IVOA Data model for raw radio telescope data – note for discussion
Version 0.1

IVOA Note
2003-10-16

This version:

http://www.ivoa.net/internal/IVOA/IVOADMInterferometryWP/NOTE-RDM-2003-10-16.pdf

Latest version:

http://www.ivoa.net/internal/IVOA/IVOADMInterferometryWP/NOTE-RDM-2003-10-16.pdf

Previous versions:

Editors:

P. Lamb

Authors:

P. Lamb

R. Power

Abstract

A simple data model is described whose intention is to allow the description of raw uncalibrated data from single-dish (single- and multi-beam) radio telescopes and single-site radio interferometers. Allowance is made for beam forming by phased arrays within any of the telescope.

The data model does not include modeling of Very Long Baseline Interferometry data, nor of calibrated data, but it is intended that it should, if possible.

Status of this document

This is a Note for discussion at the IVOA Interoperability Meeting Strasbourg 16-17 October 2003.
The first release of this document was not yet released.

This is an IVOA Note 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 Note 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/docs/.

Acknowledgments

Acknowledgements are due to Anita Richards, Jodrell Bank, for assistance with the understanding of interferometry data, and to Australia Telescope researchers and staff Dave McConnell, Robin Wark and Jessica Chapman for assistance with Australia Telescope Compact Array data.

Contents

1 Introduction

The data model is intended to allow the description of raw uncalibrated data from single-dish (single- and multi-beam) radio telescopes and single-site radio interferometers. Allowance is made for beam forming by phased arrays within any of the telescope.
The data model does not include modeling of Very Long Baseline Interferometry data, nor of calibrated data, but it is intended that it should, if possible.
The model is in an early stage of development. Several issues are noted as needing more work, and probably more that need to be covered. Notes in *italics* refer to areas where it is acknowledged that more work is required. As a general note, there are important issues about data accuracy that are not addressed by the present model.
The model is intended to capture a description of the data stored in radio telescope archives and available from data services. It describes the data, not the telescope, though of course, many issues of the telescopes characteristics and configuration will be reflected in the characteristics of the data.

2 Overview

The data model consists of three main parts:

1. The classes `Project`, `ObservationSet`, `Observation`, and `Proposal` and its associated classes, which contain data describing the scientific purpose of the observations, and collect them into related sets; an `Observation` is a set of `Scans` taken in a single observing session with identical telescope configuration, but with a variety of telescope pointings and observed sources (including calibration sources). A `Project` is the central administrative class of the data model – each `ObservationSet` is associated with a single `Project`. .An `ObservationSet` is the set of `Observations` made in a

single observing session. Each `Project` is associated with a set of `Proposal`s that give a summary of the purpose of the observations during the lifetime of a `Project`. These classes are described in more detail in Section 3, and the UML Class diagram is in Figure 4.

2. The classes associated with `InstrumentConf` describe the telescope(s) used and their configuration. These classes are described in Section 5 and their class diagram is in Figure 5.
3. The remaining classes associated with `Observation` describe the data in the scans, and the classification of the sources observed, and the relationships between them (e.g. which phase calibrator with which target source). See Section 6 and Figure 6.

3 Project structure

This section describes the classes used to model the processes that lead to observing time being granted; the data model diagram is in Figure 4.

3.1 Project

Contains an identifier code for the project and the project title. Observations carried out in the project are contained in the associated class `ObservationSet`. Information about proposals that led to the granting of observing time is in the associated class `Proposal`.

3.2 Proposal

Contains the `title` of the project, a short `description` of the project (e.g. the proposal abstract) and the `scheduleType`, which contains the scheduling type for the project[3][4]:

`normal` `knownTransient` `targetOfOpportunity` `exploratory`

`normal` projects are scheduled by the observatory's standard rules for time assignment. `knownTransient` projects are for the observation of transient phenomena that are in general predictable, but not in their specific detail (NAPA projects on ATNF telescopes). `targetOfOpportunity` projects are for unexpected or unpredicted events, eg. supernovae in nearby galaxies and other extreme events. `exploratory` projects are for extensions of time to allow for additional unexpected time extending the observations of a project (e.g for VLA proposals [4]).

Further details are in the associated classes `ProjCategory`, `ProjType`, `Author`, `Institution`, `Country`, `ProposalContact` and `Term`.

3.3 ProjCategory

The class contains a scientific category for the project, as for example in the observing proposals for the VLA and for the ATNF telescopes [1][2]. A `Project` may be in (reference) more than one category.The values for the field `name` are drawn from a controlled vocabulary:

`instrumental`	`terrestrial`	`solarSystem`	`stellar`
`galactic`	`magellanicClouds`	`extragalactic`	

3.4 ProjType

This class contains the type of observation being made in the project; the types here are taken from the proposal forms for the ATNF and NRAO[1][2]. The type generally concerns the manner in which the data is collected, or the data product intended to be produced. The values for the field `name` are drawn from a controlled vocabulary:

`astrometry`	`bandwidthSynthesis`	`circularPolarisation`	`continuum`
`engineering`	`fillerTime`	`highTimeResolution`	`imaging`
`instrumental`	`linearPolarisation`	`mapping`	`monitoring`
`mosaic`	`multibeam`	`pointSource`	`phasedArray`
`polarimetry`	`snapshot`	`solar`	`spectroscopy`
`survey`	`timeBinning`		

3.5 Author

Contains the `familyName` and `initials` of authors of the proposal. The class is in two associations with `Project`; `PrincipalInvestigator`, which names the single PI for the proposal, and a second association that names any other proposal authors.

3.6 ProposalContact

`ProposalContact` is an association class that indicates the author for contact for the proposal, and provides their e-mail address.

3.7 Classes Institution and Country

Give the author's affiliation to an institution (or more than one institution), and that instution's country (or more than one for multinational institutions). In `Institution`, `code` is an abbreviated name for the institution (e.g. ATNF), and `name` is the institution's full name (e.g. Australia Telescope National Facility). It would be desirable if `code` was drawn from a standard vocabulary, but this may be difficult to achieve. In `Country`, `code` is the ISO 3166[5] two-letter country code (e.g. GB), and `name` is the ISO 3166 official short name in English (e.g. United Kingdom).

3.8 Term

The observing term or scheduling block for the proposal, giving its (optional) `name`, `startDate` and `endDate`. `endDate` is *exclusive* of that date (e.g. ATNF term JANT2003, 2003-01-01, 2003-5-01).

3.9 ObservationSet

An `ObservationSet` refers to a set of `Observations`. Each `ObservationSet` contains the set of observations made at a single observing session for the project. An `ObservationSet` will normally contain sufficient data to be processed into its intended data product(s), but higher quality results may be achieved by combining data from `Observations` in several `ObservationSets` (e.g. by combining data for the same source obtained with different interferometer array configurations).

3.10 Observation

An `Observation` is a set of scans that form the smallest set of data in a project that might reasonably be processed into a data product. The scans in the `Observation` are made with the same instrument configuration, and should include any scans necessary for the calibration of the data product. The `Observation` has associated with it data describing the instrument configuration with the aim of permitting the searching for the best relevant data for a user's needs. The `Observation` also has a set of scans associated, and the relationships that describe the calibration sources and targets, and their interrelationships.

4 Instrument configuration model

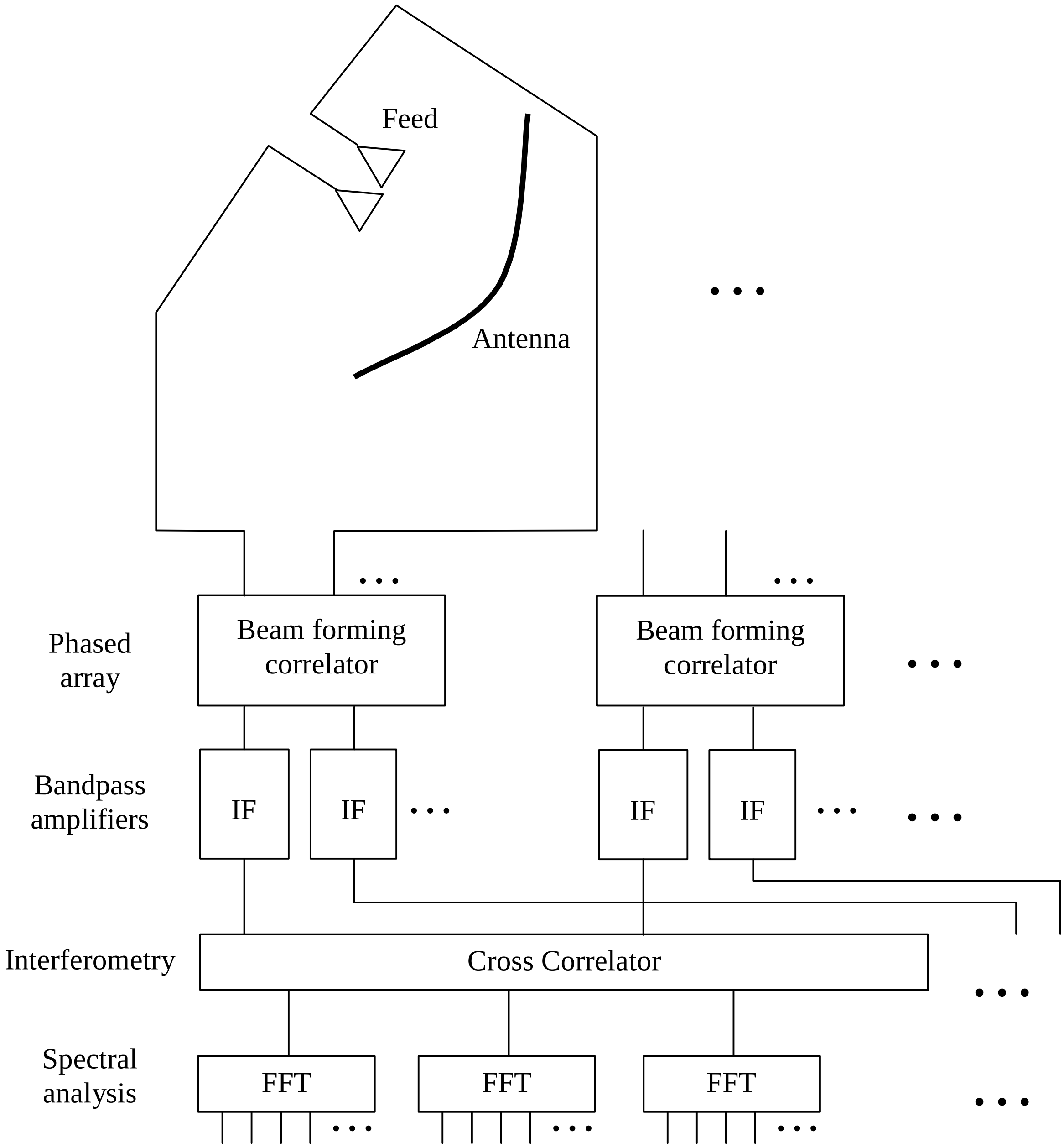

Figure 1 Instrument model

The instrument model in Figure 1 is intended to be capable of representing a wide variety of radio telescope configurations. Each antenna system supports multiple feeds (e.g. single dish multibeam, cylindrical reflector phased array). Sets of feeds (from the same antenna, as shown here, or from different antennas) may pass into a beam-forming correlator in phased arrays; systems with no beam-forming correlator may be represented by a correlator with a single input and output. Multiple distinct beams may be formed in the phased array; each beam is amplified and band-pass filtered by the IF stages. The outputs of the IF stages feed a cross-correlator for radio interferometry; if the system is not being used for interferometry, then the cross-correlator may be represented by a single input and output. The cross-correlators feed a bank of spectrum analysers to produce spectral visibilities (or spectral time series for single-beam instruments). The instrument model is drawn as cross-correlation first, then frequency analysis. This should not be taken that frequency then cross-correlation (FX) instruments are excluded.

For VLBI instruments, the antennas may be very widely spaced, and on-line cross correlation may not be possible; this instrument model represents logical data flows, and so VLBI can be represented.

The model can be instantiated to model single dish, single feed telescopes, phased array telescopes, interferometers built from single-dish, single feed telescopes, and phased array interferometers.

5 Instrument data model

This section describes the classes used to model observing instruments described in Section 4 and Figure 1; the data model diagram is in Figure 5.

5.1 InstrumentConf

Each `Observation` is associated with a set of `InstrumentConf`s that specify the instruments being used for the observation. There will typically be only a single `InstrumentConf`, except for VLBI observations. Associated with each `InstrumentConf` is an `Instrument` class providing the basic instrument details (name, location), and a set of `ArrayConf`s specifying the array of antennas in the instrument.

5.2 Instrument

This class provides the `shortName`, `name`, `locationName` and a brief `description` of the instrument being used in an observation. The `URL` field allows the instrument's Web site to be accessed.

5.3 Location

`Location` gives the instrument site's `latitude` and `longitude` in the specified lat/long `datum`, and its `height` relative to the `heightDatum`. The location information is intended to be for the observatory site, rather than an accurate location of the instrument's antenna. Instruments on the same site may have the same `Location` here.

5.4 ArrayConf

The class represents the configuration of the antenna array used by the instrument. The configuration may be named in `name`; the configuration of the individual antennas is specified through the associated `AntennaConf`s.

5.5 AntennaConf

`AntennaConf` allows the specification of the instrument's set of antennas and feeds, and their locations, and of the telescope beams that are formed by the antennas directly, or by the use of the antennas in phased arrays. The antenna configuration may be named in `name`.

5.6 Feed

Class `Feed` specifies the set of feed horns or other devices which collect RF energy from the antenna. The feed `polarisation` is one of `L`, `R`, `X`, `Y`. The `Feeds` are associated with a set of `BeamConfs` that represent the beams formed by the feeds, or combinations of feeds. Note that a single feed horn with dual polarizations constitutes two `Feeds`. *Feed requires some way of specifying its offset from the pointing direction of the antenna.*

5.7 AntennaPosition

An `AntennaConf` references a set of `AntennaPositions` that associate antennas with precise locations – antennas are not assumed to be at fixed positions. Each `AntennaPosition` references an `Antenna` description, an `AntennaLoc` and a `Station`. `Stations` are positions in a reconfigurable array where antennas can be connected into the array's cabling infrastructure. However, an antenna at a station need not be at exactly the station location, and frequently is not. `Station` locations are typically given in some convenient local coordinate system aligned with the layout of the array. Antenna locations must be specified in a way that accurate interferometer baselines can be derived from them; they are often given in geocentric Cartesian coordinates.

5.8 Antenna

This class gives the antenna `name`, `mount` type, `majorAxis`, `minorAxis`, and `effectiveArea`. The mount type is one of `equatorial` or `altAz`. The antenna location specifications are insufficient to allow orbiting antennas to be modelled. The minor axis is set to zero for cylindrical parabolic reflectors, both axes are set to zero if the antenna has no reflector. *Are there other useful attributes of the antenna that should be modeled?*

5.9 Station

`Stations` are positions in a reconfigurable array where antennas can be connected into the array's cabling infrastructure. The class names the station, and refers to its location as an `AntennaLoc`. The location of a Station may be different from the location of an antenna positioned at it, and may be described in a different datum.

5.10 AntennaLoc

`AntennaLoc` defines the location of antennas and antenna stations in the model. The for antenna locations is typically geocentric Cartesian coordinates, but if it is a lat/long specification, `x` is used for latitude, `y` for longitude and `z` for height relative to the reference ellipsoid of the `datum`. Heights above Mean Sea Level (above the geoid) are not well modeled by height and `datum` alone.

5.11 BeamConf

The `BeamConf` class expresses the relationship between physical `Antennas` and `Feeds` and telescope beams. The relationship between `BeamConf` and `Feed` permits the expression of a variety of ways of forming beams from feeds. In a single-dish, single-feed telescope, each `Feed` results in a single `BeamConf` per `Receiver`. A phased array will require a `BeamConf` per `Receiver` for each beam formed from a set of feeds. A single-dish multi-beam feed will produce one `BeamConf` per `Receiver` per

Feed, unless it is being used as a phased array, in which case the number of beams produced will also depend on the phased array configuration.
A BeamConf is associated with a set of SynthBeamConfs that are produced as a result of interferometry.

5.12 Beam

A telescope beam, defined by its beamMajor and beamMinor FWHP (full width half power), and sensitivity.
As with Feeds, there is a need here to describe an offset pointing direction for any beam that is not in the main pointing direction of the telescope.

5.13 Receiver

A model of the bandpass filtering component of the instrument, giving its skyCentreFrequency, IFCentreFrequency and bandwidth. The skyCentreFrequency is the frequency of incoming radiation, uncorrected for any motion. The baddwidth accounts for the bandpass in the receiver's intermediate frequency stages and any further restriction in the bandwidth by later bandpass filtering – it is the bandwidth presented to the inputs of the correlators. This later filtering may shift the from its nominal position at the output of the IF stage. The IFCentreFrequency is the center frequency of the fixed –frequency bandpass filters in the receiver's intermediate frequency stages. *Is IFCentreFrequency useful?*

5.14 SynthBeamConf

In a non-interferometer instrument, SynthBeamConf is in a one-to-one association with the corresponding BeamConf and the corresponding fields in its associated SynthBeam have identical values to its BeamConf, and its PolProduct simply reflects the polarization of the feed. In interferometer telescopes, the associated BeamConf and PolProduct define the characteristics of the synthesized beam.
In all cases the associated Spectrum and Velocities classes define the spectral analysis carried out on the data.

5.15 SynthBeam

In non-interferometer telescopes, the beamMajor, beamMinor and sensitivity are of the corresponding Beam.
In interferometer telescopes, the beamMajor, beamMinor and sensitivity are of the synthesised beam. nVisibilities is zero in non-interferometry instruments, and non-zero in interferometers, and contains the number of observed visibilities, not the maximum number that the telescope may use. *Is this reasonable?*

5.16 Spectrum

The class defines the characteristics of spectrum analysers connected to the beam's data stream and the reference data deeded to interpret them. numChannels, chanSeparation, freqResolution, define the basic channel characteristics., refChanNum and refFreq define the velocity frame of reference, the channel number for the reference frequency and the reference frequency for the spectrum. restFreq, molecule and transition provide information about the spectral line being observed. velRefFrame is taken from a controlled vocabulary:

topocentric	earthMoonBarycentric	heliocentric
solarSystemBarycentric	localStandardOfRest	galactocentric
localGroupBarycentric	virgocentric	microwaveBackground

5.17 Velocities

Where spectra have been processed into velocity channels, the class provides the corresponding data for the velocity channels as the `Spectrum` class. In addition, the velocity definition (`radio`, `optical`) is defined in `velType`.

5.18 PolProduct

For non-interferometers, reflects the polarization of the `Feed`, and the value of `product` is taken from (L, R, X, Y). For interferometers, `product` is taken from (LL, RR, LR, RL, XX, YY, XY, YX).

6 Observed data model

6.1 RawScan

Class `RawScan` contains a description of the data in a single unprocessed, uncalibrated scan. `numDataGroups` is the total number of visibility sets recorded in the scan, `startTime` and `endTime` are the start and end (*sidereal?*) times for the scan. `dataGroupIntTime` is the integration time for each data group (if the integration time varies in the scan, then a suitable representative integration time) and `totalIntTime` is the total integration time for all visibilities in the scan. `Ta`, is the the antenna temperature and `Tsys` (system temperature) for the source observed.

6.2 ScanType

`ScanType` indicates the type of observation being made in the scan. It is taken from the controlled vocabulary [refs to come]:

`point`	`focus`	`cal`	`skyDip`
`holography`	`onTheFly`	`onOff`	`positionSwitch`
`raster`	`cross`	`mosaic`	`pulsar`
`tiedArray`	`frequencySwitch`	`dwell`	`beamSwitch`
`track`	`dopplerTrack`		

A `RawScan` may be associated with multiple `ScanTypes`.

6.3 Pointing

The class defines the name and location of an instrument pointing. The location is in Right Ascension and Declination for the given equinox.

- *Queries should be possible in other coordinate systems.*

6.4 ScanPointing

Associates the sources in a `Scan` with the `Scan`, and gives the role of the source in the `Observation`. A source in a scan may have multiple roles. The `sourceRole` field is taken from a controlled vocabulary following the usage in the B/Merlin catalog in the CDS Vizier database[6]:

`target`	Source of interest – phase referenced
`tarX`	Source of interest – not phase referenced
`ampCal`	Amplitude calibrator
`phiRef`	Phase reference
`polCal`	Polarisation leakage calibrator
`paCal`	Polarisation angle calibrator
`bpCal`	Bandpass calibrator

6.5 Target

The class associates a set of `Pointing`s with an `Observation` as targets.

6.6 TargetPath

In scan modes where the telescope's beam(s) move relative to fixed sky positions, `TargetPath` represents a sequence of source positions that represent the path of the beam; the beam path is taken to trace out a sequence of celestial great circle arcs between the `Pointing`s in the `TargetPath`.

6.7 Classes AmpCal, PolCal, PACal, BPCal

The class associates respectively an amplitude, polarization leakage, polarization angle or bandpass calibrator with an observation. Note that there may be more than one source in an observation that can take any of these roles; the associated source is the recommended source for the role.

6.8 PhiRef

`PhiRef` associates a target source with phase calibrator source in an `Observation`. There may be more than one phase calibrator source suitable for use with a target; the one indicated in `PhiRef` is the recommended calibrator.

6.9 CatalogSource

In some data sets calibrator sources are sometimes referred to by non-standard names, or there may be more accurately known location for the source than was recorded in the raw data for the observation. `CatalogSource` allows a calibrator source to be associated with a source name and location drawn from an astrometric catalog or catalog of calibrator sources.

7 Notes on reading the data model diagram

The data model diagram uses the conventions for class models in the Unified Modelling Language (UML).

7.1 UML Classes

In the UML convention, Classes represent data types with data fields and operations (methods); in this model, all data fields are publicly accessible, and Class operations are not used. Classes appear as rectangles in the model with the class name at the top of the rectangle, and its data fields and their types listed below.
The data types used in classes in this document are: `integer`, `float`, `string` and `date`. Dates are taken to include time-of-day. For the purposes of this document, sizes of data types are not considered; they are simply assumed to be large enough to accommodate the value's range and precision.

7.2 UML associations

The lines drawn between classes in the UML class diagram represent associations – the two classes at either end of the line are related to each other logically.

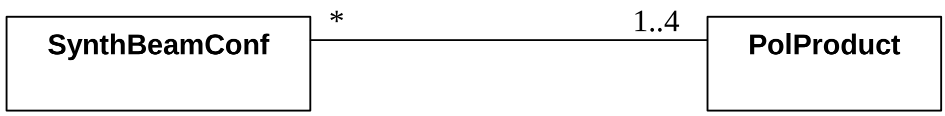

Figure 2 Multiplicity in UML associations

The number or number range at the end of the association line is the multiplicity of instances of the class at that end of the association that may be associated with a single instance of the class at the other end.
In Figure 2 any instance of `SynthBeamClass` may be associated with 1 to 4 instances of `PolProduct` class (a synthesised beam may have 1 to 4 different polarization products). An instance of `PolProduct` may be associated with any number of `SynthBeamConf` instances; a single instance of a `PolProduct` (e.g. with `product` = XX) may be referred to by any number of `SynthBeamConf` instances.
The asterisk * is a wildcard meaning “any number”. 1..* means 1 or more; * standing 2 alone means 0..*.

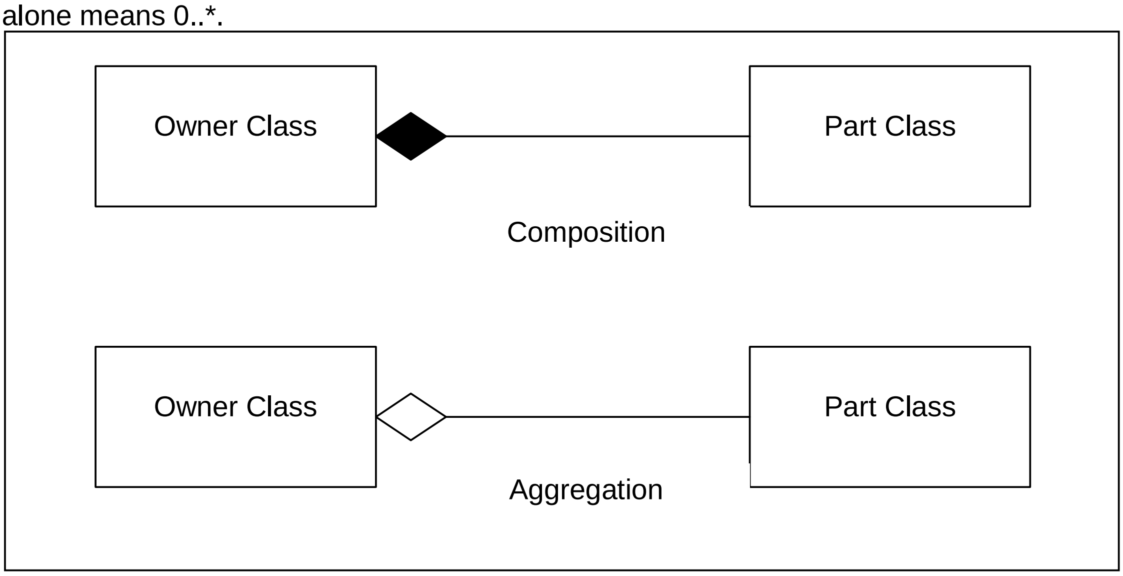

Figure 3 Composition and aggregation

A diamond at the end of a UML association indicates composition or aggregation, and the class at the end of the association with the diamond is the owner in the association and cannot have multiplicity greater than 1. When the diamond is filled, composition is indicated. Composition is the tightest form of association. if the owner is deleted, then the part must be deleted. When the diamond is not filled it indicates that the owner may not be deleted without first deleting the part; but this may not be done automatically, since other associations may be involved. When an explicit multiplicity is not given for a composition or aggregation owner, it is 1..1, not 0..1.

8 Changes from V0.0

This is the initial version of the document.

9 References

[1] ATNF, *ATNF Proposal Application Forms*, http://www.atnf.csiro.au/observers/apply/form.html, Australia Telescope National Facility Web.

[2] NRAO, *The Very Large Array Telescope Proposal Forms*, http://www.nrao.edu/administration/directors_office/tel-vla.shtml, National Radio Astronomy Observatory Web.

[3] ATNF, *Target of Opportunity and NAPA Information*, http://www.atnf.csiro.au/observers/apply/too_apply.html, Australia Telescope National Facility Web.

[4] NRAO, *Rapid Response Science*, http://www.vla.nrao.edu/astro/prop/rapid/, National Radio Astronomy Observatory Web.

[5] ISO, *English country names*, http://www.iso.ch/iso/en/prods-services/iso3166ma/02iso-3166-code-lists/list-en1.html, ISO-3166, International Standards Organisation.

[6] CDS, *VizieR Service*, http://vizier.u-strasbg.fr/viz-bin/VizieR, Centre de Données astronomiques de Strasbourg Web

10 Appendix

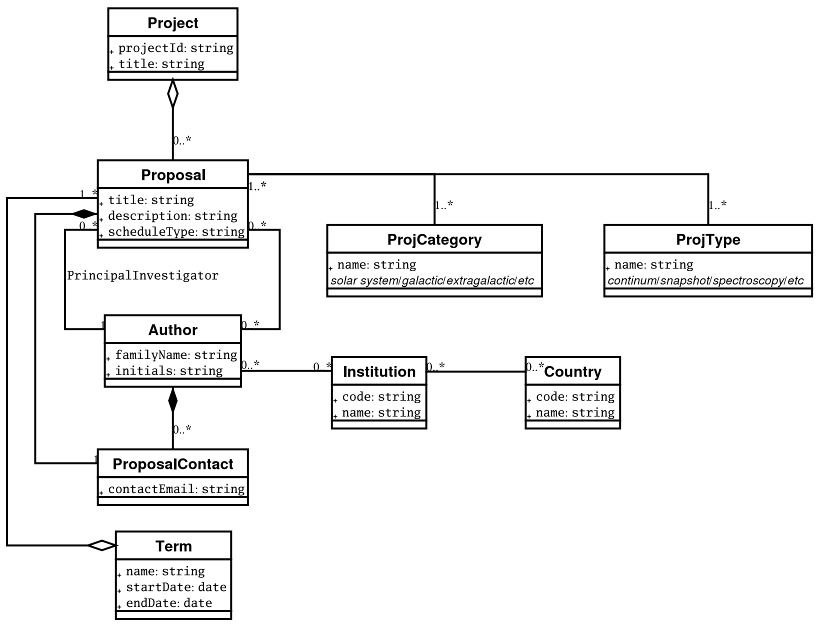

Figure 4 Project structure

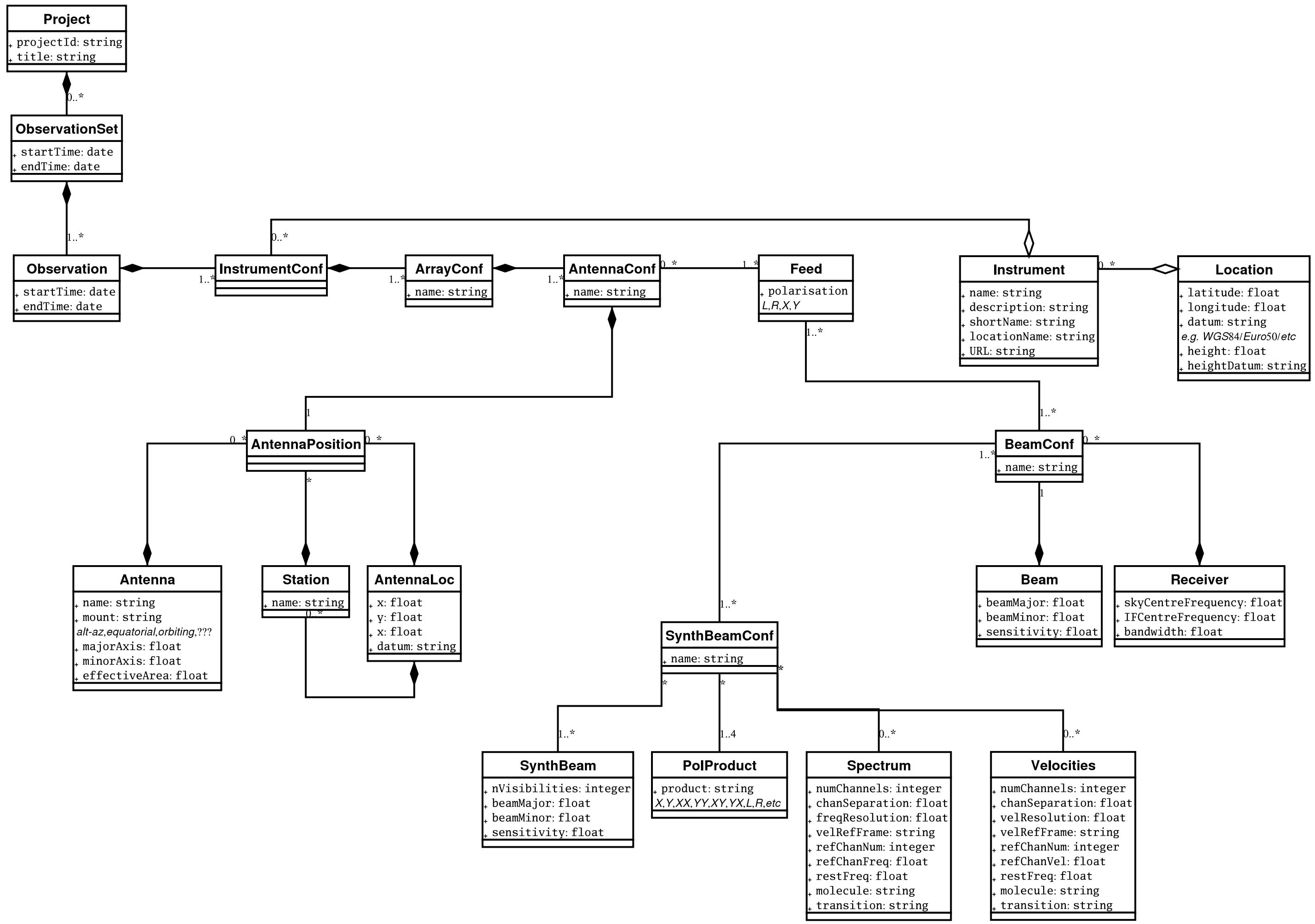

Figure 5 Instrument data model

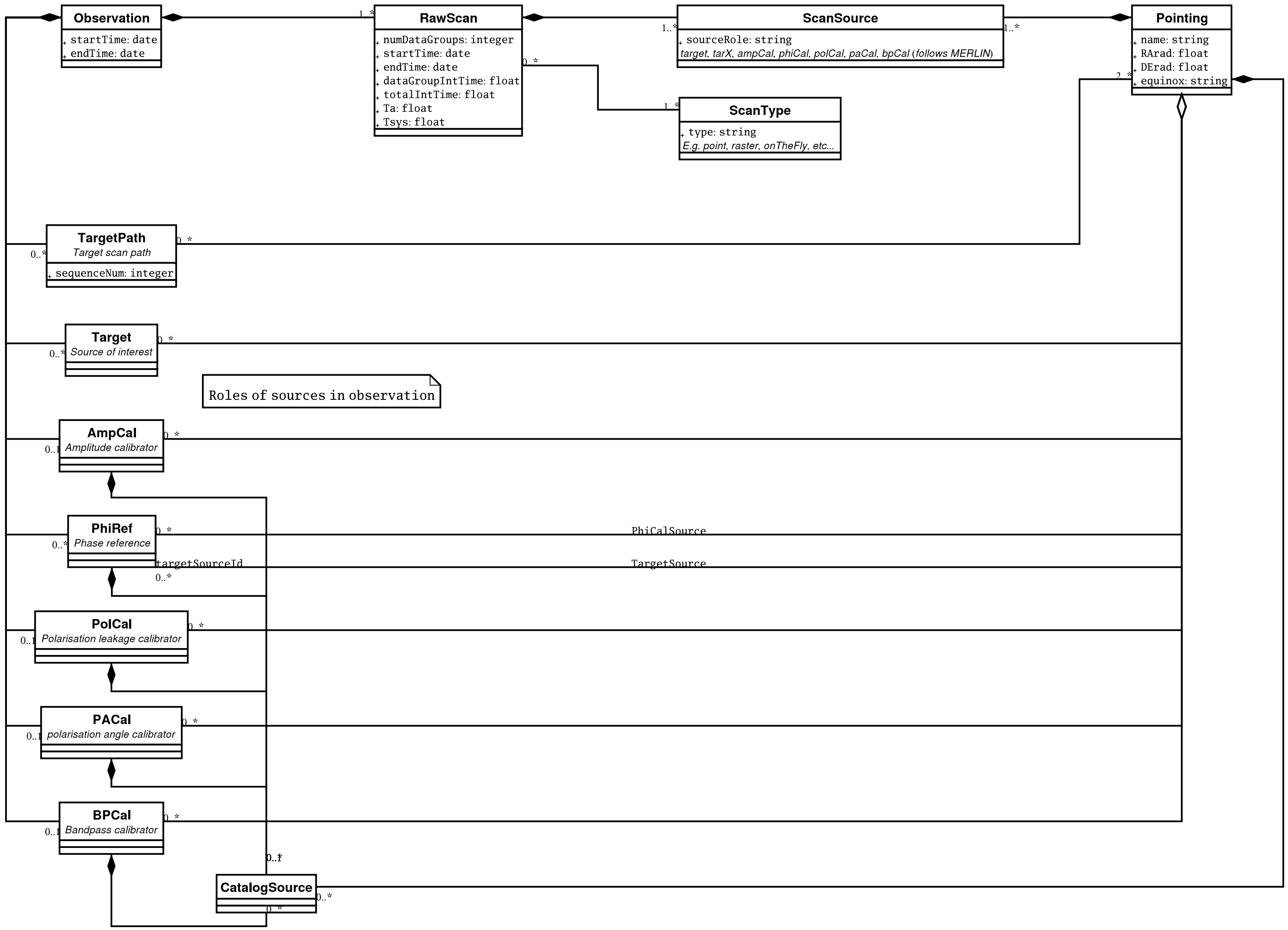

Figure 6 Observed data model