

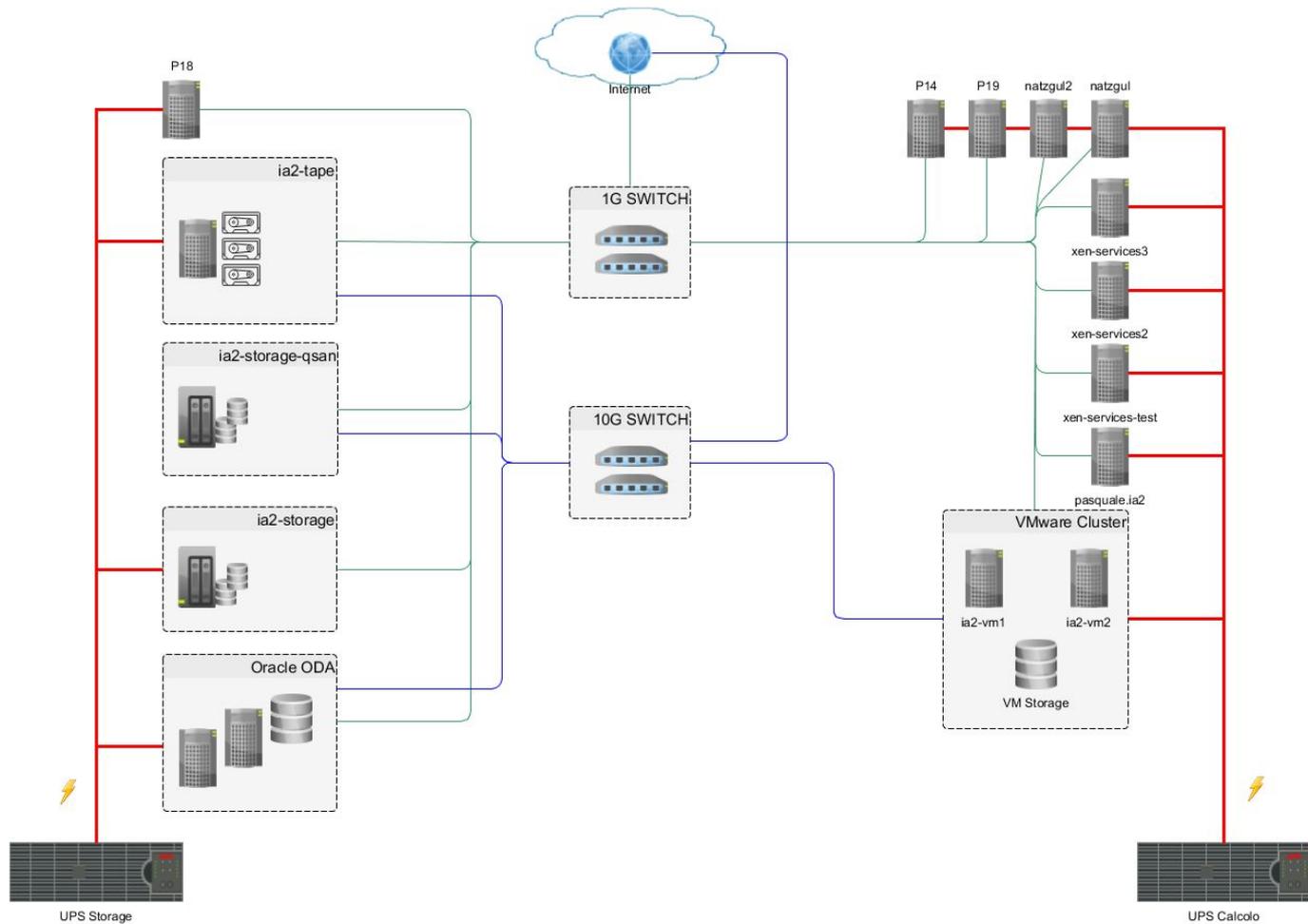
# INAF Astronomical Data: IA2 project

*Cristina Knapic*  
in behalf of IA2 team

# Summary

- IA2 status:
  - infrastruttura
    - storage on line;
    - preservazione;
  - progetti attualmente gestiti e in procinto di partire;
  - servizi web;
- progetti formativi:
  - progetti internazionali;
  - progetti H2020;
- necessità:
  - interazione con piattaforme di calcolo;
  - interazione con altri DC;
  - servizi di visualizzazione/elaborazione/condivisione;
  - supporto all'utenza;
  - PERSONALE!!

# IA2 status : Infrastruttura



# Storage on line

- **Hardware IA2 @ TS:**
  - 500 TB (300 used + 200 free TB)
  - backup : 100 TB for VMs
- **Hardware @ other sites:**
  - IRA : 60 TB on new machine (x testing)
- **Hardware owned by others:**
  - IRA : 40 TB Radio Distributed Archive
  - SRT : 1 TB (pulsar testing machine)
  - Serra La Nave : 500 GB on site
  - LBT : 12 TB upgraded 1TB /y Full LBT Archive
  - Asiago : 500 GB on site
- **Bandwidth:** 10Gb/s GARR



# Storage off-line (asynchronous calls)

- IA2 tape library @ Ts:
  - HPE MSL LTO-6 6250 con 200 TB
  - in valutazione l'acquisto di una T950 HPE LTO-8 da 1.25 PB espandibile a 12.5 PB sullo stesso rack



- Tape library @ other sites:
  - IRA : previsto passaggio della HPE MSL 6250 da Ts a Bo

# IA2 status: Progetti attuali

## Dati da Telescopi / simulazioni:

- TNG : all instruments
- LBT : all instruments except LBTI
- Asiago Observatory : all instruments
- Serra La Nave
- ExoClimates (simulations)
- Intrigoss (simulations)
- Prisma
- Radio (Medicina, Noto, SRT)
- BaSTI
- Svas
- MWA mirror (150 TB)

## Dati da nuova strumentazione:

- LBT : Shark - Vis and Shark - Nir
- TNG : GianoB
- Osservatorio del Chianti
- MS per ALMA
- SRT : Pulsar
- .....

[IA2 mette a disposizione di tutti i progetti partecipati INAF la propria expertise e la propria capacità di storage per qualsiasi necessità di archiviazione dati, dal “save the bit” alla gestione di archivi alla pubblicazione dei dati pubblici e privati tramite servizi ed interfacce web.](#)

# WEB: Servizi, portali e VO compliancy (1/3)

Ogni archivio è di default dotato di :

- web portal per l'accesso a dati pubblici e privati;
- servizi Virtual Observatory compliant:
  - TAP : default
  - SSAP (1.1)
  - SIAP (v2.0)
  - Cone Search

Remote Authentication Portal

Account Management

eduGAIN

Use the eduGAIN Logo to Login or Register to the RAP facility if you belong to an eduGAIN IGP.

G f in

Use these Logos to Login or Register to the RAP facility with your social identity

X.509

Use the X.509 Logo to Login with your personal certificate (IGTF and TERENA-TACAR, are allowed).

IA2

Use the IA2 Logo to Login if you have an account provided by IA2 or self registered

Need help? Please read our [User guide](#) and [FAQ](#).

[Privacy policy](#)

TNG Archive

Name resolver: m87 Resolve

RA: 12:30:49.42 Dec: +12:23:28.0 Radius (arcmin): 14

Filename: [input]

Obsv. Date: From [input] To [input]

Obs. Type: Select... Instrument: Select... Program: [input]

Object: [input]

Exp. time: From [input] To [input]

Night date: From [input] To [input]

Airmass: From [input] To [input]

Single instrument search

Search Reset

Create fail from selected VOTable

Rows displayed: 20

File name	Obsv. Date	Obs. Type	Instrument	Program	Object	Policy
LRS.2013-04-12T00-15-41.250.Its.gz	2013-04-12T00:14:17.840	OBJECT	LRS	ALTDOT1	M87	FREE
LRS.2013-04-12T00-17-50.875.Its.gz	2013-04-12T00:16:27.488	OBJECT	LRS	TEST	M87	FREE
LRS.2013-04-12T00-19-34.966.Its.gz	2013-04-12T00:18:11.560	OBJECT	LRS	TEST	M87	FREE
LRS.2013-04-12T00-21-16.577.Its.gz	2013-04-12T00:19:52.910	OBJECT	LRS	TEST	M87	FREE
LRS.2013-04-12T00-33-40.559.Its.gz	2013-04-12T00:31:17.061	OBJECT	LRS	TEST	M87	FREE
LRS.2013-04-12T00-23-09-110.Its.gz	2013-04-12T00:21:45.358	OBJECT	LRS	TEST	M87	FREE

IA2 - Italian Center for Astronomical Archives

Single Sign On Services

Home FAQ Privacy Policy

## Logging in into IA2 services

You can access IA2 services through the RAP (Remote Authentication Portal) application. RAP allows you to login using eduGAIN credentials, X.509 certificates and most popular social networks.

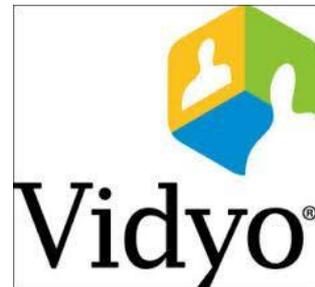
We recommend to authenticate using one of these modes but, if you don't have any of the listed account types, it is possible to perform a self-registration.

List of IA2 services using Single Sing On:

- Telescopio Nazionale Galileo (TNG) portal
- Asiago Astrophysical Observatory portal
- RAP Account Management
- Grouper (groups management)

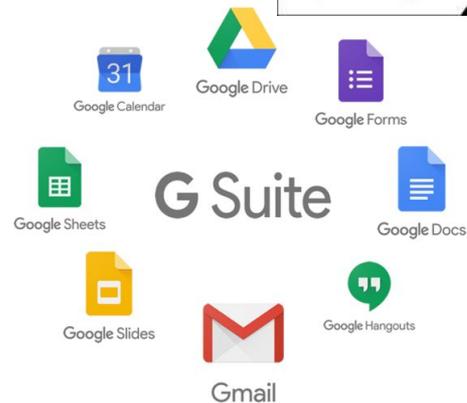
## Account linking

# WEB: Servizi, portali e VO compliancy (2/3)



**REDMINE**

flexible project management



# WEB: Servizi, portali e VO compliancy (3/3)

## Dati science ready: TNG example

- dati ridotti con pipeline DRS pubblicati sul portale TNG a seconda della policy del programma osservativo (da 1 a 3 anni);
- tutti gli osservatori degli strumenti HarpsN e GIANO-B hanno la possibilità di ri-ridurre i dati utilizzando in maniera interattiva le pipelines installate (DRS e Gofio) e mantenute aggiornate su YABI. Al momento questi dati sono proprietari.

Yabi : piattaforma per la gestione dei flussi di riduzione dati.

Vantaggi:

- software sempre aggiornato all'ultima release delle pipelines;
- pipelines privata, dati potenzialmente pubblici;
- uso intuitivo ed interattivo delle pipeline;
- utente non necessita di potenza di calcolo

Svantaggi:

- Gestione locale dei dati (al momento non vengono re-ingestiti in archivio per questioni di privacy)
- sistema non integrato con RAP



# Progetti formativi

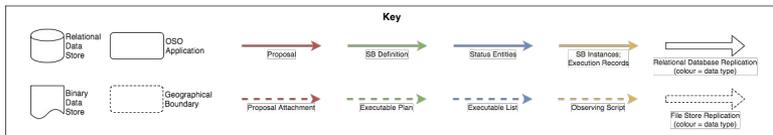
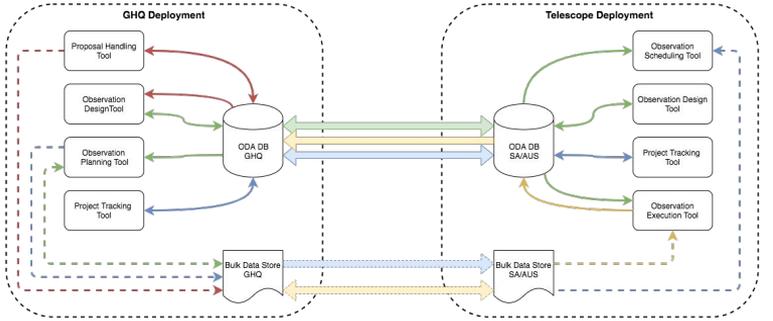
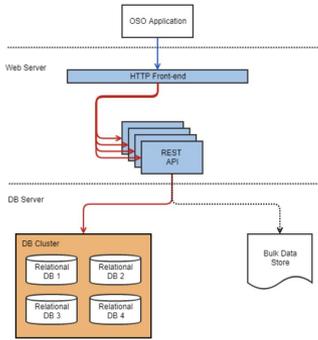
Nel tempo lo staff e i collaboratori IA2 hanno preso parte a molti progetti nazionali, internazionali ed Europei da cui sono stati tratti notevoli insegnamenti. Molte delle scelte dei recenti sviluppi di IA2 sono dipese dall'esito di queste sperimentazioni. Se ne citano le maggiori:

- SKA.TM - Observation Management, Observation Data Archive e Authentication, Authorization and Auditing;
- PoC Google
- INDIGO-DataCloud: OneData, container Docker e Ophidia
- Asterics: sistemi di workflow, AAI
- AENEAS: centri dati distribuiti, sistemi di interfacciamento utenti e AAI
- Collaborazione con Egi-Engage (CADC): VOSpace
- collaborazione TANGO: sistemi di controllo distribuiti
- ...

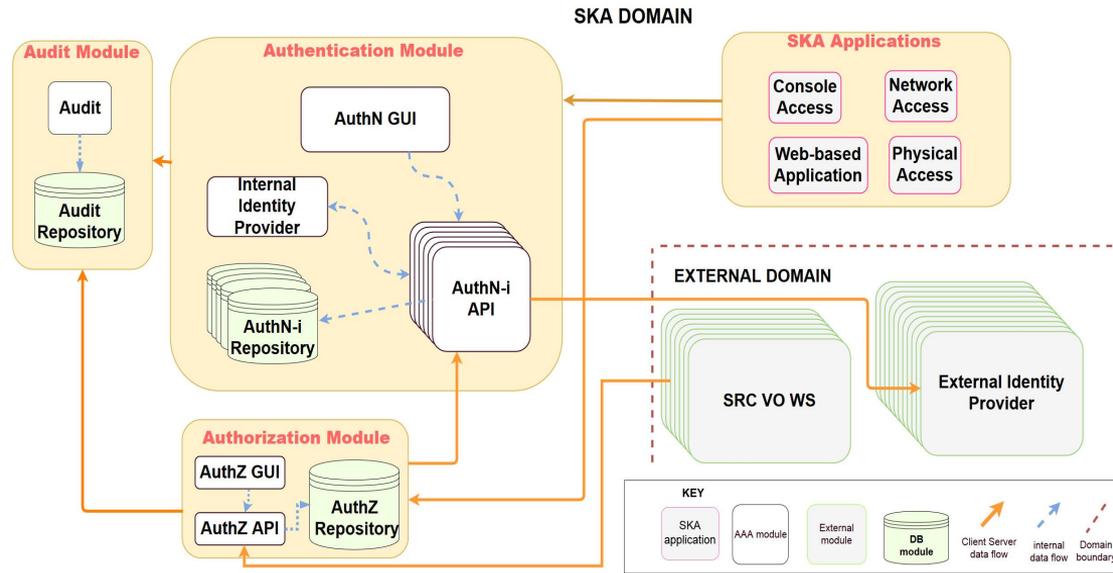


# Progetti formativi: SKA (2/2)

Google Cloud Spanner



## SKA AAA System architecture

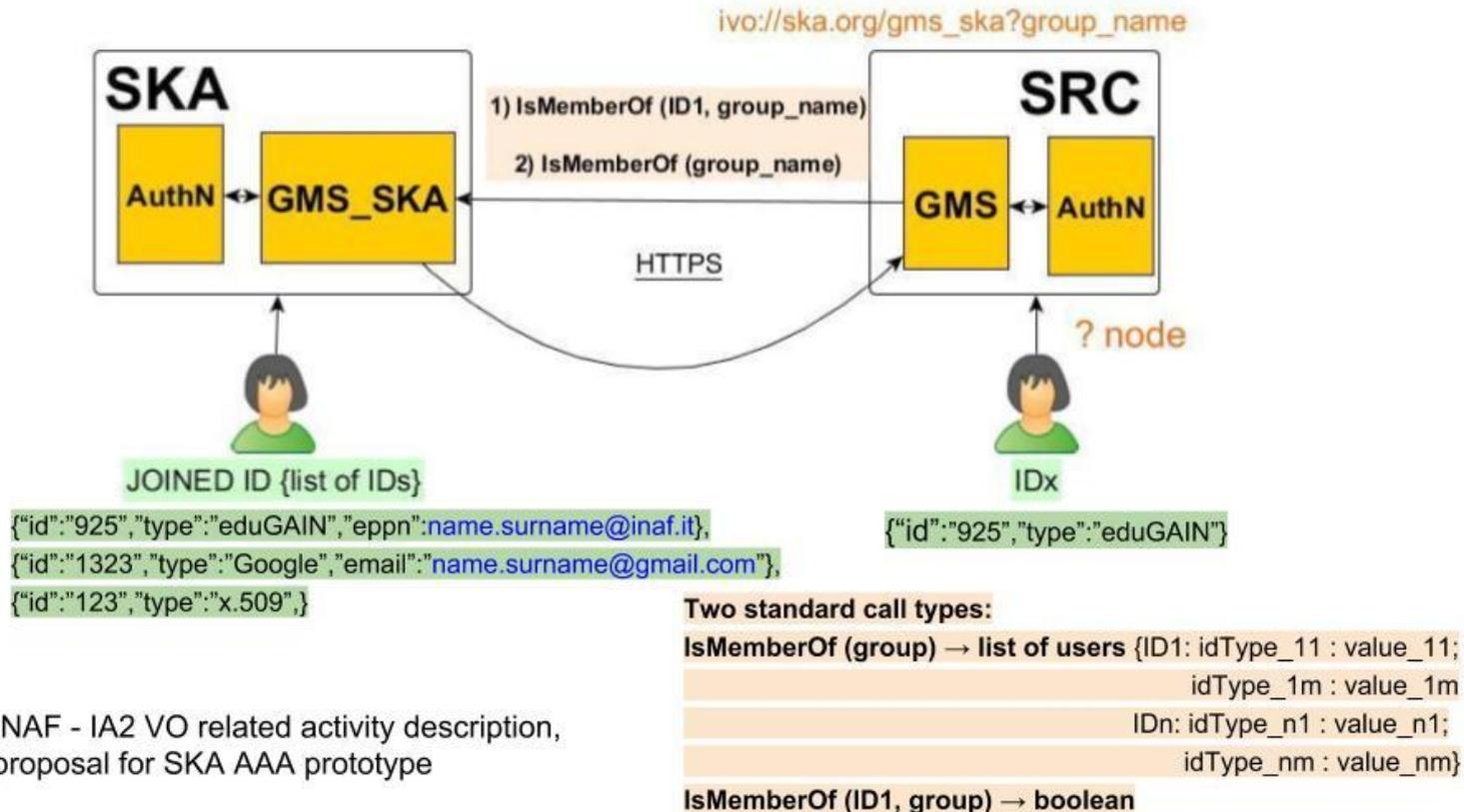


Importanza del tracciamento della proprietà dei prodotti dalle proposal ai dati scientifici tramite sistemi di Obs Mgt con AAI

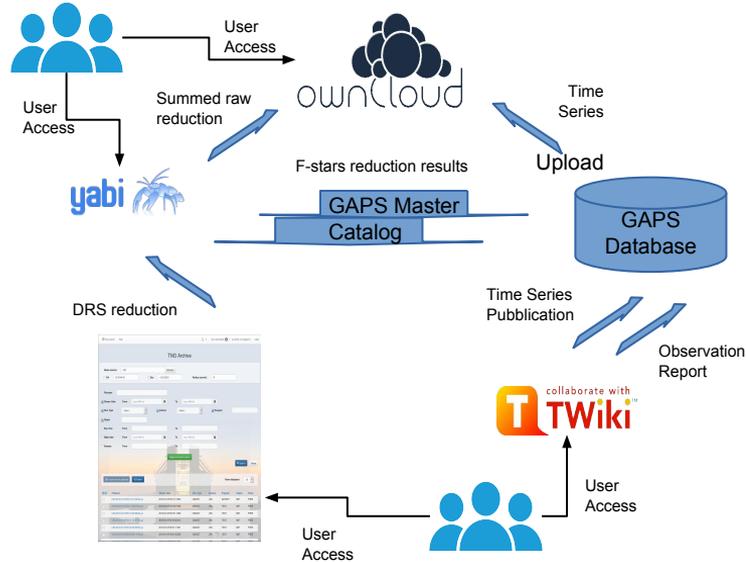
# Progetti formativi: AENEAS, ASTERICS & CADC

## 1. Multi-protocol AuthN - 2) Account Linking

Courtesy of C. Knapic - G. Jerse -- INAF - OATs



# Necessità: piattaforme di calcolo (1/2)



- I dati IA2 applicano da sempre il concetto dell'archivio distribuito, acquisendo o gestendo dati in 3 continenti;
- L'esperienza di GAPS ci ha portato notevole know-how. Consapevoli del fatto che i dati da telescopi dell'era 2020 hanno altre dimensioni, vogliamo essere pronti per integrare i servizi di storage con il calcolo, ovunque esso venga effettuato.
- I portali IA2 hanno già integrato sistemi di autenticazione e autorizzazione utente allo stato dell'arte. I stessi paradigmi sono stati applicati per la prototipazione in SKA e in AENEAS, con sorprendenti risultati.

# Necessità: piattaforme di calcolo (2/2)

IA2 supporta una serie di servizi per lo sharing e l'accesso ai dati scientifici dei telescopi a partecipazione italiana, ai dati di simulazione e ai dati educational.

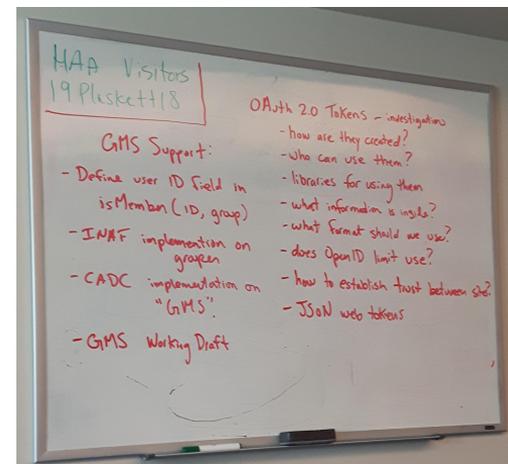
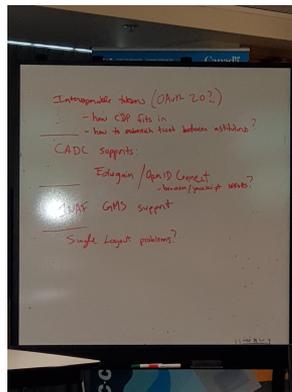
Necessità di interoperabilità con i centri di calcolo derivanti da:

- Pipelines;
- visualizzazione;
- mosaici;
- multi-messenger;
- data mining e machine learning su cataloghi;
- image/pattern recognition;
- .....

Ricordiamoci che le infrastrutture di calcolo sono distribuite geograficamente e possono essere infrastrutture cloud private o pubbliche. Primi test con i PoC Google.

# Necessità: interazioni con altri DCs

- IA2 ha interagito con CADC per poter avviare una sperimentazione di DC distribuito grazie alla volontà di configurare uno user space compatibile con l'implementazione CADC di VOSpace.



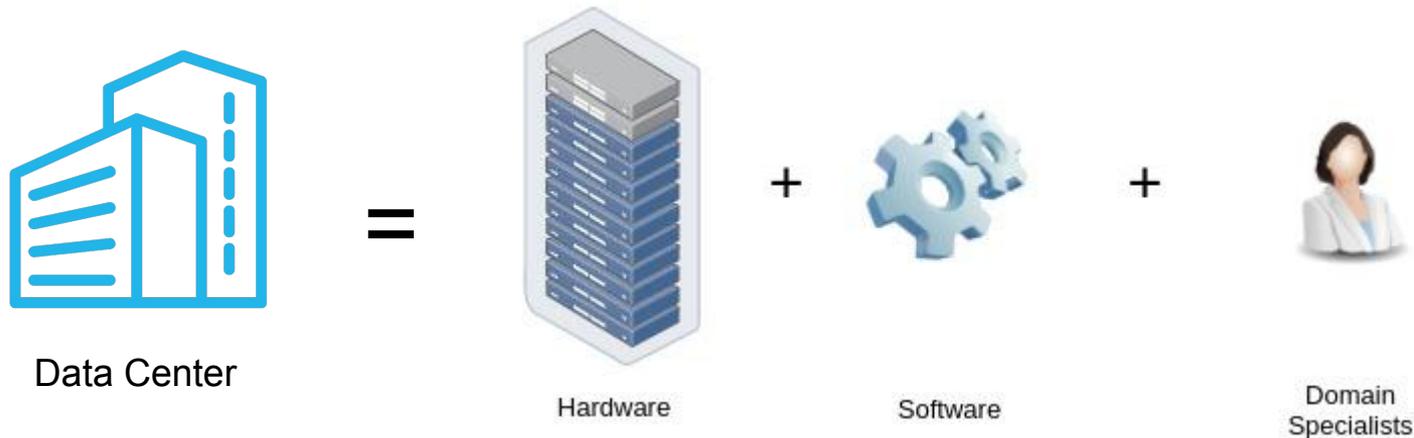
- IA2 ha preso contatti con SSCDC per la condivisione delle expertise
- IA2 è stata contattata dallo SKAO per contribuire in prima battuta alla gestione dei dati SKA simulati ed eventualmente alla prima prototipazione degli SKA Regional Centers che forniscano i servizi necessari previsti per l'utenza.
- Qualunque infrastruttura -che fornisca servizi all'utenza astronomica- intenda collaborare con IA2 è benvenuta!

# Necessità: servizi di visualizzazione / elaborazione / condivisione

La mole dei dati dei telescopi dell'era 2020 non permette all'utente di elaborare l'informazione scientifica sul desktop. Soltanto alcuni centri di eccellenza INAF forniscono supporti per calcolo e servizi di visualizzazione che vengano incontro alle necessità degli utenti. Rendere accessibili al calcolo i dati degli archivi IA2 in maniera programmatica e strutturata, minimizzando il traffico dati, laddove possibile, incrementerebbe la fruibilità dei dati e incentiverebbe la produttività scientifica.

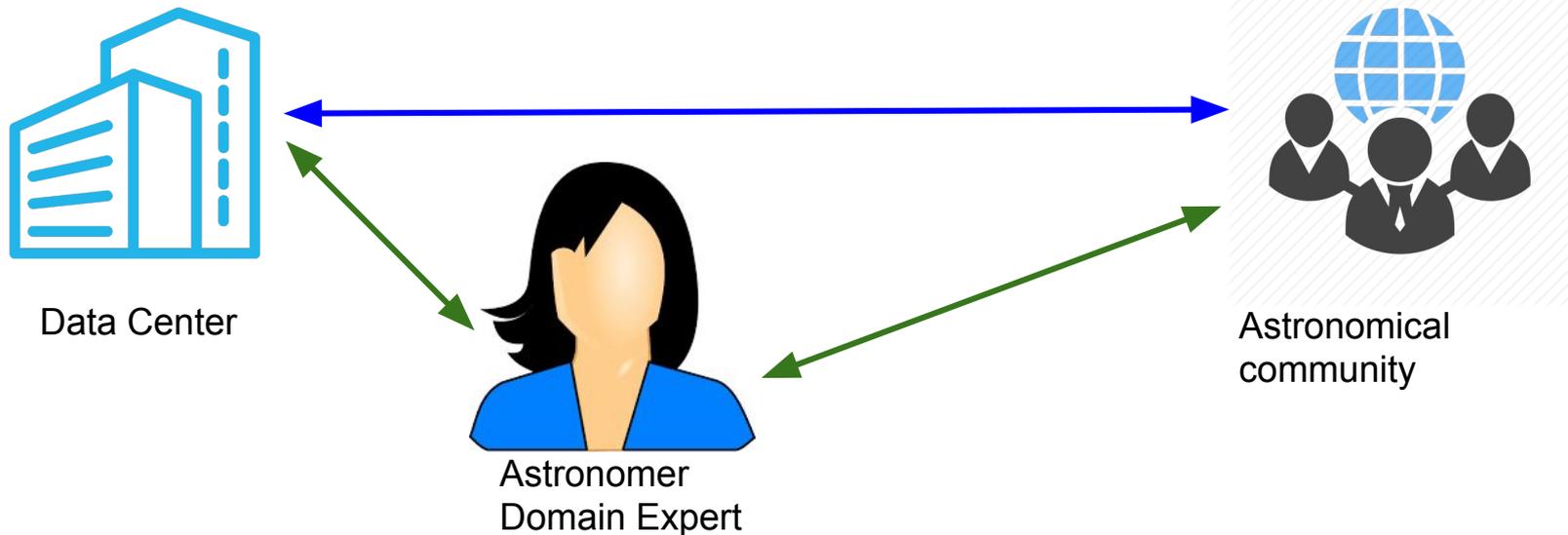
Questo esperimento si configurerà come una sfida nei prossimi anni per INAF perchè necessiterà di risorse e collaboratività da tutta la parte tecnologica (e scientifica!) dell'ente.

# Necessità: supporto all'utenza (1/2)



Come da esperienza acquisita lato data provider, è indispensabile la collaborazione con esperti di dominio e parte scientifica per la corretta **identificazione dei modelli** prima e per la **definizione delle specifiche** del servizio poi. In questo rientrano anche i modelli di pubblicazione dei dati (Virtual Observatory, FAIR policies, Open Data, RDA....). Non dimentichiamo il supporto all'utenza per il retrieval dei dati!

# Necessità: supporto all'utenza (2/2)



- Esperienza ARC ed AENEAS: emerge la necessità per alcuni telescopi di fornire supporto anche per la corretta estrazione dell'informazione scientifica dai dati;
- esperienza CADC: l'utente deve avere a disposizione uno user space ove caricare eventuali prodotti (dati o pipelines o cataloghi) per la cross-correlazione di diversi prodotti scientifici;
- DOI :per articoli e data set!! I dataset andrebbero resi pubblici.

# Personale

Responsabile: [C. Knapic](#)

Ingestione: A. Bignamini, E. Londero

Workflow e sistemi collaborativi: A. Bignamini

Interfacce e servizi web: S. Zorba

Sistemistica: M.Sponza

Sviluppo software: M.Vela Nunez

Collaboratori tecnici:

IRA: F. Tinarelli, F. Bedosti..

OATs: G. Taffoni, S. Bertocco

OATo: R. Morbidelli, A. Volpicelli..

Collaboratori scientifici:

Radio: A. Zanichelli, M. Nanni, M.Stagni, A. Possenti,....

LBT: D. Thompson, O.Khun,...

TNG: M.Lodi, A.Harutyunyan, I. Pagano....

Asiago: L.Tommasella, S. Benetti,..

Prisma: D. Gardiol, A. Carbognani, D.

Barghini...

LBT- Shark: J. Farinato, F. Pedichini

Collaboratori Virtual Observatory:

Responsabile: [M. Molinaro](#)

Sviluppatori: R. Butora, N.F. Calabria

Collaborazioni con altri DCs:

CADC: S. Gaudet, B. Major, P.Dowler.....

IRA-ARC: M. Massardi, E. Liuzzo, A. Burkutean,...

In prospettiva

SSDC: G. Polenta, M. Perri, M. Giardino

# Conclusioni

## IA2v1.0

- infrastruttura matura per la long preservation dei dati astronomici
- Ottima gestione dati scientifici raw;
- Portali e servizi VO compliant;
- Archivi science ready per dati TNG - HARPN, in progress per LBT

## IA2v2.0

- creazione dello spazio utente
- connessione con le infrastrutture di calcolo
- gestione dei dataset e dei prodotti scientifici tramite DOI e loro pubblicazione
- necessità: personale!

*Spero che questa visione sia condivisa o comunque stimoli la discussione a livello di Direzione Scientifica ed ente tutto*

*Grazie per l'attenzione!*

