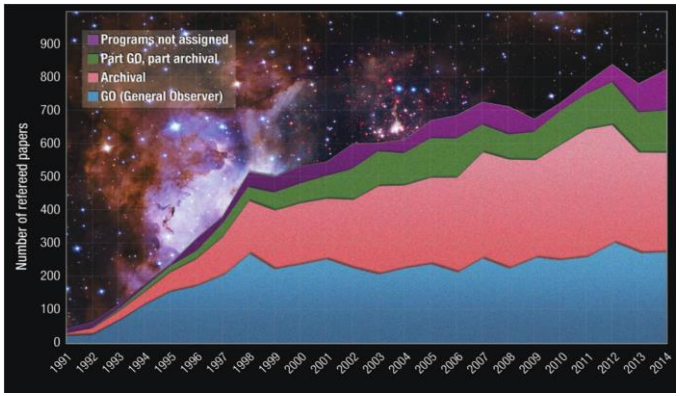
Archives and Virtual Observatory

R. Smareglia
Head of ICT & Science Data Management office

Rome – IT-UKR meeting – 22-23 March 2018

Data Archive: Why it's important



HST Newsletter: "*At the present time, approximately **half of the refereed publications** based on Hubble observations are derived purely **from archival data**, and, every year, this number is slightly higher than the number of publications based on new observations. .... the Hubble Archive has become a goldmine for the astronomical community....*"

Archive are:

- Data Management
- Data Curation
- Data Preservation

Archive are not:

- Data Sharing

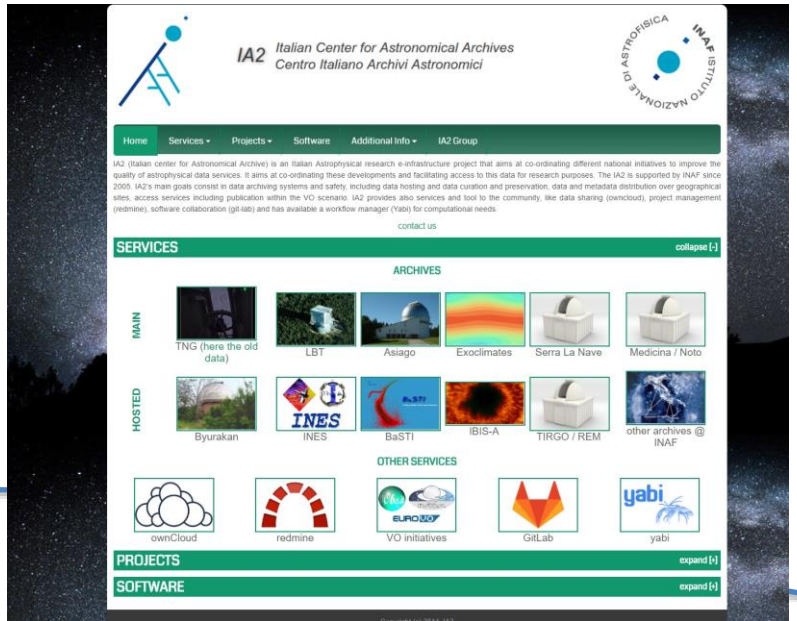
Archives @ INAF

- Archives ..
  - Archives (> 50) .. preservation
  - IA2
    - LBT, TNG, ..
    - Ground base
- International projects
  - IVOA
  - Open Universe
  - EOSC
  - Open Science

The Italian Astronomical Archives (IA2)
ia2.inaf.it

Aims to:

- Support the Italian (and International) community in the data storage, curation and preservation;
- Support the astronomical user community in the data retrieval (Web Interfaces and VO services);
- Support the astronomical user community in the data reduction and sharing (User Space)
- Support the user community in the collaboration tool usage.



"IA2 currently is the only e-infrastructure in INAF (from PT 2017-2019)"

The data
storage, curation and preservation

Telescope's data handled/hosted (ex):

- TNG : all instruments
- LBT : all instruments except LBTI
- Asiago Observatory : all instruments
- Serra La Nave
- Radio (Medicina, Noto, SRT)

- Prisma (all sky camera)

- MWA mirror (150 TB)

- ExoClimates (simulations)
- BaSTI

The data:
storage, curation and preservation

Hardware IA2:

on line :

500 TB

backup : 200 TB

"off line" : 200 TB with expansion to 5 PB

Bandwidth: 10Gb/s GARR

Hardware owned by partners:

IRA : 40 TB Radio Distributed Archive

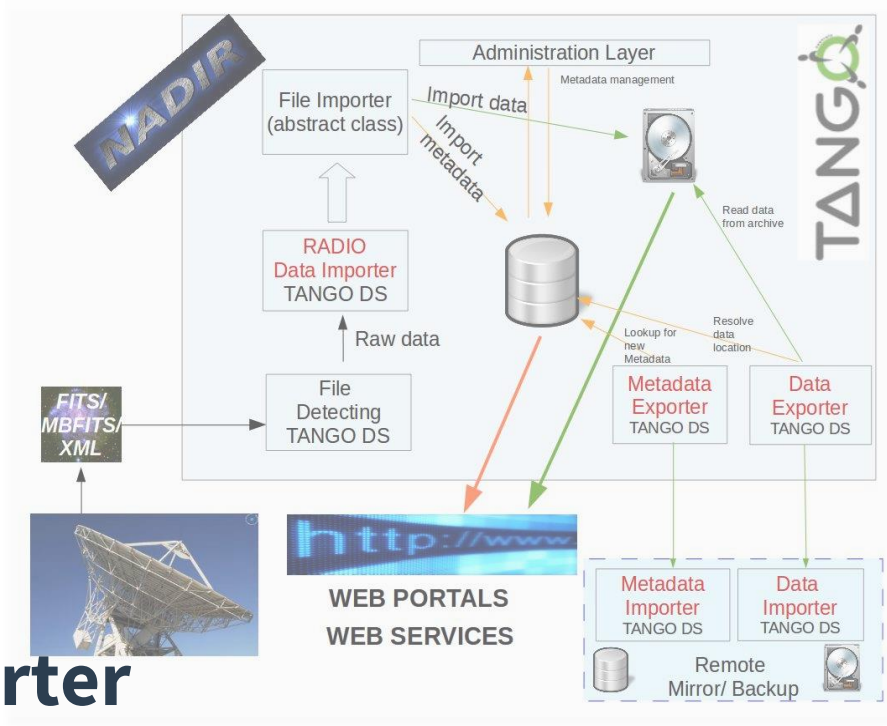
Serra La Nave : 500 GB on site

LBT : 12 TB upgraded 1TB /y Full LBT Archive

Asiago : 500 GB on site

The data storage, curation and preservation

- **Software**
  - **NADIR**
    - **Preprocessor**
    - **Fits Importer**
    - **Radio Data Importer**
    - **Meta and Data Exporter/Importer**
    - **Data Distribution / Radio Data Distribution**
    - **Administration Interface**



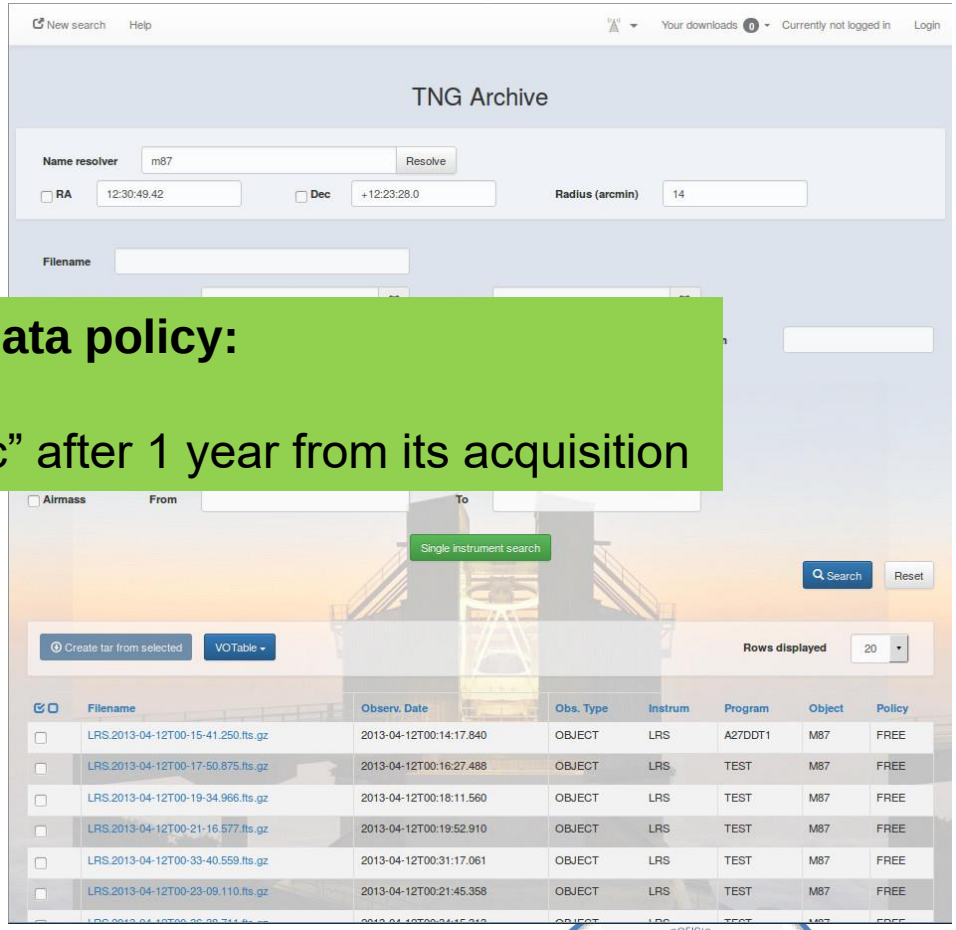
Data Retrieval

IA2 offer several services for the data retrieval:

- Web portals
- Virtual
  - TA
  - SS
  - SIA
  - ConeSearch
- Help Desk
  - users help desk (~ 10 e/m)
  - provider help desk (~ 2 e/m)

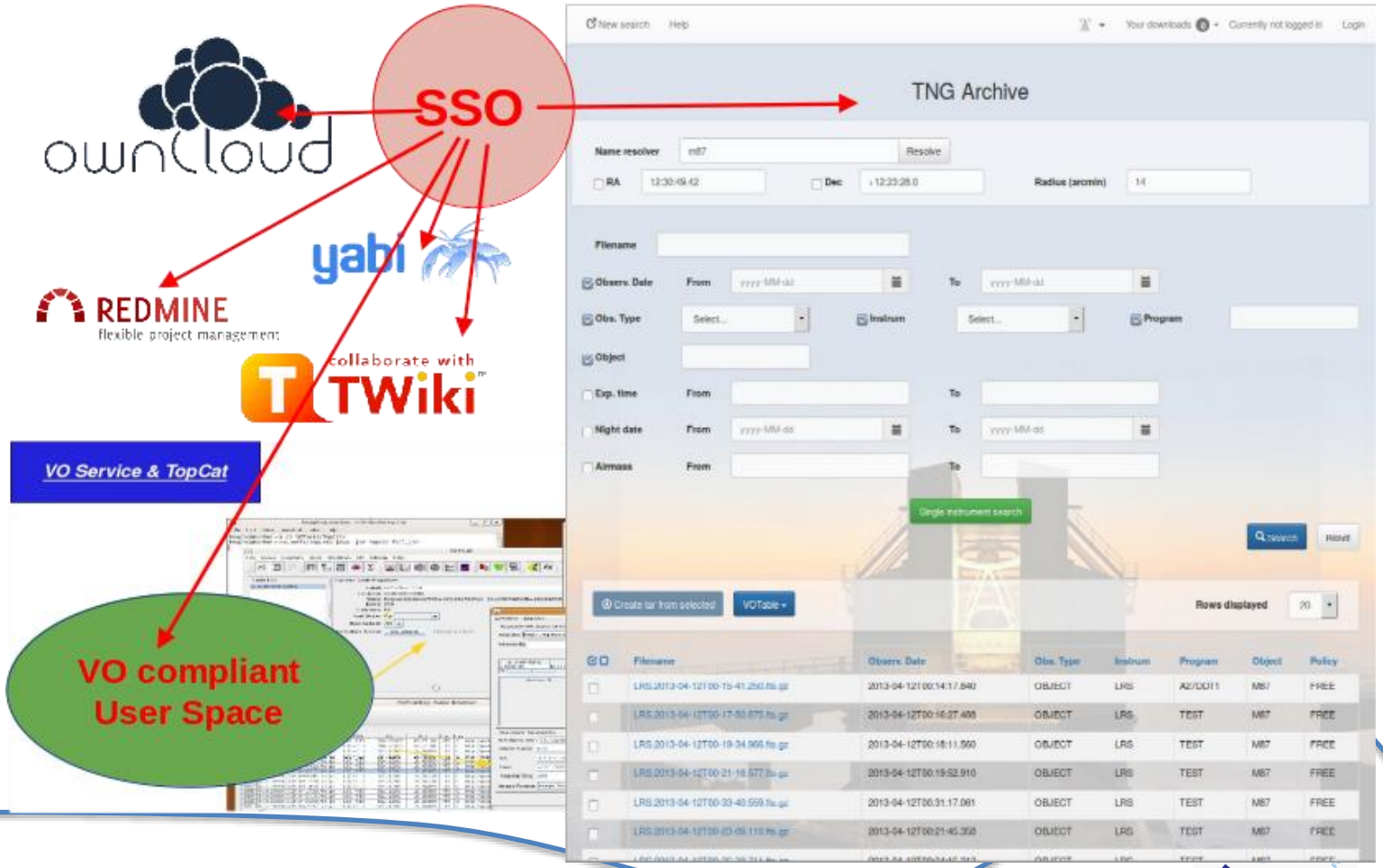
INAF data policy:

~ all INAF raw data is “public” after 1 year from its acquisition



SSO access

LDAP INAF ← → IDEM ← → EduGain



Interoperability: Open Access - Open Science

Open Access and Open Science is one of the MUST of the EU/H2020 funding project policy

- *The **<u>European Open Science Cloud</u>** (EOSC) pilot project, in which INAF is involved, will support the first phase in the development as described in the EC Communication on European Cloud Initiatives [2016].*

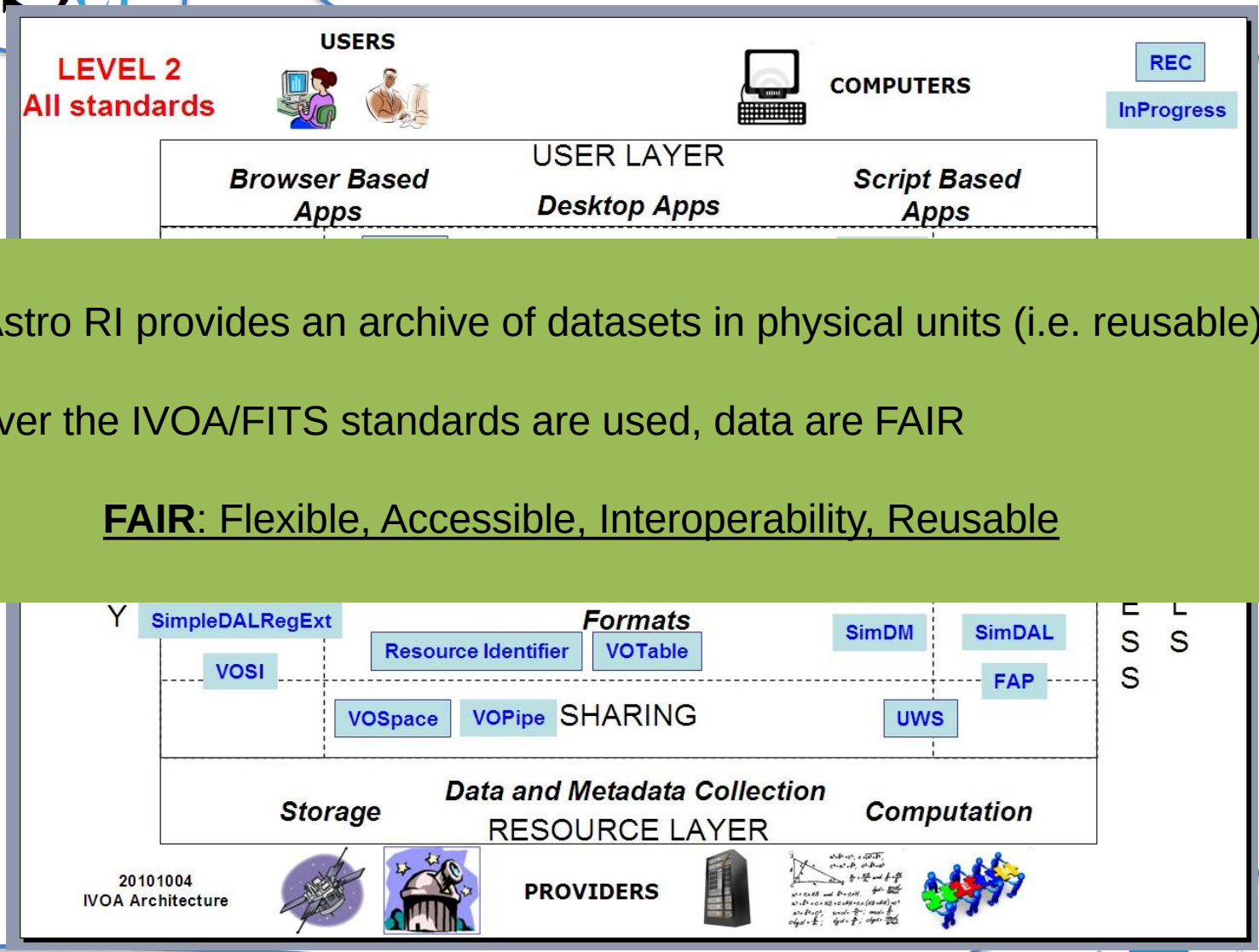
  - *It will establish the governance framework for the EOSC and contribute to the development of European open science policy and best practice;*
  - *It will develop a number of pilots that integrate services and infrastructures to demonstrate interoperability in a number of scientific domains; and*
  - *It will engage with a broad range of stakeholders, crossing borders and communities, to build the trust and skills required for adoption of an open approach to scientific research .*

Interoperability: IVOA

- *"Allow **astronomers** to interrogate multiple data centers in a seamless and transparent way"*
- *"Give **data centers** a standard framework for publishing and delivering services using their data."*

+

- Give **data centers** a standard framework for integrating interoperable authorization
- Allow **astronomers** to interrogate multiple data centers in a seamless and transparent authenticated way

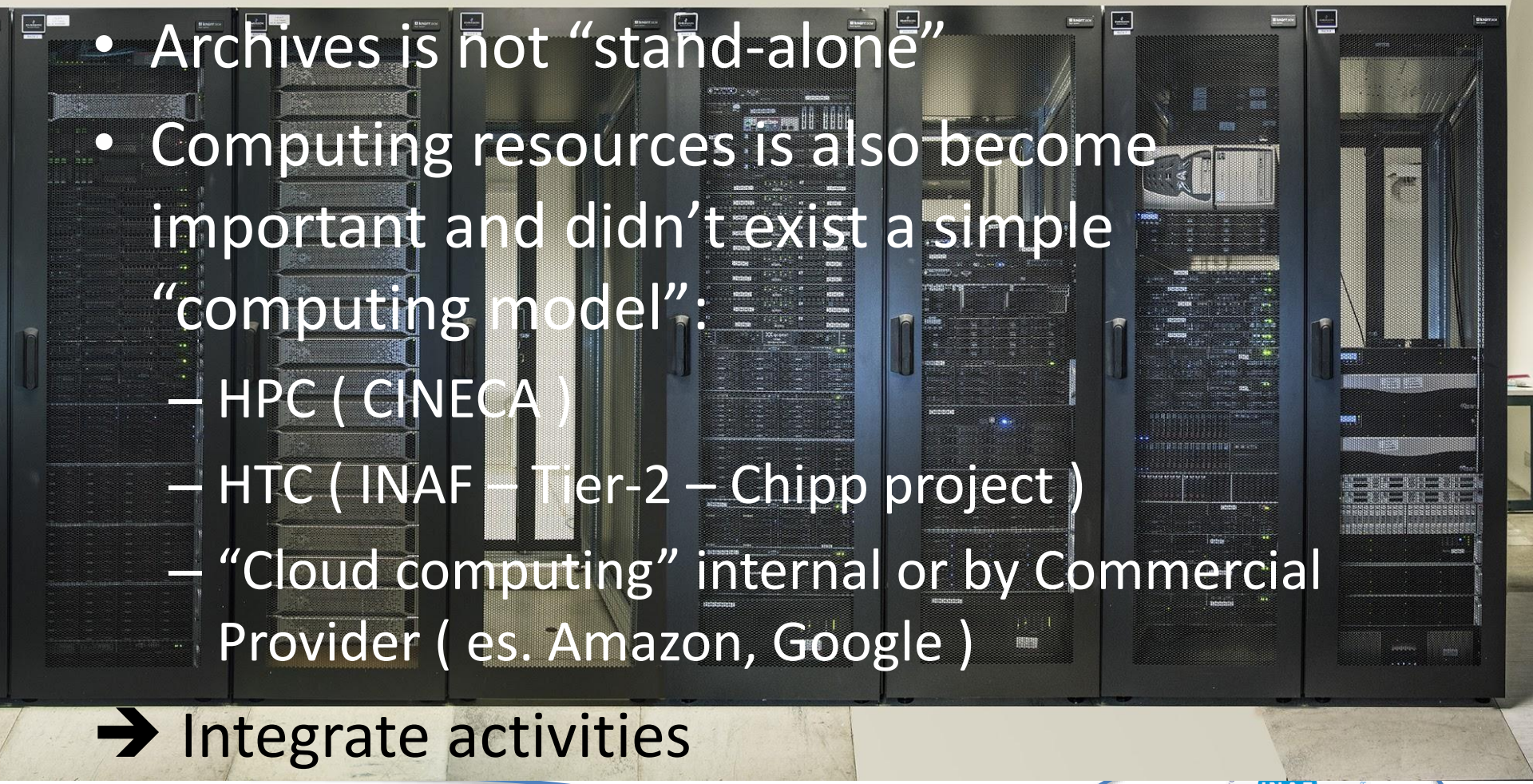


Each Astro RI provides an archive of datasets in physical units (i.e. reusable);

whenever the IVOA/FITS standards are used, data are FAIR

FAIR: Flexible, Accessible, Interoperability, Reusable

“Computing Facility”



- Archives is not “stand-alone”
- Computing resources is also become important and didn’t exist a simple “computing model”:
  - HPC (CINECA)
  - HTC (INAF – Tier-2 – Chipp project)
  - “Cloud computing” internal or by Commercial Provider (es. Amazon, Google)

➔ Integrate activities

Intellectual Propriety: DOI

- Digital Object Identifiers: persistent, globally unique, resolvable
- Can be assigned to **publications, data, software**
- Example: `10.18727/0722-6691/5000`

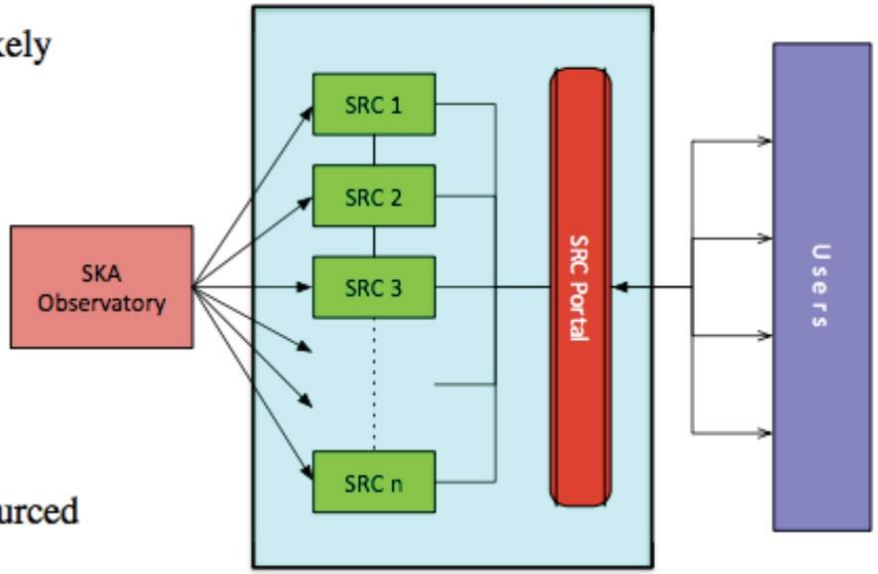
- Resolvable by prepending https://doi.org/
- Citable (unambiguously)
- As a URL, avoids link rot
- Machine-readable when cited
- Repositories offer (linked) metadata (for humans and machines)

SKA – Regional Center

Advanced European Network of E-infrastructures
for Astronomy with the SKA AENEAS - 731016

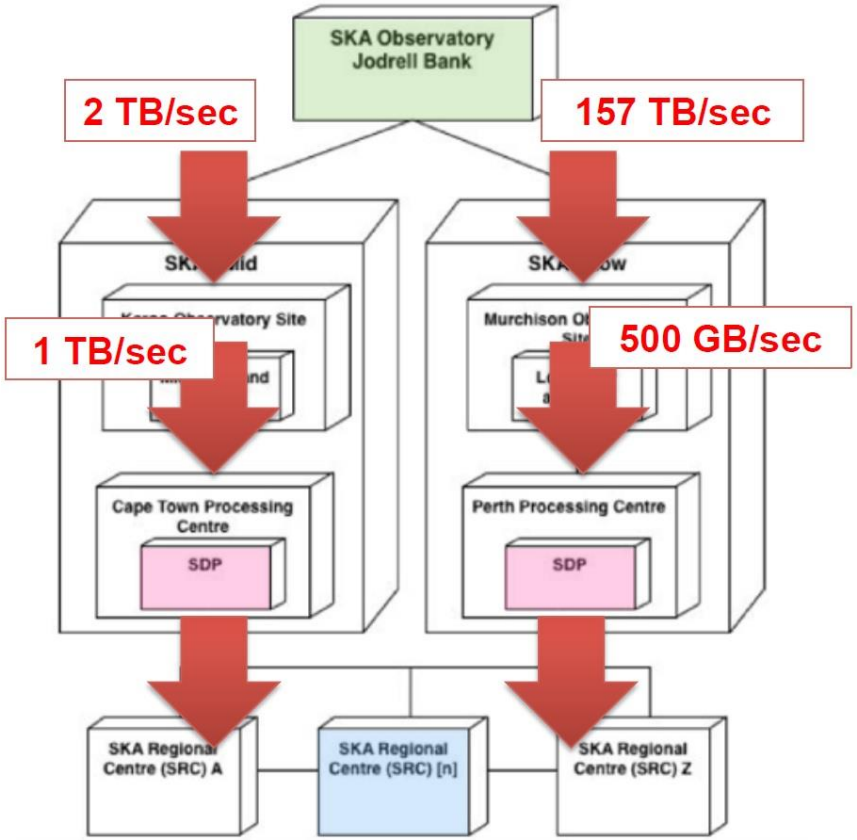
SKA Regional Centres

- Science Data Centres (SDCs) will likely host the SKA science archive
- Provide access and distribute data products to users
- Provide access to compute and storage resources for users
- Provide analysis capabilities
- Provide user support
- Multiple regional SRCs, locally resourced



SKA – Regional Center

Advanced European Network of E-infrastructures
for Astronomy with the SKA AENEAS - 731016



CENTRAL SIGNAL PROCESSING

SCIENCE DATA PROCESSING

REGIONAL DATA CENTRE


Thanks for your attention!