

The Italian Astronomical Archive (IA2): current status and future perspectives



Elisa Londero
on behalf of the IA2 team
elisa.londero@inaf.it



Laboratorio Spettroscopia INAF
Rome, June 11 2019

IA2 services: ia2.inaf.it



IA2 *Italian Center for Astronomical Archives*
Centro Italiano Archivi Astronomici



Home Services ▾ Projects ▾ Software Additional Info ▾ IA2 Group

ABOUT US

collapse [-]

IA2 (Italian center for Astronomical Archive) is an Italian Astrophysical research e-infrastructure project that aims at co-ordinating different national initiatives to improve the quality of astrophysical data services. It aims at co-ordinating these developments and facilitating access to this data for research purposes. The IA2 is supported by INAF since 2005. IA2's main goals consist in data archiving systems and safety, including data hosting and data curation and preservation, data and metadata distribution over geographical sites, access services including publication within the VO scenario. IA2 provides also services and tool to the community, like data sharing (owncloud), project management (redmine), software collaboration (git-lab) and has available a workflow manager (Yabi) for computational needs.

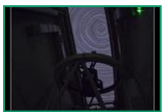
[contact us](#)

MAIN ACTIVITIES

collapse [-]

TELESCOPE ARCHIVES & SIMULATIONS

MAIN



TNG



LBT



Asiago

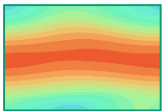


SVAS

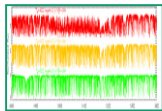


RADIO

MAIN



Exoclimates



INTRIGOSS



LBT OLD

HOSTED



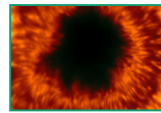
Byurakan



INES



BaSTI



IBIS-A



other archives @ INAF

OTHER SERVICES



ownCloud



redmine



VO initiatives



GitLab



yabi

Data cycle:

- data providers
- collection
- distribution
- curation
- long term preservation
- data publication:
 - web interfaces
 - VO
- data analysis
- collaborative tools

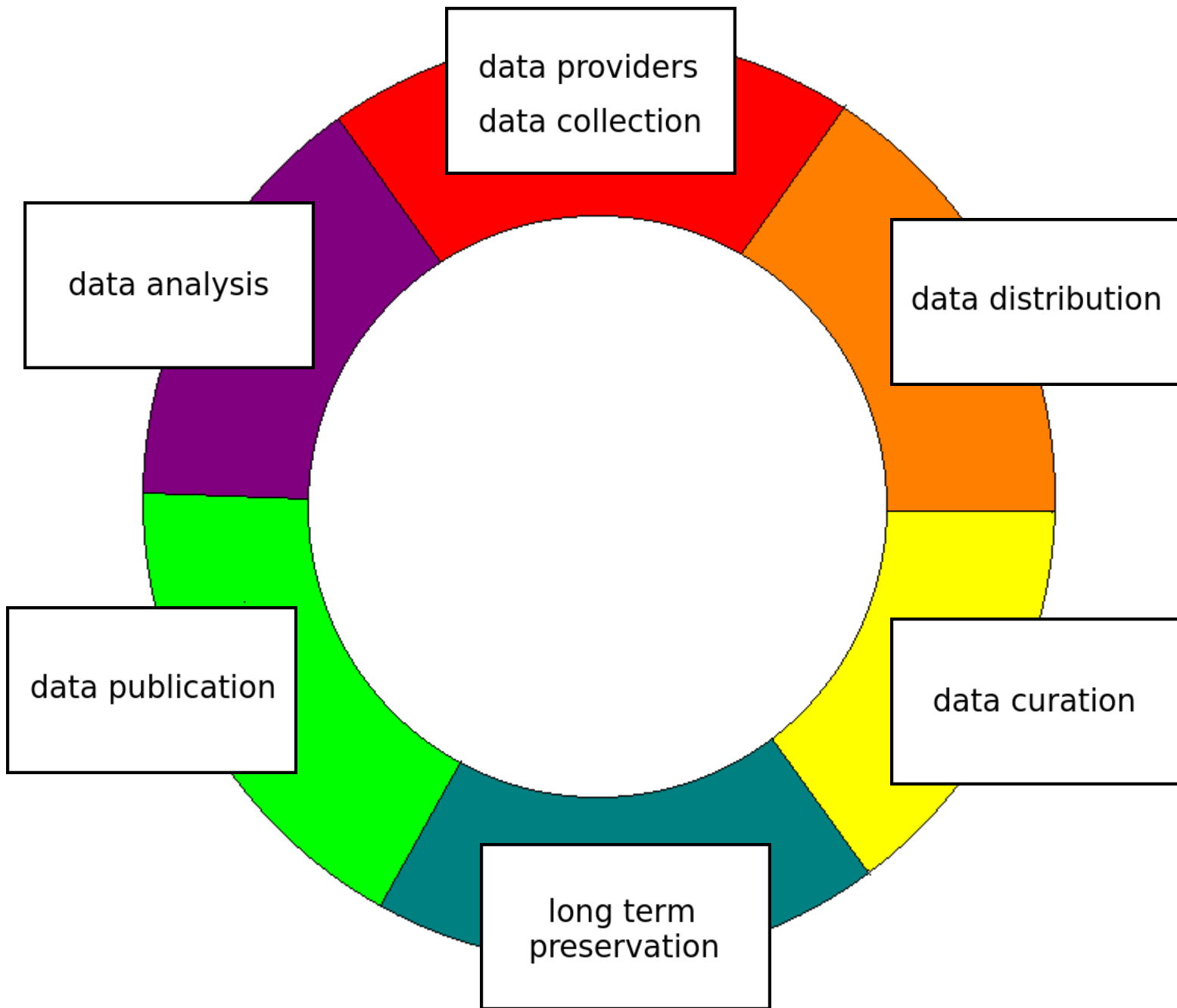
Future developments:

- DOIs for datasets
- access portal
- user space

PROJECTS

expand [+]

Data life cycle



Data providers

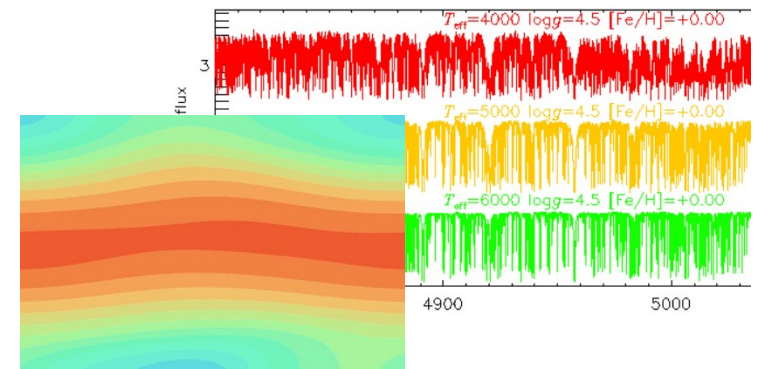


Telescopes:

- Large Binocular Telescope
- Telescopio Nazionale Galileo
- Sardinia Radio Telescope
- Medicina Radio Observatory
- Asiago telescopes
- SVAS telescope

Simulations:

- Exoclimates
- Synthetic spectra



Data formats:

- FITS
- MBFITS
- XML
- txt
- tar.gz



All Sky cameras:

- PRISMA project (fireballs)

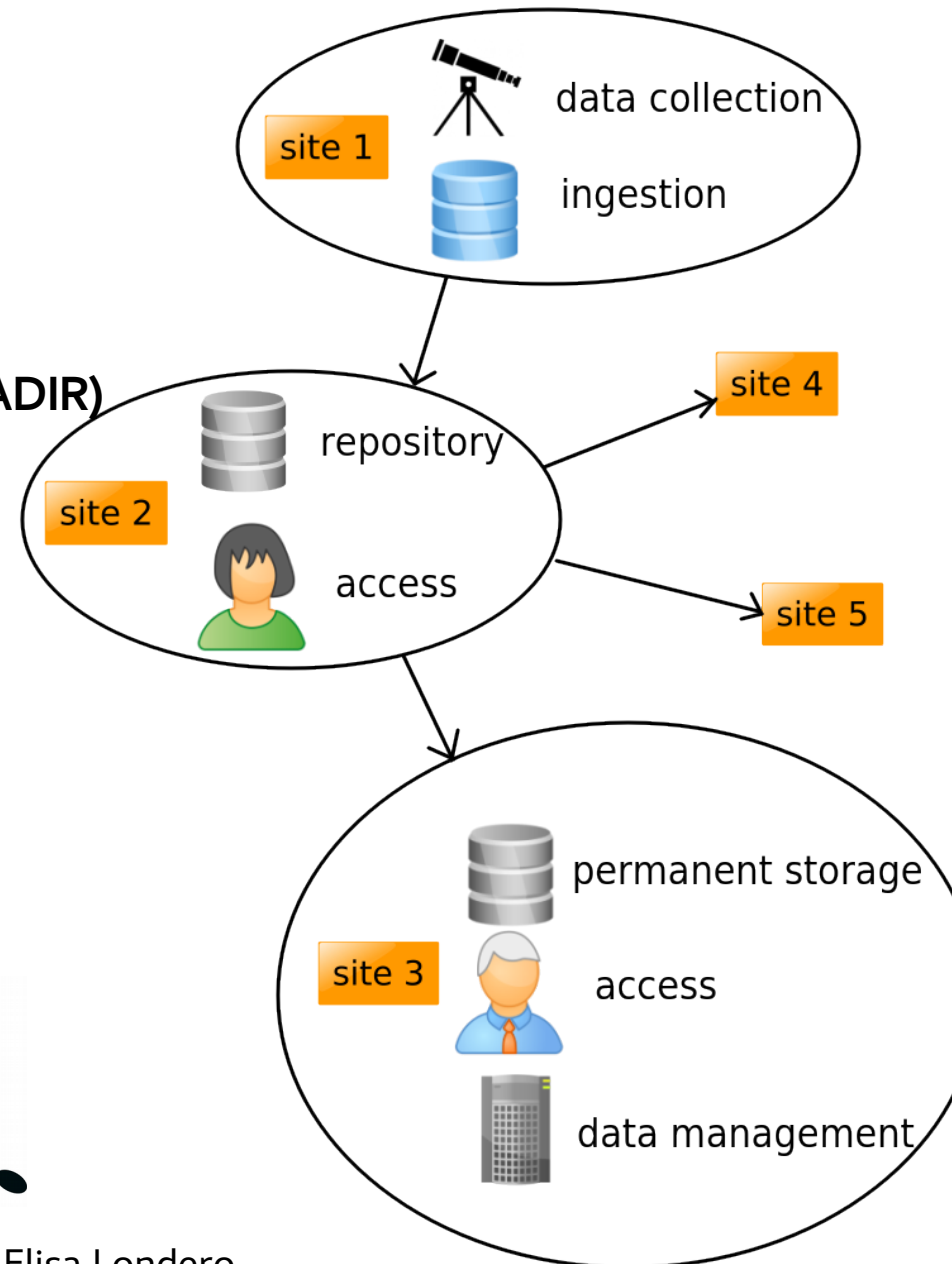
Data distribution

Distributed archive:

- spread among several sites
- remote control and configuration
- monitoring/logging anomalous conditions
- security issues

New Archiving Distributed InfrastructuRe (NADIR)

- flexible
- modular
- configurable
- easily deployable
- efficient monitoring
- event logging
- error tracking
- based on TANGO controls



Important definitions

Data collection: it is the procedure of collecting, measuring and analyzing information from a variety of sources to get a complete and accurate picture of an area of interest in a standardized and Manner.

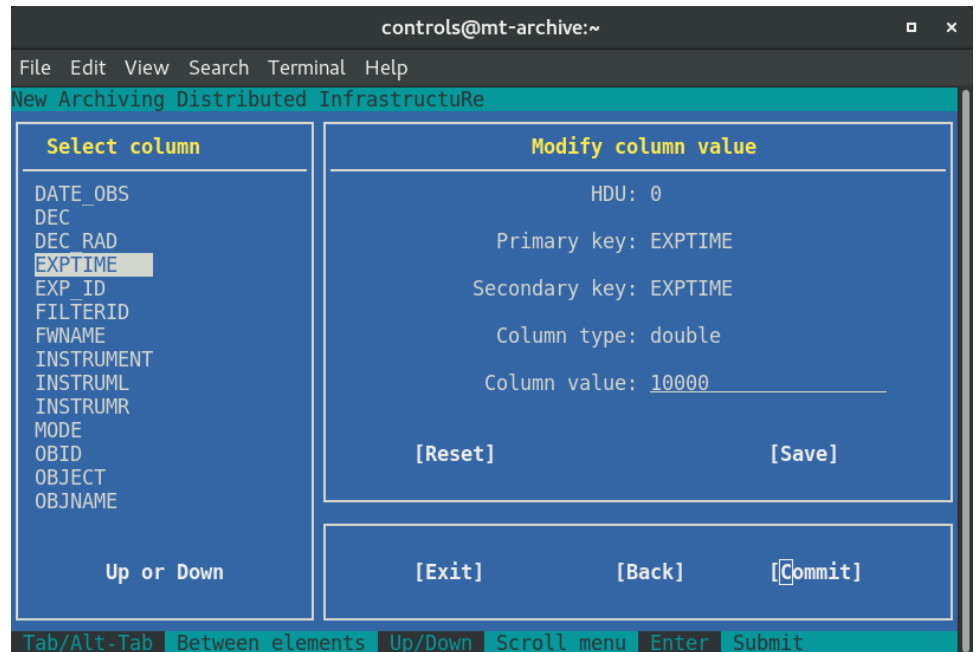
Data model: it shows the logical structure of a database, including the relationships and constraints that determine how data can be stored and accessed.

Dataset : collection of related, discrete items of related data that may be accessed individually or in combination or managed as a whole entity.

Data curation

NADIR Administration Interface:

- amend wrong metadata
- add missing metadata
- updates the DB
- changes propagated to the remote sites
- no second version
- ncurses library → lightweight
- available on data ingestion machines



Checksum calculation:

- calculated on data ingestion machines
- stored in the DB
- re-checked when the file reaches the permanent storage
- detect file corruption through the ingestion process
- easier file integrity checks

Data curation:

- metadata selection by querying the database
- creation of new “clean” tables
- support from scientific staff needed
- final goal: publication of catalogs in the VO

Reference persons: Martina Vicinanza, Marina Nunez and Elisa Londero

Long term preservation

ON DISK:

Hardware IA2 (Trieste):

- 800 TB (100 TB VMs)

Hardware IA2 (other sites):

- 60 TB - IRA
- 60 TB - SRT telescope

Hardware owned by others:

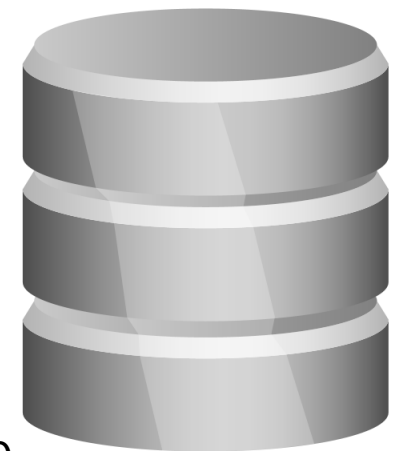
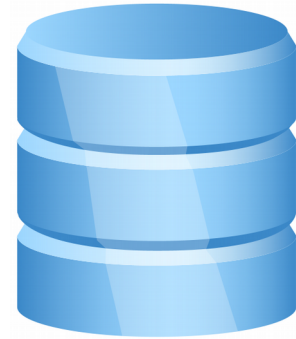
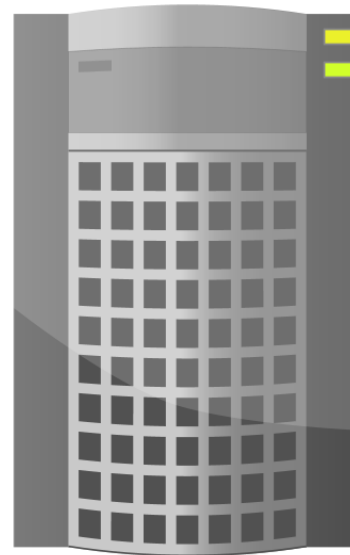
- 40 TB - IRA
- 1 TB - SRT telescope
- 500 GB - Serra la Nave telescope
- 12 TB - LBT telescope
- 500 GB - Asiago telescope

Bandwidth: 10Gbit/s GARR

ON TAPE:

Hardware IA2 (Trieste)

- 200 TB: old tape library. Model: HPE MSL LTO-6 6250
- 8 PB: new tape library (starting from 1.2 PB). Model: IBM TS4500



Data publication: web interfaces

LBT Archive

Name resolver

RA **Dec** **Radius (arcmin)**

File name

Observ. Date **From** **To**

Exposure time (s) **From** **To**

Filter **Instrum.**

Object **Obs. Type**

PI name **Partner** **Program**

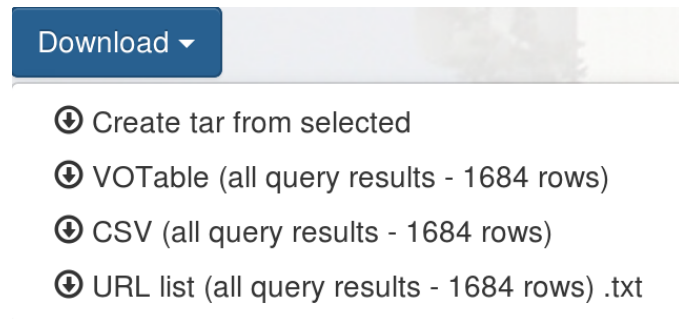
Rows displayed

Total results: 0

Web interfaces: new features

Single sign-on:

- same credentials to access multiple services
- access through Remote Authentication Portal (RAP)
- archive interfaces, Redmine, OwnCloud, Indico, GitLab

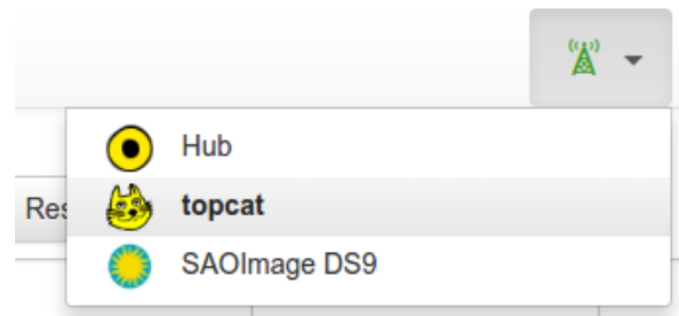
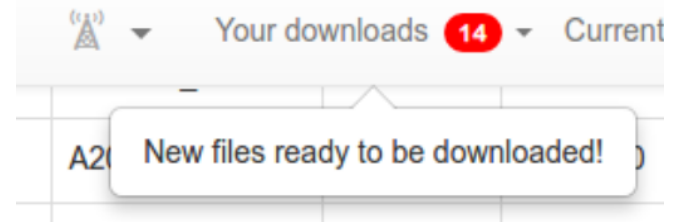


Downloadable products:

- Tar archives
- VO tables
- CSV files
- URL lists

Asynchronous product generation:

- other operations on the portal possible at the same time
- based on Universal Worker Service (IVOA)



SAMP integration:

- Simple Application Messaging Protocol (IVOA)
- send images or metadata to clients

Developer and reference person: Sonia Zorba

Data publication: VO compliant services

IA2 supports the publication in the VO:

- >30 registered services already available
- programmatic access to the archive



IVOA recommendations integrated in the portals:

- Table Access Protocol service (TAP)
- Universal Worker Service (UWS)
- Simple Application Messaging protocol (SAMP)

Further efforts to be VO compliant:

- TAP Schema MANager (TASMAN) → create/edit TAP schema
- transition from MySQL to PostgreSQL → cone search
- development of Simple Image Access Protocol (SIAP v.2), in test phase
- development of a DataLink service

Reference person: Marco Molinaro

Workflow management: yabi

What is yabi:

- workflow management system
- deployed across diverse disciplines
- drag and drop workflow creation
- used for re-reduction purposes
- available for TNG users (HARPN and GIANO-B)
- two pipelines available: drs and gofio



Advantages:

- make proprietary pipelines usable
- always up to date version
- web-based: no need to install or configure
- easy access to reduced data

Future plans:

- RAP integration (single sign-on)
- dataset reingestion (DOI)
- publication in a datasets archive

Reference person: Andrea Bignamini

Other services



- Infrastructure: IA2
- Support: ICT

Reference person: Massimo Sponza



Another step forward: DOIs for datasets



<https://www.ict.inaf.it/idoi>

What are Digital Object Identifiers

- persistent identifiers
- unique identification of an object
- flexible: metadata can change in time

Why to use DOIs

- to identify datasets
- to make data citable
- to give increased visibility to the author
- to foster the **findability** of the resource

Our plan:

- develop a metadata schema
- create a metadata form to be filled in by the author
- mandatory and recommended metadata keys
- develop and maintain a web interface for the data collection and form upload
- make the data publicly available through a web portal

Future perspectives: data access portal

Remote Authentication Portal

Image Credits & Copyright: Colombari/E.Recurt

Account Management



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



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



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



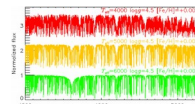
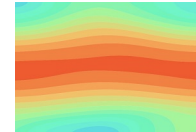
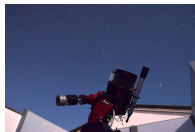
Use the IA2 Logo to Login if you have an account provided by IA2 or self registered

One login...

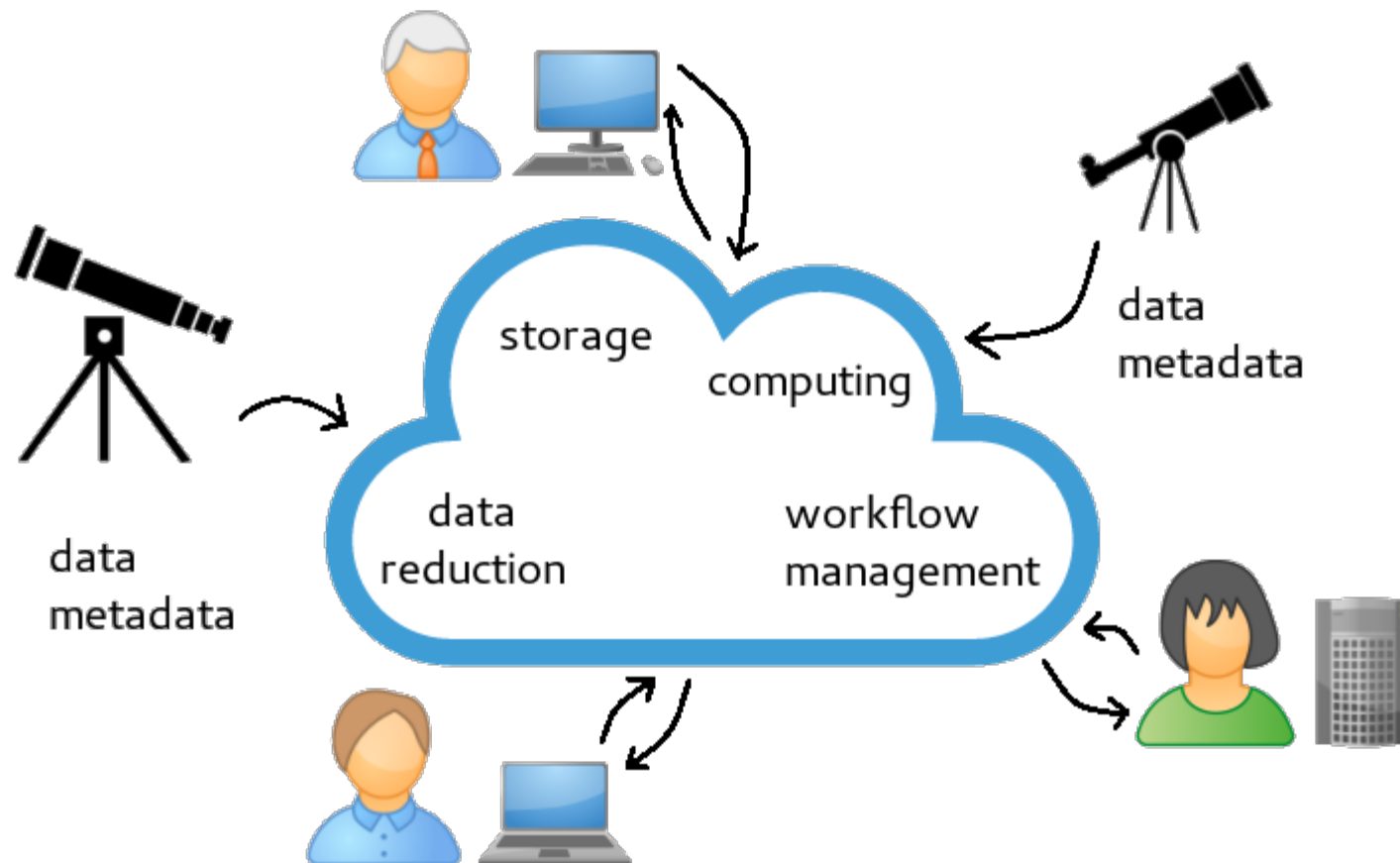
Username

Password

...access to all the resources



Future perspectives: user space



Services to be offered:

- storage capability
- data curation and preservation
- workflow management
- data publication (web portals/VO)
- access to HPC/HTC resources (pilot pr. CHIPP)

Advantages:

- quick access to big data
- bring software to the data
- boost collaboration through data sharing

Conclusions

- Infrastructure is in place and ready to be used
- Storage, web portals, VO services, workflow management, collaborative tools
- Contact us to solve your data problem
- Collaborate with us suggesting new features for our services
- 17-18-19 June 2019: INAF Science Archives and the Big Data Challenge



IA2 *Italian Center for Astronomical Archives*
Centro Italiano Archivi Astronomici

