IVOA Obscore Extension for Radio data Version 1.0

IVOA Note 2022-10-14

Working group

Data Model Working Group

This version

http://www.ivoa.net/documents/ObsCoreExtensionForVisibilityData/20221014

Latest version

http://www.ivoa.net/documents/ObsCoreExtensionForVisibilityData

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 Note expressing suggestions from and opinions of the authors. It is intended to share best practices, possible approaches, or other perspectives on interoperability with the Virtual Observatory. It should not be referenced or otherwise interpreted as a standard specification.

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

Contents

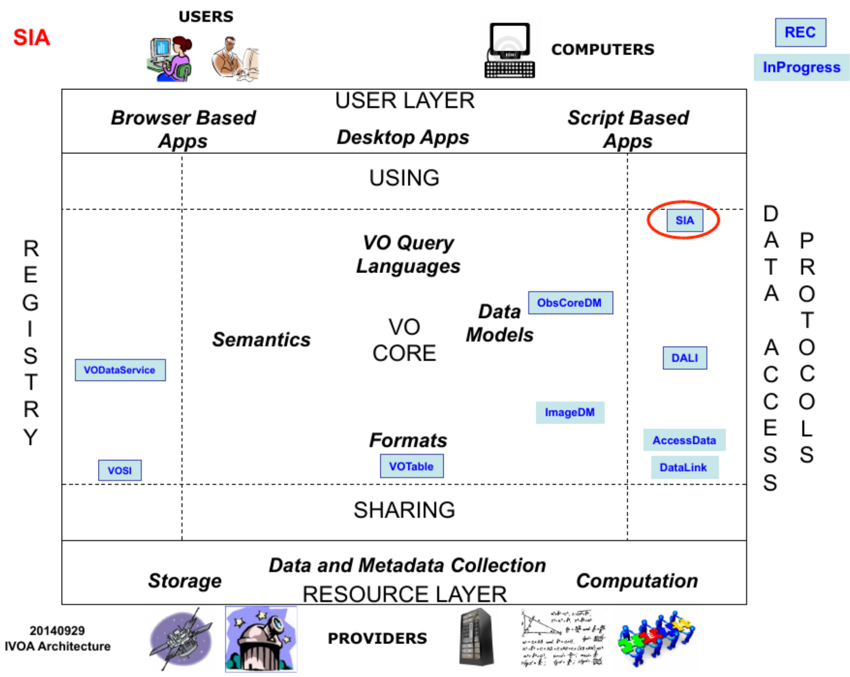
Acknowledgments

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 Bonnarel et al., 2011) 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 format or whatever) 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.

Moreover radio data show some dependency of the spatial field of view and resolution with wavelength. Generally only a typical value for these two numbers 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).

For the same reason the resolution power quantity doesn't make much sense for the whole dataset. Only a typical one can be given It would be better to introduce the constant absolute resolution in frequency unit.

2.1 Single dish and beam forming data

Typically Single Dish data gather emission in a central beam for each spectral sample in the spectral band and polarization. They provide spectra (the special case where observers span the time axis, record time stamps and thus produce dynamical spectra is tackled in an other document). In a spectrum case the spatial "field of view" and spatial "resolution" may be considered as the same concept because it's not possible to distinguish two sources inside the beam. Another perspective on this is to consider that spatial resolution doesn't make sense for spectra.

However in many cases more sophisticated observing modes allow to span wider parts of the sky. Fig 1 shows some of these modes.

It is important to add these observing scan modes in the dataset description in order to allow astronomers to better understand in advance the structure of the data which will be retrieved in the typical SDFITS format (or any other adapted format such as set of related FITS tables or whatever).

Observing configuration features such as multi-feed receptors available on the telescope (allowing multi spectra observation at the same time) should also be given.

In the case of beamforming data obtained with an array of antennae, "beamforming" should also be considered as an observing configuration feature.

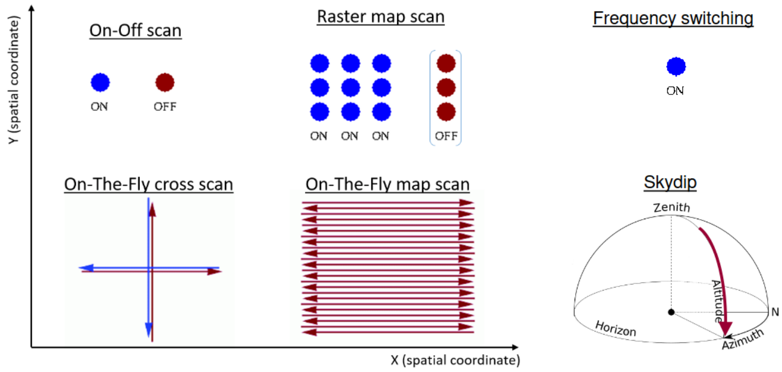


Figure 1: Single Dish Observation Sky scan modes

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 antennae and the minimum and maximum distance between antennae of the array.

3 ObsCore attributes definition valid for radio data

For visibilities 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

Visibility 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 will be given by λ/D where λ is the mid value of the spectral range and D is the diameter of the telecope or the largest diameter of the array antennae or telescopes.

3.4 s_resolution

In the case of single dish this is the beam size inferred from the wavelength and 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.

3.5 s_region

For single dish data it will strongly depnd of the scanning mode. 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 case of single dish data it is generally a flux or flux density in Jy.

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

TBC

3.8 t_resolution

TBC

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, s_resolution_min, s_resolution_max are estimated like the typical values (see subsections 3.3 and 3.4) where λ is replaced by the minimum and maximum wavelength of the spectral ranges 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 and t_exp_max are added because of strong variation in the individual timestamps duration.

4.4 Observation modes, Observation configuration and instrumental parameters

They all give a predefined raw idea of resolution, field of view, maximum angular scale and sampling quality.

4.5 uv coverage and dirty beam map

These uv_coverage_map and s_resolution_beam_dirty parameters 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 joigned 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.AngResol;stat.min	arcsec	radio
s_resolution_max	Angular resolution, longest baseline and min frequency dependant	Char.SpatialAxis. Resolution.Bounds. Limits.HiLim	pos.AngResol;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
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_exc	excentricity of uv distribution	Char.UVAxis.Coverage.Bounds.Excentricity	stat.fourier;pos		interferometry

uv_distribution_fill	filling factor of uv distribution	Char.UVAxis. Coverage.Bounds. FillingFactor	stat.fourier;pos		interferometry
ant_number	number of antennae in array	Provenance.ObsConfig. Instrument.Array. AntNumber	instr.baseline;meta.number		interferometry, beamforming
instrument_ant_min_dist	minimum distance between antennae in array	Provenance.ObsConfig. Instrument.Array. MinDist	instr.baseline;stat.min	m	interferometry
instrument_ant_max_dist	maximum distance between antennae in array	Provenance.ObsConfig. Instrument.Array. MaxDist	instr.baseline;stat.max	m	interferometry
instrument_ant_diameter	diameter of telecope or antennae in array	Provenance.ObsConfig. Instrument.Array. Diameter	instr	m	radio
instrument_feed	number of feeds	Provenance.ObsConfig. Instrument.Feed	instr		single dish
sky_scan_mode	scan mode among on-off,raster map, on the fly map	Provenance. Observation. sky_scan_mode	instr		single dish

uv_distribution_map	uv distribution map	Char.UVAxis.Sampling.Sensitivity.Map	stat.fourier;pos	interferometry
s_resolution_beam_dirty	dirty beam	Char.SpatialAxis.Resolution.Variability.DirtyBeam.Map	pos.AngResol	interferometry

Table 1: ObsCore radio data extension parameters

References

Dowler, P., Bonnarel, F., Michel, L. and Demleitner, M. (2015), 'IVOA datalink', IVOA Recommendation 17 June 2015.
http://www.ivoa.net/documents/DataLink/

Louys, M., Bonnarel, F., Schade, D., Dowler, P., Micol, A., Durand, D., Tody, D., Michel, L., Salgado, J., Chilingarian, I., Rino, B., de Dios Santander, J. and Skoda, P. (2011), 'Observation data model core components and its implementation in the Table Access Protocol, version 1.0', IVOA Recommendation.
http://www.ivoa.net/documents/ObsCore/20111028/REC-ObsCore-v1.0-20111028.pdf