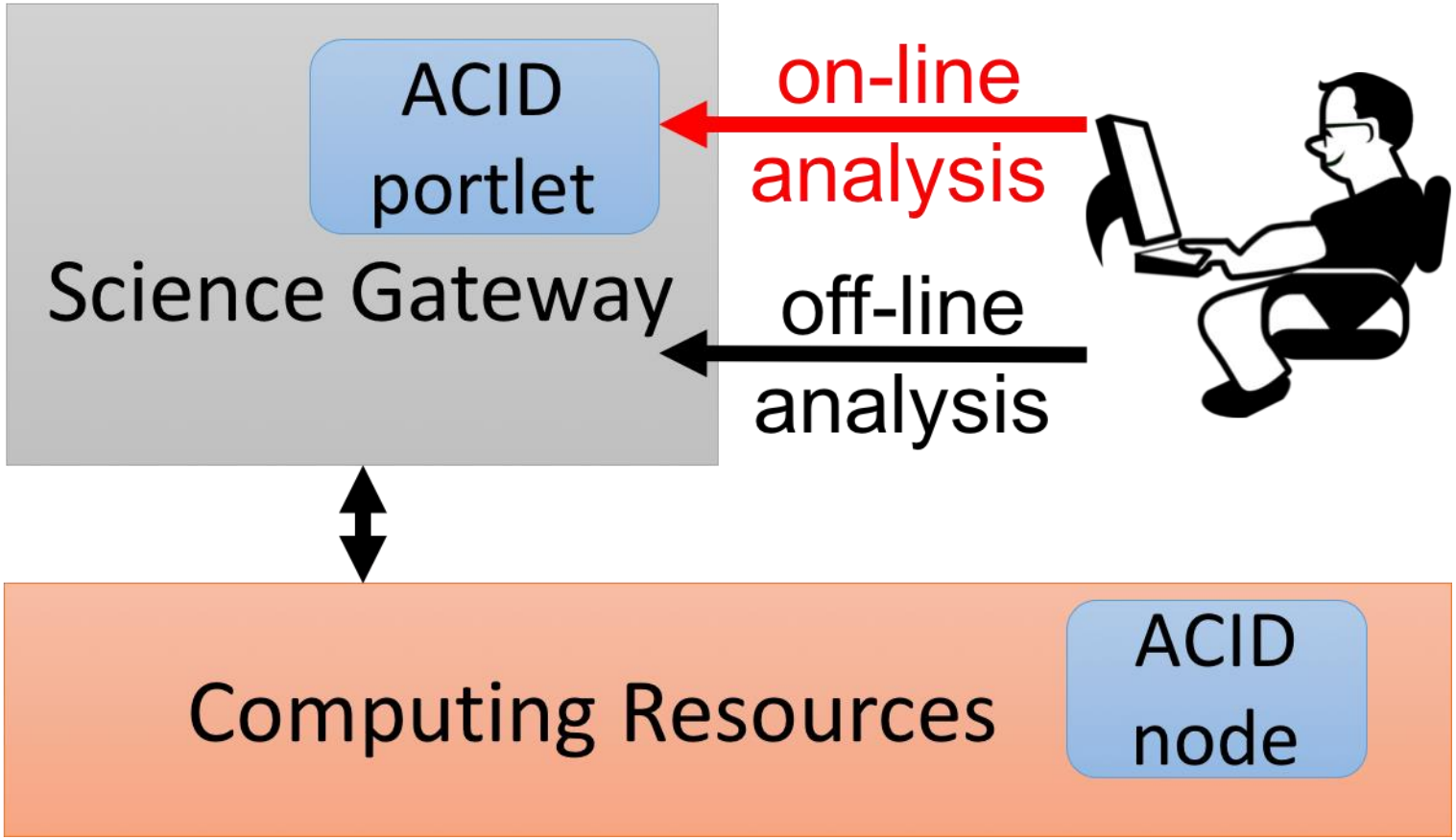
- On the science gateway two workflows have been implemented: BINNED and UNBINNED running on to the INAF Astrophysical Observatory of Catania.
- The workflows have been designed to set the input datasets and the parameters to run the process into the **InputSet** job so that only the entry job is configured and then the parameters are passed to the other jobs automatically.
- The separation of the Fermi processing into different jobs within the workflow allowed us to exploit the **full parallelization** of the computations within the configured DCIs.
- Finally the **OutputFileSet** job collects all the jobs output files and send them also to the ownCloud server hosted into the ACID environment.

The Interactive Virtual Desktop:

Astronomical & physics **C**loud **I**nteractive **D**esktop

The Interactive Virtual Desktop:

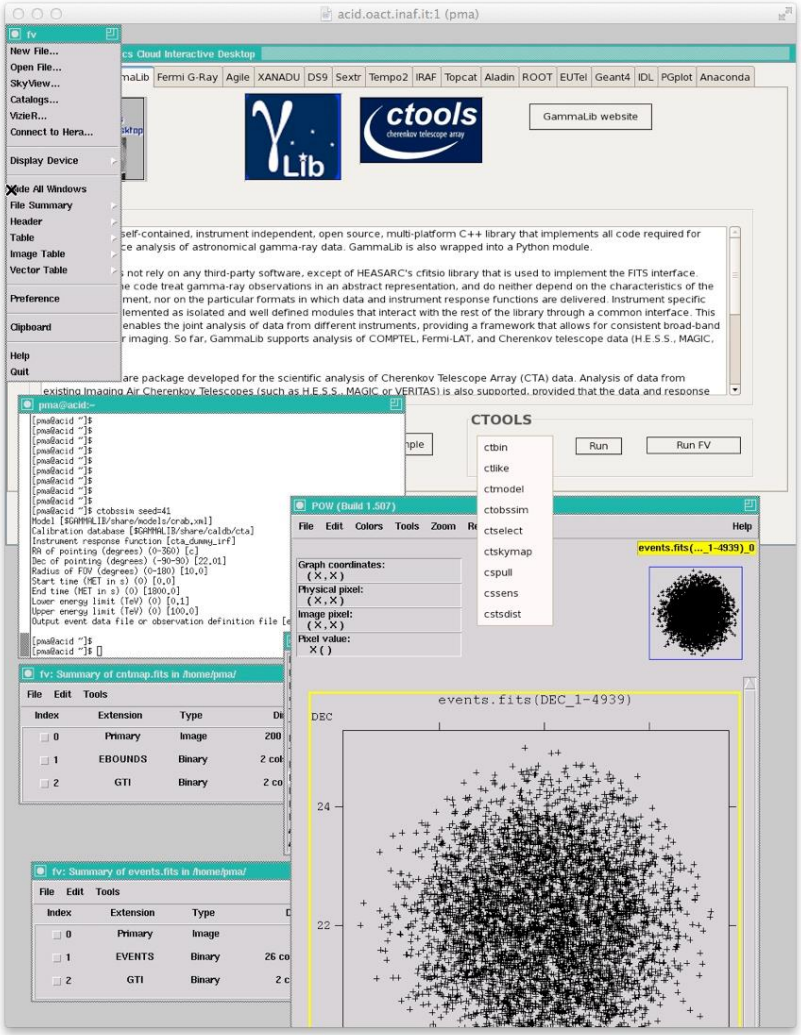
Astronomical & physics **C**loud **I**nteractive **D**esktop



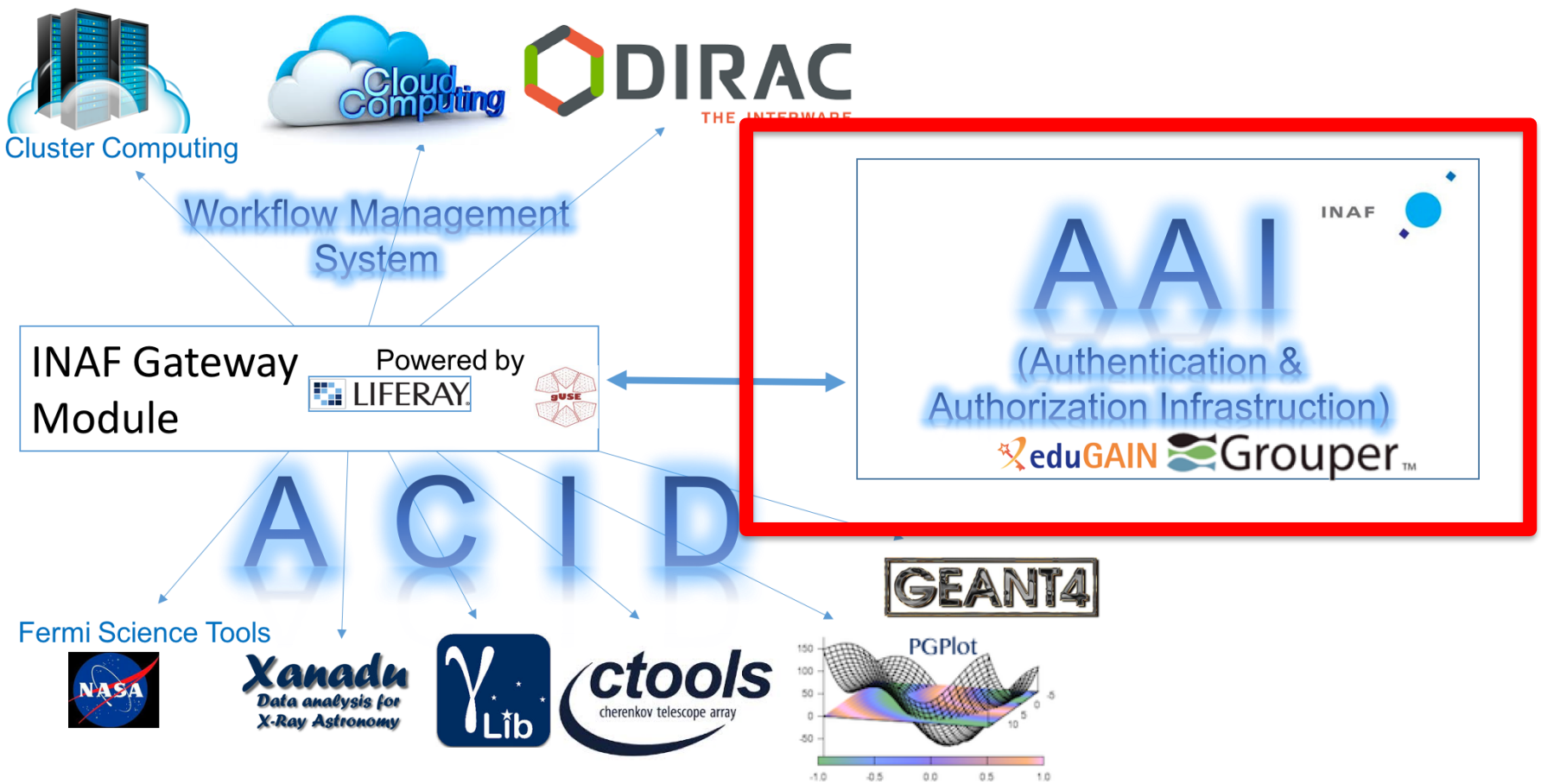
The Interactive Virtual Desktop:

Astronomical & physics **C**loud **I**nteractive **D**esktop

- The VNC-based User Interface runs on the ACID servers and allows the usage of any native GUI of the software launched remotely.
- ACID is independent of the O.S. and does not require any VNC client installation on the user device, except for iOS and Android devices that require the installation of a VNC client (such as the VNC Viewer application).
- ACID uses *ownCloud* to easily share data between the user device and the ACID server.



Authentication & Authorization Infrastructure

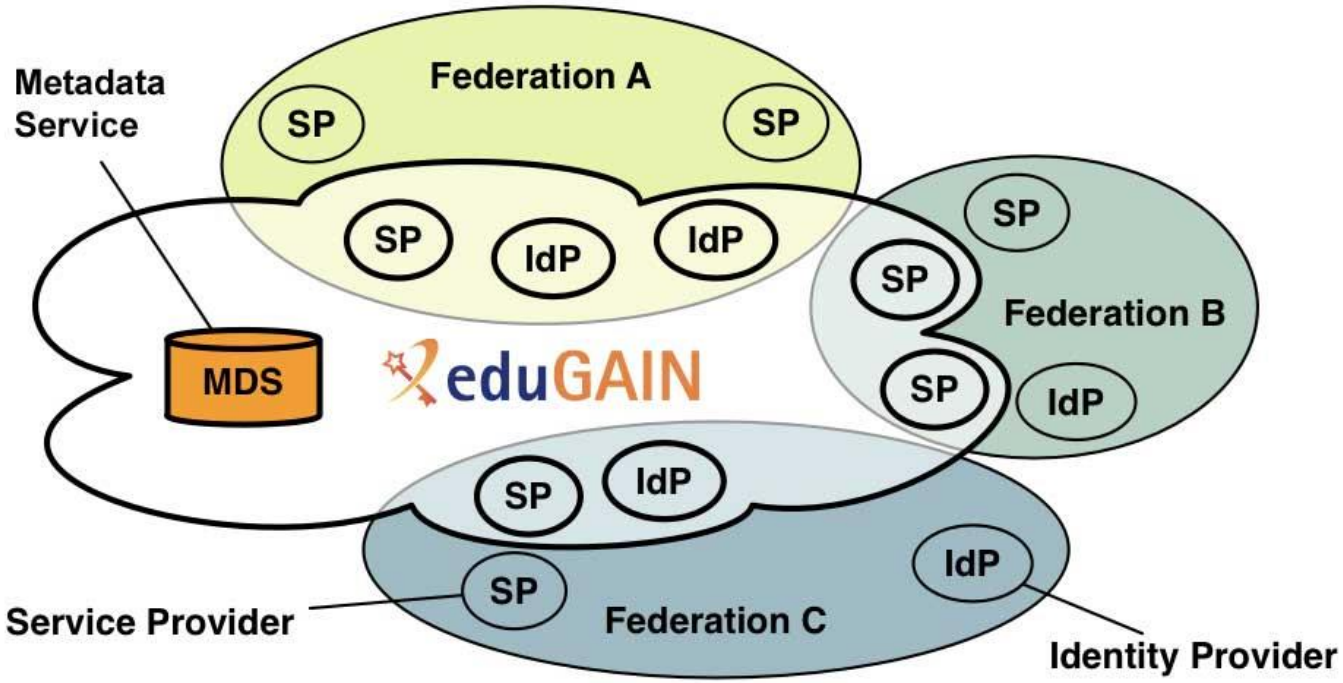


Authentication & **A**uthorization **I**nfrastructure

The CTA consortium is an experimental scientific collaboration consisting of over 1200 members working in 32 countries from 200, mostly academic, institutes.

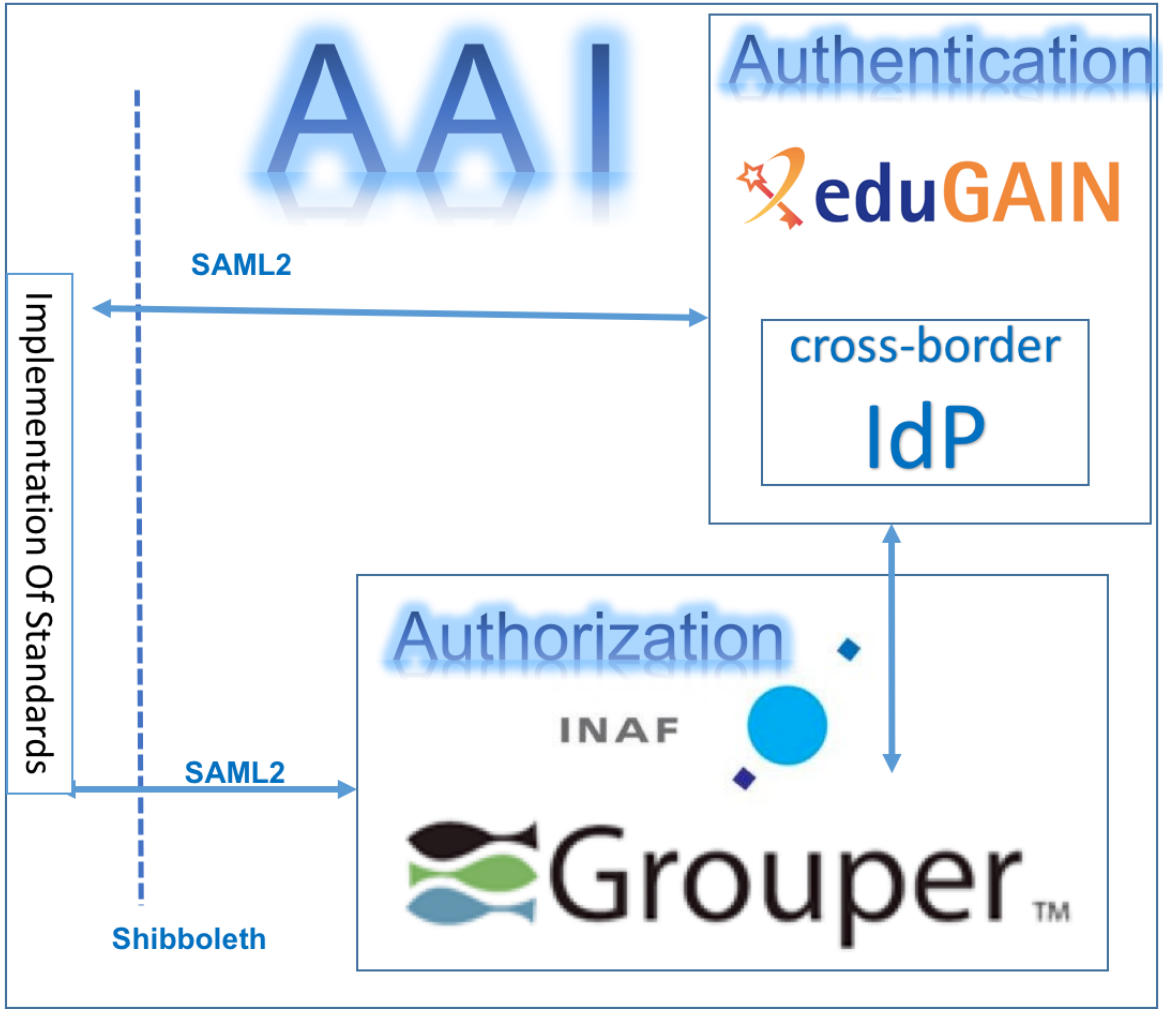
The geographical location of consortium members leads to the need of a pervasive **Federated Identity Management network**.

Authentication is based on **eduGAIN**



- ✓ Federations adhere to a common lightweight technical and policy infrastructure.
- ✓ Each national federation publishes a trust registry in the form of a metadata file.
- ✓ Each federation sends its registry to eduGAIN, and eduGAIN combines them into a unique Metadata Service.

Authentication & Authorization Infrastructure



On top of the eduGAIN federated authentication infrastructure the **authorization** solution is based on **Grouper**

Authorization with Grouper™

- Grouper keeps the **membership affiliation consistent** across multiple applications allowing to create and manage **groups**.
- Groups are used within each CTA application (e.g. a science gateway, the archive user interface or the project management portal) to track an **individual role**, or to determine which users are authorized to access the resources.
- If groups are managed separately in each application, keeping the membership list consistent across these services becomes very difficult.
- Grouper provides a way to define a group once and use that **group across multiple applications** managing it at a single point. The single point of control implies that, once a person is added or removed from a group, the group-related privileges are automatically updated in all of the collaborative applications.

CTA AAI Pilot @ AARC2

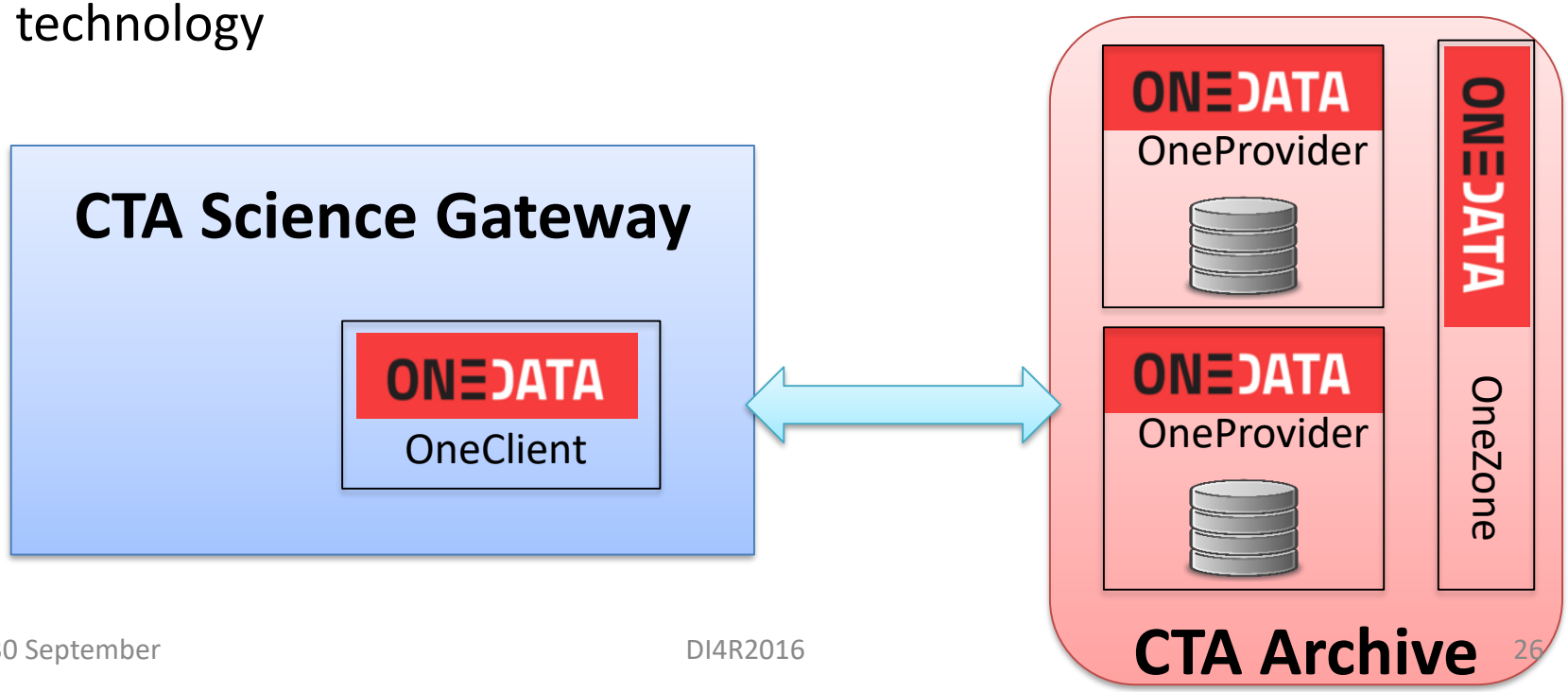
- The INAF CTA AAI pilot has been selected as one of the new use cases for the **AARC2 Project** (H2020 starting 2017).
- Navigate the route towards a unified, inter-operable AAI for Research and Education that improves **collaboration**, **support data intensive research** and **reduce the overall cost** delivery for all participants.
- Enable researchers to access all services available for their work using one set of credentials (**SSO across e-infrastructures**)
- Propose **sustainability models**, to be deployed in an effective way.

Conclusions

- We have introduced a workspace tailored to the requirements of the CTA community. It consists of: a science gateway module based on the Liferay framework endowed of a workflow management system and embedding a web-desktop environment (ACID) and provides an authentication and authorization infrastructure.
- Wide adopted **standards** (such as SAML 2.0 and Shibboleth 2.0) and **open-source** technologies (such as WSPGRADE/gUSE and Grouper) have been adopted. This aims at enlarging the developer community and improving the sustainability of the workspace during the whole CTA lifetime.
- The proposed solution provides an highly **flexible ecosystem** in order to tailor a product suitable to the present and future requirements of the CTA community.
- The next steps within this work are foreseen to be focused on the integration of the proposed workspace with the other modules and services of CTA (e.g. focusing on solutions to provide **messaging protocols** between the different modules).
- We will give support for the integration with the developed AAI to the other modules.

Future works

- Connection to the CTA Archive
- A CTA Archive pilot is being built within the INDIGO-DATACLOUD, a project funded by EU HORIZON 2020 Call: EINFRA-2014-2 Topic: EINFRA-1-2014 - “Managing, preserving and computing with big research data” -> See related Poster and Lightning talk!
- Connect to a federation of storage (hundreds of PB) using OneData technology



References

- Costa, A., Sciacca, E., Becchini, U., Massimino, P., Riggi, S., Sanchez, D., & Vitello, F. (2016). **An Innovative Workspace for The Cherenkov Telescope Array**. In *8th International Workshop on Science Gateways (IWSG)*.
- Costa, A., Massimino, P., Bandieramonte, M., Becciani, U., Krokos, M., Pistagna, C., Riggi, S., Sciacca, E. & Vitello, F. (2015). **An Innovative Science Gateway for the Cherenkov Telescope** Array. *Journal of Grid Computing*, *13*(4), 547-559.
- Massimino, P., Costa, A., Becciani, U., Vitello, F., & Sciacca, E. (2014, June). **ACID: an interactive desktop for CTA science gateway.** In *Science Gateways (IWSG), 2014 6th International Workshop on* (pp. 55-60). IEEE.

Documentation

- Documentation on INAF CTA Science Gateway:

http://cta-sg.oact.inaf.it/web/guest/documentation

- Demo of the Fermi Use Case on YouTube:

https://youtu.be/Qru6joO-Vw8