The INAF Radio Telescopes Data Archive

A geographically distributed infrastructure towards Big Data and FAIRness



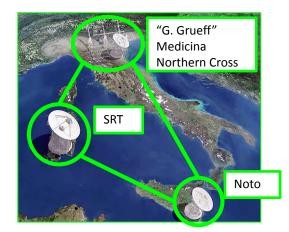
V. Galluzzi, A. Zanichelli (INAF Radio Data Archive WG Coordinator), C. Knapic (INAF-IA2 Coordinator), F. Bedosti, A. Bignamini, M. Burgay, R. Butora, A. Fara, M. Molinaro, A. Orlati, S. Righini, M. Stagni

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

V. Galluzzi – Archives and Long Term preservation Session at USC VIII–General Assembly (Galzignano Terme, 14-18 Nov 2024)

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



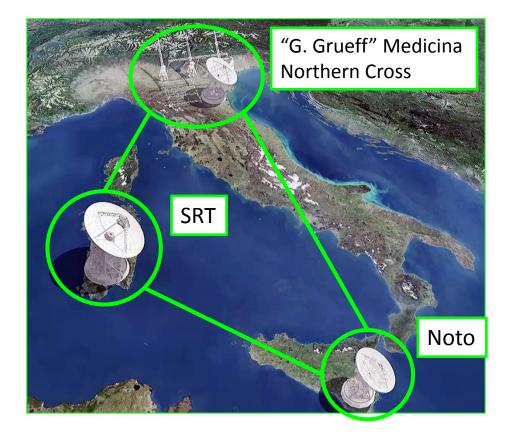
INAF Radio Telescopes

Observing modes:

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

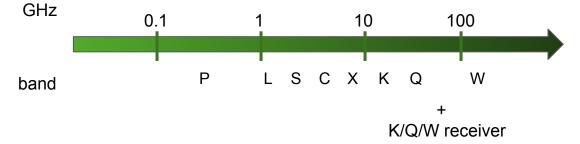
Northern Cross :

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



INAF Radio Telescopes

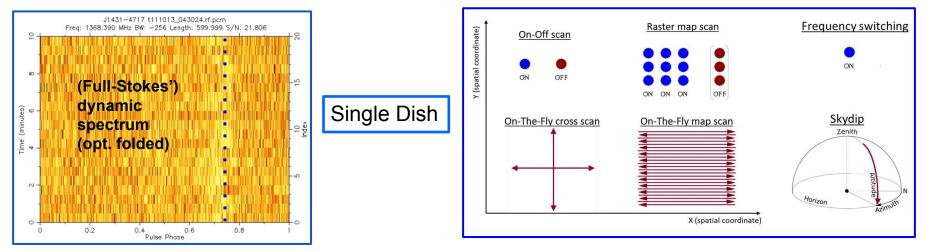
- State-off-the-art instrumentation available at the INAF radio telescopes allow observed bandwidths up to 2 GHz with up to 65k spectral channels and time resolution down to 100 microsec.
- This translates in data rates up to hundreds GB/hr, posing challenges for several archival operations.

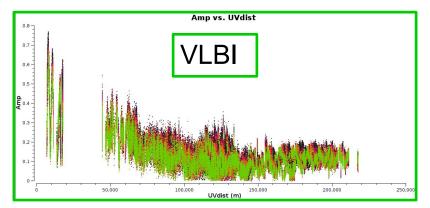


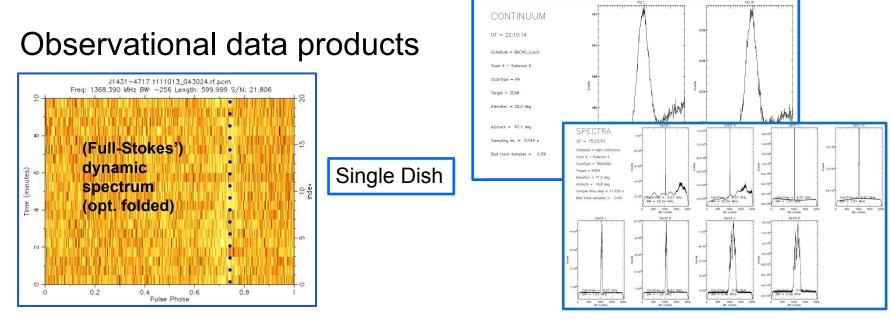
Thanks to high-frequency receivers available in the next future, INAF telescopes are going to offer bands from P to W (i.e. from 305 MHz up to 116 GHz).

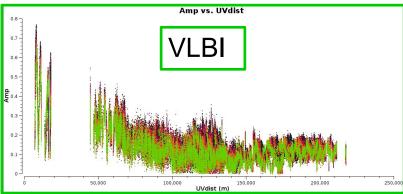
Up-to-date information about Calls for Proposals, offered instruments (e.g. receivers and back-ends), and access to the Data Archive are available through the page at the URL: <u>https://www.radiotelescopes.inaf.it/</u>

Observational data products









Observational data formats

Image: State of the state o					F۱٦	٢S	-11	DI	
<pre><?xml version='1.0' encoding='UTF-8'?> <database></database></pre>	🕷 👩 fy: S	Summary of 1313C.0.bin000).source0000.	FITS in /tmp ira/home2/az			••		
<pr0jid>1313f</pr0jid>		Tools							Hel
<telescop>VLBIT</telescop> <piname>Marcello Giroletti</piname>	Index	Extension	Туре	Dimension		View			
<0BJECT>0212+735 0BJECT <ra>30.2918926036</ra> <dec>73.8257282611</dec>	0 []	Primary	Image	0	Header	lm	ige	1	Fable
<pre><start>2014-03-28 14:19:43</start> <antennas>Nt-Md-Mc</antennas></pre>	⊒ 1	ARRAY_GEOMETRY	Binary	7 cols X 2 rows	Header	Hist	Plot	All	Select
<freq>4966.49</freq> <085 NODE>16x8MHz 085 MODE	_ 2	SOURCE	Binary	26 cols X 4 rows	Header	Hist	Plot	All	Select
<rate>512.0</rate> <ontine>4430</ontine>	_ 3	ANTENNA	Binary	13 cols X 18 rows	Header	Hist	Plot	All	Select
	□ 4	FREQUENCY	Binary	6 cols X 1 rows	Header	Hist	Plot	All	Select
<projid>1313f</projid> <telescop>VLBIT</telescop> <piname>Marcello Giroletti</piname>	_ 5	INTERFEROMETER_MODEL	Binary	20 cols X 210 rows	Header	Hist	Plot	All	Select
<0BJECT>0340+362 0BJECT <ra>45,7247090019</ra>	□ 6	CALC	Binary	11 cols X 5 rows	Header	Hist	Plot	All	Select
<pre><dec>36.370119325</dec> <start>2014.03.28 14:19:43</start></pre>	□7	MODEL_COMPS	Binary	21 cols X 210 rows	Header	Hist	Plot	All	Select
<antennas>Nt·Md·Mc</antennas> <freq>4966.49</freq>	0	UV_DATA	Binary	13 cols X 8410 rows	Header	Hist	Plot	All	Select
<0BS_MODE>16x8MHz 0BS_MODE <rate>512.0</rate> <0NTIME>4430 0NTIME	<u> </u>	SYSTEM_TEMPERATURE	Binary	10 cols X 0 rows	Header	Hist	Plot	Ali	Select
	🗌 10	PHASE-CAL	Binary	17 cols X 378 rows	Header	Hist	Plot	All	Select
<projid>1313f</projid> - INSERT		1,1 Top							

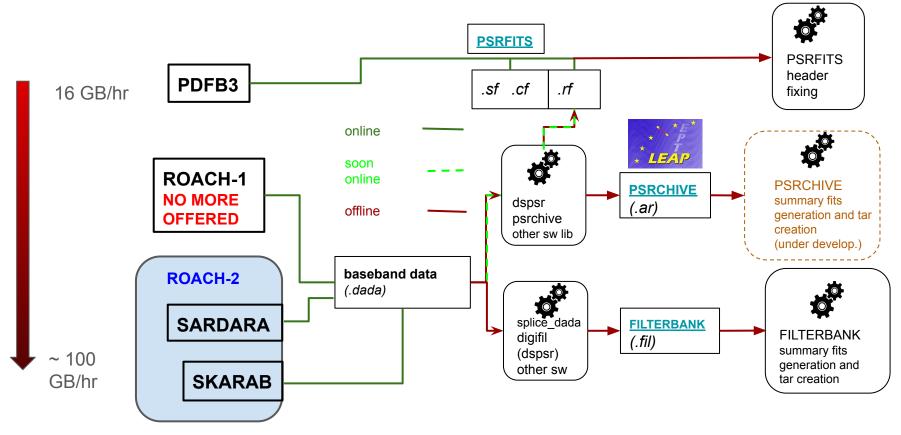
	fv: Summary of 2	20190409-1846	51-34-18-B0355+54.rf.gz	in /homes/vga	lluzzi/Do	wnloads		2122	RCHIV
e Edit	Tools								
ndex	Extension	Туре	Dimension		View	,		חפח	LITC
□ 0	Primary	Image	0	Header	Ima	age		POR	FITS
□ 1	HISTORY	Binary	25 cols X 1 rows	Header	Hist	Plot	All	Select	
_ 2	PSRPARAM	Binary	1 cols X 14 rows	Header	Hist	Plot	Ali	Select	
] 3	T2PREDICT	Binary	1 cols X 43 rows	Header	Hist	Plot	All	Select	
□ 4	SUBINT	Binary	18 cols X 8 rows	Header	Hist	Plot	All	Select	
<u> </u>	DIG_STAT	Binary	2 cols X 8 rows	Header	Hist	Plot	Ali	Select	
<u> </u>	DIG_CNTS	Binary	3 cols X 8 rows	Header	Hist	Plot	Ali	Select	
□ 7	BANDPASS	Binary	3 cols X 1 rows	Header	Hist	Plot	All	Select	

 \square

					File Er	it Tool	s					•
					Search f	ar:		<u>+</u>	Find	Case sensitive?	No	
					SIMPLE BITPIX NAXIS EXTEND COMMENT COMMENT	1		8 / n 0 / n T / F ole Inage Tra paice', volum by S. Righi	umber of umber of ITS data maport S e 376, p ni. H. B	e conform to FIT bits per data data axes set may contain Nystem) format i bage 359, bibcod artolini & A	exten exten defin e: 200. Drlati	sions ned in 'Astronos 1ASA . 376. 3598
		۰ I		10	IMARY ADER	Í	st o tsy sect tabl	output stands a column in a measurement tion table ha le the flux colu	rd for 1 data tab for eac a been s an in se	Italian radiotel ole raplaced wit ch input of each plitted into tw oction table	section table	Tant table, it r on
FITS					DATA] ENSION EADER		f ni SEX OB1	w table to s WO TaBLE 10.50 33' / = 'BET060 ' / N	tore pos file cre / Name ame of t ame of t	y header keywor sition of subrif sation date (YYY of the project the observer the station	Lector 7-MM-D	07hh:ma:ss UT)
• • (*)	v. Summary of 20140508-105	TO STAL MARK	Extension 2	EXI	ENSION EADER		ode 5 = 5 ode	= 0.68928357 650. / H 1 / N 2 / T 4 / N 5 = 'CCB / S	9621821 wight of otal run umber of / Keyw ource id	aber of sections bytes of a dat word that identi lentifier	be oit rs) fies t	e (radiane) he receiver
File Edit			Extension 3		ENSION		ffee	t = 0. / S	ource ra	17 / Source righ / Source declina sdial velocity offset in hori offset in horiz	contal	frame
Index	Extension	Туре		1	DATA		1010	on Offset = 0	/ Long	nitude offset in le offset in equ ide offset in qu	equat	orial frame 1 frame
L 0	Primary	lmage			SCANID	~	at (ffset = 0. /	Latitud can Ider	se offset in gal	sctic :	frame
1	SECTION TABLE	Binary	5 cols X 2 rows	Header	HIERARCH	Sched.	leName	- '02T060 20	ubscan 1 140308-1	Identifier 105028'		
2	RF INPUTS	Binary	9 cols X 2 rows	Header	HIERARCH	Logril SubSca	eNane : mType :	. 95T060_201	40308-10 / descri	15930.log" lbes the scan ty	pe bas	ed on telescope
3	FEED TABLE	Binary	4 cols X 1 rows	Header	7	_	_				_	-J.
	DATA TABLE	Binary	12 cols X 364 rows	Header	Hist	Plot	All	Select				
_ 4			2 cols X 364 rows	-			All	Select				
_ 4 _ 5	ANTENNA TEMP TABLE	Binary	2 COIS A 364 YOWS	Header	Hist	Plot						

Data file Header size (bytes) Data size (bytes) Data type Telescope Datataking Machine Source Name Source RA (J2000) Source DEC (J2000) Frequency of channel 1 (MHz Channel bandwidth Number of channels Number of channels	: 349 : 1231175680 : filterbank (topoc : SRT : ROACH-1 : B0329+54	
Beam number Time stamp of first sample (M Gregorian date (YYYY/MM/DI Sample time (us)	:0 JD):58749.46050):2019/ :64.00000	09267316 09/23
Number of samples Observation length (minutes) Number of bits per sample Number of IFs	: 1923712	Filterbank

Pre-ingestion procedures: example for time-domain data



The archival system

C New search Help

Simple search

File name

Name resolver

RA

Obs date

Object name

From

hh:mm:ss.ss

SAMP broadcast

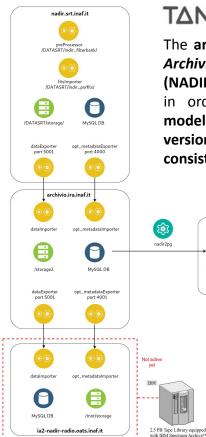
0.0

=

Login

Reset

20 🔻



TANGA

archivio-web.ira.inaf.it

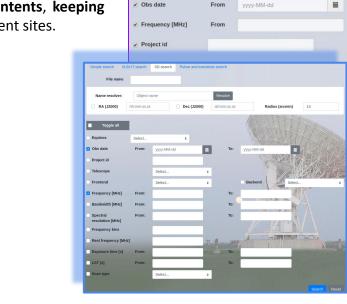
PostgreSQL

portal web

application

/storage

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.



Your files o - Currently not logged in VLBI-IT search SD search Pulsar and transients search

Radius (arcmin)

yyyy-MM-dd

To

To

Resolve

dd:mm:ss.ss

Dec

Remote Authentication Portal GMS (Group Membership Service) G f in ie the eduGAIN or OrcID I Use these Logos to Login or Register to the RAP facility with Use the IA2 Logo to Login if yo Login or Register to RAP facility have an account provided by IA2 is with war Institutional account your social identity self registered O Need help? Please read our User guide and EAQ.

Rows displayed

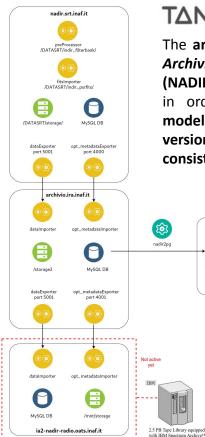
The archival system

C New search Help

SAMP broadcast

Your files o - Currently not logged in

Login



TANGA

archivio-web.ira.inaf.it

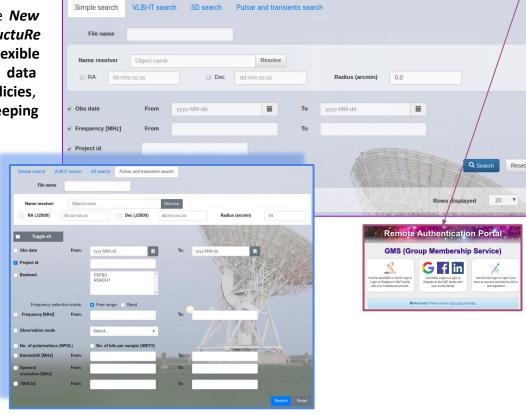
PostgreSQL

portal web

application

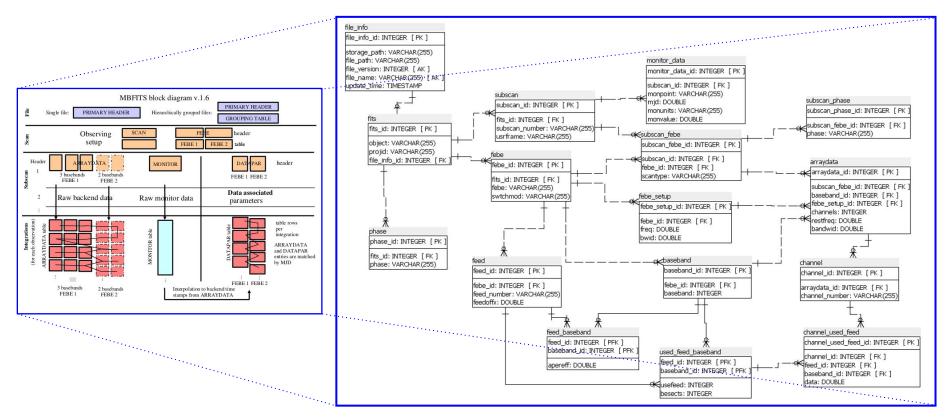
/storage

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

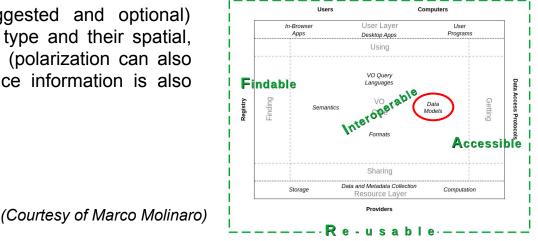
• The metadata of an observation are all written in the header of the primary HDU of a FITS file (in case of FILTERBANK or PSRCHIVE, we produce an accompanying FITS file containing only a primary header PSRFITS-like).

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	-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 Findable, Accessible, Interoperable and Reusable. 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 (<u>ObsCore DM</u>)
- 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



ObsCore DM

Column Name	Unit	Туре	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

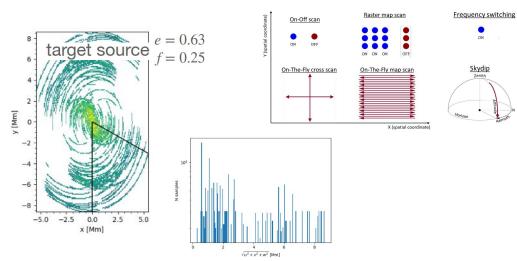
Energy/spectral characterization

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

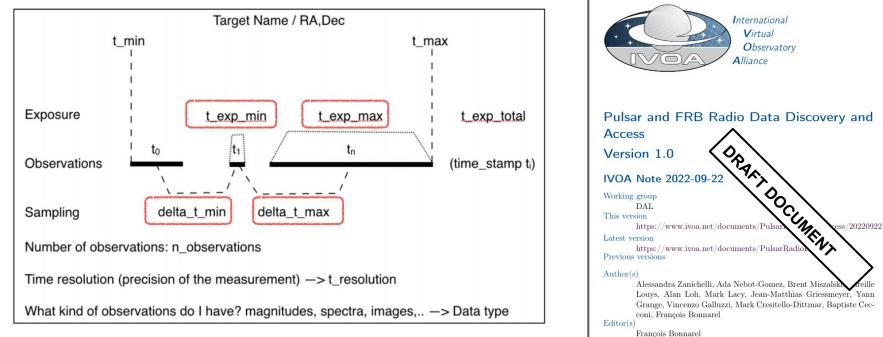


	International Virtual Observatory Alliance
IVOA	Obscore Extension for Radio data
Versio	on 1.0
IVOA I	Proposed Recommendation 2024-06-14
This version h Latest ver h	Data Model Working Group on ittps://www.ivoa.net/documents/ObsCoreExtensionForRadioData/20240614 sion ittps://www.ivoa.net/documents/ObsCoreExtensionForRadioData
Previous v	versions
l M Editor(s)	Trançois Bonnarel, Mireille Louys, Baptiste Cecconi, Vincenzo Gal- uzzi, Yan Grange, Mark Kettenis, Mark Lacy, Alan Loh, Mattia Jancini, Peter Teuben, Alessandra Zanichelli
I	rançois 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