*I*nternational

*V*irtual

*O*bservatory

*A*lliance

# Astronomical Source Data Model

# Version 0.1

## *IVOA Document-Type 2016 April 30*


**This version:**

0.1-20160430

**Latest version:**

http://www.ivoa.net/Documents/latest/latest-version-name

**Previous version(s):**



**Author(s):**

Jesús Salgado
Markus Demleitner
Gerard Lemson


---

## Abstract


Catalogues of astronomical objects make up an important part of the Virtual Observatory (VO) and, historically, the most important type of data in astronomy. Almost since the beginning of the VO, the Simple Cone Search (SCS) standard provided a simple access protocol that relies in a very simplified data model that allows the inspection of this kind of data.

The IVOA Source Data Model defines a set of metadata common in the definition of astronomical source catalogues, extending and formalising the data model

implied by SCS with the overall goal of enabling a greater degree of automation in processing such catalogues.

This standard derives requirements on the source data model from a set of use cases and then discusses the data model derived in turn from the requirements so it can be used to homogenize the description of the metadata for different catalogues, mark up metadata into IVOA DAL protocol responses or to help in the definition of the common metadata and structure that any astronomical catalogue could/should have.

Two main classes will define the core of present DM; Source, linked to a single/multiple detections within a certain observation, and Astronomical Object that could incorporate intrinsic fields of the real entity.

## Status of This Document

This is a Document-Type. The first release of this document was 2016 April 30.

*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 http://www.ivoa.net/Documents/.*

## Acknowledgements

"Ack here, if any"

# Contents

# 1 Introduction

From ancient times, one of the more important outputs of science is the production of astronomical catalogues. Star catalogues were produced by Babylonians, Greeks, Chinese and Arabs among others. When the observation techniques and classification techniques were improved, catalogues started to include also other fainter astronomical sources like galaxies, quasars, etc

Originally, the typical metadata that were included in these catalogues were: an index, a sky position and, if possible, the magnitude of the object, as a comparison in flux with another source or calibration.

When more than one detection is possible, it was also possible to add to the metadata the description of the movement of the sources on the celestial sphere (proper motion) that can be considered the transverse velocity as seen from the center of mass of the Solar System. Using, in general, spectroscopy data from the source, the radial velocity of the source as seen from the Solar System can be also evaluated. Doing that, the phase space of every source can be set.

The first IVOA protocol to provide support to catalogues of astronomical sources was the Simple Cone Search standard (SCS). This protocol was very early in the development of the Virtual Observatory introducing a simple but effective model consisting of:

- A string-valued identifier (compulsory)
- A ``main'' position, given as ICRS Right Ascension and declination in two floating-point fields (compulsory)
- Arbitrary further information (optional)

The individual items are identified through Unified Content Descriptors (UCDs). To enable limited machine interpretation of the further information, SCS relies on the mechanisms defined by VOTable, in particular again UCDs, but also units, description, etc.

While SCS has served the VO community well over more than a decade, several important use cases are not sufficiently covered by it, and the limited amount of annotation precludes many applications that require machines to have a deeper understanding of the catalogue's contents.  This standard intends to supply the missing elements.

We begin by sketching the motivating use cases and derive requirements from them in the requirement section.  Based on this analysis, in data model we give the UML based definition of the SourceDM components.  We conclude with considerations on phasing in SourceDM usage and possibilities to automatically infer rudimentary SourceDM annotation from legacy VOTable annotation.

# 2 Role within the VO Architecture

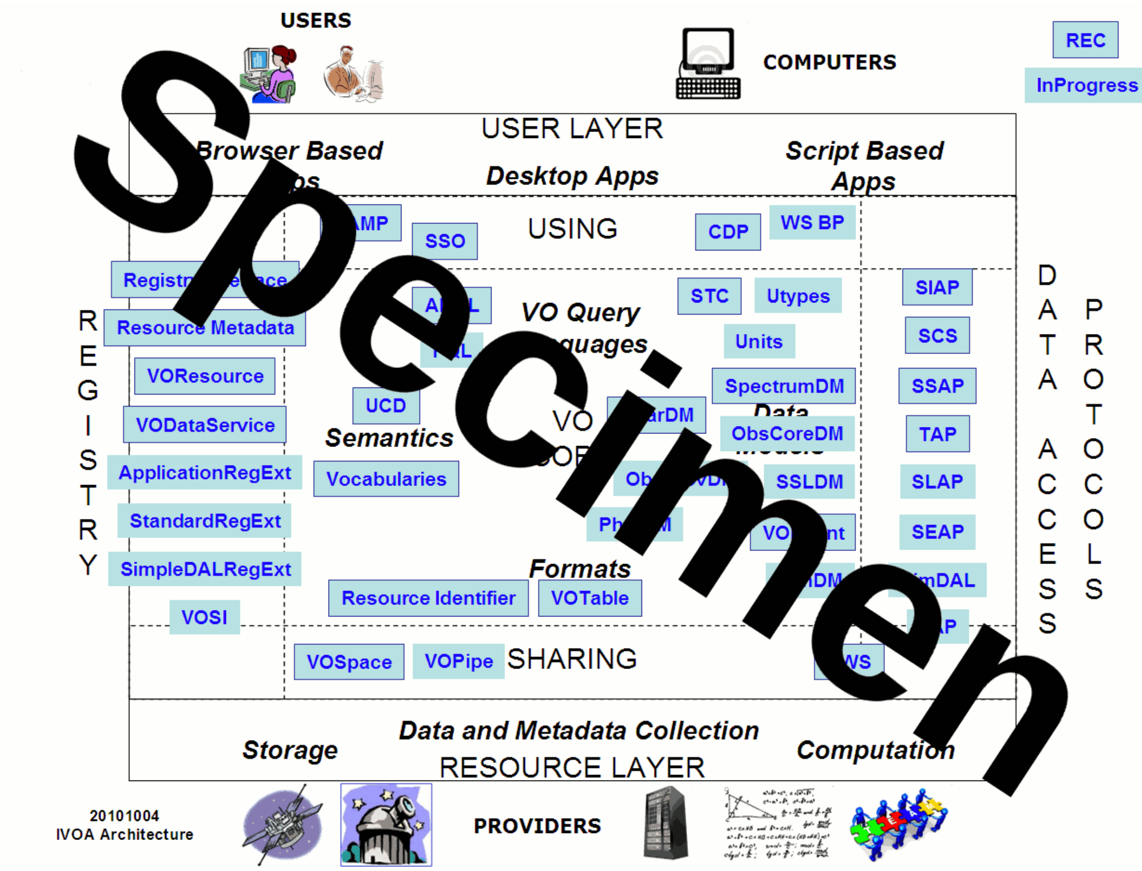




Figure 1: Architecture diagram for SourceDM

Fig 1 shows the role SourceDM plays within the IVOA architectur. It depends on the following standards:

- **VO-DML:** SourceDM is written in VO-DML, a standard language to define data models, essentially providing an easy-to-process subset of the common modelling language UML sufficient to express data models useful in the Virtual Observatory context.
- **Mapping**: The Mapping document defines how to actually annotate catalogues written in VOTable (or possibly other formats in the future) with the concepts defined in this document.
- **PhotDM**: The Photometry data model is re-used by SourceDM to express photometric quantities.

# 3 Use Cases

Since a full model for an astronomical source would presumably result in a model of a large part of astronomy and physics, it is important to be guided by use cases when delimiting the scope of the model and formulating requirements on the model.  This subsection will list the use cases the community agreed upon.

We use a ``want to'' form for the use cases, with two parties involved: Data providers who annotate their catalogues, and authors of programs or libraries (``clients'') dealing with the annotated data.

## *3.1 Standard Locations*

Client authors want to to take SourceDM-annotated tables and easily provide plots of the locations of the contained objects in charts or over astrometrically calibrated images, preferably without having to implement a large number of coordinate transforms; this is essentially what tools can do right now with the ICRS positions given in SCS responses.

## *3.2 Identities*

Catalogue annotators want to express whether or not two instances (``rows'') refer to the same physical object on the sky.  Client authors want to be able to match known-identical sources, for instance to aggregate information from several different catalogues.

## *3.3 Native Frames*

Catalogue annotators want to give coordinates in their catalogue's native frames and for whatever epoch is convenient for them in order to achieve maximum precision, avoid hard-to-control error propagation in transformation requiring additinal information, and keep provenancesimple. This becomes more important as we have more and more data noticeably distant from the J2000 epoch most of today's major surveys are close to, and more and more data precise enough that for a large number of objects proper their motions since J2000 become relevant for daily work.

## *3.4 Precise Astrometry*

Client authors want to recognise all information in catalogues relevant to precision astrometry (e.g., distances or radial velocities as necessary in foreshortening calculations).  This also involves the position of the observer, at least for a few standard reference positions.  Of course it would be beneficial to be able to express actual observer positions, but in order to keep the model manageable, we suggest deferring this to provenance.

### 3.5 Crossmatching

Client authors want to implement smart crossmatching schemes between different catalogues. This includes reducing catalogues to common reference frames and epochs, but also unifying non-positional source characteristics (e.g., photometry). In the presence of crossmatching schemes based on principled statistical approaches, they want to identify data aiding them in the estimation of the relevant distributions.

### 3.6 Combining Data

Following a crossmatch (or just within a single table), tool authors want to aggregate various pieces of information into new data points or products.

Examples include:

- Colours or even SEDs from photometry points
- Proper motions or even orbits from positional measurements
- Time series from, say, radial velocity measurements

### 3.7 Automatic Errors

Where catalogues contain the necessary information, tool authors want to provide sound error estimates for derived quanities as discussed in section 3.6

### 3.8 Annotation of Existing Catalogues

Catalogue annotators do not want -- and indeed, in general cannot change catalogue data. SourceDM must therefore be flexible enough to enable the meaningful annotation of a large majority of the catalogues already in the VO.

### 3.9 Standard Properties are Annotated

Client authors want to find scientifically relevant source properties easily. These include:

- Source type
- Position
- Radial Velocity
- Proper motion
- Parallax/distance
- Classification
- Photometry
- Redshift

# 4 Requirements

This section derives concrete requirements from the use cases collected in section 3. Some requirements are formulated using ``should'' phrases. ``Should'' as used in such requirements is not to be understood in the sense of RFC2119.

## *4.1 General Requirements*

Most of the use cases of course require basic metadata necessary to interpret physical quanitites, in particular their units.

“Identities” requires a concept of a physical object that a source refers to. It also means that in relational representations of collections of SourceDM instances, the physical object's identifier should not be the primary key.

Regarding ”Standard properties are annotated”, UCDs may be sufficient to annotate the concepts listed. Thus, at this point we derive no requirements from this use case.

## *4.2 Requirements on Space and Time Metadata*

“Standard locations” use case suggests that SourceDM instances should maintain the SCS practice of including ICRS positions wherever possible. Given \usecase{precise-astrometry}, it is highly desirable that these positions are nevertheless properly annotated, in particular with information on the epoch for which the position is pertinent, and where available the reference position.

“Crossmatching” use case means that the following spatial concepts should be present in the model (even though of course many catalogues cannot fill all of them):

- Positions (given the context, spherical positions are probably sufficient in the model)
- Derivatives of positions (i.e., proper motions)
- Errors in the positions and their derivatives
- Epochs of the positions

In view of “Precise Astrometry” use case together with “Native frames” use case, SourceDM should enable specifying frames with enough detail that sufficiently capable clients can use native frames for crossmatching. Less capable clients could still crossmatch at least partially using the standard locations from

“Standard locations” use case is that a single source should admit multiple positional specifications. In order to not leave the set of SourceDMrepresentable instances in “Combining data” use case, SourceDM should allow that these multiple specifications refer to different epochs.

“Precise astrometry”use case requires the third dimension, i.e.,

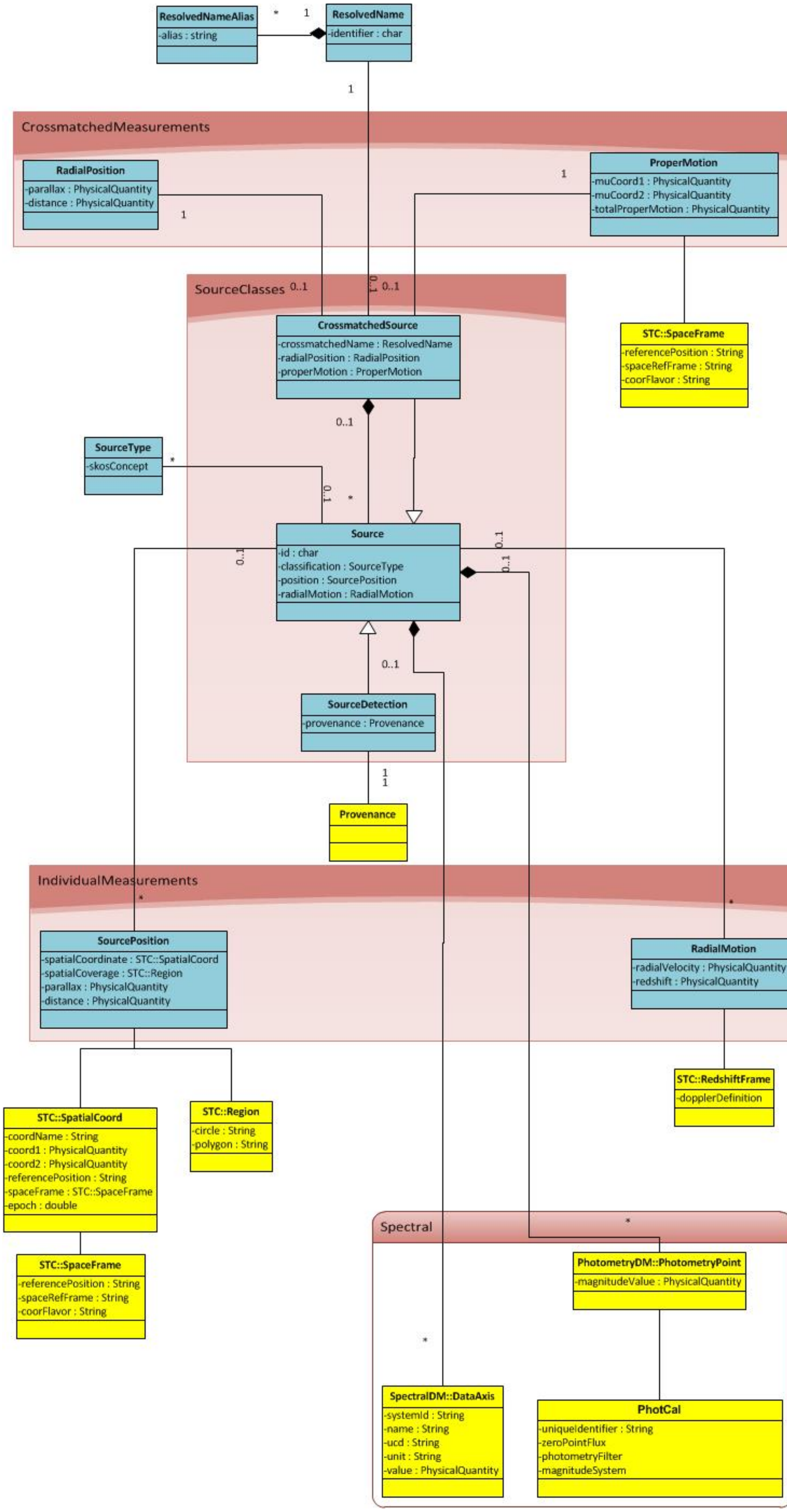
- Distance
- Its derivative
- Errors on both quantities

Taking “Annotation of existing catalogues” into account, various forms of distance specification need to be admitted; in particular, many important catalogues give parallaxes rather than distances. Also note that the derivative of the distance, the radial velocity, will typically given in spatial coordinates even if the distance is given as (angular) parallax.

In extragalactic astronomy, redshift is frequently used as a proxy for distance, and it is of course also a measure for the radial velocity. We still propose to not allow cosmological redshifts for distance or the derivative of the distance. In the nearby universe, the presence of a random motions (which might indeed be the actual objectiv a a catalog) makes this undesirable, in the more distant universe the interpretation of the term distance becomes difficult and should be avoided in favour of the redshift itself. Of course, where authors actually give redshift-based distances, these would be annotated as such.

One consequence of the “Identities” use case is that SourceDM should allow statements of the type ``this row corresponds to source *identifier-cat1* in catalogue 1 and to source *identifier-cat2* in catalogue 2''.

“Automatic-errors” use case requires suffient information on the errors. In the astrometric case, this in particular concerns the mean epochs, i.e., the epoch at which the covariance matrix of positions and proper motions becomes diagonal. The mean epochs in general are different on two coordinates.

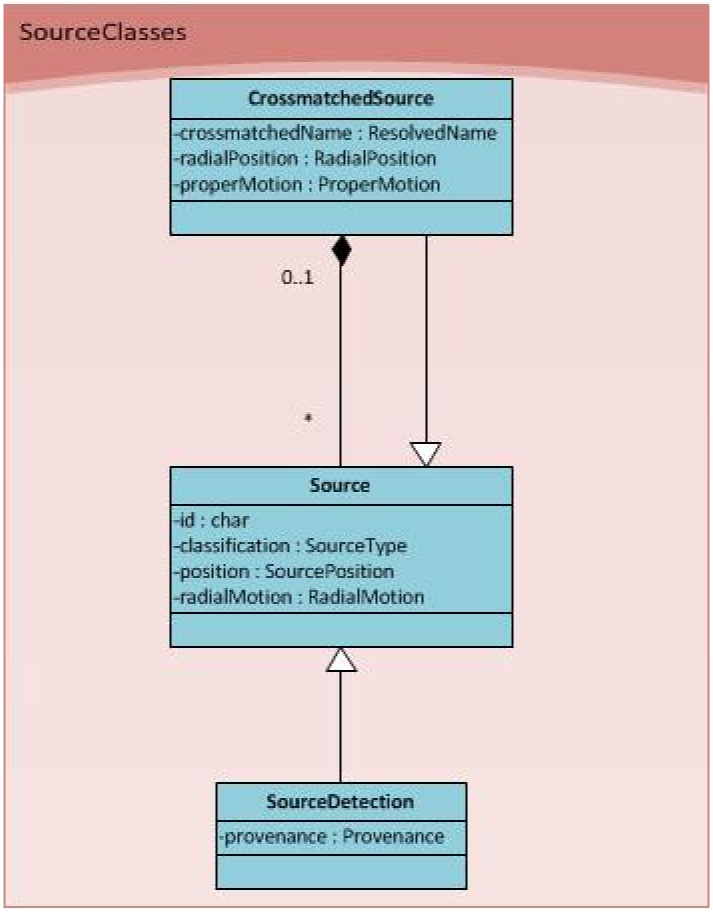


ResolvedNameAlias
-alias : string
ResolvedName
-identifier : char
CrossmatchedMeasurements
RadialPosition
-parallax : PhysicalQuantity
-distance : PhysicalQuantity
ProperMotion
-muCoord1 : PhysicalQuantity
-muCoord2 : PhysicalQuantity
-totalProperMotion : PhysicalQuantity
SourceClasses
CrossmatchedSource
-crossmatchedName : ResolvedName
-radialPosition : RadialPosition
-properMotion : ProperMotion
STC::SpaceFrame
-referencePosition : String
-spaceRefFrame : String
-coorFlavor : String
SourceType
-skosConcept
Source
-id : char
-classification : SourceType
-position : SourcePosition
-radialMotion : RadialMotion
SourceDetection
-provenance : Provenance
Provenance
IndividualMeasurements
SourcePosition
-spatialCoordinate : STC::SpatialCoord
-spatialCoverage : STC::Region
-parallax : PhysicalQuantity
-distance : PhysicalQuantity
RadialMotion
-radialVelocity : PhysicalQuantity
-redshift : PhysicalQuantity
STC::RedshiftFrame
-dopplerDefinition
STC::SpatialCoord
-coordName : String
-coord1 : PhysicalQuantity
-coord2 : PhysicalQuantity
-referencePosition : String
-spaceFrame : STC::SpaceFrame
-epoch : double
STC::Region
-circle : String
-polygon : String
STC::SpaceFrame
-referencePosition : String
-spaceRefFrame : String
-coorFlavor : String
Spectral
PhotometryDM::PhotometryPoint
-magnitudeValue : PhysicalQuantity
SpectralDM::DataAxis
-systemId : String
-name : String
-ucd : String
-unit : String
-value : PhysicalQuantity
PhotCal
-uniqueIdentifier : String
-zeroPointFlux
-photometryFilter
-magnitudeSystem

### 4.3 Data model summary

Present model have three main classes:

- **Source:** This is a mother class that contains the basic support for any element inside a source catalogue. That implies an identifier, a position and, possibly, a classification using a SKOS element.
- **SourceDetection:** Considered as a detection (or a crossmatched set of detections) of a statistically significant radiation excess above the background in a given sky position. This class would be the main one for most of the astronomical catalogues as it is the output result of the observational effort. That implies the link to IVOA observational DMs (e.g. Provenance or Characterization) could be needed to properly describe the observation where the source were detected.
- **CrossmatchedSource:** Some properties (e.g. proper motions and parallaxes) require a cross-identification of source detections. This is the case of different main catalogues of astronomy (e.g. Gaia) where different transient observations of different detections are consolidated in a single instance adding extra information. Also, metadata can be combined from different catalogues. This kind of objects are called crossmatched sources in the present DM.




Support for main coverage properties on the phase space (position and velocity) is divided into two group:

- **Individual Measurements:** These properties can be extracted from individual source detection. Two main axes are defined here: Position and, possibly, RadialMotion. This later one is usually an indirect measurement extracted from spectroscopic information.


- **Crosmatched Measurements:** These properties are derived after a crossmatch from different source detections from one or more catalogues. Two main axes are defined here (not compulsory): Proper motion and parallax/distance. Also, that requires a structure for the identifiers that could cover a main identifier and a set of aliases. In some cases, the radialPosition could be also calculated from individual measurements by using classification mechanisms and apparent versus total modeled magnitude (e.g. Cepheids)


Other related catalogues will be connected to characterize a Source class:

- **IVOA STC DM (Space/Time Coordinates Data Model):** This IVOA standard DM will be used to describe coordinates and velocities of the sources in a quite flexible and detailed manner.
- **IVOA PhotDM (Photometry Data Model):** Individual magnitudes and fluxes observed for these sources can be easily described by this DM
- **IVOA SpectralDM:** A more powerful and detailed description of the spectral behavior, including evolution of the emission in time, can be done through this DM
- **IVOA Char DM (Characterization Data Model)**

Other IVOA DMs will be linked through the Source DM but, due to the lack of a recommended version at this stage, the description of the elements used will be more described into the present standard to make it more self-consistent:

- Provenance DM

As the presence of errors for scientific magnitudes is a must, most of the magnitudes described in the present data model are of type PhysicalQuantity that is an element used on different IVOA data models and it is considered a common DM of the IVOA profile.

# 5 Source Data Model detailed description

## *5.1 Source*

Mother class that contains the basic support for any element inside a source catalogue. That implies an identifier, a position and, possibly, a classification using a SKOS element.

### 5.1.1 Source.id – Data Type: char

Unique identifier for this detection within the source detections catalogue.

### 5.1.2 Source.classification – Data Type: SourceType

Classification of the type of object with validity within the source catalogue. In order to maintain certain freedom for this classification, this is done through Simple Knowledge Organization System Reference (SKOS) concepts and capture similiraties on heterogenous thesauri, taxonomies, classification schemes that represent the source type classification within catalogues.

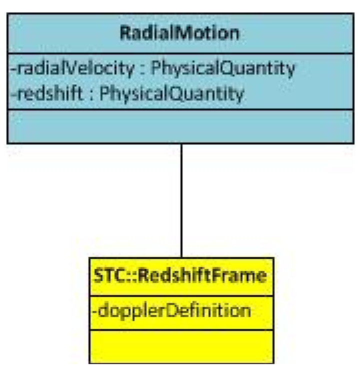
### 5.1.3 Source.position – Data Type: SourcePosition

Object describing the source position in the sky in the 3D space. It is described by a spatial coordinate object, reused from STC, a spatial coverage (specially for extended sources). See IVOA STC document for details.


### 5.1.4 Source.radialMotion – Data Type: RadialMotion

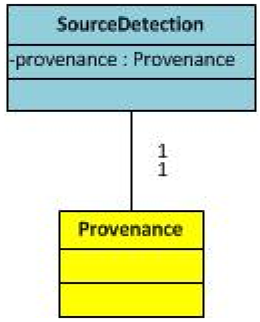
A Source object COULD contain a radialMotion associated object with that contains the calculation of the tangential velocity of the source detection. Usually done through spectroscopic calculations.

RadialMotion is usually expressed in catalogues through a radial velocity or through the redshift value and, also, it needs to have a Doppler definition qualifier (from STC) that could be one of: OPTICAL, RADIO or RELATIVISTIC




## *5.2 SourceDetection*

Considered as a detection a statistically significant radiation excess above the background in a given sky position.  This class would be the main one for most of the astronomical catalogues as it is the output result of the observational effort.

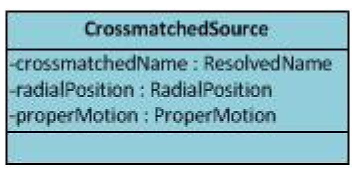



It should contain a connection to the ongoing Provenance DM to describe facility/instrument/configuration details/code used for extraction of the measurement.

## 5.3 CrossmatchedSource

As a consolidation of different source detections (from one of more catalogues), general entities called source are generated. For these entities, other quantities can be defined, e.g. the proper motions (tranversal velocity of the sources on the sphere) that imply different detections and a crossmatch of the different source detections.

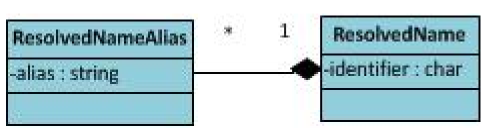
Catalogues of this kind of entities can be considered the other type of main catalogues in astronomy.




It contains a resolved name structure (with main identifiers and aliases), a possible radial position calculation

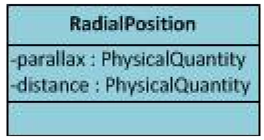
### 5.3.1 Crossmatched.crosmatchedName – Data Type: ResolvedName

Unique identifier for this detection within the source catalogue. As a crossmatched source is a result of a combination of different detections or sources from different catalogues, the structure contains a main identifier and a set of aliases.



### 5.3.2 Crossmatched. RadialPosition – Data Type: RadialPosition

A crossmatched source COULD contain a radial position describing the measurement of the distance to the object. Distance that can be expressed in two ways: as a distance and as a parallax.




Usually, the distance is described as a parallax in most of the catalogues as the conversion from parallax to distance implies certain models for the error conversion, although, in a first approximation:

$$d = 1/\rho$$

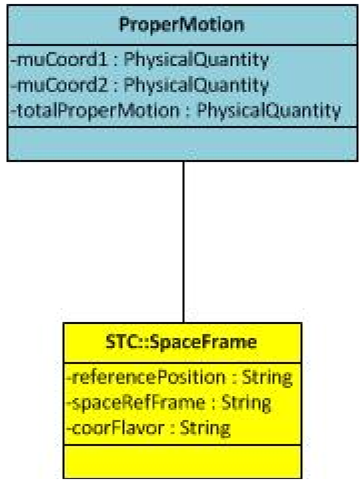
Where $\rho$ is the parallax angle expressed in arseconds that produce a distance in parsecs.

### 5.3.3 Crossmatched.properMotion – Data Type: ProperMotion

A crossmatch source COULD contain a properMotion element describing the velocity on the sky (imaginary fixed background) as seen from the center of mass of the Solar System, as compared to the imaginary fixed background of the more distant stars. It is usually express as a vector of two components, describing the variation of the position on certain sky coordinates (in general right ascension and declination), corrected by declination and expressed on mas/year.

$$pm_{ra} = \frac{\Delta ra}{\cos(\delta)}$$

$$pm_{\delta} = \Delta\delta$$




## Appendix A: Example of mapping of known catalogue

TODO

## Appendix B: Interoperability example

TODO

## References

[1] R. Hanisch, *Resource Metadata for the Virtual Observatory* , http://www.ivoa.net/Documents/latest/RM.html
[2] R. Hanisch, M. Dolensky, M. Leoni, *Document Standards Management: Guidelines and Procedure* , http://www.ivoa.net/Documents/latest/DocStdProc.html