The INAF Radio Telescopes Data Archive

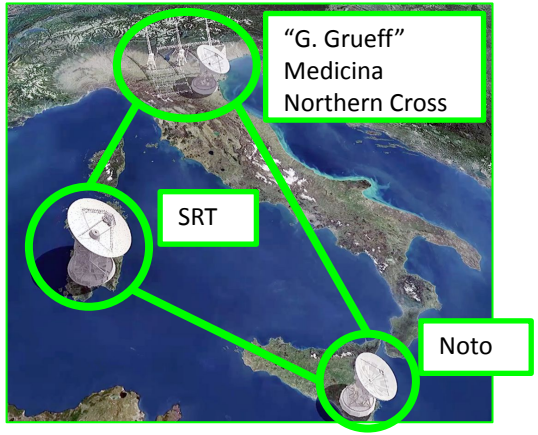
A geographically distributed infrastructure towards Big Data and FAIRness

V. Galluzzi, A. Zanichelli (INAF Radio Data Archive WG Coordinator), C. Knapic (INAF-IA2 Coordinator), F. Bedosti, A. Bignamini, M. Burgay, R. Butora, A. Fara, M. Molinaro, A. Orlati, S. Righini, M. Stagni

Acknowledgements: *N. Calabria, G. Coran, C. Urban, M. Vicinanza*

Summary

1. The INAF radio telescopes
2. Observation data products and formats
3. Data archiving workflow
4. Internal Data Models
5. Mapping onto ObsCore DM
6. Conclusions



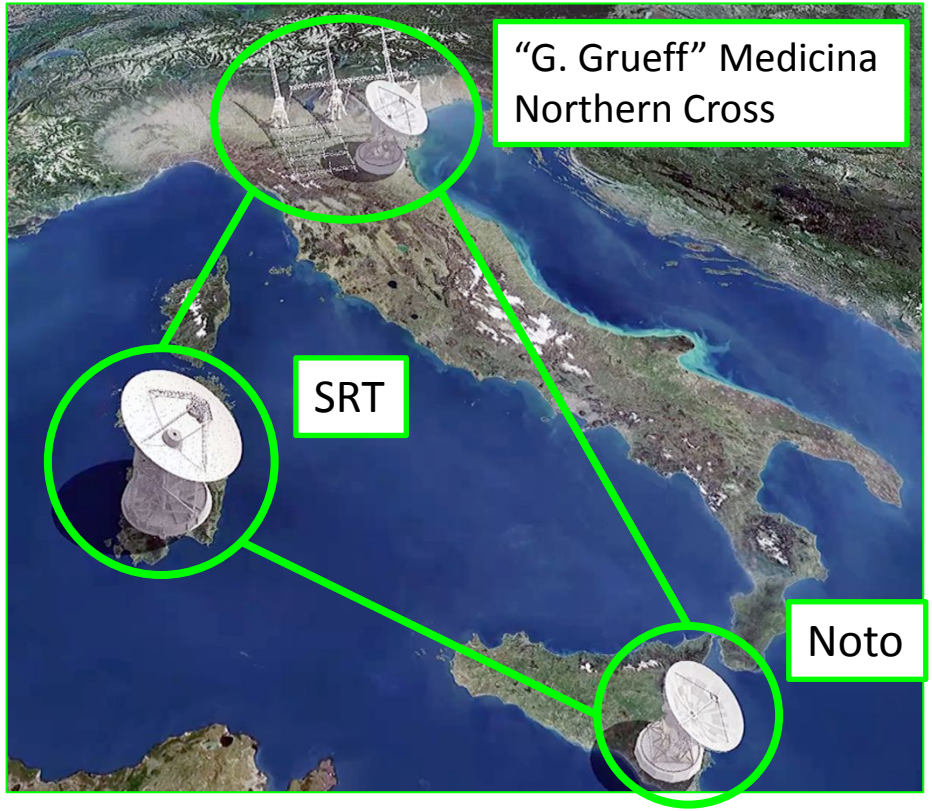
INAF Radio Telescopes

Observing modes:

- Single-dish
- VLBI (EVN, IVS, EAVN)
- VLBI-it: the «EVN lite» concept applied (+software correlator)

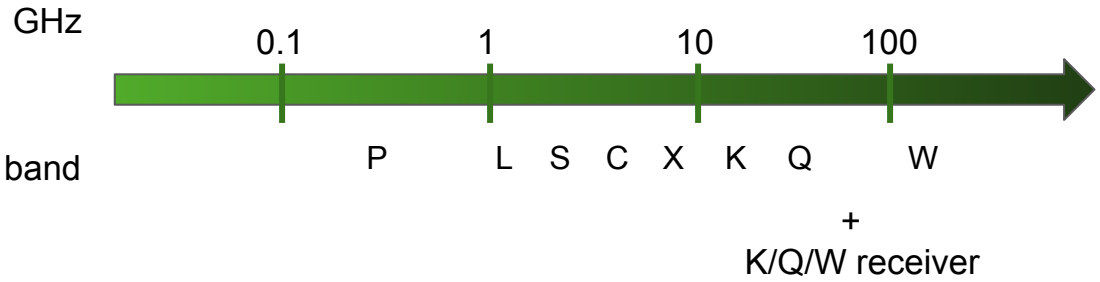
Northern Cross :

- Space debris monitoring
- Fast Radio Burst (FRB) search and monitoring



INAF Radio Telescopes

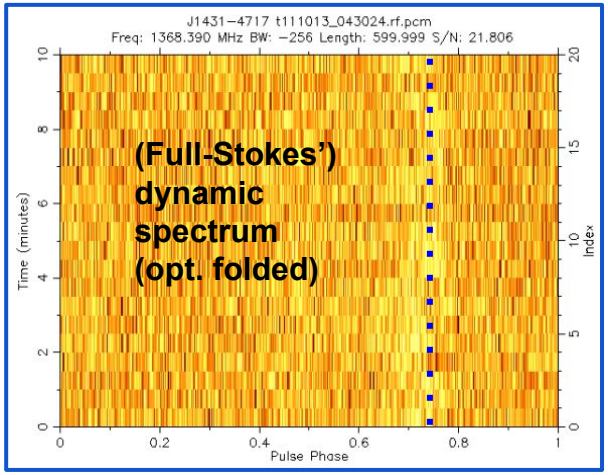
- State-off-the-art instrumentation available at the INAF radio telescopes allow observed bandwidths up to 2 GHz with up to 65k spectral channels and time resolution down to 100 microsec.
- This translates in data rates up to hundreds GB/hr, posing challenges for several archival operations.



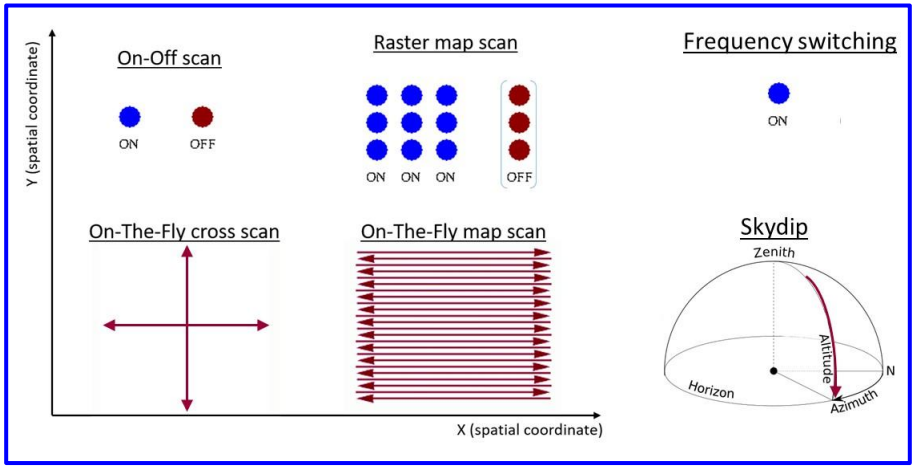
Thanks to high-frequency receivers available in the next future, INAF telescopes are going to offer bands from P to W (i.e. from 305 MHz up to 116 GHz).

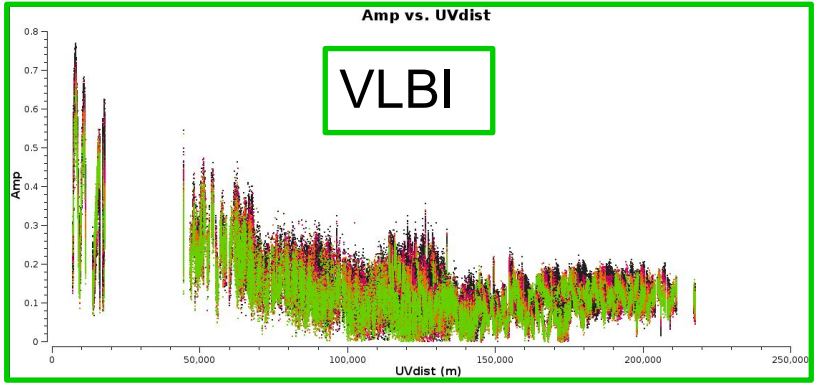
Up-to-date information about Calls for Proposals, offered instruments (e.g. receivers and back-ends), and access to the Data Archive are available through the page at the URL: https://www.radiotelescopes.inaf.it/

Observational data products

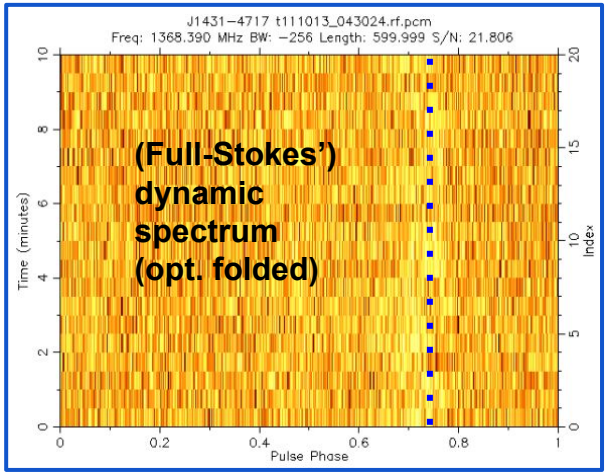


Single Dish

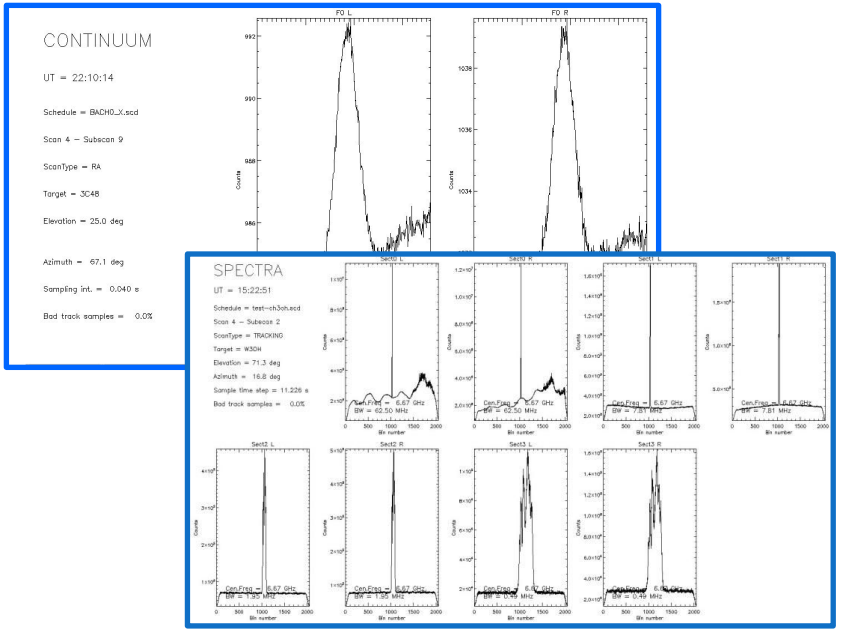




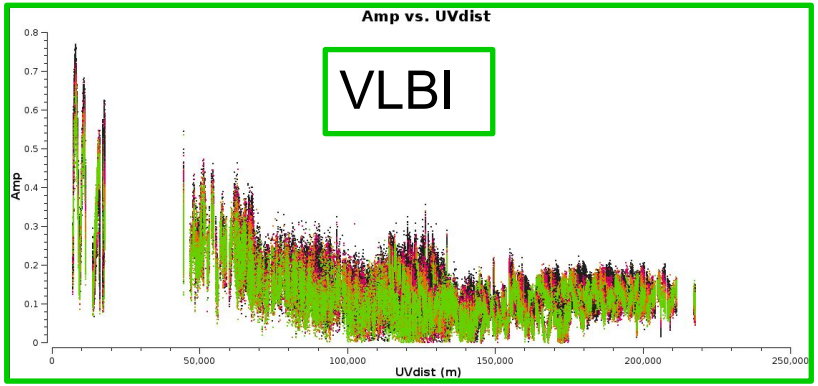
Observational data products



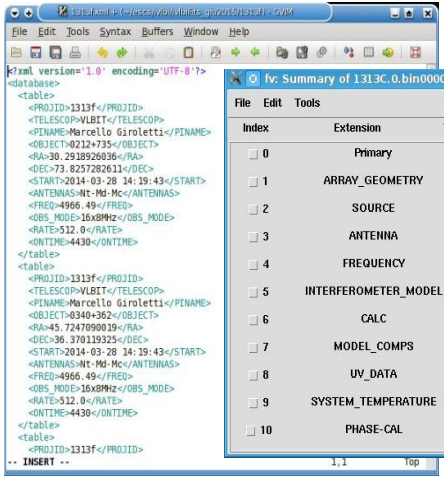
Single Dish



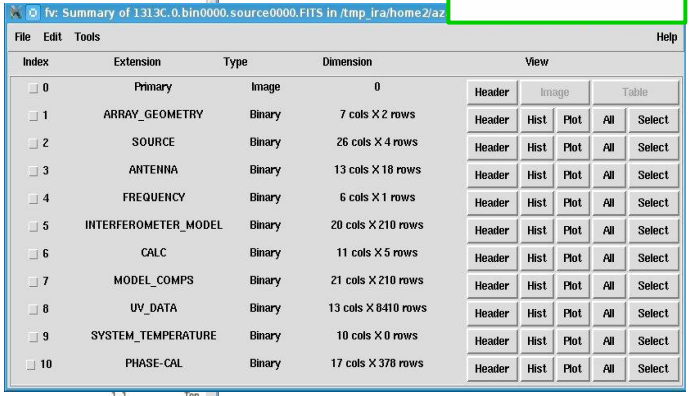
VLBI



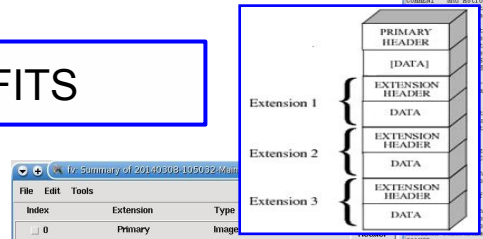
Observational data formats

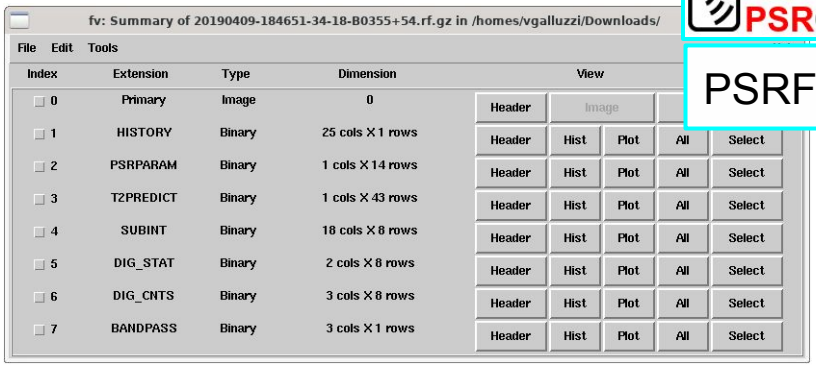


FITS-IDI



FITS





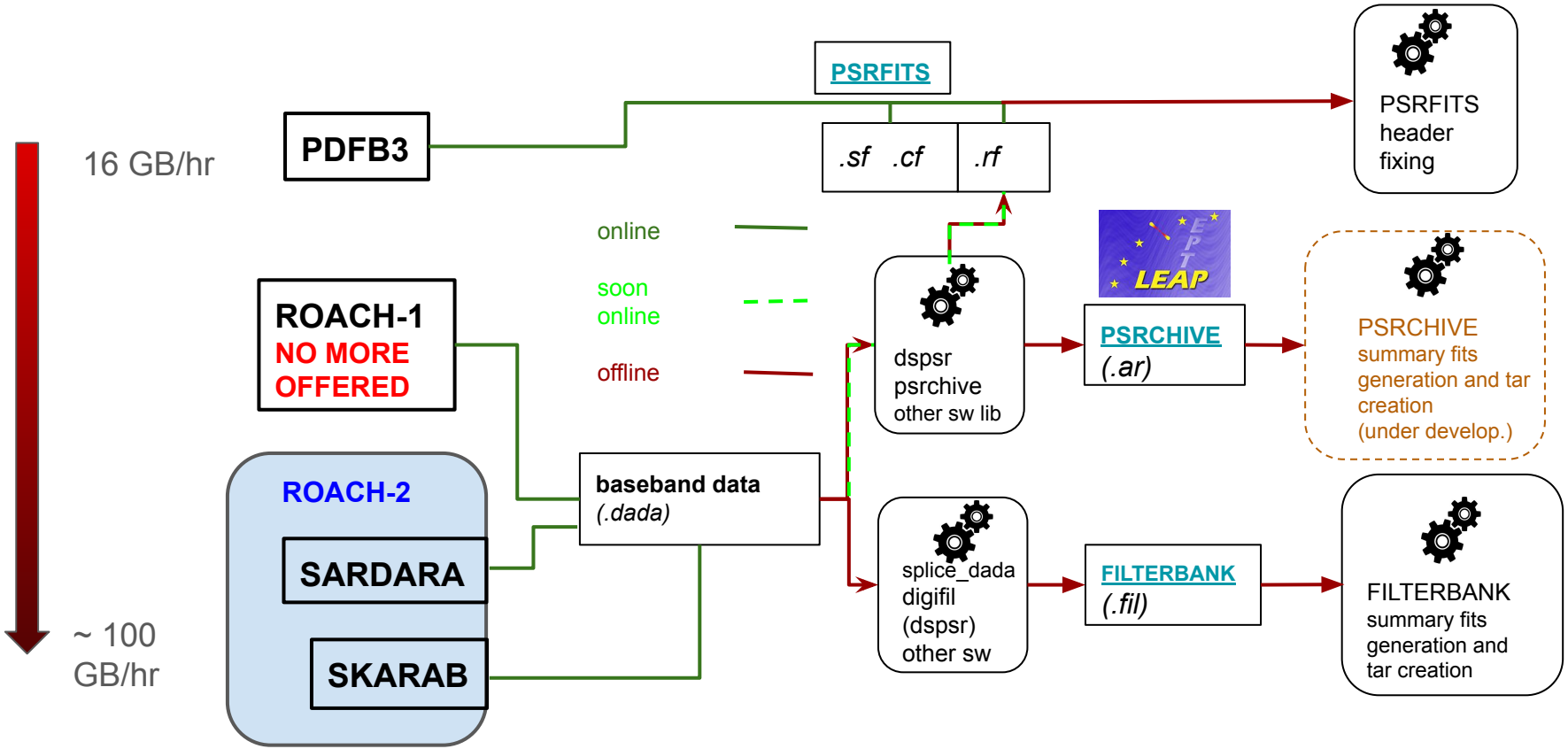
PSRCHIVE

PSRFITS

```
Data file                          : 20190923-110249-17-19-B0329+54.fil
Header size (bytes)                : 349
Data size (bytes)                  : 1231175680
Data type                          : filterbank (topocentric)
Telescope                          : SRT
Datataking Machine                 : ROACH-1
Source Name                        : B0329+54
Source RA (J2000)                  : 03:32:59.4
Source DEC (J2000)                 : +54:34:43.6
Frequency of channel 1 (MHz)       : 359.937500
Channel bandwidth            (MHz) : -0.125000
Number of channels                 : 640
Number of beams                    : 0
Beam number                        : 0
Time stamp of first sample (MJD)   : 58749.460509267316
Gregorian date (YYYY/MM/DD)        : 2019/09/23
Sample time (us)                   : 64.00000
Number of samples                  : 1923712
Observation length (minutes)       : 2.1
Number of bits per sample          : 8
Number of IFs                      : 1
```

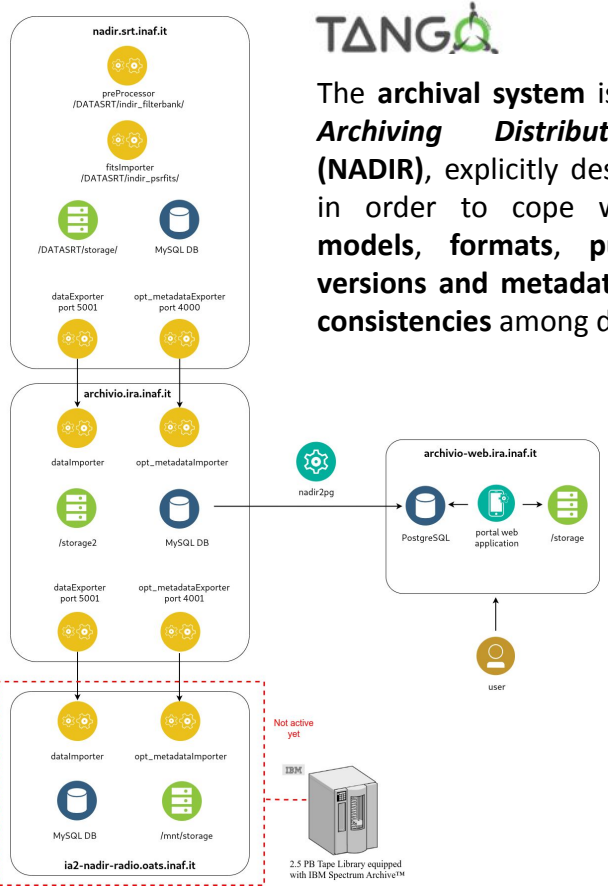
Filterbank

Pre-ingestion procedures: example for time-domain data



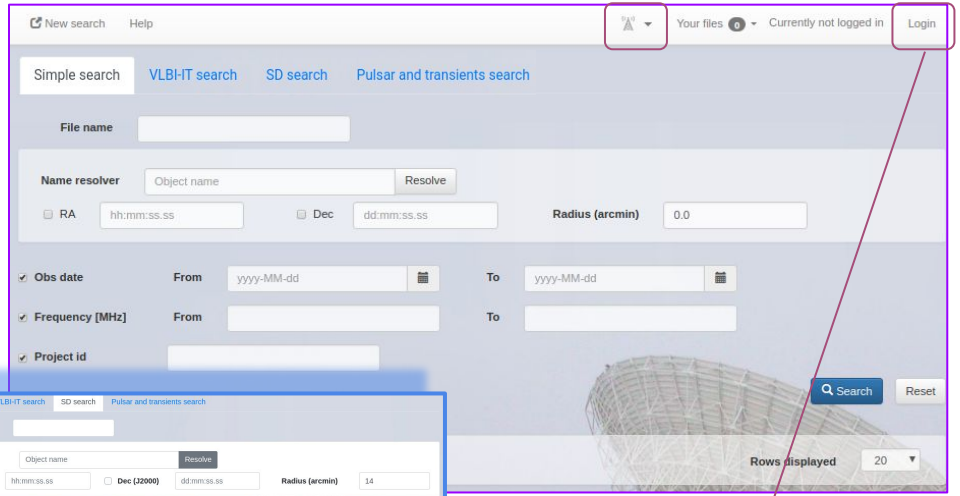
The archival system

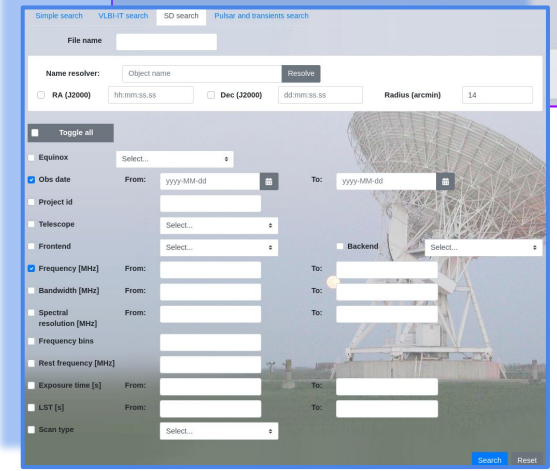
SAMP broadcast

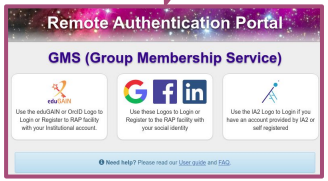


TANGO

The **archival system** is based on the ***New Archiving Distributed InfrastructuRe* (NADIR)**, explicitly designed to be **flexible** in order to cope with **evolving data models**, **formats**, **publication policies**, **versions and metadata contents**, **keeping consistencies** among different sites.

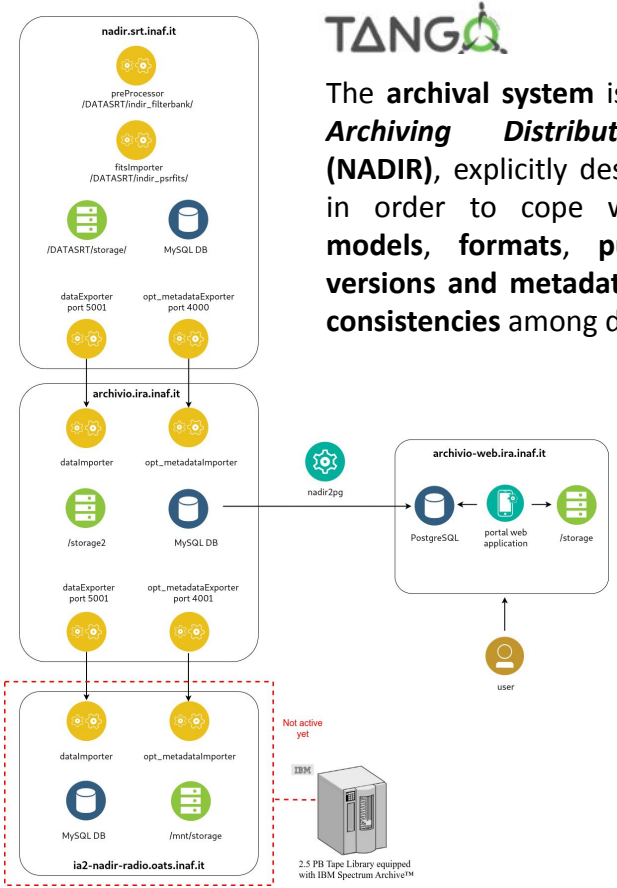






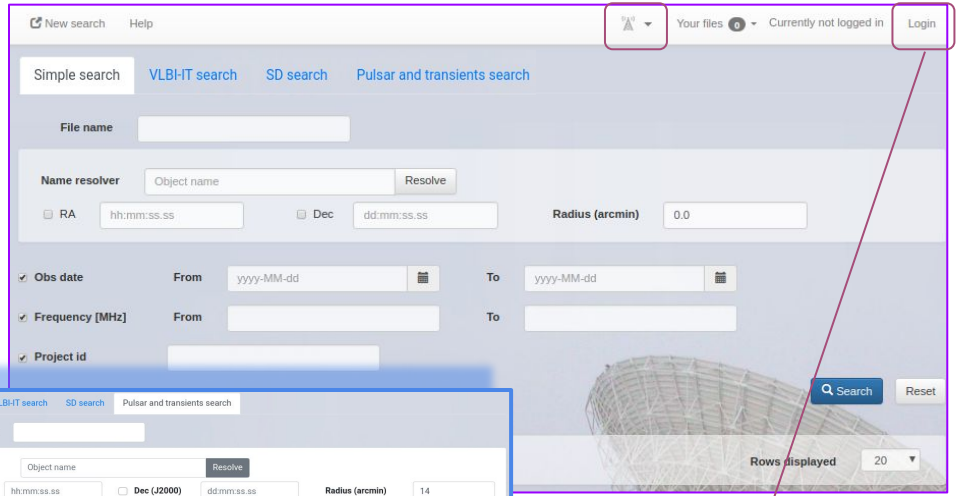
The archival system

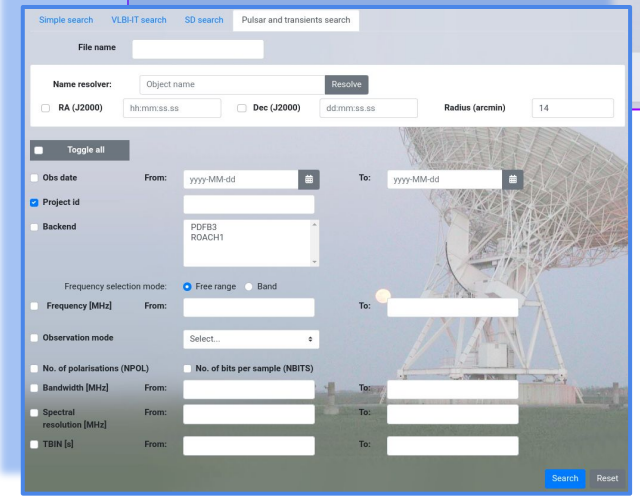
SAMP broadcast

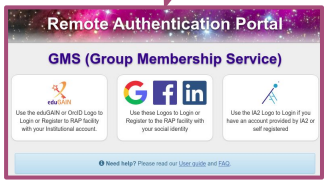


TANGO

The **archival system** is based on the ***New Archiving Distributed InfrastructuRe* (NADIR)**, explicitly designed to be **flexible** in order to cope with **evolving data models**, **formats**, **publication policies**, **versions and metadata contents**, **keeping consistencies** among different sites.

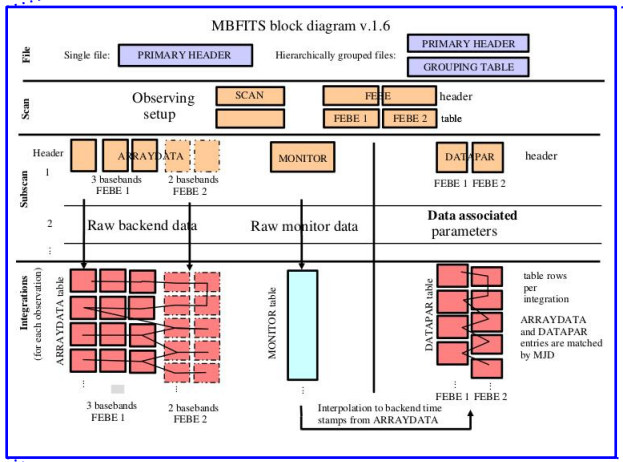


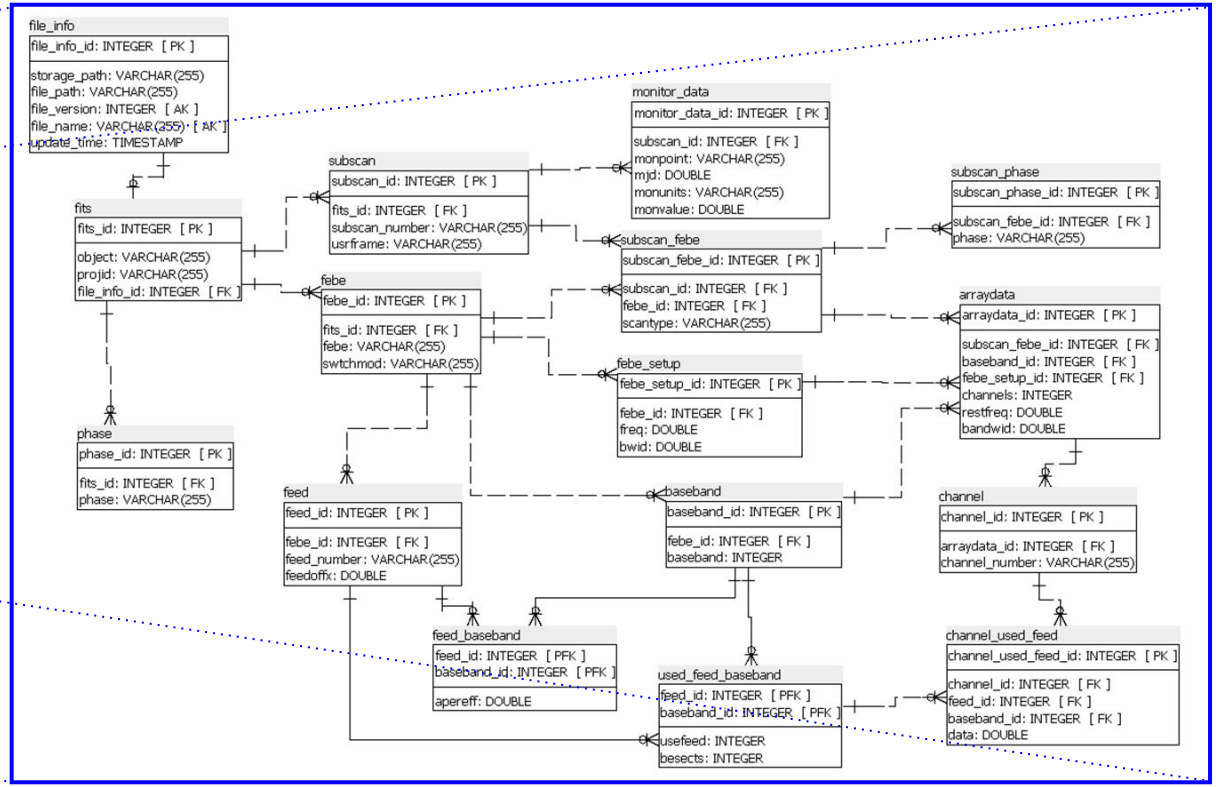




Internal data model: SD (no-time domain) and VLBI

- There is a unique internal data model for Single Dish and VLBI observations encompassing 8 relational entities (tables), mapped from the MBFITS (meta)data structure.





Internal data model for time-domain data

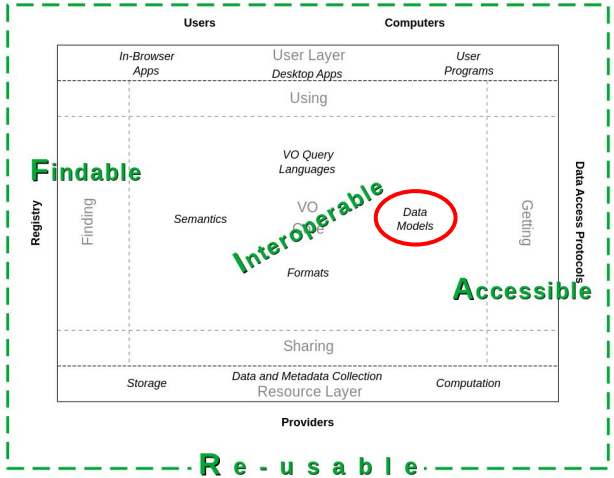
- The metadata of an observation are all written in the header of the primary HDU of a FITS file (in case of FILTERBANK or PSRCHIVE, we produce an accompanying FITS file containing only a primary header PSRFITS-like).

ID	column_name	type	HDU	keyword 1	keyword 2	description
1	TELESCOP	varchar	0	TELESCOP	TELESCOP	Telescope name
2	DATE_OBS	varchar	0	DATE-OBS	DATE-OBS	Date of observation (YYYY-MM-DDThh:mm:ss UTC)
3	OBSERVER	varchar	0	OBSERVER	OBSERVER	Observer name(s)
4	OBS_MODE	varchar	0	OBS_MODE	OBS_MODE	(PSR, CAL, SEARCH)
5	BACKEND	varchar	0	BACKEND	BACKEND	Backend ID
6	RA_C	varchar	0	RA	RA	Right ascension (hh:mm:ss.ssss)
7	DEC_C	varchar	0	DEC	DEC	Declination (-dd:mm:ss.sss)
8	EQUINOX	double	0	EQUINOX	EQUINOX	Equinox of coords (e.g. 2000.0)
9	PROJID	varchar	0	PROJID	PROJID	Project name
10	OBSFREQ	double	0	OBSFREQ	OBSFREQ	Centre frequency for observation [MHz]
11	OBSBW	double	0	OBSBW	OBSBW	Bandwidth for observation [MHz]
12	SCANLEN	double	0	SCANLEN	SCANLEN	Requested scan length (E) (N.B.: diff. from MBFITS) [s]
13	SRC_NAME	varchar	0	SRC_NAME	SRC_NAME	Source or scan ID
14	NPOL	int	0	SUBINT.NPOL	SUBINT.NPOL	Nr of polarisations
15	TBIN	double	0	SUBINT.TBIN	SUBINT.TBIN	Time per bin or sample [s]
16	NBITS	int	0	SUBINT.NBITS	SUBINT.NBITS	Nr of bits/datum (SEARCH mode 'X' data, else 1)
17	CHAN_BW	double	0	SUBINT.CHAN_BW	SUBINT.CHAN_BW	Channel/sub-band width [s]
18	OBSDATAFORMAT	varchar	0	OBSDATAFORMAT	OBSDATAFORMAT	Data format of the observation

Towards Virtual Observatory: ObsCore Data Model

- We want to make the INAF radio telescope data **F**indable, **A**ccessible, **I**nteroperable and **R**eusable. To this aim, we are going to publish the INAF Radio Data Archive as a VO resource.
- IVOA identified *Observational Data Model Core Components* and turned them into metadata to describe data products generated by astronomical observations (*ObsCore DM*)
- These metadata, implemented as a tabular view deployable through TAP, allow interoperable discovery and access of observational data in VO registered archives (ObsTAP*)*
- Core components and tabular interface were derived from a collection of use cases

- Core metadata (mandatory, suggested and optional) characterise datasets by product type and their spatial, temporal, energy observable axis (polarization can also be described); minimal provenance information is also made available



(Courtesy of Marco Molinaro)

ObsCore DM

Column Name	*Unit*	*Type*	*Description*
dataproduct_type	unitless	String	Logical data product type (image etc.)
calib_level	unitless	enum integer	Calibration level {0, 1, 2, 3, 4}
obs_collection	unitless	String	Name of the data collection
obs_id	unitless	String	Observation ID
obs_publisher_did	unitless	String	Dataset identifier given by the publisher
access_url	unitless	String	URL used to access (download) dataset
access_format	unitless	String	File content format (see in App. BB.5.2)
access_estsize	kbyte	integer	Estimated size of dataset in kilo bytes
target_name	unitless	String	Astronomical object observed, if any
s_ra	deg	double	Central right ascension, ICRS
s_dec	deg	double	Central declination, ICRS
s_fov	deg	double	Diameter (bounds) of the covered region
s_region	unitless	String	Sky region covered by the data product (expressed in ICRS frame)
s_xel1	unitless	integer	Number of elements along the first spatial axis
s_xel2	unitless	integer	Number of elements along the second spatial axis
s_resolution	arcsec	double	Spatial resolution of data as FWHM

Spatial characterization

ObsCore DM

Column Name	*Unit*	*Type*	*Description*
t_min	d	double	Start time in MJD
t_max	d	double	Stop time in MJD
t_exptime	s	double	Total exposure time
t_resolution	s	double	Temporal resolution FWHM
t_xel	unitless	integer	Number of elements along the time axis
em_min	m	double	Start in spectral coordinates
em_max	m	double	Stop in spectral coordinates
em_res_power	unitless	double	Spectral resolving power
em_xel	unitless	integer	Number of elements along the spectral axis
o_ucd	unitless	String	UCD of observable (e.g. phot.flux.density, phot.count, etc.)
pol_states	unitless	String	List of polarization states or NULL if not applicable
pol_xel	unitless	integer	Number of polarization samples
facility_name	unitless	String	Name of the facility used for this observation
instrument_name	unitless	String	Name of the instrument used for this observation

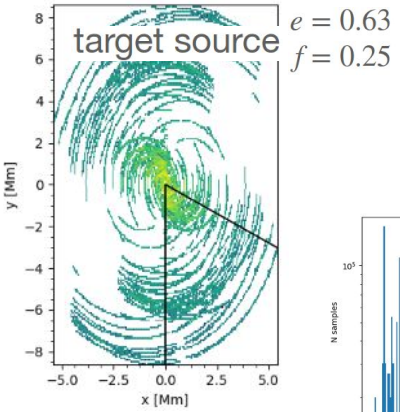
Time characterization

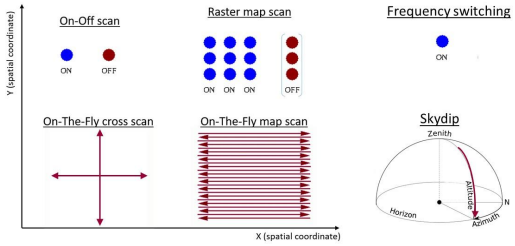
Energy/spectral characterization

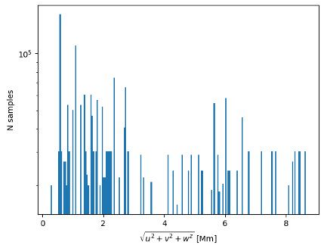
Polarization characterization

Mapping radio data onto *ObsCore DM - I*

- ObsCore DM is not sufficient for describing radio data (both SD and interferometric ones):
  - missing data products types (e.g. spatial profile)
  - scanning strategies (e.g. for SD observations)
  - variability of spatial coverage/resolution with wavelength/frequency
  - uv-coverage characterization







International Virtual Observatory Alliance

IVOA Obscore Extension for Radio data
Version 1.0

IVOA Proposed Recommendation 2024-06-14

Working Group
Data Model Working Group

This version
https://www.ivoa.net/documents/ObsCoreExtensionForRadioData/20240614

Latest version
https://www.ivoa.net/documents/ObsCoreExtensionForRadioData

Previous versions

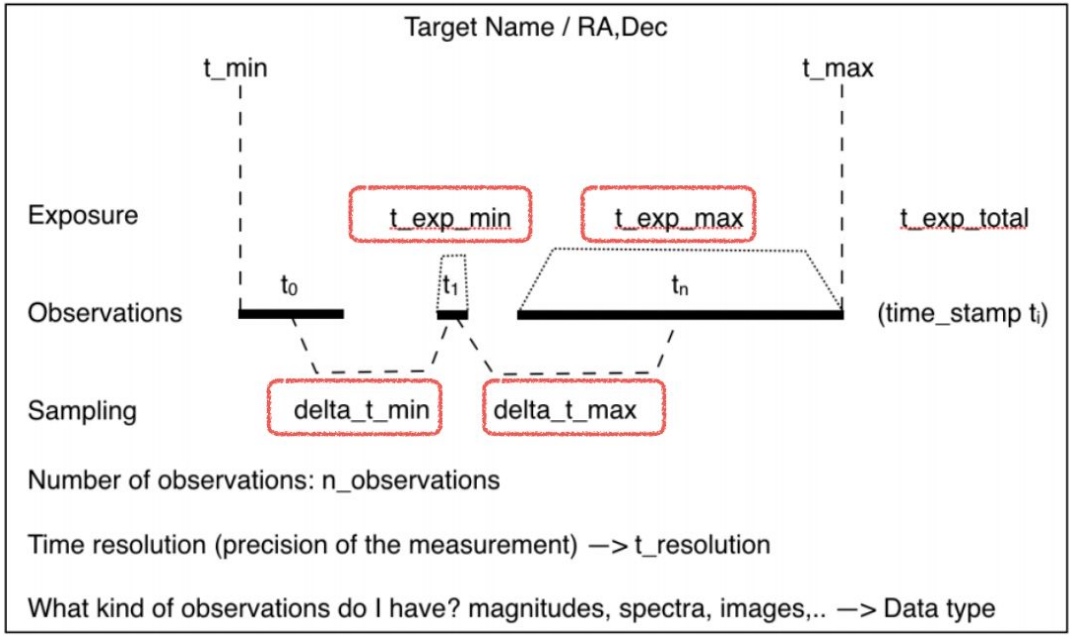
Author(s)
François Bonnarel, Mireille Louys, Baptiste Cecconi, Vincenzo Galluzzi, Yan Grange, Mark Kettenis, Mark Lacy, Alan Loh, Mattia Mancini, Peter Teuben, Alessandra Zanichelli

Editor(s)
François Bonnarel, Mark Kettenis , Mireille Louys

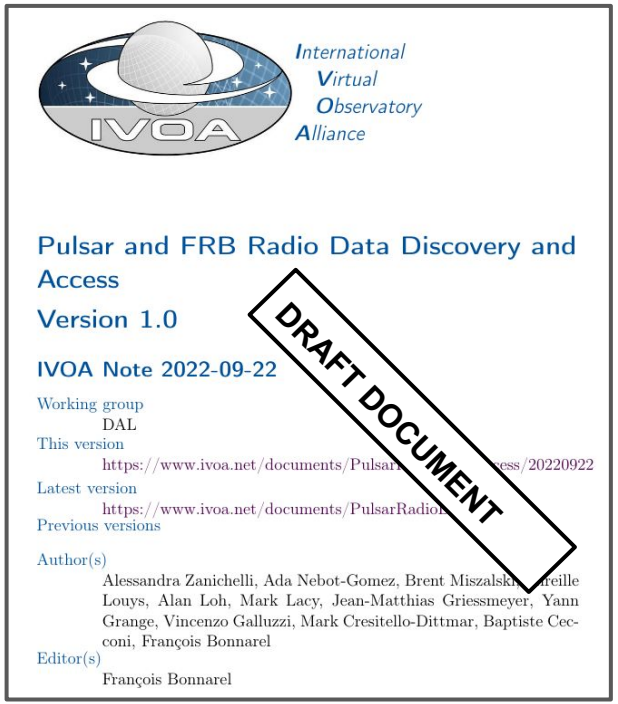
Document prepared by the IVOA *Radio Interest Group* (RIG)

Mapping radio data onto *ObsCore DM* - II

- For time-domain data, PSRCHIVE is the most comprehensive and most used for data exchange. PSRFITS depends on the telescope/data provider.



(Courtesy of Ada Nebot - CDS)



Concluding remarks

- Italian radio telescopes are undergoing a phase of significant updates and refurbishment (active surface, multi-beam, multi-band receivers, *Next Generation Croce del Nord*)
- Thanks to new instrumentation data rates are up to hundreds of GB/hr
- The INAF radio data archive currently contains SD and pulsar data from Medicina and SRT observed between 2018 and 2020.
- INAF Radio Data Archive WG is contributing to many IVOA activities:
  - ObsCore DM extension for radio data
  - discovery of time-domain data
  - activity in Semantics WG (e.g. for dataproduct_type/sub-type, o_ucd)

… Ongoing/future activities:

- *Complete the archiving operations for data observed so far*
- *Completing OS updates and migration to pyNADIR*
- *Improving long-term preservation capacity and robustness (e.g. exploiting the IA2 tape library)*
- *Implementation of new VO services for discovery, access and retrieval of Italian radio telescopes data*
- *Provide support for advanced data products generation through science platforms and dedicated user space*