IVOA Obscore Extension for Radio data Version 1.0

IVOA Working Draft 2023-09-29

Working Group
Data Model Working Group
This version
https://www.ivoa.net/documents/ObsCoreExtensionForRadioData/20230929
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

Abstract

This is a proposed extension to the Obscore specification for description of radio data.

Status of this document

This is an IVOA Working Draft for review by IVOA members and other interested parties. It is a draft document and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use IVOA Working Drafts as reference materials or to cite them as other than “work in progress”.

A list of current IVOA Recommendations and other technical documents can be found at https://www.ivoa.net/documents/.

Contents

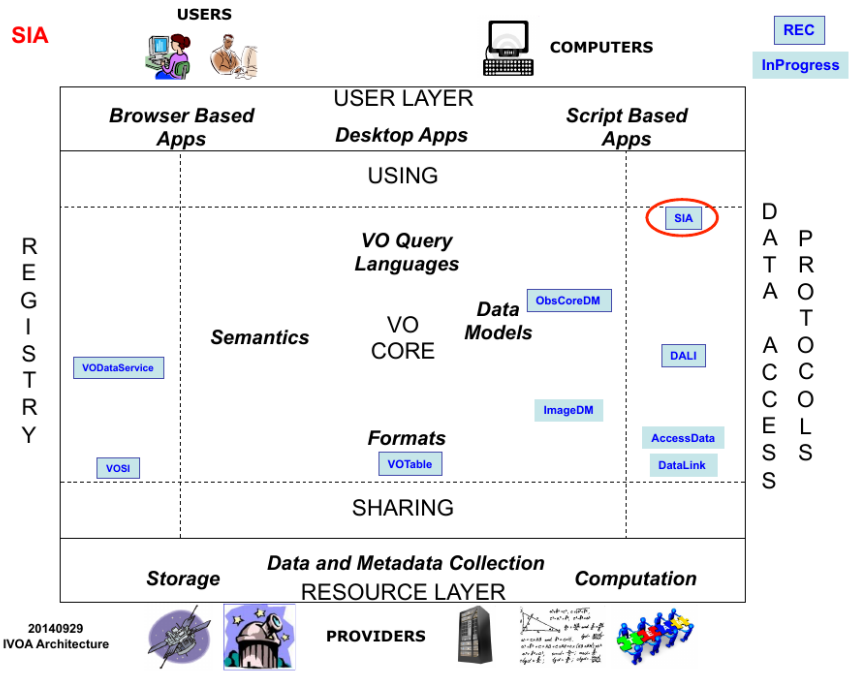
Acknowledgments

The authors would like to thank all the participants in DM-WG and Radioastronomy-IG discussions for their ideas, critical reviews, and contributions to this document. We acknowledge also the support of ESCAPE (European Science Cluster of Astronomy and Particle Physics ESFRI Research Infrastructures) funded by the EU Horizon 2020 research and innovation program (Grant Agreement no 824064).

1 Introduction

ObsCore specification (Louys and Tody et al., 2017) defines both a minimal datamodel to describe datasets and a table consistent with the model which can be served by TAP services. It has been successful to define a lot of data discovery services in astronomy.

The emergence of the Radioastronomy Interest Group in the IVOA in April 2020 confirmed the strong interest of the radio astronomy community to distribute their data in the VO. Many teams now distribute their data using VO standards[1]. While reduced radio data products, such as images or spectral cubes, are mostly covered by the ObsCore model, the lower level observational data (interferometric visibilities, single dish data in SDFITS, filterbank or whatever other specific formats) require additional description parameters. For interferometry, this was already exposed in 2010 when Anita Richards wrote a note called "Radio interferometry data in the VO" [2] which is still worth reading. Among other goals, the current specification tries to cover the needs exposed in this note. With the expansion of large radio astronomy projects such as LOFAR, NenuFAR, the future SKA, etc. and the emergence of interesting research topics matching data in all electromagnetic regimes, the Virtual Observatory framework can facilitate a wider access to radio data for experts and non-specialists radio astronomers in order to support collaborations in multi-wavelength, multi-messenger astronomy.



---

[1] https://ivoa.net/documents/Notes/RadioVOImp/index.html

[2] https://wiki.ivoa.net/internal/IVOA/SiaInterface/Anita-InterferometryVO.pdf

The scope of this Radio ObsCore extension in the IVOA is very close to ObsCore itself. Its goal is to add new specific features to the existing ObsCore metadata tables and expose them in the ObsTAP TAP_SCHEMA.

2 Radio data specifities from the Data Discovery point of view

On the lower end of the radio spectrum, radio astronomers generally make use of frequencies for designating the spectral ranges of their observation. The standard ObsCore attributes em_min em_max are in wavelength and are not really convenient. That's why we should also provide their translation into frequencies.

Receivers with a (ultra)wide bandwidth, up to tens of GHz, are nowadays commonly used for both interferometric and Single Dish (SD) radio observations. Given that the spatial field of view and resolution linearly depend on wavelength, these quantities may significantly vary across the observed bandwidth in a radio observation. Generally only a representative value (for instance the median) for these two parameters can be given. It is noticeable that this is the case for any measuring system allowing a large interval of λ/D (where λ and D are the wavelengths and the measuring system aperture scale).

Similarly, the resolution power quantity, commonly provided to describe optical spectroscopic data, does not make much sense in the radio domain and it is generally not used. To properly represent radio data it would be necessary to introduce a new ObsCore term for the absolute spectral resolution in frequency unit, for which a representative value for each observation can be given.

Modern radio instrumentation offer the possibility of n different spectral windows within the same observation with significant separation or different resolutions. Such observations may be represented at the highest granularity as many entries in an ObsCore Table. However it's up to data provider to decide which level of granularity is best adapted in order to optimize data discoverability and ease data access, depending on the scientific content of the observation (see Sect. 2.1 for an example).

2.1 Single dish data

Single Dish observations can be done with different types of receiving systems. Typical frontends are mono-feed, multi-feed and phased array feed (PAF), the last two suitable to efficiently span wider parts of the sky. Data can be acquired by various backend systems providing either the total intensity (integrated over the whole available bandwidth) or the spectroscopic/spectropolarimetric intensity (acquired in each spectral channel/sample).

Data are saved as raw counts resulting from the digitisation of the voltage signal measured by the receiving system. Commonly-used SD data formats are registered FITS standard conventions (FITS, SDFITS and MBFITS) or unregistered conventions like the standard FITS-based format delivered by the INAF radio telescopes.

The combination of telescope, frontend and backend permits the realization of various observing strategies characterized by specific spatial and/or spectral patterns. Typical SD observing strategies are: on source, frequency switching, ON-OFF observations, raster or on-the-fly (OTF) mapping, raster or OTF cross-scan, skydip calibrations, see Fig 1. For each spatial position in the observation, SD data gather emission for any of the spectral samples in the given frequency band and polarization. If multi-feed/PAFs are used, a set of spatial positions are simultaneously measured. Scan modes should be described in ObsCore in order to allow astronomers to better understand the structure of the data which will be retrieved.

Spatial resolution in the SD case is intended as the beam size. This holds true for any type of receivers, since also for multi-feed/PAF ones the spatial resolution is ruled by the size of the individual beam.

Contrary to what usually happens for interferometric observations, in the case of INAF radio telescopes a SD observation (scan) contains only one scientific target. In any case, each target in an observation is listed as a separate entry in an ObsCore Table sharing the same obs_id.

Complex frequency setups are possible in the same observation, as already mentioned in Sect. 2.

The ObsCore definition of t_resolution as the minimal interpretable interval between two points along the time axis (being it an average or representative value) is generally not applicable for SD data. Typically time is not an independent variable in SD measurements, it can be saved together with spatial/spectral/intensity information as a timestamp associated to each data sample. Even in the case of on-source tracking, time information in SD data is not intended for time domain studies.

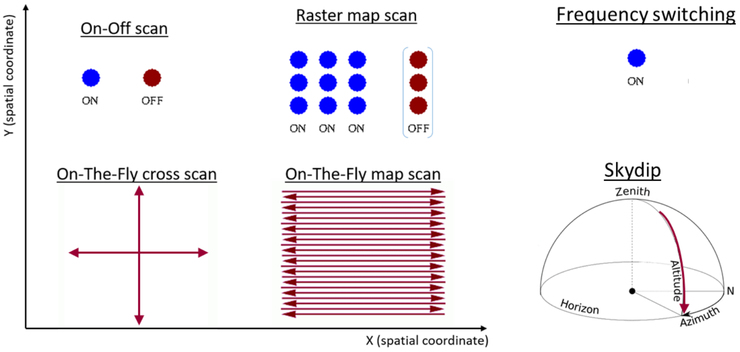


Figure 1: Single Dish Observation scan modes

2.2 Visibility data

Visibility data are sets of complex numbers corresponding to the amplitude and phase of correlation coefficients measured between pair of antennas (i.e., a baseline), at a given time, a given wavelength or polarisation. The visibility data are a sparse representation of the observed sky. The visibility data sets can be processed to obtain interferometric images, through inverse Fourier algorithms. Each visibility measurement corresponds to an interferometric fringe system on the sky.

The imaging algorithms include a calibration step allowing to set the center of the reconstructed image, setting this direction as a phase reference. The visibilities are then usually represented in a spatial frequency plane, called the uv plane, whose orientation is perpendicular to phase reference direction. The instantaneous PSF (Point Spread Function) of an interferometer is the Fourier transform of all baselines sampled in the uv plane. Hence, the quality of the reconstructed images are directly related to the set of baselines used for the measurements.

Visibility data are usually organised as sets of matrices for various phase references (i.e., pointing, or fields) and configuration of the baselines, such as their distances and orientations. Such matrices may or may not be regularly sampled in time, wavelength and polarisation.

As for any other observation product described with ObsCore, the description may be split into several records in the ObsCore table, when ObsCore parameters cannot represent the variety of the observation results coverage (e.g., if there are several observed “fields”, requiring different s_ra and s_dec value, or various groups of spectral bands, etc.)

We consider that consistent ObsCore records as described above defines datasets with a dataproduct type set to “visibility”.

Contrary to what occurs with direct imaging observations, the PSF of the interferometer is filtering spatial scales (large scales, when the small baselines are insufficiently sampled; and vice versa for small scales with long baselines). For large spectral ranges, the variations of the field of view and the spatial resolution along the axis may become so large that the typical value cannot be sufficient to characterize the dataset. Ranges of values for such parameters are required to accurately describe such datasets.

The quality of the data strongly depends from the distribution of the visibility measurements in the uv plane : the more complete the uv sampling plane, the better the reconstructed image. The uv plane distribution can be characterized by several numbers. The minimal and maximum distance between measurements in the uv plane provide assessments for spatial resolution and largest angular scale. Beside this a uv plane filling factor of the distribution will allow to predict the quality of reconstruction of the image in the distance plane (sky). Eventually, the ellipticity of the distribution is a measure of the distortions that can affect the reconstruction.

Radioastronomers also check the quality of the visibility data by looking at some maps of the data structure. The uv coverage map can show how complete and regular is the sampling in the *uv* plane and give an hint of resolution and maximum angular scale. The visualisation of the dirty beam, which is the Fourier transform of the *uv* sampling function gives an hint of the intrinsic quality of possible reconstruction, as maps they are not queriable. So it is questionable if links to these maps are to be exposed in the extension table or only via a DataLink service.

If none of these *uv* characterization features are available to be exposed in the service we can still predict ranges of some of those by using some kind of facility description. Important features are the antenna diameter (or maximum antenna diameter), the number of antennas and the minimum and maximum distance between antennas of the array.

In addition to these specifities most of the scan modes shown on figure 1 also apply to some interferometry observations and should be described.

3 ObsCore attributes definition valid for radio data

For radio data some of the definitions on Obscore datamodel elements need to be adjusted to fit the peculiarity of medata for datasets partition, uv space, etc.

3.1 obs_id

Astronomers usually know what they identify as a single observation : a complex set of measurements made in a given sequence of time. obs_id should define unambiguously each observation.

3.2 obs_publisher_did

Radio data observations can be split in several subparts with homogeneous spatial, time, spectral coverage intervals, spectral resolution, etc. Each part can be described by a single ObsCore dataset and has its own obs_publisher_did. It has to be unique in the Virtual Observatory domain.

3.3 s_fov

A typical value for the field of view size is to be computed on the observation by taking into account the sky scan geometry and receiver type in use. s_fov coincides with the instantaneous field of view λ/D only for pointed observations (for instance, an ON in the SD case) obtained with a mono-feed receiver. In this case, λ is the mid value of the spectral range and D coincides with the telescope diameter (SD case) or the largest diameter of the array antennae or telescopes (interferometric case).

3.4 s_resolution

In the case of single dish using mono- or multi-feed/PAF receivers this is the beam size inferred from the wavelength and telescope diameter. In the case of interferometry, a typical value for the spatial resolution will be given by λ/L where λ is the mid value of the spectral range and L is the longest distance in the *uv* plane. For beamforming applied to SD s_resolution is set by the size of one individual electronically-formed beam, while in the interferometric case it is ruled by the maximum distance among the stations.

3.5 s_region

For single dish data it will strongly depend on the scanning mode and type of receiver in use. This shape will be the typical contour of the detectable beam for interferometry. Of course it cannot be accurate.

3.6 o_ucd

In the current UCD vocabulary (UCD1+ controlled vocabulary - Updated List of Terms Version 1.5) there appear to be no primary words suitable to describe raw SD data. o_ucd=phot.flux.density does not seem appropriate, since the single dish measured quantity is expressed in raw counts coming from the digitisation of a voltage signal generated in the receiver chain by the incoming electromagnetic field. Further discussion on o_ucd is ongoing within the Semantics WG.

In the case of visibility data the "observable" is a complex number representing Fourier coefficients of the image Fourier transform. Its UCD string is *stat.fourier*.

3.7 t_exptime

Total duration of the observation for the given dataset/ObsCore entry. Example: in case of multiple targets, t_exptime will be computed for each source and written in the appropriate ObsCore Table entry.

3.8 t_resolution

Not applicable for single dish data (see Sect. 2.1). For interferometric observations it is the integration time set at the correlation level.

4 ObsCore extension for radio data

Table 1 shows the parameters we propose to add to ObsCore in order to better describe radio single dish and visibility data. The last column indicates if the attribute is useful for all radio datasets or only for visibilities, beamforming, or single dish data. Two options can be considered in the TAP or SIA services descriptions:

- adding the new data model elements directly into the main ObsCore table
- providing an extra table for these, named ivoa.visibilities for instance, which will be joined to the main table.

4.1 spatial parameters

s_fov_min, s_fov_max are estimated like the typical value (see subsection 3.3). In the case of SD pointed observations with mono-feed receivers and the majority of interferometric observations the minimum and maximum λ values in the spectral range(s) will be used in the formula. In the case of mapping scans or multi-feed/PAF receivers s_fov_min and s_fov_max are derived as the minimum and maximum sizes of the circular region encompassing the covered area.

s_resolution_min, s_resolution_max are estimated like the typical value (see subsection 3.4) where λ is replaced by the minimum and maximum wavelength of the spectral range(s). The size D is the telescope diameter for

SD observations and the largest distance in the uv plane. Beamforming may represent an exception to this rule, see 3.4.

In the case of interferometry, the s_maximum_angular_scale is estimated as λ/l where λ is the typical wavelength and l is the smallest distance in the uv plane.

4.2 uv parameters

These parameters are valid for interferometry only.

uv_distance_min and uv_distance_max are evaluated by fitting an ellipse on the visibilities present in the uv plane.

To compute the ellipse's eccentricity of the UV distribution a principal component analysis (PCA) with 2 components is performed over the data points sampling the UV plane to select the main axis of data scattering. The first component is used to rotate the distribution of UV in a way that the major variation of the distribution is leaning towards the x axis of a bi dimensional xy Cartesian plane. The major axis length and the minor axis length of the ellipse are therefore defined as the semi distance between the most positive point along the x/y axis and the most negative point among the y axis. For instance, if the range of the rotated UV will cover on the $x \in [-10, 10]$ the major axis distance would be 10, a similar procedure is done on the y axis. This procedure allows the definition of the UV distribution eccentricity:

uv_distribution_exc) computed as follows:

$$uv\_distribution\_exc = \sqrt{1 - \frac{b^2}{a^2}} \tag{1}$$

where a is the major axis length and b is the minor axis length. The filling factor of the UV plane (hereafter uv_distribution_fill) is computed as the average number of samples found in a $N^{uv}_{samples}$x$N^{uv}_{samples}$ equi-spaced grid enclosing the rotated ellipse. In formulas, the boundaries of a cell (i,j) are defined by the boundaries

$$u \in [u_{min} + \frac{u_{max} - u_{min}}{N^{uv}_{samples}} \cdot i, u_{min} + \frac{u_{max} - u_{min}}{N^{uv}_{samples}} \cdot (i+1)] \tag{2}$$

and

$$v \in [v_{min} + \frac{v_{max} - v_{min}}{N^{uv}_{samples}} \cdot j, v_{min} + \frac{v_{max} - v_{min}}{N^{uv}_{samples}} \cdot (j+1)] \tag{3}$$

where u_{max}/v_{max} are the respective maximum u/v of the uv sample and u_{min}/v_{min} is the minimum u/v of the uv sample.

Given the above boundaries the number of samples within a cell (i,j) will be $n^{uv}_{i,j}$ and uv_distribution_fill will be then computed as

$$uv\_distribution\_fill = \frac{\sum_{i=1}^{N^{uv}_{samples}} \sum_{j=1}^{N^{uv}_{samples}} n^{uv}_{i,j}}{(N^{uv}_{samples})^2}, \tag{4}$$

in the preliminary analysis $N^{uv}_{samples} = 1000$.

4.3 time parameters

t_exp_min, t_exp_max and t_exp_mean are added in the case of variation in the individual timestamps duration. This is usually not the case for SD data and these parameters will be set to NULL.

4.4 Observation modes and instrumental parameters

These parameters are intended to describe the main telescope(s) characteristics for both SD antennas and interferometers. Such instrumental characteristics give an approximate idea on the spanned angular scales, field of view, etc.

Parameters instrument_ant_number, instrument_ant_min_dist and instrument_ant_max_dist are related to interferometers only while instrument_ant_diameter, instrument_feed are valid also for SD. We note that instrument_feed could be changed to instrument_beam to account for the number of beams in the case of a beamforming/PAF receiver.

The scanning strategy adopted in an observation is described by the parameter sky_scan_mode. We note that it would be better to adopt one of the following solutions: either use scan_mode to include both spatial and frequency scanning modes, or leave sky_scan_mode for spatial scanning while a second parameter frequency_scan_mode should be added for frequency scanning stategy (see Sect. 2.1). Scanning parameters are applicable to both SD and interferometry.

4.5 uv coverage and dirty beam map

Parameters uv_distribution_map and s_resolution_beam_dirty are intended to be url to files containing these maps. Implementers may want to avoid adding url columns to the ObsCore table.

In that case DataLink (Dowler and Bonnarel et al., 2015) may provide a solution. The semantics FIELD in the {link} response will contain #auxiliary for links to this map while the content_qualifier FIELD could contain the utype defined here in this ObsCore extension.

5 How to implement the extension in a TAP service

The ObsCore extension for radio (including or not visibility data) described above SHOULD not be added to the main ObsCore table. An extension table called "radioObscore" SHOULD be added to the same schema instead. The two tables will be joined in an extended ObsTAP ADQL query. A single dataset in each observation will be associated to a single row in ObsCore. It will be identified by a unique obs_publisher_did. This obs_publisher_did can be used as a foreign key to join the main table and the extension

In the registry, the two service tables will be described in the tableset of the service. They will show respectively the ObsCore Model utype an the radioObscore Model utype.

column name	definition	utype	ucd	unit	validity
`s_resolution_min`	`Angular resolution, longest baseline and max frequency dependant`	Char.SpatialAxis. Resolution.Bounds. Limits.LoLim	pos.angResolution;stat.min	arcsec	radio

`s_resolution_max`	`Angular resolution, longest baseline and min frequency dependant`	`Char.SpatialAxis.`
		`Resolution.Bounds.`
		`Limits.HiLim`
			pos.angResolution;stat.max	arcsec	radio

`s_fov_min`	`field of view diameter, min value, max frequency dependant`	`Char.SpatialAxis.`
		`Coverage.Bounds.`
		`Extent.LowLim`
			phys.angSize;instr.fov;
			stat.min
				deg	radio

`s_fov_max`	`field of view diameter, max value, min frequency dependant`	`Char.SpatialAxis.`
		`Coverage.Bounds.`
		`Extent.HiLim`
			phys.angSize;instr.fov;
			stat.max
				deg	radio

`s_maximum_angular_scale`	`maximum scale in dataset, shortest baseline and frequency dependant`	`Char.SpatialAxis.`
		`Resolution.Scale.`
		`Limits.HiLim`
			phys.angSize;stat.max	arcsec	interferometry

`f_min`	`spectral coverage min in frequency`	`Char.SpectralAxis. Coverage.Bounds Limits.LoLim`	em.freq;stat.min	Mhz	radio
`f_max`	`spectral coverage max in frequency`	`Char.SpectralAxis. Coverage.Bounds Limits.HiLim`	em.freq;stat.max	Mhz	radio
`t_exp_min`	`minimum integration time per sample`	`Char.TimeAxis. Sampling.Extent LoLim`	time.duration;obs.exposure; stat.min	s	radio
`t_exp_max`	`maximum integration time per sample`	`Char.TimeAxis. Sampling.Extent HiLim`	time.duration;obs.exposure; stat.max	s	radio
`t_exp_mean`	`average integration time per sample`	`Char.TimeAxis. Sampling.Extent HiLim`	time.duration;obs.exposure stat.mean	s	radio
`uv_distance_min`	`minimal distance in uv plane`	`Char.UVAxis. Coverage.Bounds. Limits.LoLim`	stat.fourier;pos;stat.min	m	interferometry
`uv_distance_max`	`maximal distance in uv plane`	`Char.UVAxis. Coverage.Bounds. Limits.LoLim`	stat.fourier;pos;stat.max	m	interferometry

`uv_distribution_ecc`	`eccentricity of uv distribution`	`Char.UVAxis.Coverage.Bounds.Eccentricity`	stat.fourier;pos		interferometry
`uv_distribution_fill`	`filling factor of uv distribution`	`Char.UVAxis.Coverage.Bounds.FillingFactor`	stat.fourier;pos;arith.ratio		interferometry
`instrument_ant_number`	`number of antennas in array`	`Provenance.ObsConfig.Instrument.Array.AntNumber`	meta.number;instr.param		interferometry, beamforming
`instrument_ant_min_dist`	`minimum distance between antennas in array`	`Provenance.ObsConfig.Instrument.Array.MinDist`	instr.baseline;stat.min	m	interferometry
`instrument_ant_max_dist`	`maximum distance between antennas in array`	`Provenance.ObsConfig.Instrument.Array.MaxDist`	instr.baseline;stat.max	m	interferometry
`instrument_ant_diameter`	`diameter of telecope or antennas in array`	`Provenance.ObsConfig.Instrument.Array.Diameter`	instr.param	m	radio
`instrument_feed`	`number of feeds`	`Provenance.ObsConfig.Instrument.Feed`	instr.param		radio

sky_scan_mode	scan mode (on-off, raster map, on-the-fly map,...)	Provenance. Observation. sky_scan_mode	instr.param	radio
uv_distribution_map	uv distribution map	Char.UVAxis. Sampling. Sensitivity.Map	stat.fourier;pos;meta.ref.url	interferometry
s_resolution_beam_dirty	dirty beam	Char.SpatialAxis. Resolution. Variability.DirtyBeam. Map	pos.angResolution;instr.beam; meta.ref.url	interferometry

Table 1: ObsCore radio data extension parameters proposal

References

Dowler, P., Bonnarel, F., Michel, L. and Demleitner, M. (2015), 'IVOA DataLink Version 1.0', IVOA Recommendation 17 June 2015, arXiv:1509.06152.
http://doi.org/10.5479/ADS/bib/2015ivoa.spec.0617D

Louys, M., Tody, D., Dowler, P., Durand, D., Michel, L., Bonnarel, F., Micol, A. and IVOA DataModel Working Group (2017), 'Observation Data Model Core Components, its Implementation in the Table Access Protocol Version 1.1', IVOA Recommendation 09 May 2017.
http://doi.org/10.5479/ADS/bib/2017ivoa.spec.0509L