ObsCore mapping for INAF pulsar/FRB data

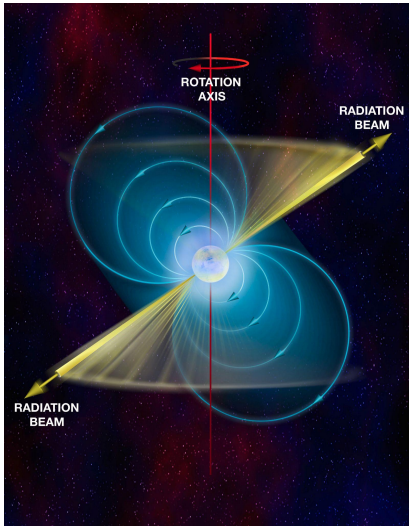
ongoing activities & future plans about pulsar/FRB data

Coordinators: *A. Zanichelli, C. Knapic (IA2)*

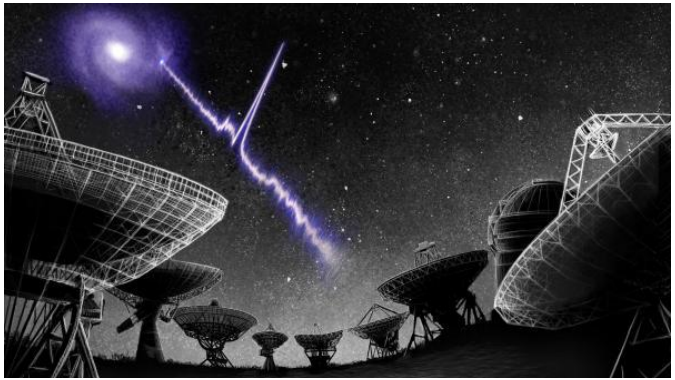
People: *V. Galluzzi, M. Molinaro, N. F. Calabria, A. A. R. Fara, M. Burgay*

Summary

1. Pulsar/transients observations and data formats
2. Data archiving workflow
3. Data model
4. ObsCore mapping proposal
5. Summary of current activities and future plans



Credit: B. Saxton, NRAO/AUI/NSF



Credit: Astron/Jive/Evn, Daniëlle Futselaar (artsource.nl)

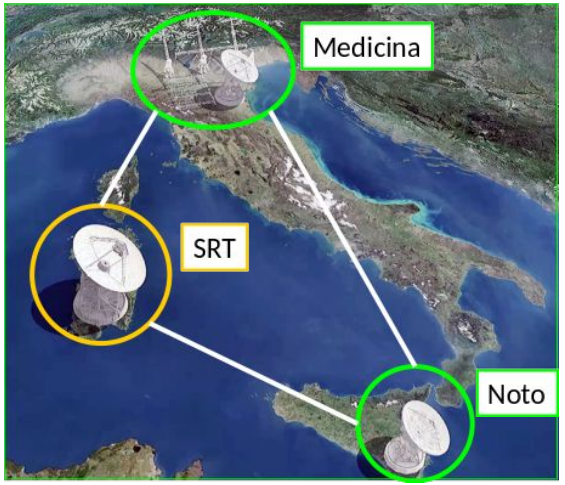
Pulsar/transients observations

BACK-ENDS

Name available bandwidths (MHz)	Type
TP 250, 680, 1200 (C and K bands only), 2000 (C and K bands only)	Analog total power
XARCOS Narrow band spectrometer; up to four (for single-feeds) simultaneous bandwidths: 0.5, 2.0, 7.8, 62.5	Spectro-polarimeter
SARDARA(e) 420, 1000 (L-band only, no f-track), 1500	Spectro-polarimeter
DFB3 1024	Correlator for pulsars
ROACH1(f) 128 (8 x 16 MHz)	Baseband recorder
DBBC 512	VLBI

RECEIVERS

RF band (GHz)	Type	Offered for
P 0.30-0.36(c) **L** 1.3-1.8	Dual frequency coaxial feed, cryo-cooled	VLBI, single-dish
C-high 5.7-7.7	Single-feed, cryo-cooled	VLBI, single-dish
K 18-26.5	7-feed, cryo-cooled	VLBI, single-dish

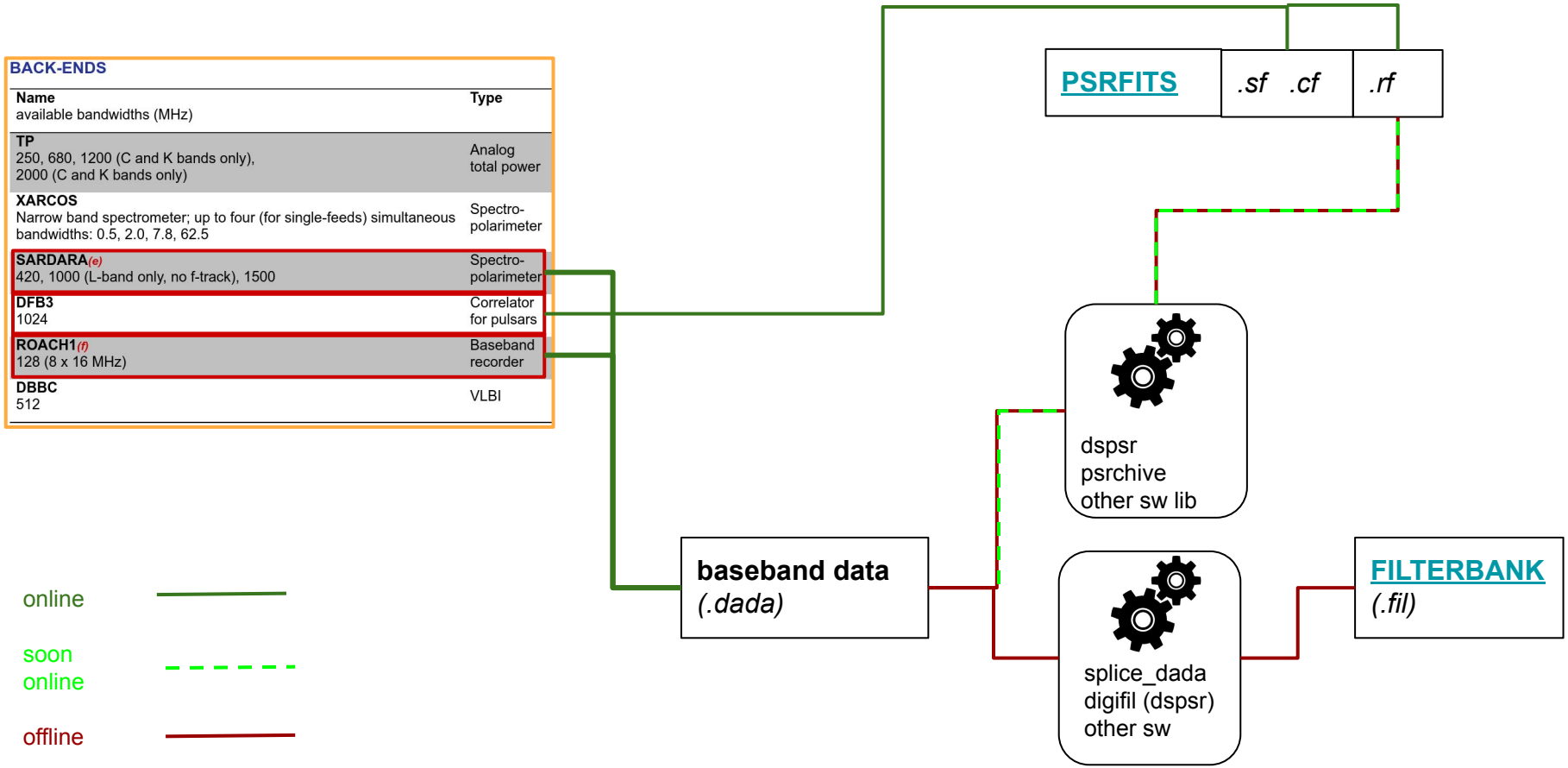


- 2 regular call for proposals per year (deadlines in April and October)
- ~ 2.1 TB raw data collected during the period 2018-2020 with PDFB3 (~ 470 files published so far)
- SARDARA spectrometer: effective data rate 20-30 GByte/hour
- on average, 150 MB to 10 GB per data file in folding mode acquisition, 3 to few hundreds GBs in search mode
- proprietary period of 1 yr (except from long programmes and/or particular requests from PIs)
- For long-term preservation, 2.5 PB tape library by IBM (equipped with LTO 8 cartridges)
- More information available at https://www.radiotelescopes.inaf.it/

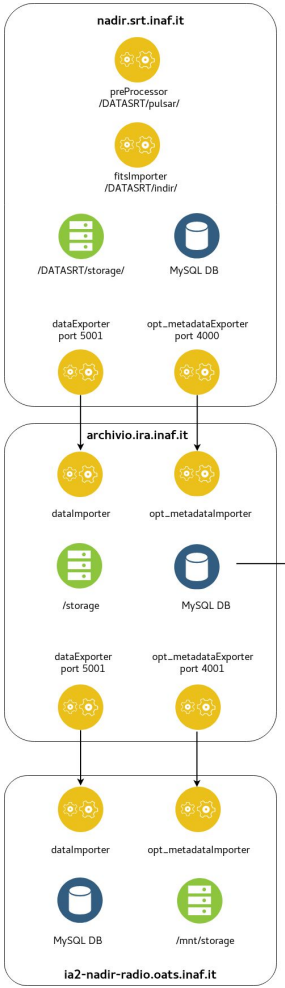
Pulsar/transients data formats

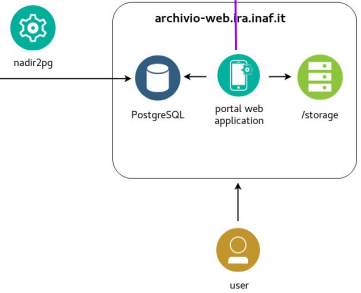
BACK-ENDS

Name available bandwidths (MHz)	Type
TP 250, 680, 1200 (C and K bands only), 2000 (C and K bands only)	Analog total power
XARCOS Narrow band spectrometer; up to four (for single-feeds) simultaneous bandwidths: 0.5, 2.0, 7.8, 62.5	Spectro- polarimeter
SARDARA*(e)* 420, 1000 (L-band only, no f-track), 1500	Spectro- polarimeter
DFB3 1024	Correlator for pulsars
ROACH1*(f)* 128 (8 x 16 MHz)	Baseband recorder
DBBC 512	VLBI



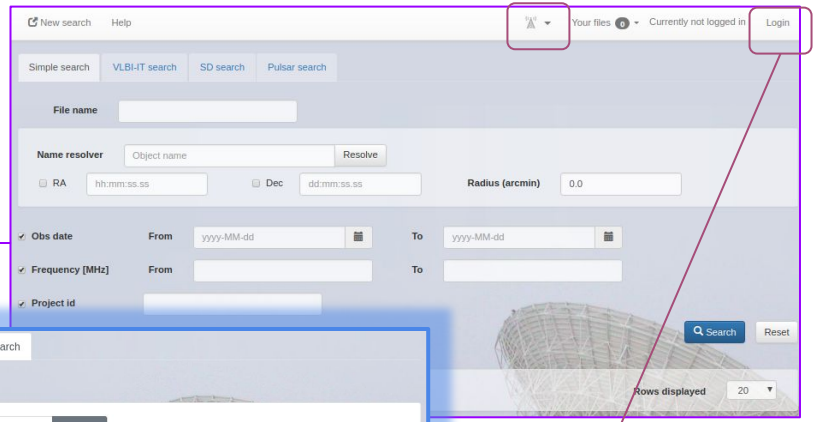
The archival system

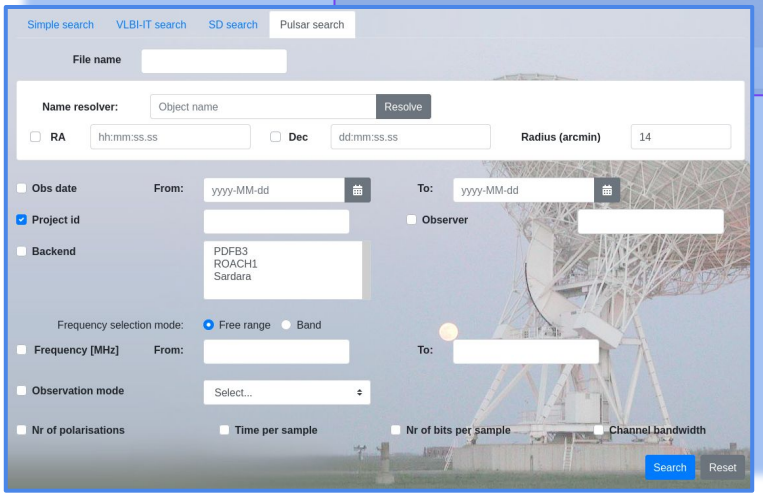


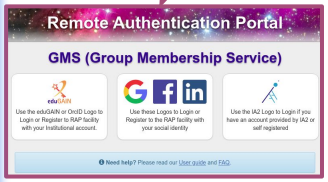


The **archival system** is based on the ***New Archiving Distributed InfrastructuRe*** **(NADIR)**, explicitly designed to be **flexible** in order to cope with **evolving data models, formats, publication policies, versions and metadata contents, keeping consistencies** among different sites.

SAMP broadcast







Pulsar data model

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

```
SIMPLE  =                    T / file does conform to FITS standard
BITPIX  =                    8 / number of bits per data pixel
NAXIS   =                    0 / number of data axes
EXTEND  =                    T / FITS dataset may contain extensions
COMMENT   FITS (Flexible Image Transport System) format is defined in 'Astronomy
[MORE COMMENT LINES]
HDRVER  = '4.0               '   / Header version
FITSTYPE= 'PSRFITS           '   / FITS definition for pulsar data files
DATE    = '2018-12-07T17:59:10' / file creation date (YYYY-MM-DDThh:mm:ss UT)
OBSERVER= 'Marta Burgay      '   / Observer name(s)
PROJID  = '34-18   '             / Project name
TELESCOP= 'SRT     '             / Telescope name
ANT_X   =              4865206.61169 / [m] Antenna ITRF X-coordinate (D)
ANT_Y   =               791927.345226 / [m] Antenna ITRF Y-coordinate (D)
ANT_Z   =              4035152.25842 / [m] Antenna ITRF Z-coordinate (D)
FRONTEND= 'Our Frontend'     / Rx and feed ID
NRCVR   =              2 / Number of receiver polarisation channels
FD_POLN = 'CIRC    '         / LIN or CIRC
FD_HAND =                    1 / +/- 1. +1 is LIN:A=X,B=Y, CIRC:A=L,B=R (I)
FD_SANG =             45. / [deg] FA of E vect for equal sig in A&B (E)
FD_XYPH =             23. / [deg] Phase of A^* B for injected cal (E)
BACKEND = 'PDFB3   '             / Backend ID
BECONFIG= 'pdfb4_1024_512_2048' / Backend configuration file name
BE_PHASE=                    -1 / 0/+1/-1 BE cross-phase:0 unknown,+/-1 std/rev
BE_DCC  =                    0 / 0/1 BE downconversion conjugation corrected
BE_DELAY=                    0. / [s] Backend propn delay from digitiser input
TCYCLE  =             10. / [s] On-line cycle time (D)
OBS_MODE= 'PSR     '         / (PSR, CAL, SEARCH)
DATE-OBS= '2018-12-07T17:59:20' / Date of observation (YYYY-MM-DDThh:mm:ss UTC)
OBSFREQ =             6206. / [MHz] Centre frequency for observation
OBSBW   =             512. / [MHz] Bandwidth for observation
OBSNCHAN=                    2048 / Number of frequency channels (original)
CHAN_DM =                    0. / [cm-3 pc] DM used for on-line dedispersion
SRC_NAME= 'B0355+54'         / Source or scan ID
COORD_MD= 'J2000   '         / Coordinate mode (J2000, GAL, ECLIP, etc.)
EQUINOX = '2000.000'         / Equinox of coords (e.g. 2000.0)
```

```
RA      = '03:58:54.717'     / Right ascension (hh:mm:ss.ssss)
DEC     = '+54:13:13.727'    / Declination (-dd:mm:ss.sss)
BMAJ    =  0.0547856912665163 / [deg] Beam major axis length
BMIN    =  0.0547856912665163 / [deg] Beam minor axis length
BPA     =             0. / [deg] Beam position angle
STT_CRD1= '03:58:54.717'     / Start coord 1 (hh:mm:ss.sss or ddd.ddd)
STT_CRD2= '+54:13:13.727'    / Start coord 2 (-dd:mm:ss.sss or -dd.ddd)
TRK_MODE= 'TRACK   '         / Track mode (TRACK, SCANGC, SCANLAT)
STP_CRD1= '03:58:54.717'     / Stop coord 1 (hh:mm:ss.sss or ddd.ddd)
STP_CRD2= '+54:13:13.727'    / Stop coord 2 (-dd:mm:ss.sss or -dd.ddd)
SCANLEN =            139.810 / [s] Requested scan length (E)
FD_MODE = 'FA      '         / Feed track mode - FA, CPA, SPA, TPA
FA_REQ  =            0.2 / [deg] Feed/Posn angle requested (E)
CAL_MODE= 'SYNC    '         / Cal mode (OFF, SYNC, EXT1, EXT2)
CAL_FREQ=                    0. / [Hz] Cal modulation frequency (E)
CAL_DCYC=                    0. / Cal duty cycle (E)
CAL_PHS =            0. / Cal phase (wrt start time) (E)
STT_IMJD=            58459 / Start MJD (UTC days) (J - long integer)
STT_SMJD=            64760 / [s] Start time (sec past UTC 00h) (J)
STT_OFFS=            0.1312937734375 / [s] Start time offset (D)
STT_LST =            86399.8683467561 / [s] Start LST (D)
STT_DATE= '2018-12-07'
STT_TIME= '17:59:20.000'
HIERARCH OBSDATAFORMAT = 'PSRFITS ' / Data format of the observation
HIERARCH SUBINT.NPOL =     4 / Nr of polarisations
HIERARCH SUBINT.TBIN = 0.00015272180983952 / [s] Time per bin or sample
HIERARCH SUBINT.NBITS =    1 / Nr of bits/datum (SEARCH mode 'X' data, else 1)
HIERARCH SUBINT.CHAN_BW = 0.25 / [MHz] Channel/sub-band width
COMMENT This file has been verified and finalized via the
COMMENT pre-processor script checkfits_radio.sh (version 3.7)
[MORE COMMENT LINES]
```

Pulsar data model

- The metadata of an observation are all written in the header of the primary HDU of a FITS file (in case of FILTERBANK, we produce an accompanying FITS file containing only a primary header).
- Then, the metadata are mapped onto a flat table (datamodel_pulsar) in a MySQL db (metadata_pulsar).

id	column_name	column_type	fits_key_hdu	fits_key_pri	fits_key_sec	
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.sss UTC)
3	SRC_NAME	varchar	0	SRC_NAME	SRC_NAME	Source or scan ID
4	OBSERVER	varchar	0	OBSERVER	OBSERVER	Observer name(s)
5	OBS_MODE	varchar	0	OBS_MODE	OBS_MODE	PSR, CAL, SEARCH)
6	BACKEND	varchar	0	BACKEND	BACKEND	Backend ID
7	RA_C	varchar	0	RA	RA	Right ascension (hh:mm:ss.ssss)
8	DEC_C	varchar	0	DEC	DEC	Declination (-dd:mm:ss.sss)
9	EQUINOX	double	0	EQUINOX	EQUINOX	Equinox of coords (e.g. 2000.0)
10	PROJID	varchar	0	PROJID	PROJID	Project name
11	OBSFREQ	double	0	OBSFREQ	OBSFREQ	[MHz] Centre frequency for observation
12	**OBSBW**	**double**	**0**	**OBSBW**	**OBSBW**	**[MHz] Bandwidth for observation**
13	**SCANLEN**	**double**	**0**	**SCANLEN**	**SCANLEN**	**[s] Requested scan length (E) (N.B.: diff. from MBFITS)**
14	**NPOL**	**int**	**0**	**SUBINT.NPOL**	**SUBINT.NPOL**	**Nr of polarisations**
15	**TBIN**	**double**	**0**	**SUBINT.TBIN**	**SUBINT.TBIN**	**[s] Time per bin or sample**
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**	**[MHz] Channel/sub-band width**
18	**OBSDATAFORMAT**	**vachar**	**0**	**OBSDATAFORMAT**	**OBSDATAFORMAT**	**Data format of the observation**
19	***FRONTEND***	***vachar***	***0***	***FRONTEND***	***FRONTEND***	***Frontend ID***

ObsCore mapping proposal

OBSCore keyword (mandatory are in boldface)	**ObsCore Unit (Type)**	**Value (expressions are based on fields of the table pulsar; db: metadata_pulsar)**	**Comments**
dataproduct_type	Unitless (string, from an IVOA controlled list)	**Dynamic Spectrum**, *i.e.* "Consecutive spectral measurements through time,organized as a time series of successive spectra of the same target or field of view.", if NPOL =1 OR **Cube** (i.e.a full Stokes radio data cube), if NPOL > 1 ALTERNATIVELY: **Dynamic Spectrum** in any case	
dataproduct_subtype	Unitless (string, from an INAF controlled list)	**“search mode”** if OBS_MODE=”SRCH” **“folded mode”** if OBS_MODE=”PSR” **“cal mode”** if OBS_MODE=”CAL”	
calib_level	Unitless (integer, following an IVOA standard)	**1,** if OBSDATAFORMAT = “PSRFITS”) or **0**, if OBSDATAFORMAT= “FILTERBANK”	Considering that FILTERBANK is not a standard data format, strictly speaking.
obs_collection	Unitless string (from an INAF controlled list)	INAF-radio, pulsar and transients ALTERNATIVELY: INAF-radio, time domain	

OBSCore keyword	ObsCore Unit (Type)	Value (expressions are based on fields of the table pulsar; db: metadata_pulsar)	Comments
obs_id	Unitless string (regular expression)	TELESCOP_yyyy-mm-ddThh:mm:ss-PROJID-SRC_NAME E.g.: 20190114-175309-48-19-FRB180814	yyyy-mm-ddThh:mm:ss is parsed from DATE_OBS (which is given following ISO 8601, as yyyy-mm-ddThh:mm:ss.sss)
obs_publisher_did	Unitless (string, regular expression following an IVOA standard)	ivo://ia2.inaf.it/radio/pulsar-transients/<obs_id> E.g.: http://radioarchive.inaf.it/files/radio/pulsar/20190928-002435-34-19-B1929%2B10_R	The same observation may be published elsewhere (e.g. in case of data acquired by SRT as a part of a project, EPTA) (N.B.: substitute "%2B" to "+")
access_url	Unitless (string, regular expression following a W3C standard)	http://radioarchive.inaf.it/files/radio/pulsar/<file_name> E.g.: http://radioarchive.inaf.it/files/radio/pulsar/20190928-002435-34-19-B1929%2B10_R.cf.gz	The URL is composed by a fixed string (i.e. http://radioarchive.inaf.it/files/radio/pulsar/) + the file name of the .tar.gz payload (N.B.: substitute "%2B" to "+")
access_format	Unitless (string, from an IVOA controlled list)	application/x-fits-psrfits-gzip if OBSDATAFORMAT="PSRFITS" application/x-tar if OBSDATAFORMAT="FILTERBANK"	A dataset/observation can be only retrieved as a .gz (PSRFITS obs.) or .tar (FILTERBANK obs.)
access_estsize	kbyte (integer)	Produced from the archival process	It is an estimated size, so there is no need to be very accurate on this.

OBSCore keyword	ObsCore Unit (Type)	Value (expressions are based on fields of the table pulsar; db: metadata_pulsar)	Comments
s_ra	degree (double)	RA_RAD*180.0/π ALTERNATIVELY, you can use a function to convert from sexagesimal RA (hh:mm:ss.ssss)	
s_dec	degree (double)	DEC_RAD*180.0/π ALTERNATIVELY, you can use a function to convert from sexagesimal DEC (dd:mm:ss.sss)	
s_fov	degree (double)	if TELESCOP eq 'SRT' then (*19.7*/((OBSFREQ-OBSBW/2.0)/1000.0))/60. else if TELESCOP *[...]*	N.B.: The first factor changes for each TELESCOP value. (*): In case of multi-feed observations, we need to revise the formula
s_region	Unitless string (syntax following an IVOA standard)	CIRCLE('ICRS',<s_ra>,<s_dec>, <s_fov>/2.0)	ALTERNATIVELY, a more stringent formula could adopt the central frequency only. (*)
s_xel1	Unitless (integer)	NULL	not applicable
s_xel2	Unitless (integer)	NULL	not applicable
s_resolution	arcsec (double)	NULL	not applicable

OBSCore keyword	ObsCore Unit (Type)	Value (expressions are based on fields of the table pulsar; db: metadata_pulsar)	Comments
t_min	day (MJD)	**UTC_To_MJD(**DATE_OBS**)**	DATE_OBS, *i.e.* the approximate UTC date and time of the start of an observation (DATE-OBS in PSRFITS), is filled following ISO 8601, as yyyy-mm-ddThh:mm:ss.sss
t_max	day (MJD)	**UTC_To_MJD(**DATE_OBS)+SCANLEN/86400.0	
t_exptime	s (double)	SCANLEN E.g.: 109.883 / [s] Requested scan length	N.B.: its meaning for SD is different (*i.e.* an angular arc length in the sky)
t_resolution	s (double)	TBIN E.g.: 8.72416600106808E-05 / [s] Time per bin or sample	
t_xel	Unitless (integer)	NULL	Not particularly useful; also, the outcome depends on the observation mode

OBSCore keyword	ObsCore Unit (Type)	Value (expressions are based on fields of the table pulsar; db: metadata_pulsar)	Comments
em_min	metre (double)	c/[(OBSFREQ - OBSBW/2.0)*1.0e6], where c= 299792458 m/s	OBSFREQ and OBSBW are provided in MHz, as in PSRFITS standard
em_max	metre (double)	c/[(OBSFREQ + OBSBW/2.0)*1.0e6]	
em_res_power	Unitless (double)	NULL	Not useful/meaningful for discovery
em_resol	TBD metre (or directly in Hz/MHz?) (double)	c/(CHANBW*1.0e6)	CHANBW is given in MHz, as in PSRFITS standard
em_xel	Unitless (integer)	OBSNCHAN E.g.: 1024 / Number of frequency channels (original)	

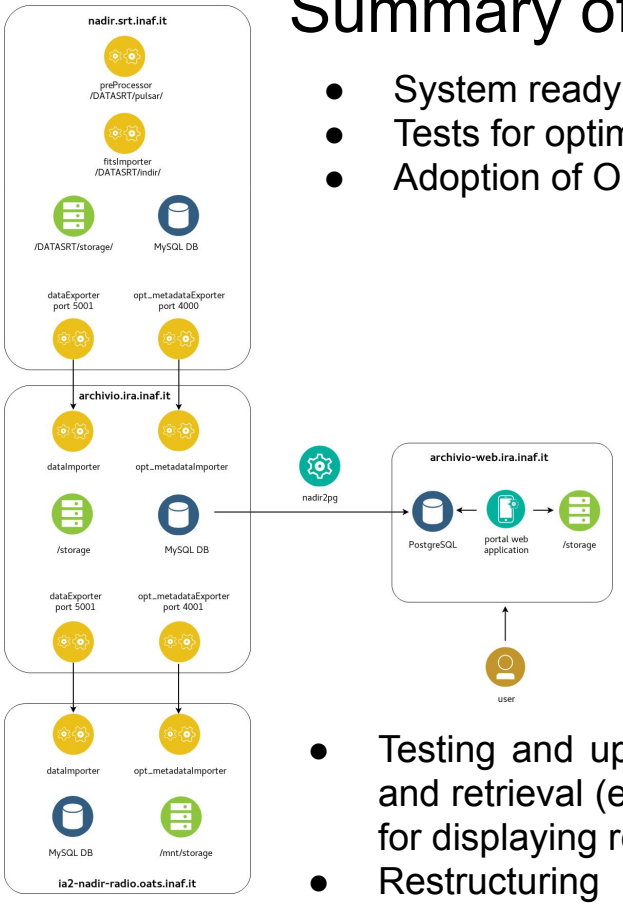
o_ucd	Unitless (string, following an IVOA controlled list)	phot.flux.density	
pol_states	Unitless (string)	Depending on NPOL. If NPOL= 1, then “I”; if NPOL=4, then “I,Q,U,V”	
pol_xel	Unitless (integer)	NPOL	

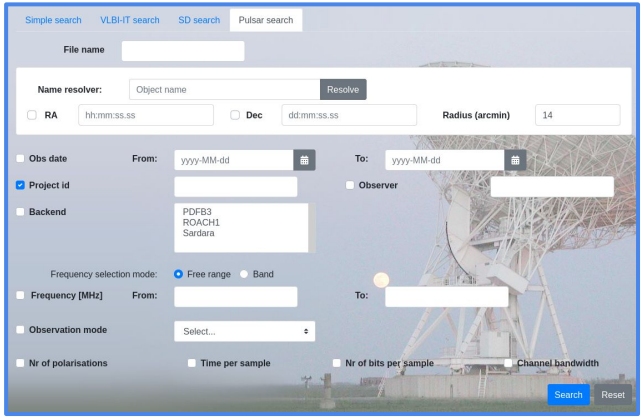
ObsCore mapping proposal

OBSCore keyword	ObsCore Unit (Type)	Value (expressions are based on fields of the table pulsar; db: metadata_pulsar)	Comments
facility_name	Unitless (string, following a list curated by INAF)	TELESCOP E.g.: SRT	
instrument_name	Unitless (string, following a list curated by INAF)	FRONTEND,BACKEND E.g. : 'LLP,PDFB3'.	
target_name	Unitless (string, following an internal convention)	SRC_NAME E.g.: 'B1929+10_R'	"_R" identifies a calibration observation (noise diode on)
target_class	Unitless (string)	TBD	A regular expression may be used for identifying known pulsars and FRBs (e.g., for FRBs, the target_name generally reports the form **FRB**yymmdd)
obs_title	Unitless (string)	TBD	

Summary of ongoing activities & future plans

- System ready to ingest FILTERBANK file
- Tests for optimizing ingestion procedures
- Adoption of ObsCore DM (and evaluating CAOM)





- Testing and upgrades of web interfaces in order to enhance data discovery, access and retrieval (e.g. additional query fields and columns for results, more readable layout for displaying results)
- Restructuring the currently available TAP service (by complementing available information and, more generally, simplifying the database schema)
- Future plans to deliver calibrated as well as more advanced data products