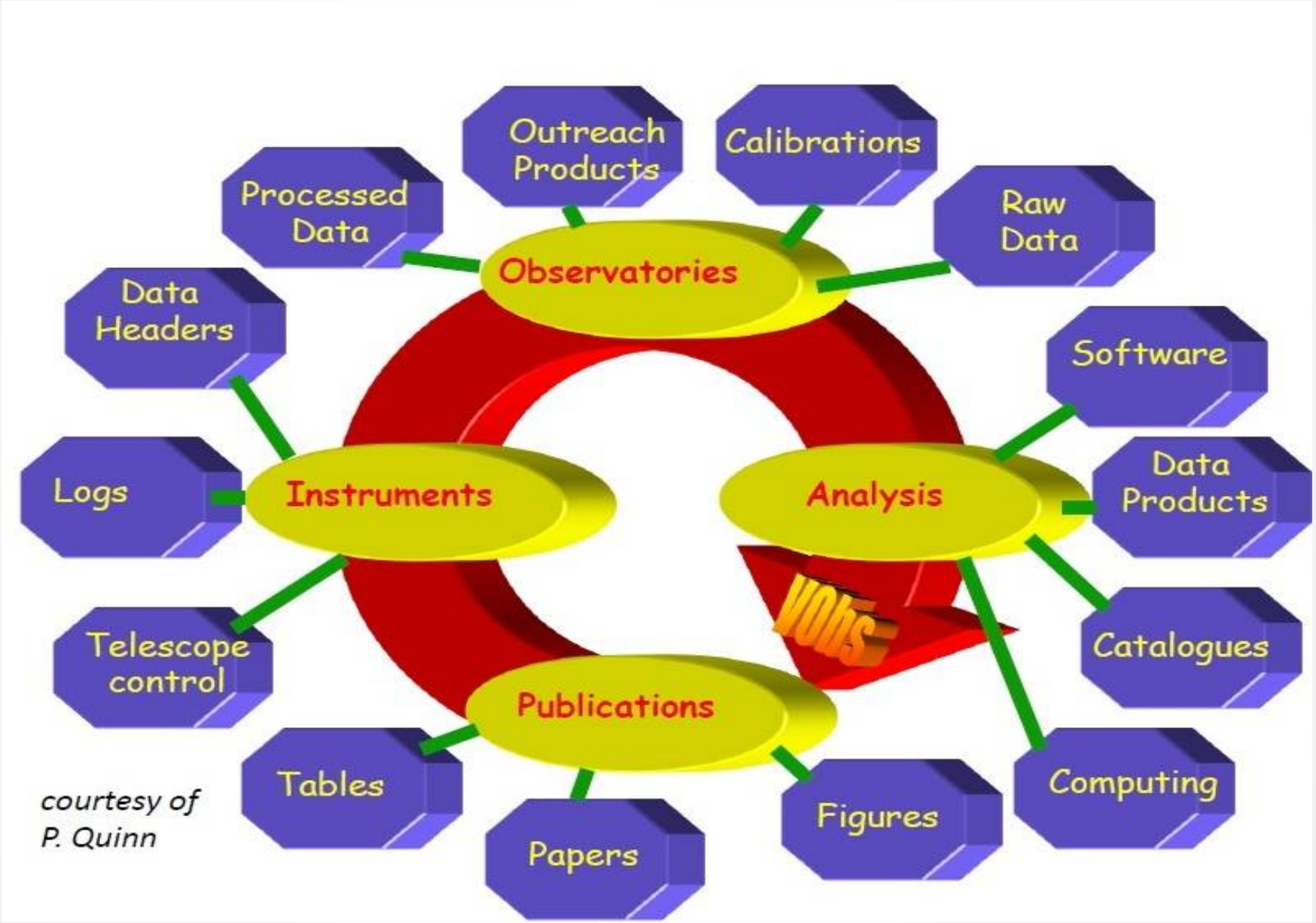
Italian Astronomical Archives - IA2

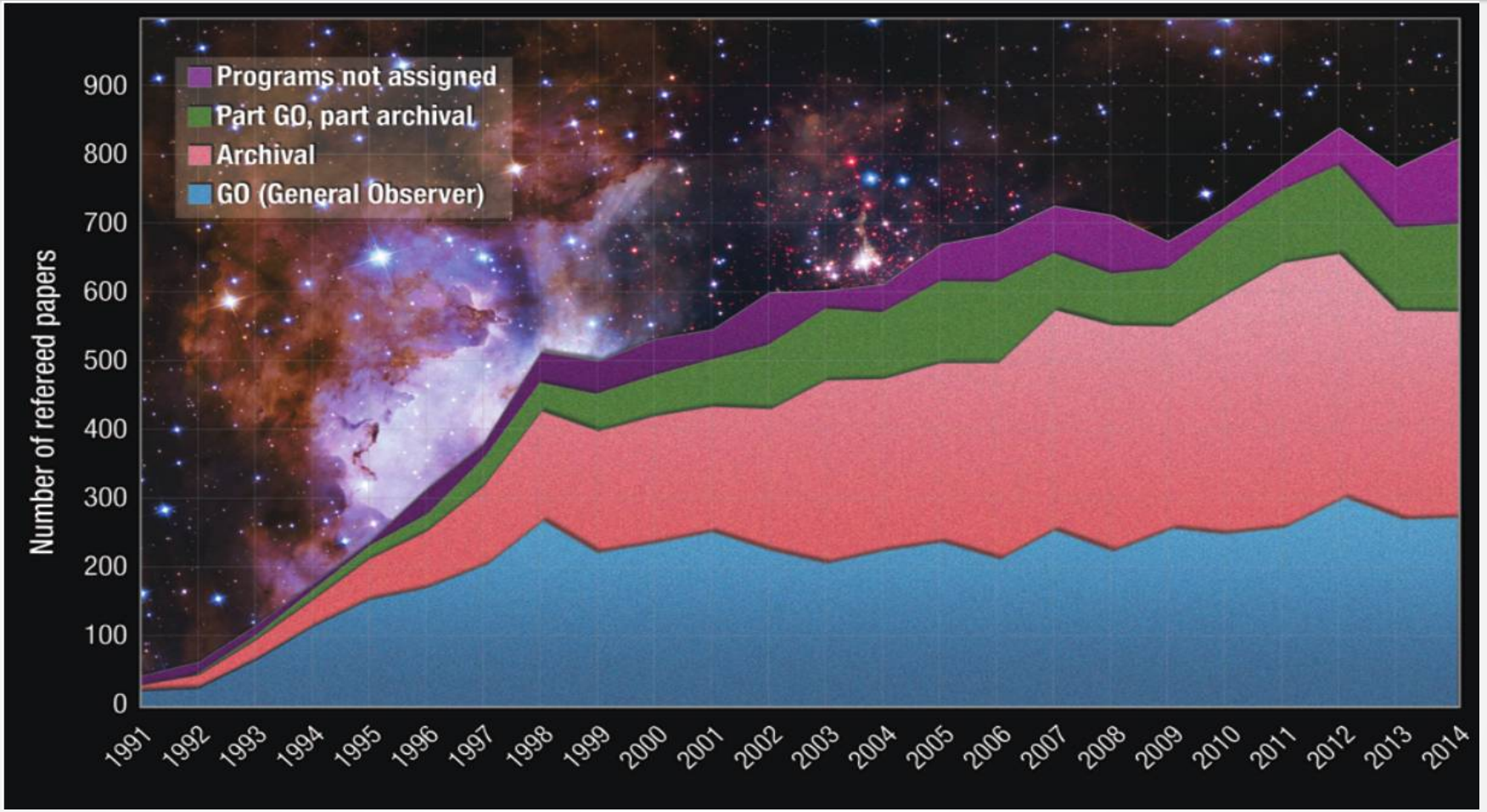
Cristina Knapic

17/11/2016 – ICT Meeting - Trieste

Lifecycle of Astronomical data



Archive importance



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....*"

IA2 components

- IA2
  - Main Goals
  - Hosted Services
    - NADIR
      - Properties and configurations
    - Workflow (Yabi)
    - Collaborative tools (Twiki, OwnCloud...)
  - Observatory Archives (LBT, TNG, Asiago, Radio....)
  - VO compliant services (SIAP, SSAP, TAP, Cone..)
  - Some examples
  - Projects hosted @ IA2 (GAPS, Pessto, ..)
- Future perspectives
  - Single Sign-On
  - VOSpace, user space

IA2 Main Goals

Centro Italiano Archivi Astronomici (IA2) main goals are :

- archiving systems
- safety
- data curation and preservation
- distribution over several geographical distributed sites
- providing services and tools (TWiki, work-flow, etc..)
- data publication in the VO

of ***Astronomical Data***

IA2 manages data of several PROJECTS.
Mainly they come from:

- TELESCOPES (raw; INAF ground based)
- SURVEYS (raw and/or calibrated)
- SIMULATIONS (ITVO)
- EU Projects (GENIUS, VIALACTEA, INDIGO)

IA2 is part of INAF-ICT

IA2 Services

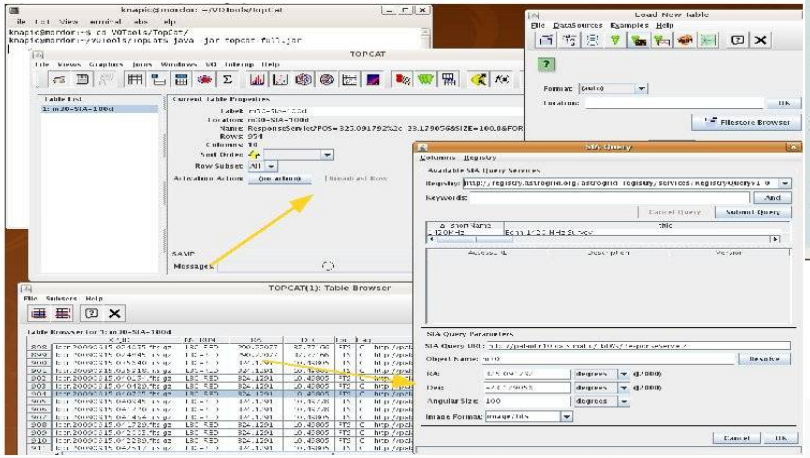
ownCloud

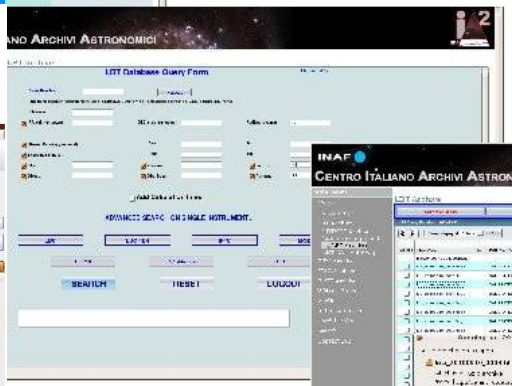
collaborate with TWiki

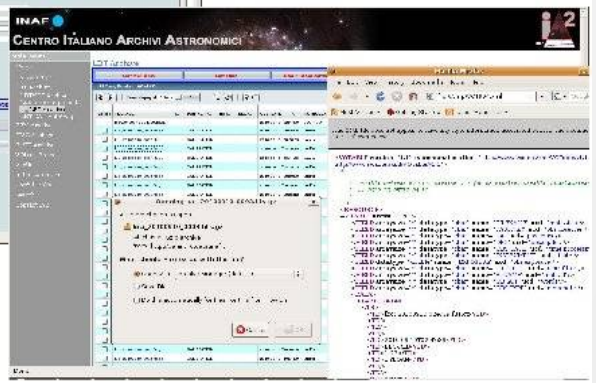
yabi

REDMINE
flexible project management

VO Service & TopCat



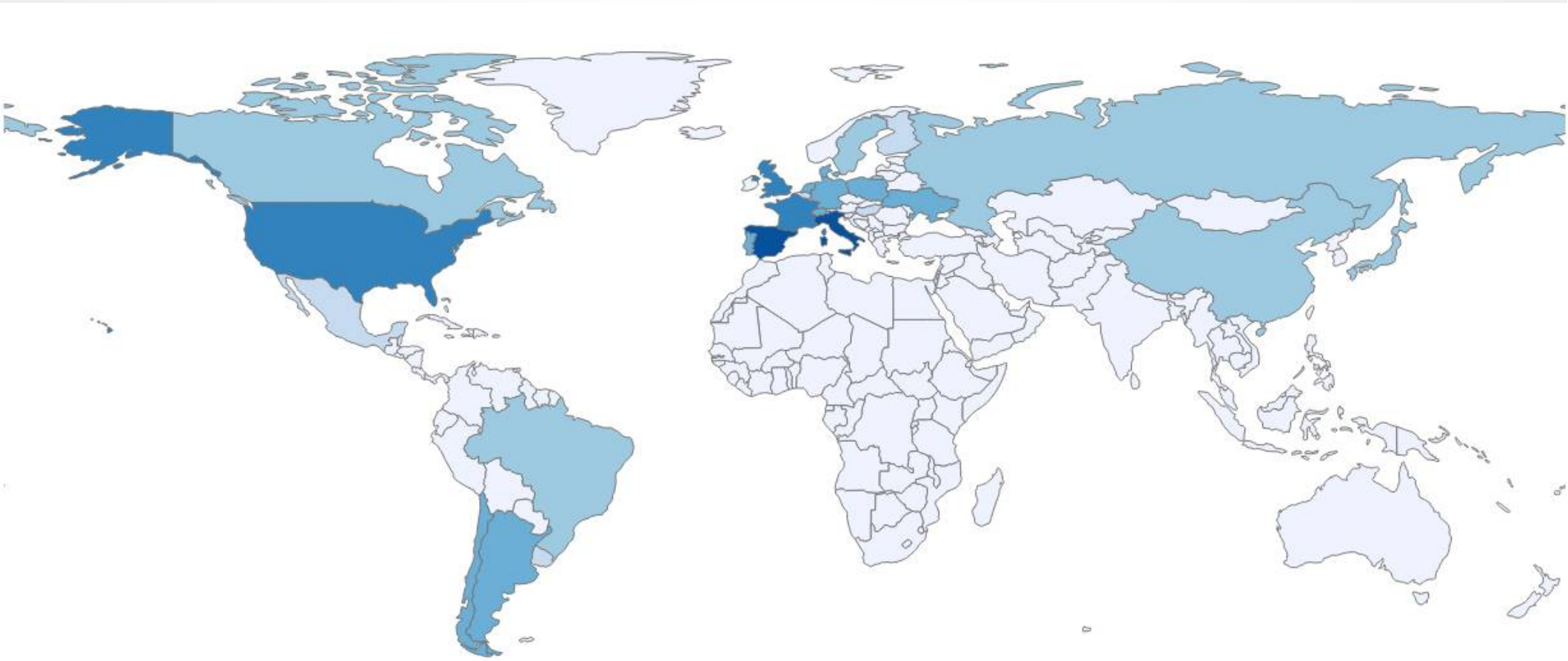




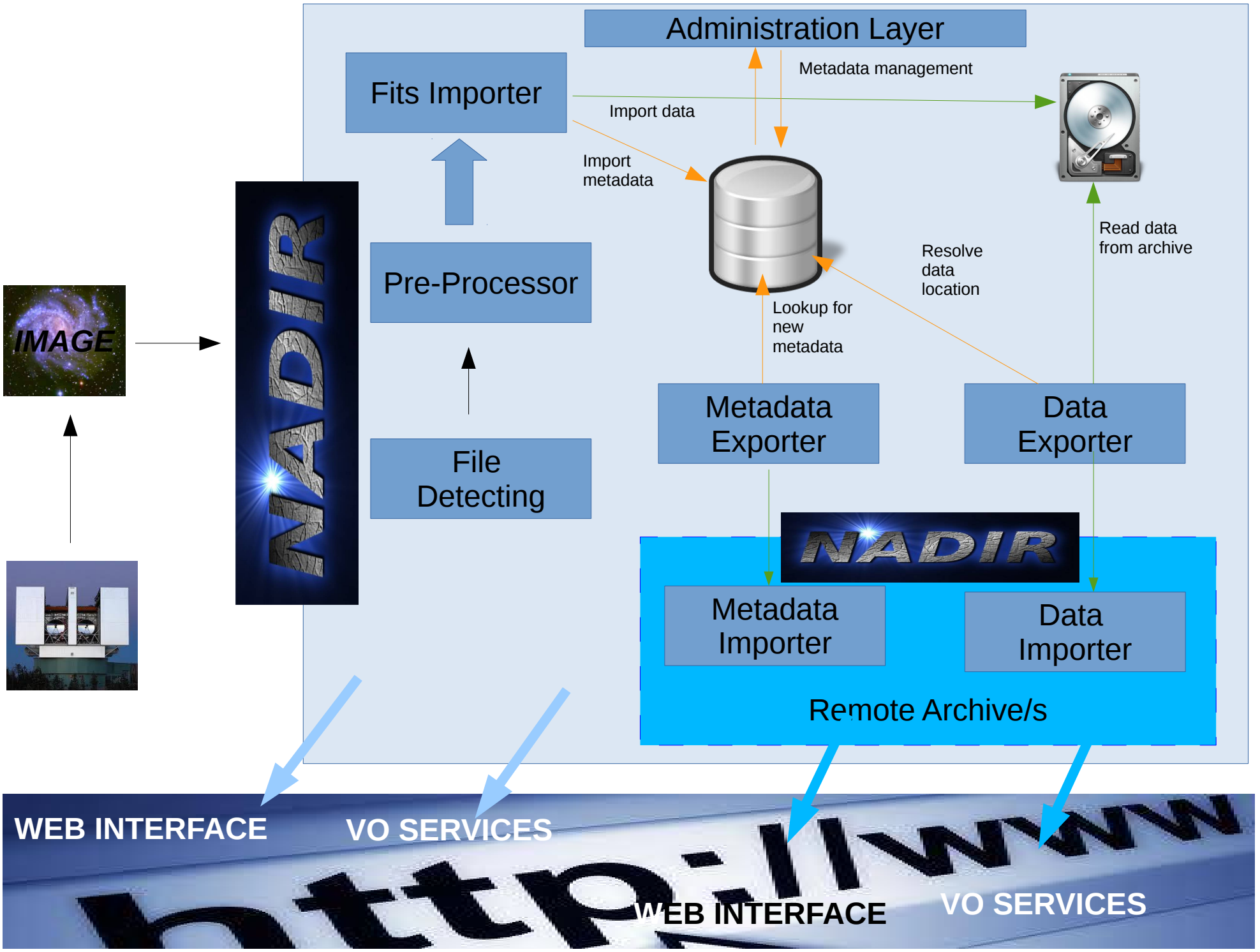
IA2 Services

Service	Protocol	Number of Accesses				
		Day	Week	Month	Trimester	Year
		Ok	Ok	Ok	Ok	Ok
aaotap	tap	0	0	0	0	4
Asiago Archive	web	7 • **Downloaded files**: 0	49 • **Downloaded files**: 187	79 • **Downloaded files**: 982	454 • **Downloaded files**: 11925	1221 • **Downloaded files**: 25355
DAbrusco2009QSO	cone	3	28	107	361	1841
DAbrusco2009QSORedshift	cone	2	24	107	352	1869
epntap	tap	38	260	1394	2622	19190
Laurino2010	cone	2	24	97	347	1551
Laurino2011_optUV_QSO	cone	2	23	104	370	2004
LBT Archive	web	74 • **Downloaded files**: 24	502 • **Downloaded files**: 162	3972 • **Downloaded files**: 1275	8812 • **Downloaded files**: 2783	31619 • **Downloaded files**: 9478
TNG Archive	web	4 • **Downloaded files**: 0	62 • **Downloaded files**: 3376	412 • **Downloaded files**: 6587	1729 • **Downloaded files**: 32159	7918 • **Downloaded files**: 130648
TNG-LRS	siap	0	0	17	339	2048
TNG-NICS	siap	0	0	45	376	2842
TNG-OIG	siap	0	0	17	337	1934
Planck_ERCSC_f353	cone	1	17	76	298	1630
Planck_ERCSC_f545	cone	1	17	75	292	1524
Planck_ERCSC_f857	cone	1	17	84	304	1719
Planck_ESZ	cone	1	17	83	314	1553
SampleCone	cone	0	0	0	0	0
SampleSIAP	siap	0	0	0	0	0
SVAS SIA	siap	2	16	139	466	3400
Tirgo	siap	2	16	137	457	4528
TNG Archive	web	4 • **Downloaded files**: 0	62 • **Downloaded files**: 3376	412 • **Downloaded files**: 6587	1729 • **Downloaded files**: 32159	7918 • **Downloaded files**: 130648
TNG-LRS	siap	0	0	17	339	2048
TNG-NICS	siap	0	0	45	376	2842
TNG-OIG	siap	0	0	17	337	1934
Vape	web	0	0	1	1	32
Vialactea Cutout	web	0	0	0	0	36167
Vialactea Search	web	0	0	0	0	34451
VIPERS SSAP	ssap	56	345	2703	3327	5137
vlkb	tap	0	0	0	286	4351

TNG Archive 2016 Map of Accesses

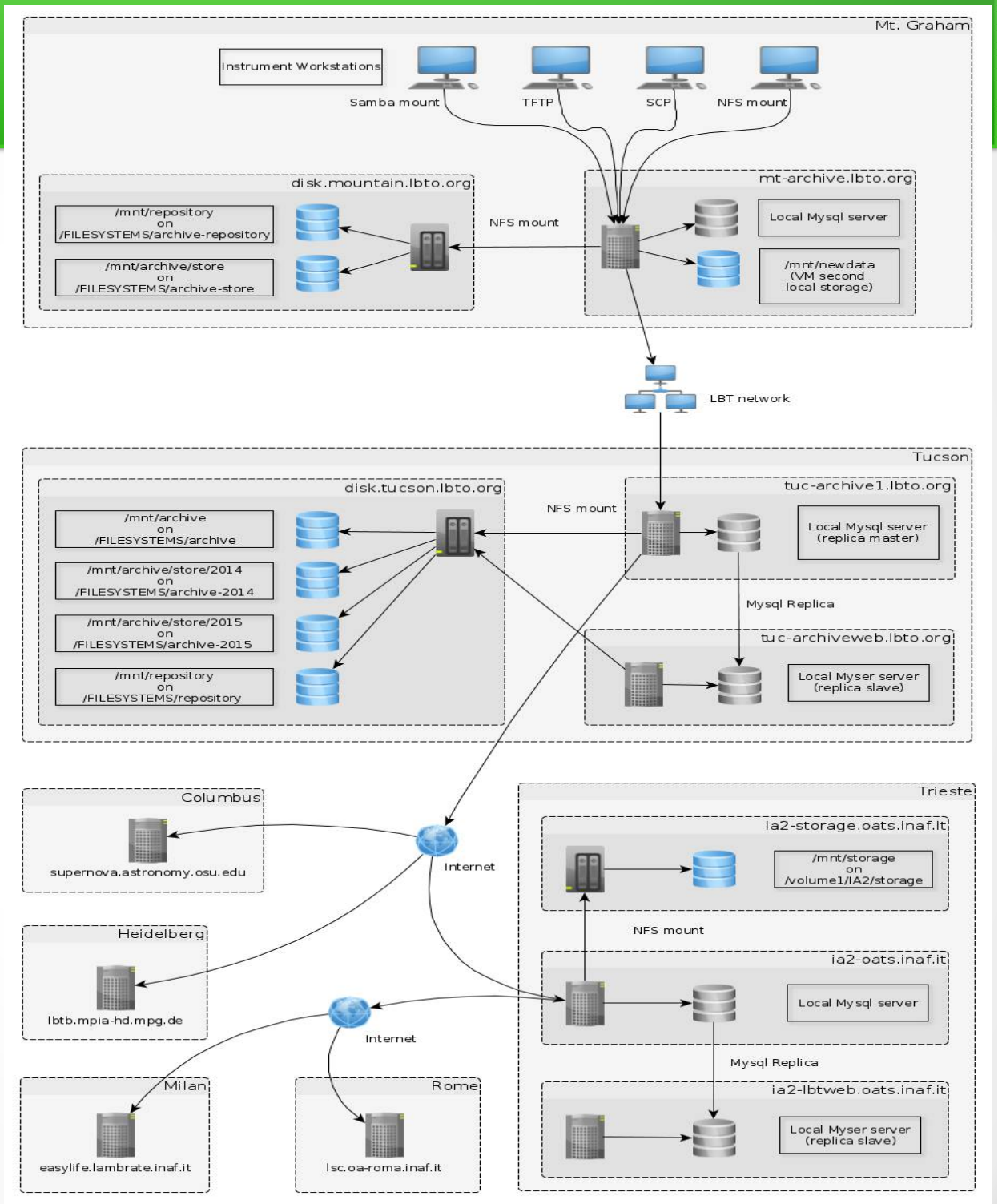


ITA: 2879	NLD: 161	DNK: 71	RUS: 26	URY: 8	FIN: 3	LTU: 1
ESP: 1367	CHE: 148	PRT: 70	SWE: 24	MEX: 6	ROU: 2	IND: 1
GBR: 517	ARG: 94	POL: 68	CHN: 24	BEL: 6	AUS: 2	ARM: 1
USA: 488	DEU: 92	UKR: 61	CAN: 17	ISR: 5	TWN: 1	
FRA: 288	CHL: 76	BRA: 28	JPN: 14	HUN: 5	MUS: 1	

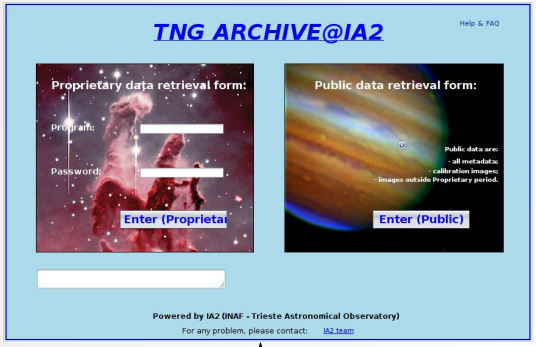

Administration Layer
Fits Importer
Metadata management
Import data
Import metadata
Read data from archive
Resolve data location
Lookup for new metadata
IMAGE
NADIR
Pre-Processor
File Detecting
Metadata Exporter
Data Exporter
NADIR
Metadata Importer
Data Importer
Remote Archive/s
WEB INTERFACE
VO SERVICES
WEB INTERFACE
VO SERVICES

LBT Data flow

- Source point of information:
  - LBT telescope site – temporary archive/repository
  - Ingestion site for standard FITS files of raw, technical or guiding data
  - Limited storage capacity
  - Full metadata DB since the first operations.
- Master archive in Tucson:
  - Full archive (metadata and data) for all partners and all data
  - In the same place, first data reduction for some data, data quality evaluation foreseen in the near future (interaction with operation team)
  - Data distribution to partners both to other LBT distributed Archive sites and to externals (data delivery, not archive)
- Distributed sites (MPIA – Heidelberg; OATs – Trieste)
  - Partial archives with proprietary data
  - Data reduction with proprietary pipelines
- Data delivery to other archives (Ohio State University – Columbus; INAF – IASF – Lambrate (Mi) ; INAF – OARm - Roma)



TNG Data Flow



TNG

Average data rate
(raw+extra) 5GB/day
(max in 2015 50GB/day)

RAW

Reduced &
Ancillary

natzgul2

Production
Front End
Database

New Front End
Database

Update every 30'

Ingested
every 20'

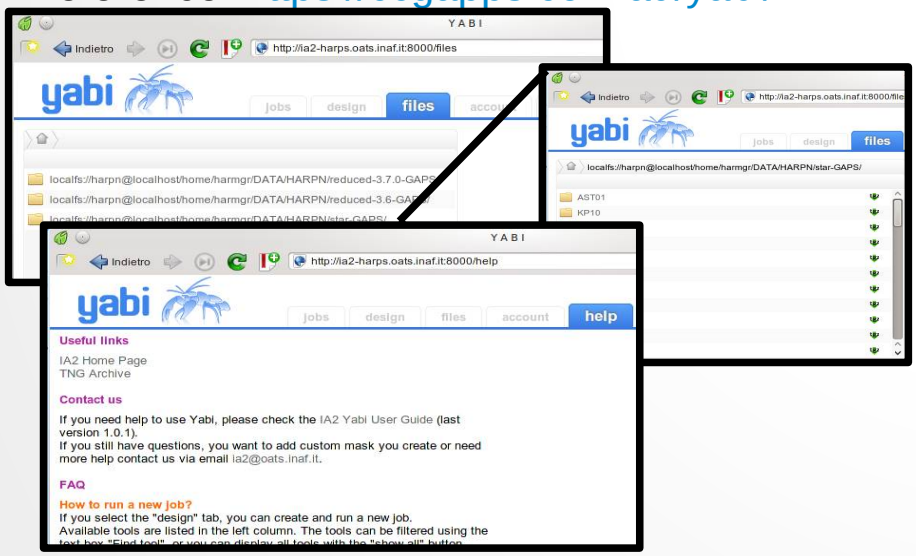
TNG Archive
Back End

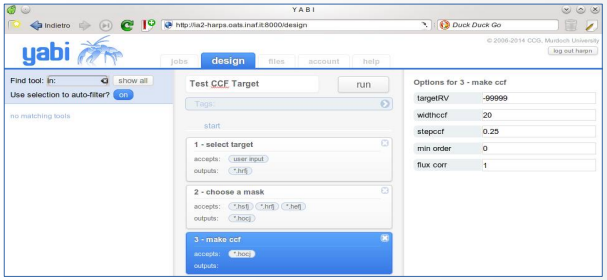
NADIR

Ingested once a
day at 14:00 CET

IA2 Workflow system : YABI

- Yabi is a 3-tier application stack to provide users with an intuitive, easy to use, abstraction of compute and data environments. Developed at the Centre for Comparative Genomics and Murdoch University, Yabi has been deployed across a diverse set of scientific disciplines and high performance computing environments.
- Yabi is a workflow system that allows accredited users to run part or full HARPN pipeline
- Yabi has a few key features:
  - simplified web based access to High Performance Computing
  - easy tool addition and maintenance
  - handling of disparate compute and storage resouces ie. PBSPro, SGE, Torque, SSH, SFTP, Amazon S3, Swift
  - easy and powerful workflow creation environment
- Reference: https://ccgapps.com.au/yabi/





Yabi is running since 2014-03-23 at IA2 with 105 total number of users. Statistics:
Total size of workflow outputs 778 GB

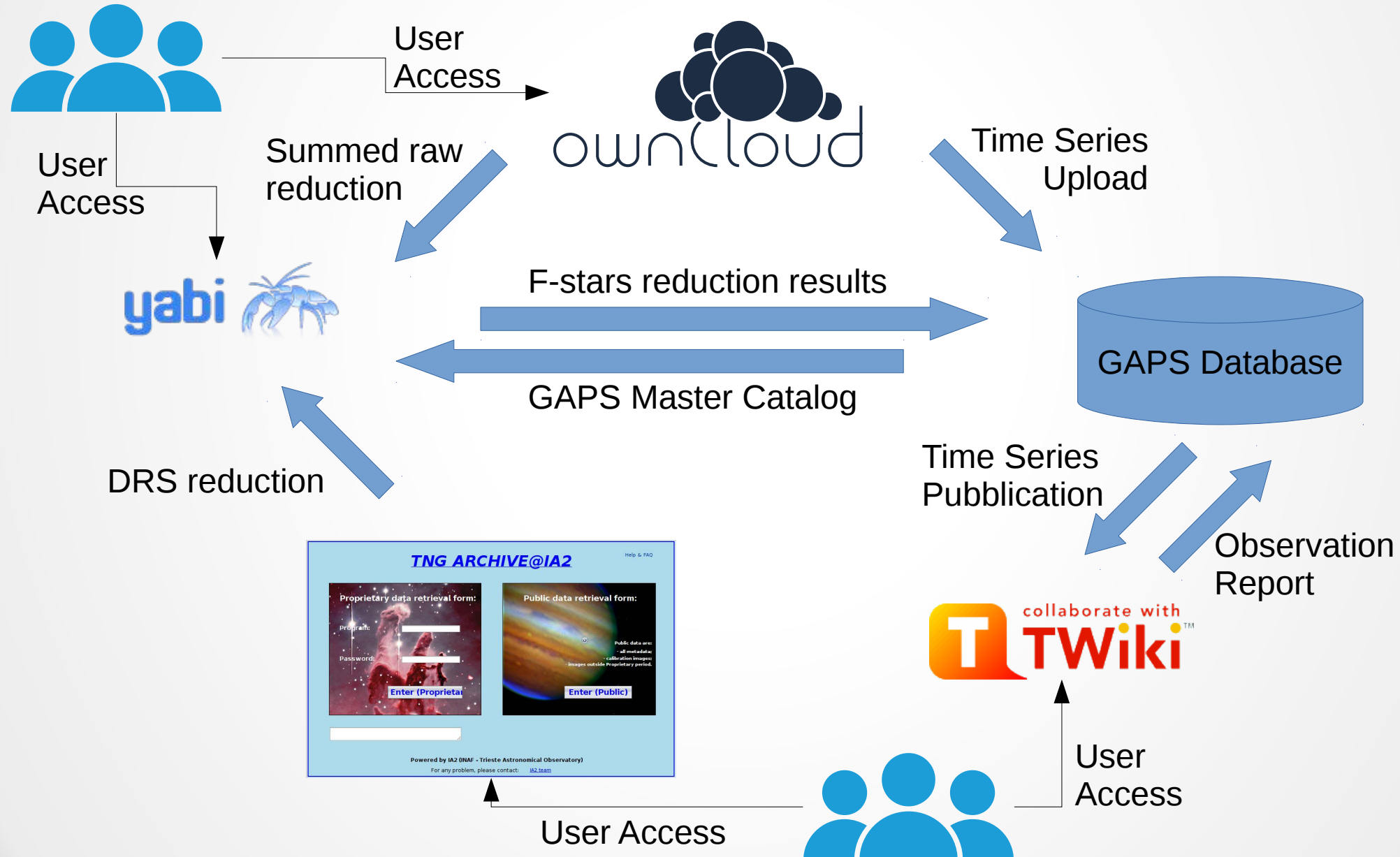
Description	Nr. Workflows	Nr. Jobs
Total	4019	11000
Last Year	1200	3217
Last Month	117	298
Last week	5	15

IA2 - TWiki

- TWiki is a Perl-based structured wiki application, typically used to run a collaboration platform, knowledge or document management system
- Number of topics: **395** static pages + **381** observational reports
- Number of users: **61** TWiki account in GAPS
- Form to submit observational reports automatically build → Store the Obsreport into a DB

collaborate with TWiki™

GAPS@IA2


User Access
User Access
Summed raw reduction
owncloud
Time Series Upload
yabi
F-stars reduction results
GAPS Database
GAPS Master Catalog
DRS reduction
Time Series Pubblication
Observation Report
TNG ARCHIVE@IA2
Help & FAQ
Proprietary data retrieval form:
Program:
Password:
Enter (Proprietar
Public data retrieval form:
Public data are:
- all metadata:
- calibration images:
- images outside Proprietary period.
Enter (Public)
Powered by IA2 (INAF - Trieste Astronomical Observatory)
For any problem, please contact: IA2 team
collaborate with TWiki
User Access
User Access

Easy access to your data

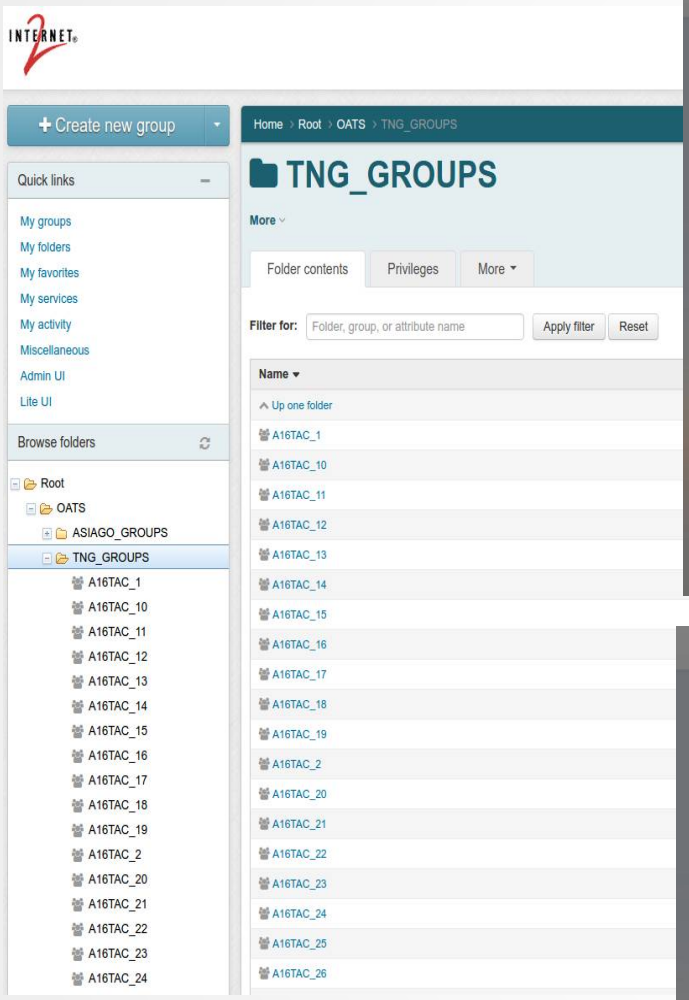
Additional tool for TNG Observatory will be a **Grouping** software. It will allow a contributed management and distribution of programs credentials and privileges. It is also the first step to the SSO (Single Sign On) technology adoption (under development).

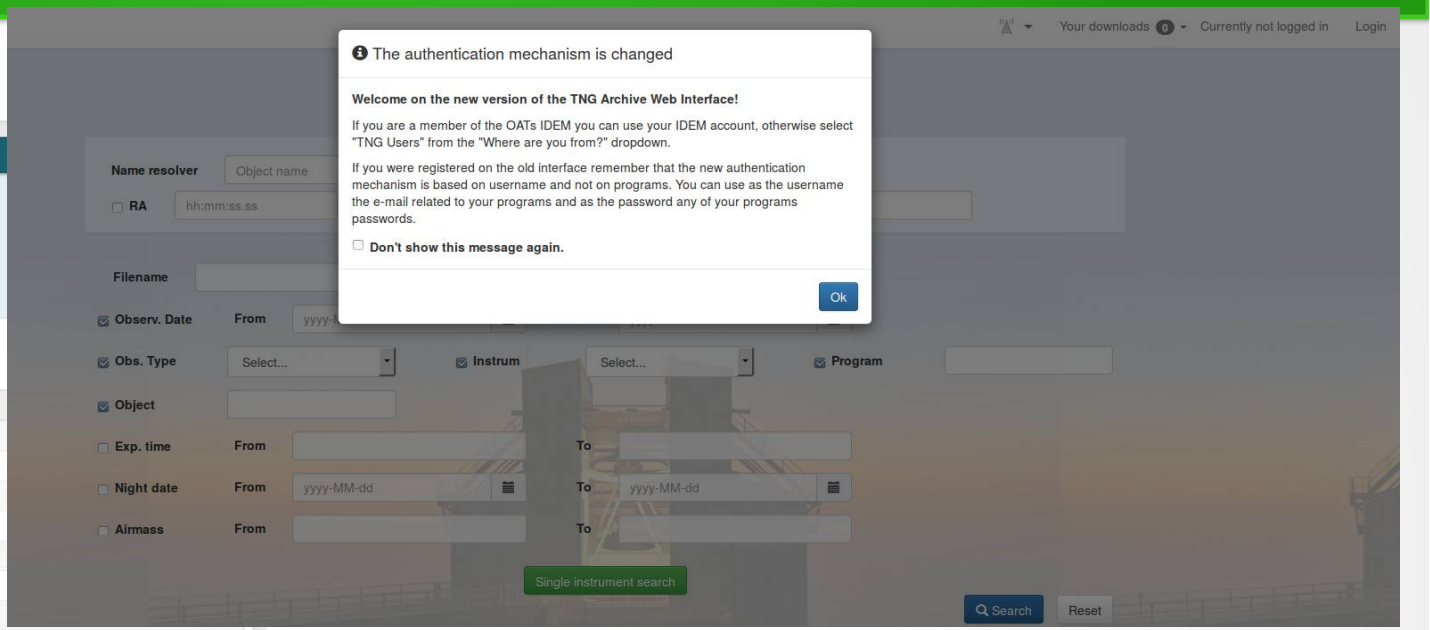
Grouper™

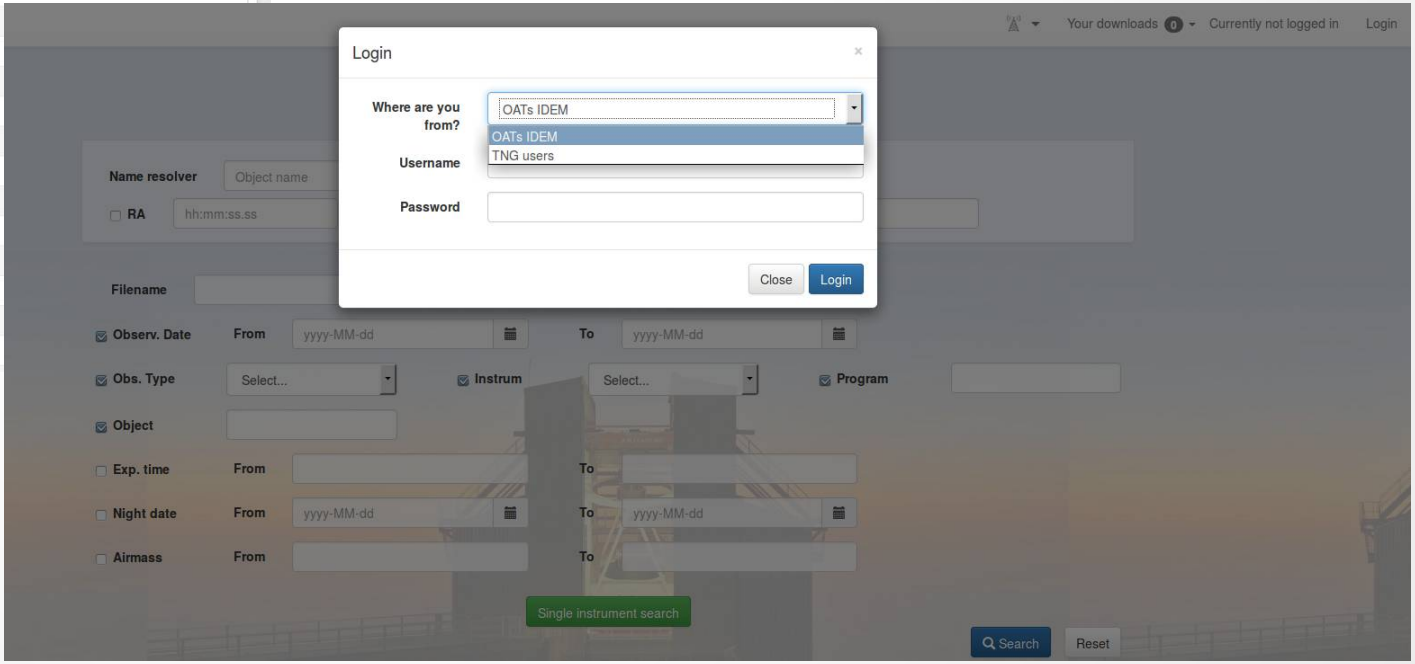
Grouper is an enterprise access management system designed for the highly distributed management environment and heterogeneous information technology environment common to Universities. It could work on top of Identity managers like IDEM (for the Italian counterpart) and is compatible with SHIBBOLET.

DOI (Digital Object Identifier):
DOI is a character string used to uniquely identify an object such as an electronic document. The ***DOI*** for a document is ***permanent***, whereas its location and other meta-data may change. Referring to an on-line document by its DOI provides more stable linking than simply referring to it by its URL, because if its URL changes, the publisher need only update the meta-data for the DOI to link to the new URL.

IA2 Archives – new approach

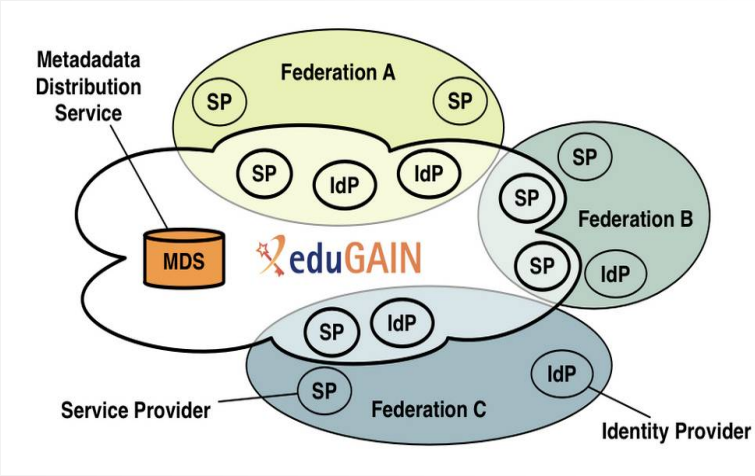






SSO activities

Authentication (SAML based)

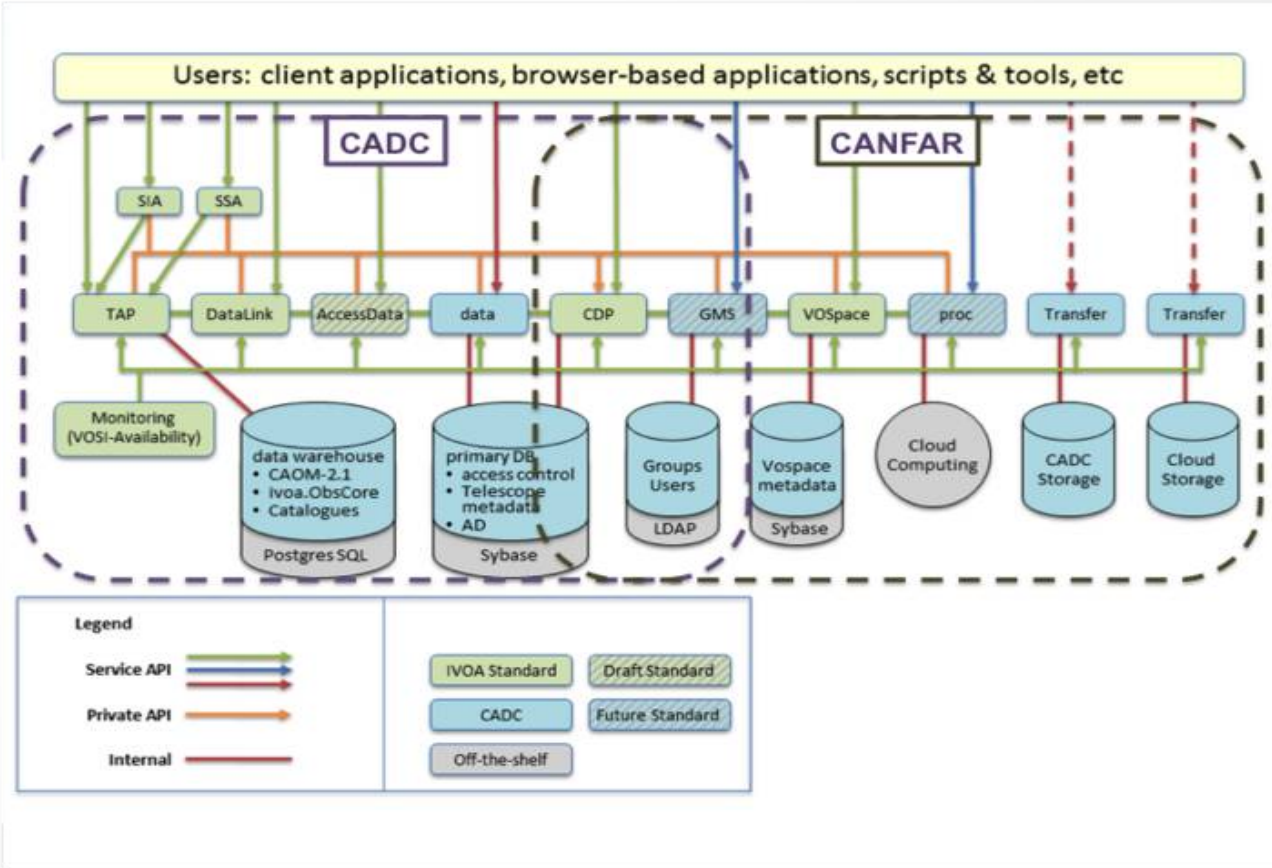


- IA2 IdP is based on a replica of the Bologna LDAP that provide Digital Identities to INAF (IDEM);
- IA2 integrate the LDAP with non IDEM identities that use to one or more IA2 Telescopes;

Authorization Internet2 application based / VO compliant based

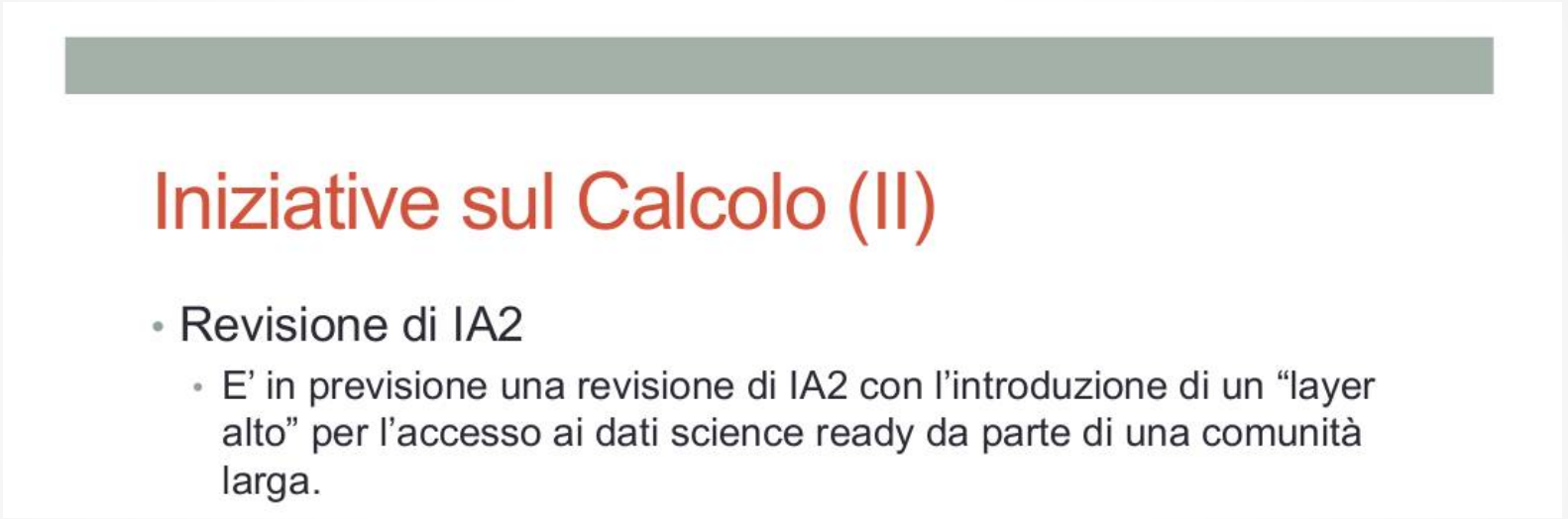
Grouper™

Future plans of interoperability:



Future plans

Estratto dal “OdG Coll Dir del 5 ottobre 16_allegato 2) verb. 32.pdf”



Iniziative sul Calcolo (II)

- Revisione di IA2
  - E’ in previsione una revisione di IA2 con l’introduzione di un “layer alto” per l’accesso ai dati science ready da parte di una comunità larga.

- Improve the number of supported Telescopes;
- Increase the number of services offered;
- Increase the number of supported data models;
- Grant access to all the offered services via SSO;
- Adopt a VOSpace compliant service to allow user space utilization;
- **Improve the user experience in order to stimulate the Astronomical Data life cycle**.

Needs

- MAN POWER!!! Too many services to be handled with a too restrict number of skilled persons;
- Professionals;
- Endorsement of activities;
- Cooperation with other data centers;

Conclusions

- Target of IA2:
  - Offer Archival services and maintain them
- Target of the archive user:
  - Use archive services
  - Have his own User Space where to make science!
  - Reporting problems
  - Suggest upgrade/necessities