IA2 : Archives & VO

Cristina Knapic

Trieste 01/04/2014

Outline

- Current goals & projects;

- Standards followed;

- LBT Archive architecture

- New project: NADIR

- NADIR Data Model

- NADIR Configurability

- A Science Gateway

- VO services @ IA2

- GAPS experience

- Data curation, access and DOI

- IA2 in data management

Centro Italiano Archivi Astronomici (IA2) main goals are :

- provide Archiving systems;
- data availability;
- data curation and preservation;
- distribution over several geographical 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)

Trieste 01/04/2014

IA2 Projects

Current status:

Project Name	Project Type	Data Type	Data Amount Archive	Data Amount DB	UI	VO	1yr User Access UI	1yr User Access VO
LBT ✓✓✓✓	Telescope	image / spectra	10TB	320MB	✓	SIA	2.5K	47k
TNG ✓✓✓✓	Telescope	image / spectra	3TB	1.2GB	✓	SIA	12k	55k
GAPS	Survey (TNG)	image / spectra	850MB			✓(TNG), TWiki, Yabi		
Asiago ✓✓✓ < 1yr	Telescopes	image / spectra	400GB	290MB	✓	(TAP)	700	(70)
PESSTO	Survey (NTT)	image / spectra	7GB	1MB	✓			
hosted services								
WGE SDSS redshifts	data mining	catalogue		8GB		SCS / TAP		11k / 170
Planck	early release	catalogue		8MB		SCS		49k
TIRGO	IR camera	image	(Arcetri)	100MB		SIA		13k
WINGS < 1yr	Survey	image / catalogue	26GB	700MB		SIA / SCS		11k / 10k
INAF-IAPS EPN < 1yr	dust particles (NASA)	catalogue		2MB		TAP		1k
ITVO	Theoretical Simulation	mixed	1TB	1MB	✓		70	

Under development:

Project Name	Project Type	Data Type	Data Amount	UI	VO	1yr User Access
RADIO	Array/antennas	Images/spectra		✓	✓	
SKA.TM.OBSMGT	Observing tools	Meta-data				

Trieste 01/04/2014

Standards followed

IA2 at the moment manage Astronomical Data mainly in FITS format (Flexible Image Transport System) for images and spectra and GADGET2 for simulations. Current projects implies also the management of Hierarchical FITS and MBFITS formats. IA2 host also survey pipeline and related products and provides support to a survey dedicated TWiki.
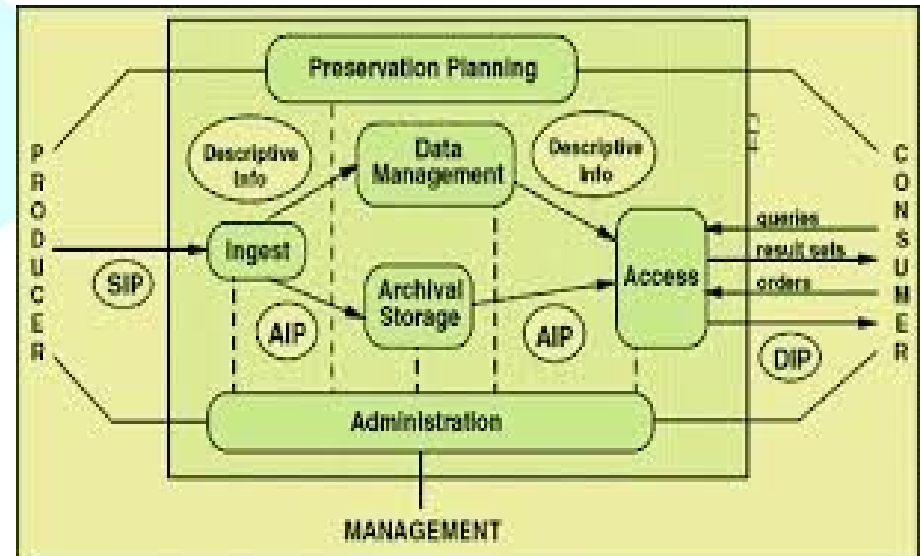
- **IA2 archives follows the directives of OAIS (Open Archival Information System) standard**:
  - ➔ data are logically split in **data descriptors** and **data content**.

Data descriptors

Data Content

- **IA2 as a service follows the IVOA standards directives** and expose several VO services and several User Interfaces VO compliant.
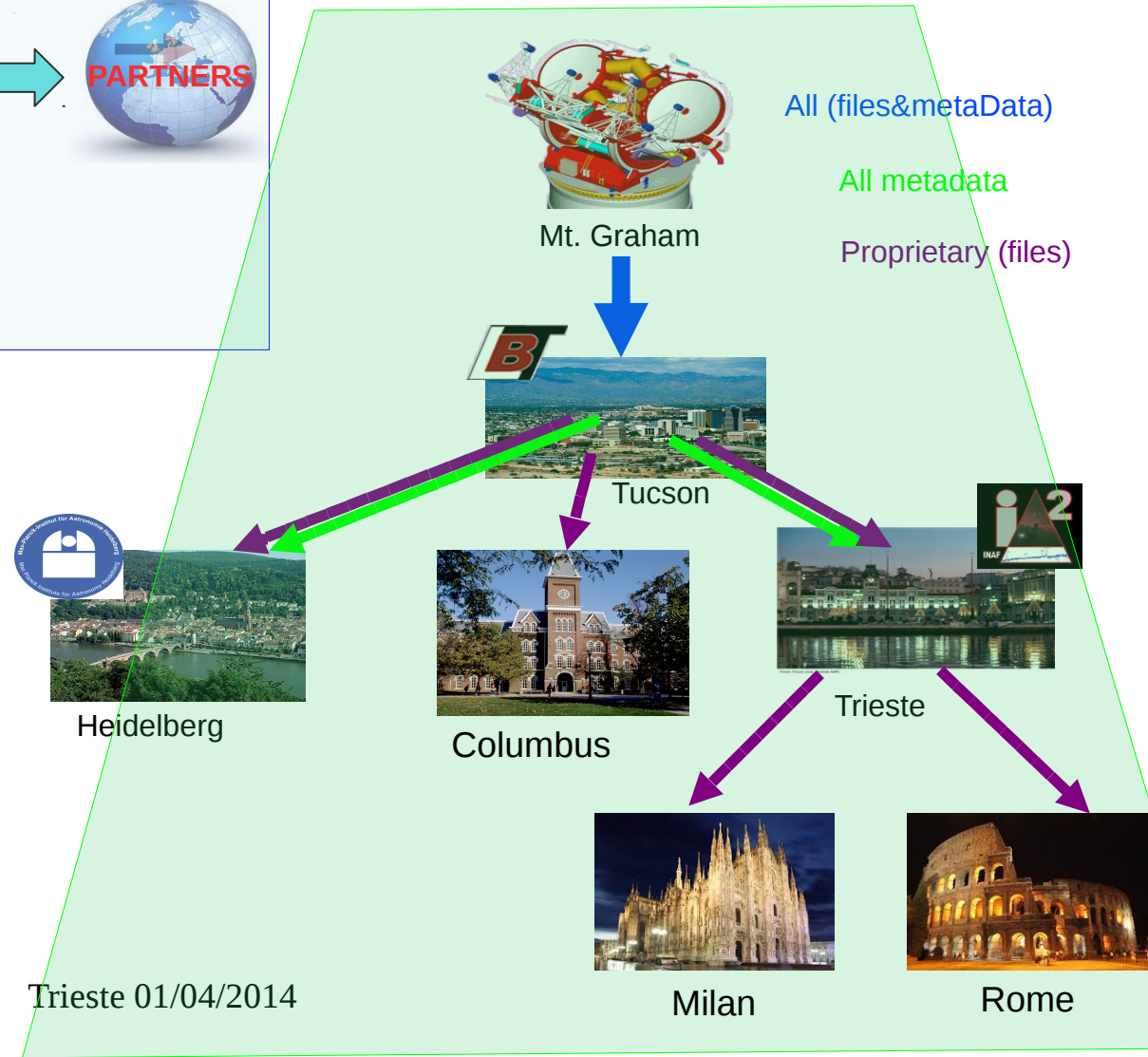
Trieste 01/04/2014

LBT @ IA2
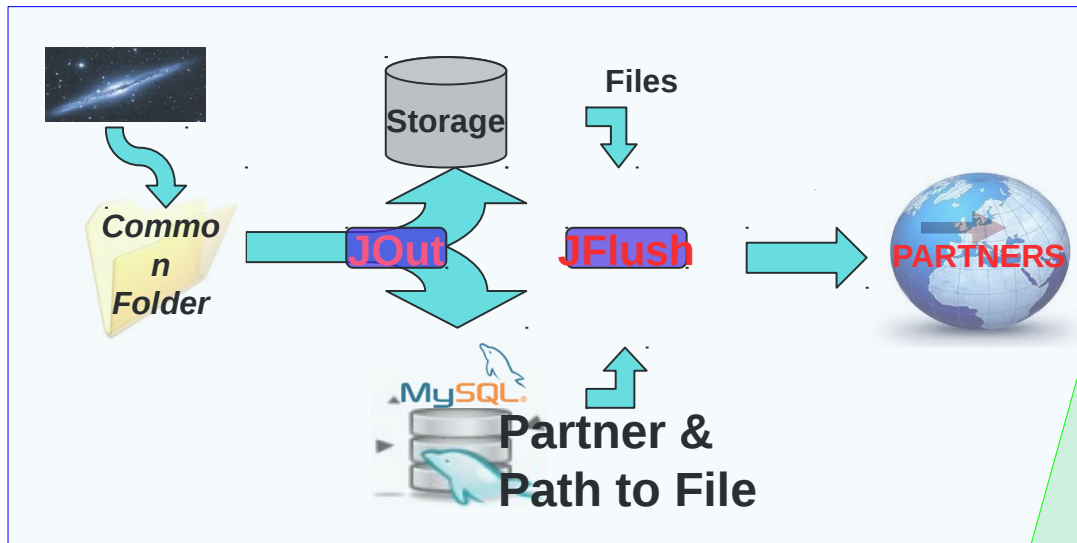
OAIS standard implementation

Files

Storage

Common Folder

JOut **JFlush** **PARTNERS**

Partner & Path to File

All (files&metaData)

All metadata

Proprietary (files)

Mt. Graham

Tucson

Heidelberg

Columbus

Trieste

Data distribution policy

Milan

Rome

Trieste 01/04/2014

IA2 Archives new Project:

NADIR build with **TANGO** Connecting things together

Issues to solve:

- Meta-data information could change in:
  - Contents;
  - Format types;
  - Keywords.
- Data formats:
  - Extensive use of ASCII and PH for calibrated data
  - FITS is not the only astronomical format (HDF5, PDS, MBFITS etc..)
  - Standards are evolving

- Consistency in meta-data content both into the file and database in case of value correction
- Data models should be, as much as possible, code independent
- Code re-usability and configurability
- Scalability in both serial and parallel ways
- Consistency over several distributed archives and secure differentiation.

**NADIR** is one configurable and flexible software that answer the challenging problem of archiving software reuse and scalability thanks also to **TANGO**.

Trieste 01/04/2014

WEB INTERFACE   VO SERVICES

Starting from FITS

FITS standard structure
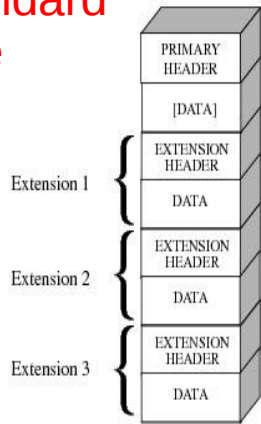
Primary HDU

Two different examples:

NADIR *DATAMODEL*

NADIR Configuration

id	host	port	user	password	schema_name	table_name	storage_path	dir_name
4	localhost	3306	user	password	lbt_metadata	warning	/mnt/storage	warning
5	localhost	3306	user	password	lbt_metadata	luci	/mnt/storage	luci
7	localhost	3306	user	password	lbt_metadata	lbc	/mnt/storage	lbc
8	localhost	3306	user	password	lbt_metadata	mods	/mnt/storage	mods
9	localhost	3306	user	password	lbt_metadata	pis	/mnt/storage	pis
10	localhost	3306	user	password	lbt_metadata	irt	/mnt/storage	irt

Instrument mapping

id	name	fits_key	fits_value	fits_date	dest_id	comment
4	Warning	NONE	NONE	NONE	4	
5	Lucifer	INSTRUME	Lucifer	DATE	5	
6	Lucifer2	INSTRUME	LUCI2	DATE	5	
7	LBCBlue	INSTRUME	LBC_BLUE	DATE_OBS	7	
8	LBCRed	INSTRUME	LBC-RED	DATE_OBS	7	
9	MODSBlue	INSTRUME	MODS1B	DATE-OBS	8	
10	MODSRed	INSTRUME	MODS1R	DATE-OBS	8	
11	Pisces	INSTRUME	PISCES	DATE	9	
12	IRT	INSTRUME	IRTC 2 - Xeva 538	DATE	10	

id	column_name	column_type	fits_key_hdu	fits_key_pri	fits_key_sec	mandatory	dest_id	comment
127	DATE_OBS	varchar		0	DATE_OBS	DATE_OBS	0	7
117	DEC	varchar		0	OBSDEC	OBSDEC	0	7
115	EXPTIME	double		0	EXPTIME	EXPTIME	0	7
110	EXP_ID	varchar		0	FILENAME	FILENAME	0	7
124	FLT_ID	varchar		0	FILTER	FILTER	0	7
129	INSTRUMENT	varchar		0	INSTRUME	INSTRUME	0	7
122	LBCOBID	varchar		0	LBCOBID	LBCOBID	0	7
111	NAXIS1	decimal		0	NAXIS1	NAXIS1	0	7
112	NAXIS2	decimal		0	NAXIS2	NAXIS2	0	7
121	OBID	varchar		0	LBCOBFIL	LBCOBFIL	0	7
119	OBJECT	varchar		0	OBJECT	OBJECT	0	7
118	OBJNAME	varchar		0	OBJNAME	OBJNAME	0	7
123	OBNAME	varchar		0	LBCOBNAM	LBCOBNAM	0	7
113	OBSERVER	varchar		0	OBSERVER	OBSERVER	0	7
126	OBS_TYPE	varchar		0	IMAGETYP	IMAGETYP	0	7
130	PARTNER	varchar		0	PARTNER	PARTNER	0	7
131	PINAME	varchar		0	PI_NAME	PI_NAME	0	7
120	PROPID	varchar		0	PROPID	PROPID	0	7
116	RA	varchar		0	OBSRA	OBSRA	0	7
128	TELESCOPE	varchar		0	TELESCOP	TELESCOP	0	7
125	TEMPERAT	double		0	CCDTEM	CCDTEM	0	7
114	USER_ID	varchar		0	LBCUSER	LBCUSER	0	7

Trieste 01/04/2014

Result for meta-data

Deployments

- **First deployment in INAF** → December 2013
Virtual Observatory **E**ducational **S**ervice **P**ublisher And **A**rchive Application

- Second deployment in INAF → March 2014
Medicina and **N**oto **R**adio **A**rchive (test bed)

- **First deployment in *USA*** → February 2014
Large **B**inocular **T**elescope **D**istributed **A**rchive
Mnt. Graham (AZ) -Tucson (AZ)
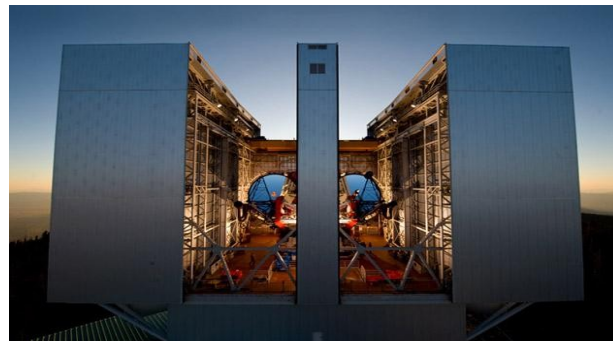
But an archive is not only that...

Archive means having raw, science ready products (calibrated data, catalogs and so on..) possibly VO compliant and/or accessible by VO services (it doesn't mater if the archives are local or remote). So:

* **data interfaces**
* **vo services**
* **work-flows / pipelines**
* **data access, curation and preservation (DOI / SSO)**

The archive front end is a Science Gateway!

IA2 on the WEB
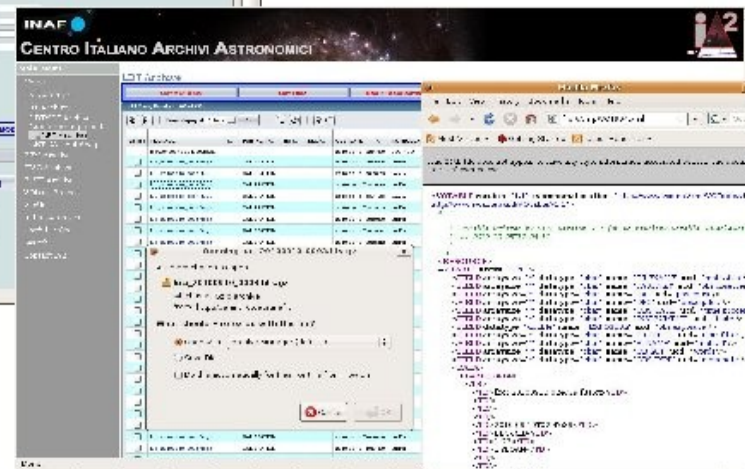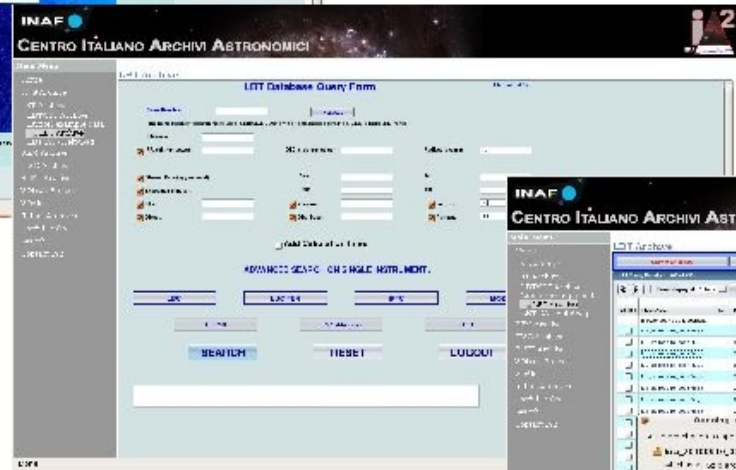
A new interface management system is under development so every hint is welcome!!

i.e.
➔ Image preview;
➔ VOTable SAMP binding;
➔ Multi-threading downloads

Trieste 01/04/2014

But an archive is not only that...

Archive means having raw, science ready products (calibrated data, catalogs and so on..) possibly vo compliant and/or accessible by VO services (it doesn't mater if the archives are local or remote). So:

* **data interfaces**
* **vo services**
* **work-lows / pipelines**
* **data curation and preservation (DOI / SSO)**

The archive front end is a Science Gateway!

Trieste 01/04/2014

IA2 on the VO

IA2 VO tools are ready to publish:

* public data (both raw and calibrated);
* catalogs;

What is publishing into the VO?
What's the VO?

Trieste 01/04/2014

Interoperable e-Infrastructure

e-Infrastructure refers to a combination and interworking of digitally-based technology (hardware and software), resources (data, services, digital libraries), communications (protocols, access rights and networks), and the people and organisational structures needed to support modern, internationally leading collaborative research be it in the arts and humanities or the sciences.

VO @ IA2

Services:
- SIAP (Simple Image Access Protocol) Services
- Cone Search services
- SSAP (Simple Spectral Access Protocol) services
- TAP (Table Access Protocol) services
- EPN TAP (EuroPlanet TAP)
- VO standard compliant Graphical User Interfaces

Resources:
- VO Services Registry hosting
- Educational VO compliant Resources
- CoSADIE VO Schools (learn the VO)

Tools:
- VODance – VO compliant data publication tool
- Powered IA2TAP – VO compliant catalogs publication tool

But an archive is not only that...

Archive means having raw, science ready products (calibrated data, catalogs and so on..) possibly vo compliant and/or accessible by VO services (it doesn't mater if the archives are local or remote). So:

* **data interfaces**
* **vo services**
* **work-flows / pipelines**
* **data curation and preservation (DOI / SSO)**
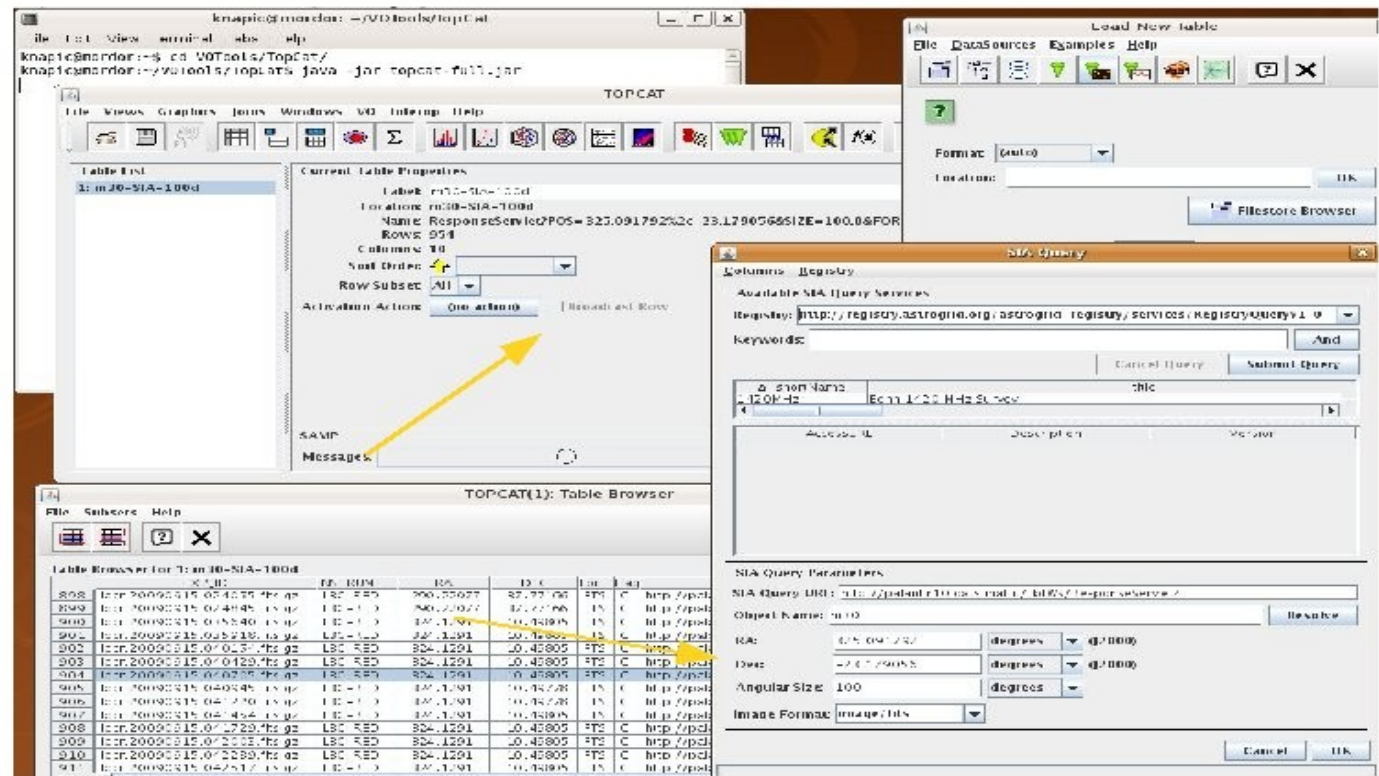
The archive front end is a Science Gateway!

Trieste 01/04/2014

Data reduction @ IA2: the GAPS experience

GAPS requests:

1. Strong **interaction** with HARPS-N@TNG private data;

2. Customizable **re-process** of GAPS data with appropriate spectral line mask and option/s;

3. Perform **queries** on additional meta-data content;

4. A **repository** where **to access, synchronize, share and search** for interesting data;

5. A flexible and collaborative **tool to manage additional info** about the project and the observations.

Trieste 01/04/2014
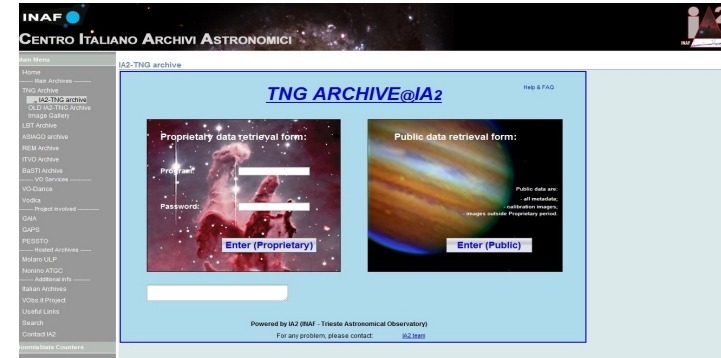
Lesson Learned with GAPS

Observation results **are not only** raw or calibrated data:

➜ Time series

➜ Catalogs

➜ New masks

➜ Alternative reduced data

➜ Scientific and ancillary data shared via Owncloud

➜ Data researchable and downloadable from wiki pages too and linked
   to night logs (collaborative tool)

References:

YABI :	https://bitbucket.org/ccgmurdoch/yabi/overview
OWNCLOUD :	http://owncloud.org/
TWIKI :	http://twiki.org/

Trieste 01/04/2014

But an archive is not only that...

Archive means having raw, science ready products (calibrated data, catalogs and so on..) possibly vo compliant and/or accessible by VO services (it doesn't mater if the archives are local or remote). So:

* **data interfaces**
* **vo services**
* **work-flows / pipelines**
* **data curation and preservation (DOI / SSO)**

The archive front end is a Science Gateway!

Trieste 01/04/2014

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 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.

References:

GROUPER : http://www.internet2.edu/products-services/trust-identity-middleware/grouper/
IDEM : https://www.idem.garr.it/en/
SHIBBOLET: http://shibboleth.net/
DOI: http://www.doi.org/

Trieste 01/04/2014

What IA2 can do:

IA2 can actually provide:
- Archiving system (raw and calibrated) with NADIR;
- Data curation;
- Graphical User Interfaces;
- VO compliant data and catalogs publication;
- Work-flow system for pipelines management;
- Data sharing area;
- Collaborative tools

What IA2 will provide in the near future:
- New Graphical User Interfaces;
- Grouping tools and SSO technology;
- Science Gateway (StarNet project collaboration) with most of the more performance visualization tools like VISIVO
- DOI system adoption

Thank you for your attention!!!

Cristina
in behalf of IA2 team

Trieste 01/04/2014

VESPA for SVAS: a test bed for EDUCATIONAL

SVAS: archive & VO
➜ Educational service for astronomical images
➜ Prototyping INAF Robotic Telescope Network data access
➜ Registry for educational resources

NADIR + **VESPA** + **VODANCE** = EDU service

VESPA is in charge to user registration, test file submission interface, data model creation and *validation* of educational (non professional) data coming from (potentially) all educational or amateur sources. NADIR is in charge of ingest and rely to VO-DANCE for the data publication in the VO.

Trieste 01/04/2014