

Data Model for Astronomical DataSet Characterisation

Version 1.0

IVOA Note

February 28, 2006

Working Group: <http://www.ivoa.net/twiki/bin/view/IVOA/IvoaDataModel>

This version: summary of the current work made on Characterisation Data Model

Editors:

Jonathan McDowell, François Bonnarel, Igor Chilingarian, Mireille Louys, Alberto Micol, Anita Richards.

Authors:

IVOA Data Model Working Group

Abstract

This document defines the high level metadata necessary to describe astronomical data sets on their physical axes, either observed or simulated, for various types of data such as 2D-images, data cubes, X-ray event lists, IFU data, etc.. The goal is to provide an abstraction which contains structured information for all these types of data so that discovery and interpretation is feasible in a general framework, and can be used jointly for science usage. The model aims at facilitating the manipulation of heterogeneous data in VO portals. A VO Characterisation instance can include descriptions of the data axes, the range of coordinates covered by the data, and information on the data sampling and resolution on each axis. These descriptions are in terms of physical variables, with the specific instrumental signature abstracted away.

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/>. A more preliminary version of this work announced on the DM list in April 2005 is available at <http://alinda.u-strasbg.fr/Model/Characterisation/characterisation.pdf>.

Acknowledgements

Members of the IVOA Data Model Working Group, including representatives of the US NVO, Astrogrid, and the AVO have contributed to the present draft. Editors would like to quote peculiarly the input given by P.Didelon, G.Lemson, A.Rots, L.Shaw, B.Thomas and D.Tody. F.Bonnarel and M.Louys acknowledge support from the French ACI-GRID project, IDHA and of the EU funded VOTech project.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 4 |
| 1.1 | What is the Characterisation Model for? | 4 |
| 1.2 | Links to other IVOA modeling efforts | 5 |
| 2 | Exploring the Characterisation concepts | 7 |
| 2.1 | Approach and motivation | 7 |
| 2.2 | Examples | 8 |
| 2.3 | Structure of Characterisation and development strategy | 12 |
| 2.4 | The Axis point of view | 12 |
| 2.4.1 | The axis concept and its attributes | 12 |
| 2.4.2 | Accuracy | 13 |
| 2.4.3 | Other Axis flags: sampling, observables, etc. | 13 |
| 2.5 | The Property point of view | 14 |
| 2.5.1 | Coverage | 14 |
| 2.5.2 | Resolution and Sampling Precision | 17 |
| 2.6 | The layers of description | 18 |
| 3 | The Model | 19 |
| 3.1 | The role and structure of the Model | 19 |
| 3.2 | AxisFrame | 20 |
| 3.3 | Errors in Characterisation: the Accuracy class | 22 |
| 3.4 | Properties and levels | 22 |
| 3.5 | Navigation in the model: by axis or by properties? | 24 |
| 3.6 | Implementing the DM with existing pieces of other models: Quantity and STC | 24 |
| 4 | XML Serialization | 26 |
| 4.1 | XML schema | 26 |
| 4.1.1 | Design of the schema | 26 |
| 4.1.2 | Building blocks of the schemata | 27 |
| 4.1.3 | Utypes generation: select one ordering strategy | 28 |
| 4.1.4 | VOTable serialisation | 28 |
| 5 | Appendix A : XML serialisation example | 34 |
| 6 | Appendix B : VOTable serialisation example | 36 |

1 Introduction

This document defines an abstract data model called "Data Set Characterisation" (hereafter simply "Characterisation"). In the first section of this document we present requirements and place the model in the broader context of VO data models. In the second section we introduce concepts and discuss their interactions. In the section called "Model" we present a formal UML class model using the concepts defined earlier. XML and VOTable serializations are presented in the last section.

1.1 What is the Characterisation Model for?

The purpose of the Characterisation Data Model is to define and organize all the metadata necessary to quantitatively -and to some extent qualitatively- describe how a given piece of data (observed or synthetic) occupies the multidimensional space it observes or tries to reproduce (simulate).

The Characterisation DM shall not describe parameters that Characterise the astronomical sources/objects observed (or simulated).¹ On the contrary, this model focuses on the multidimensional space the data are immersed in.

The multidimensional space that describes an observation often consists of the Spatial (2D), Spectral, and Temporal axes as well as the Observable (e.g. Flux, number of photons, etc.), but Characterisation is a quite general model which can support any axes - other common cases include polarization, kinematic axes, e.g. radial velocity in a radio data cube case, etc. Note that the axes for Characterisation are not limited to the axes actually explicit in the data - for instance, a two-dimensional celestial image explicitly contains just the 2D Spatial axis, but is also (implicitly) described by a Spectral axis giving the waveband of the image.

The unification of the description of concepts like *Coverage*, *Sampling*, *Resolution and Accuracy* -on each of the pertaining axes- for data products of different kinds (images, spectra, data-cubes, light-curves, visibility data, IFU data, etc..) is an obvious solution to the interoperability requirements of the VO. These descriptions abstract the physical axes of the data - so, for instance, spatial resolution expresses the level of smearing of the true sky flux distribution to the final data set, without separately distinguishing the contributions of different atmospheric, instrumental and software processing effects. The model describes the data as they are provided, not how the data got to be that way; the latter story will be encoded in a separate model called Provenance which will be elaborated in the future.

The Characterisation DM has to satisfy two types of requirements, which may overlap in some cases:

- Data Discovery requirements:

¹This aspect is to be covered by another data modeling effort: the Source Data model, now emerging.

The concepts and relationships sketched out in this model supply a description of elements to be used for requests on databases and services (in this sense Characterisation is fundamental for a VO request definition).

Given a list of observations, it shall be possible for a user to maximize the efficiency of identifying (rejecting) those observations that are relevant (not relevant) to their science case, based purely onto the description of the geometrical aspects of the observations, that is, how and with which accuracy the multidimensional space is covered and sampled.

It shall be possible for a client to generate a detailed footprint of the observation in the multidimensional space it covers.

Errors on all the axes (astrometric, photometric, spectral, temporal accuracies, etc.) may be provided.

- Data Processing/Analysis requirements:

The Characterisation Data Model shall detail the variation of sensitivity on all the axes, in order to allow detailed analysis/processing of the data.

Version 1 will fulfill all Data discovery requirements, and allow some simple automatic processing such as cross-correlation and data set comparisons. Full implementation of the Data processing/Analysis requirement will only be available with a future version of this model.

Possible applications of such a model cover the description of one observation, and more generally the description of data collections. At the observation level, it structures and completes the metadata generally conveyed by FITS keywords. It can also be applied to provide an overall description of the parameter space of a VO service (VO tools).

1.2 Links to other IVOA modeling efforts

Historically the Characterisation Model appeared as a specific and necessary class in the so-called "Observation data model", a high level description of metadata associated with observed data, considered as an "Observation" (or a collection of "Observations"). This general framework is described in an IVOA note available at

<http://www.ivoa.net/internal/IVOA/IvoaDataModel/obs23.pdf>.

A brief summary of the connection between Observation and Characterisation is given at Fig.1. Characterisation gives a physical insight to the data while DataCollection, Curation, Provenance give more instrumental or sociological information. It was decided at the Boston Interoperability meeting (2004), to put the effort on this Characterisation class first because it

appeared as the class where minimum general agreement could be reached rather easily, and for its importance with respect to the VO access protocols development. Hence, Characterisation has been upgraded to the status of an independent Data Model effort in itself.

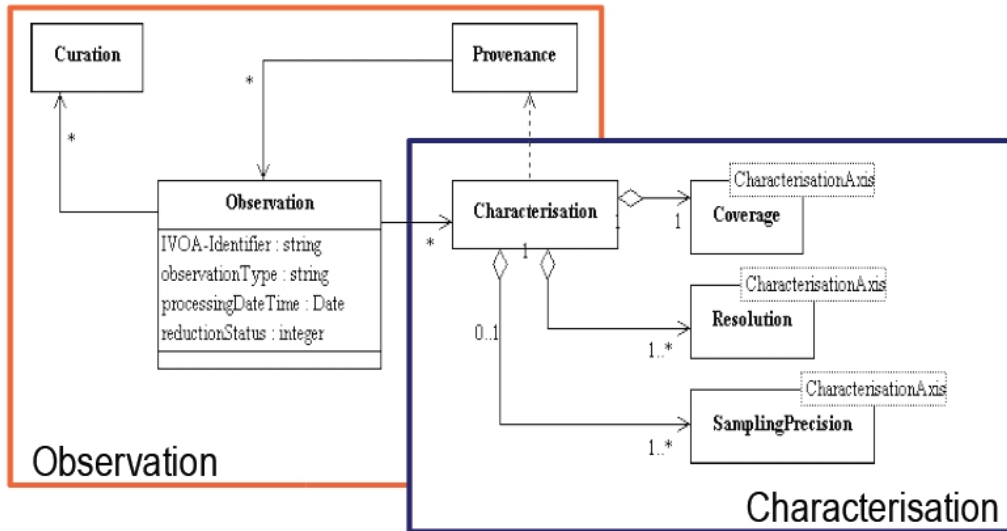


Figure 1: *Interaction between the Observation and Characterisation data models: The Characterisation DM focuses on the physical information relative to an observation. It encompasses the description of the physical axes along which the data are taken, and the essential properties of the data such as the coverage, the resolution, the sampling precision, etc... The data management part such as VO identifier, data type, data format, etc.. is handled at the level of the Observation class.*

Discussions in the DM group have also emphasized the need of a specific layer "characterising" a VO entity (such as an observation, a data collection, a service, or simulated data). This layer of abstraction was identified as necessary not only to store specific characterisation information but also to answer specific questions the user might ask about the data, before using it.

This model complements and extends to a finer level of detail some of the metadata adopted by the VO Registry (see RSM 1.1). For the description of an individual dataset, this finer level is clearly needed.

Characterisation emphasises the relations between different levels of description. In general, data query/discovery use cases, as stated above, will tend to use simplified representations, and data analysis/tool interface use cases will use the most complete representations. For example:

1. For a given VO-Event what are the observations in one selected archive that have matching coordinates (using spatial and temporal Coverage)

2. For a given galaxy which are the "CCD Mosaic" observations (or IFU) that *really* observed it (using the spatial detailed Coverage information)
3. For a given simulated spectrum (known through its Characterisation with a given sampling), which are the observed spectra with a compatible resolution e.g., matching the Shannon criterion for the simulated sample size.

2 Exploring the Characterisation concepts

2.1 Approach and motivation

While identifying the various concepts necessary to build the Characterisation DM, the notion of physical axes was clearly identified as well as the N-dimensional space the data are immersed in. Moreover some main **concepts/properties** *Coverage, Sampling, Resolution* and in a slightly different role, *Accuracy* are needed to interpret the data.

When considering a typical astronomical observation,

- *Coverage*: describes where and when the telescope was pointing, and at which wavelengths;
- *Sampling*: describes how the observation was sampled on the various axes;
- *Resolution*: describes the actual physical resolution (e.g. PSF, or LSF, etc.)
- *Accuracy*: describes the accuracy along the various axes (e.g., astrometric, spectrophotometric accuracy, etc.)

Each of those properties can be assessed for the physical axes the data are mapped to. The typical axes are the *Spatial* (2d) axis, the *Spectral* axis, and the *Temporal* axis. The model though is not limited to those axes, as it can accommodate any other (e.g. Kinematic axis, etc.).

The Accuracy property is tied to the concepts of axis and mapping process from any data coordinates to the corresponding world coordinates. Therefore it will be handled in correlation to the axis description (see section 2.3).

Moreover, these properties can be described at different levels of detail,

- *Location*: A characteristic value (a point in a N-dimensional space);
- *Bounds*: The lower and higher limits (a box in a N-dimensional space);
- *Support*: Precise definition of the domain onto which the observable is defined;
- *Sensitivity*: Numerical values indicating the variation of the response function on each of the axes.

2.2 Examples

The tables below show, in a non-exhaustive manner, some of the typical properties used to characterise the spatial, temporal and spectral domains as well as the *observable* dimension (assumed to be flux) for different kind of data sets. Table 1 shows some of the Characterisation metadata for a 2-D image, Table 2 for a 1D spectrum, Table 3 for an IFU Dataset and Table 4 for an X-ray event list.

This illustrates the concepts defined in 2.1, how they can be expressed on the different axes, and with different levels of complexity.

| Axes /Properties | SPATIAL | TEMPORAL | SPECTRAL | FLUX /OBSERVABLE |
|---------------------------|---|------------------------|---------------------------------|-----------------------------------|
| Coverage | | | | |
| Location | Central position | Mid- Time | Central wavelength | Average flux |
| Bounds | [min,max] RA,Dec or Bounding box as (center,size) | Start/stop time | Min Max Wavelength | Saturation, Limiting flux |
| Support | FOV as accurate array of polygons | time intervals (array) | Wavelength intervals (array) | |
| Sensitivity | Quantum efficiency (x,y) | | Transmission curve(λ) | Function property e.g. Linearity |
| Filling factor | Effective/ Total area | Dead time | | |
| Resolution | PSF (x,y) or its FWHM | not defined | not defined | Flux/SNR (stat error) |
| Sampling Precision | Pixel scale (x,y) | not defined | not defined | (1 ADU equivalent = Quantization) |

Table 1: *2D-Image Characterisation: A Sketch of the Property versus Axis description of the metadata involved in the description of a 2D-Image in the optical regime*

These four examples show that depending on the type of data to be characterised, some axes may be sampled and resolved, some may not.

The concepts are not always fully separable. For example, high energy missions move the telescope during the observation, leading to a time-variable mapping from detector to celestial coordinates called the ‘aspect solution’. This in turn leads to effects such as spatially variable effective exposure time.

| Properties/Axes | SPATIAL | TEMPORAL | SPECTRAL | FLUX/OBSERVABLE |
|-------------------------------|---|---------------------------|---------------------------------|-------------------------------------|
| Coverage | | | | |
| Location | Central position | Mid-Time | Central wavelength | Average flux |
| Bounds | Slit [min,max] RA,Dec or Bounding box | Start/stop time | Min Max Wavelength | Saturation, Limiting flux |
| Support | Slit as accurate array of polygons | time intervals (array) | Wavelength intervals (array) | Lowest and highest value |
| Sensitivity | Response(x,y) in the slit | | Quantum eff (lambda) | Function property e.g. Linearity |
| Filling factor | Effective/ Total area | Dead time | | |
| Resolution | not defined | not defined | LSF or its FWHM | Flux/SNR (Stat error) |
| Sampling Precision | not defined | not defined | Pixel scale in lambda | (1 ADU equivalent Quantization) |

Table 2: *Sketch of a 1D-Spectrum Characterisation*

| Properties/Axes | SPATIAL | TEMPORAL | SPECTRAL | FLUX/OBSERVABLE |
|---------------------------|--------------------------------------|------------------------|------------------------------------|----------------------------------|
| Coverage | | | | |
| Location | Central position or Bounding box | Mid-Time | Central wavelength(all spectra) | Average flux |
| Bounds | Field [min,max] RA,Dec | Start/stop time | [min,max] wavelength (all spectra) | Saturation, Limiting flux |
| Support | Union of fiber footprints on the sky | time intervals (array) | Disjoint wavelength intervals | Lowest and highest value |
| Sensitivity | Response(x,y) in the slit | | Quantum eff. (λ) | Function property e.g. Linearity |
| Filling factor | Effective/ Total area | Dead time | | |
| Resolution | PSF (x,y) or its FWHM | not defined | LSF or its FWHM | Flux / SNR (stat error) |
| Sampling Precision | Pixel scale (x,y) | not defined | Pixel scale in λ | (1 ADU equivalent Quantization) |

Table 3: *Sketch of a 3D data (IFU) Characterisation*

| Properties/Axes | SPATIAL | TEMPORAL | SPECTRAL | FLUX/OBSERVABLE |
|---------------------------|---|------------------------|--------------------------------------|-----------------|
| Coverage | | | | |
| Location | Central position | Mid- Time | Central wavelength | Average flux |
| Bounds | [min,max]RA,Dec, or Bounding box as (center,size) | Start/stop time | Min Max energy | |
| Support | FOV as accurate array of polygons | time intervals (array) | Energy filter intervals (array) | |
| Sensitivity | Quantum efficiency (x,y) | | ARF (effective area) as fn of energy | |
| Filling factor | not used | Dead time | not used | |
| Resolution | PSF (x,y) or its FWHM | Time resolution | RMF (spectral redist. matrix) | |
| Sampling Precision | Pixel scale (x,y) | Frame time | PI bin width | |

Table 4: *X-ray CCD Event List Characterisation: sketch of the Property versus Axis description*

This could be described by attributes pointing towards functional descriptions, or more likely, restricting this dependency aspect to the Sensitivity class, which is functional in essence.

One can often cut up a complex coupled description into several non-coupled pieces. These aspects will probably be tackled in a next version of the Characterisation data model.

All of these concepts are well-defined for other domains, but don't always have domain-specific names. It is a little harder to see the equivalents for theory parameters. The simulation grid gives a sampling precision, but resolution may be hard to determine without post-processing.

As will be shown below, we propose to group the first four rows of the above tables into a somewhat global Coverage Property while Sampling or Resolution are more "local" Characterisation properties.

2.3 Structure of Characterisation and development strategy

The strategy here is to give a general interpretation frame to list out all kind of metadata that could be necessary, and to make explicit the relationships they have together. The description is set according to a Property versus Axis perspective. Moreover, the different levels of details one might encounter are organised as a set of progressive description layers. In this way the possible evolution of the model can be taken into account. This makes the model expandable in three independent directions: new properties may be added as well as new axes, and if necessary new levels of description may be considered without breaking the overall structure.

2.4 The Axis point of view

2.4.1 The axis concept and its attributes

The physical dimensions along which the data are spanned are described by the concept of *axis*, for example: SPATIAL, SPECTRAL, TIME, VELOCITY, VISIBILITY, POLARISATION, OBSERVABLE (flux, radial velocity, or whatever is measured).²

The data provider is allowed to name the axis arbitrarily (often FITS data will have explicit names for the axes), although this model will recommend the above list of names for the common cases they embody. For uniqueness, data provider will be *required* to supply a UCD to describe the axis, as

²Since the VISIBILITY axis is just the Fourier transform of the SPATIAL axis, one could imagine making that explicit in the model. However, for simplicity in this version of the model, one way is to add a separate visibility axis. So, to describe a radio image and include information on the spatial frequencies present, one would have both spatial and visibility axes in the Characterisation instance. In later versions of the model, a fancier way to express the same thing may be provided .

well as the units in which the data are delivered. The UCIDs serve as our controlled vocabulary to ensure the dataset is well described; the arbitrary names provide useful labels for visualization software.

The Characterisation model is intended to describe either calibrated or uncalibrated data. For example, for extragalactic objects there exists plenty of archival spectral data which is well calibrated on the wavelength axis but not on the flux axis: it is useful for measuring redshifts but not luminosities. Similarly, some data may be well calibrated photometrically but have no astrometric calibration beyond a simple approximate sky location. Whether these partially calibrated data are useful depends on the user's science goals, and so it's useful to be able to specify such a concept in a search query. The AxisFrame object in the Characterisation model allows us to flag this information for each axis. The full details of the calibration process actually belong to the Provenance part of the Observation data Model, and is not relevant here.

2.4.2 Accuracy

The Accuracy of the data is also a concept Characterising the dataset for many purposes. Accuracy, intended as the grouping of Systematic and Statistical error, as well as a quality flag, is another piece of description attached to an axis. Astrometric accuracy is attached to the spatial axis, photometric to the Observable one, etc.. This overall accuracy estimate for the data is not to be confused with errors and accuracies provided on each of the metadata quantities used in the Characterisation DM. (See below, in the Model section 3)

2.4.3 Other Axis flags: sampling, observables, etc.

For each axis it is important to give the number of samples, a number of samples of 1 meaning that the axis is unsampled. In that case it can also be said unresolved, and the corresponding Coverage Bound element will provide actually the same information than the FWHM of the response function. If an axis is not sampled then the corresponding Sampling property in Characterisation is no longer needed. That helps in providing an optimised description of Characterisation. The same applies to the Resolution subtree.

If it is sampled, one would also like to know if the data provider considers it as undersampled or equally sampled. Two boolean quantities will be provided for that.

There is a difference between an "*Observable*" axis (whatever ucid it has: "flux", "mass", "velocity"), which represent a phenomenon measured along some other axes and these axes themselves. For instance, in a flux image of the sky, the Spatial axis is an independent axis, as is the

(implicit) Spectral axis, but the Flux axis is dependent (an ‘output’ axis) - it’s the values in the pixels. Since in more complicated types of data (and especially in simulations) more than one variable may be generated (Observable), it might be hard in general for software to figure out which axes are independent and which are dependent. A per-Axis flag will be provided to specify this.

Last but not least, the coordinates on each axis need to be referred in a specific coordinate frame, which is mandatory to describe. An Observatory Location, which may be necessary for some coordinate transformations, should also be provided if possible.

2.5 The Property point of view

The main properties that are supposed to be used by data providers and VO users are categorized under Coverage, Resolution, and SamplingPrecision.

How an observation is spread along the various axes is given by the Coverage class.

Coverage encompasses information at different granularity levels. The coarsest one describes only a single, nominal position of the VO entity. At the finest level, is a full sensitivity function, giving the absolute transmission factor for each element in the parameter space. Coverage has several intermediate levels, allowing any user/developer to find the appropriate details she/he wants to have. SamplingPrecision describes how the parameter space is scanned by the data, and Resolution gives the smearing of the sampling element between the physical and dataset parameter spaces.

2.5.1 Coverage

Several different levels of depth can be used when describing a data set. The crucial conceptual unification in our Coverage model is the realization that the field of view of an observation has a fuzzy boundary. The concept of spatial field of view presumes that there is a region in the coordinate space (e.g. celestial position) where you get data, and everywhere else you can’t see. In fact, this is just an approximation to the concept of sensitivity or transmission: you see 100 per cent in the middle of the field of view, 0 per cent far outside the field of view, but there is often a region near the boundary where the transmission changes rapidly but not instantly from 100 percent to 50 percent to zero (This is particularly true in dithering instruments such as X-ray telescopes). Similarly, a simple representation of a spectral bandpass is to say that the observation ‘covers a wavelength range from λ_1 to λ_2 ’, which is really saying that you are approximating the spectral transmission curve with a sharp-edged rectangular top-hat function. Hence, the questions ‘what range in my coordinate axes correspond to valid data?’ and ‘what is the variation in (flux) sensitivity of my observation as a function of the coordinate axes?’ are actually the same question at different levels of detail

- a unity that is obscured when you think in terms of sharp-edged pixellated FITS images. Our Coverage model provides answers to these questions at different levels of precision, with the idea that software implementations will be able to convert between the levels.

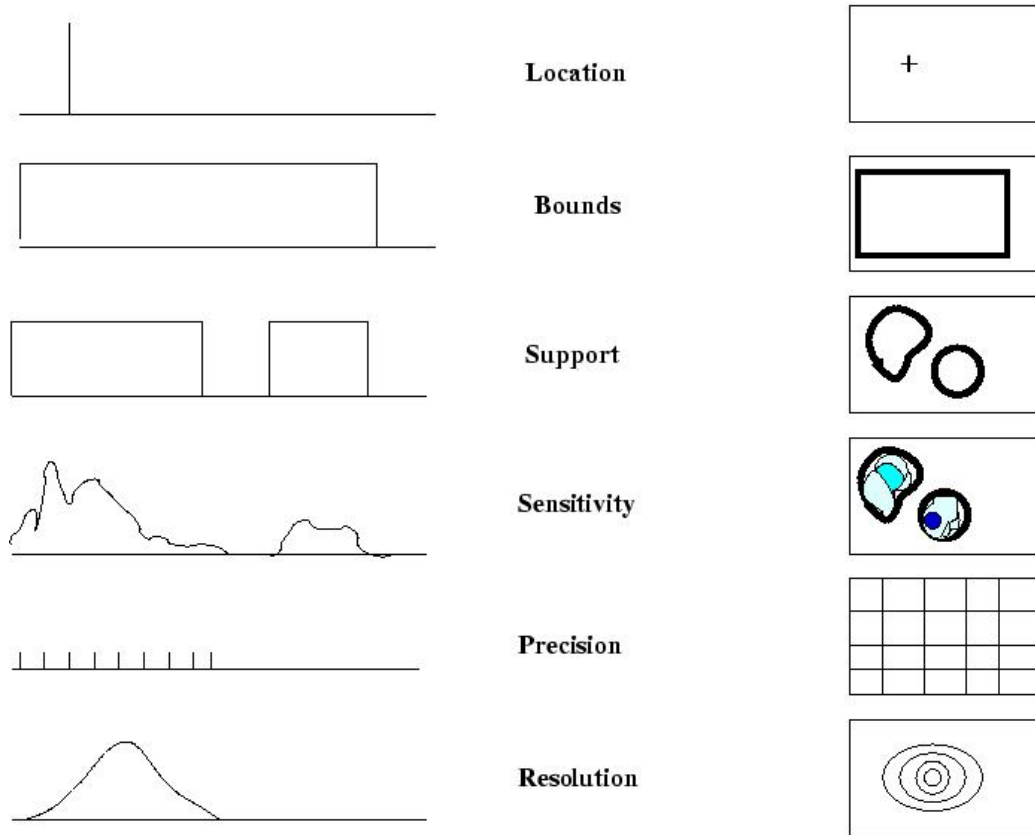


Figure 2: *Illustration of the different levels of description. left: for a 1-dimensional signal, right: for a 2D signal.*

Location The most simple Coverage description is a Location in arbitrary parameter space - for example, the statement that an image is at a particular RA, Dec, and taken at a particular wavelength and time. Here the interpretation is that the values given are fiducial values representative of the data, with no precise definition (mean, weighted median, etc.) being required.

Bounds The next level of description is the SensitivityBounds, i.e. a single range in each parameter. The interpretation is that all the valid data is guaranteed not to lie outside these bounds, but there may still be some values within the bounds for which there is no valid data. There is a slight loophole here with the word ‘valid’: for instance, if a spectral filter has a red leak, the frequency ranges in the red leak may be considered as invalid data (with quality values so marked) and outside the bounds. This satisfies the

intent of typical queries, which want to find observations which may have useful data within a given range of interest.

Support The Support component describes the more detailed context of the observation in a quantitative way. It will describe the space, time, frequency and other ranges covered by the data. Mathematically, the support of a function is the subset of its domain where the function is non-zero. In our model, this concept is slightly derived to mean the subsets of the domain where there are valid data (according to some specified quality criterion).

Note that these ranges may include the independent variables of the observational data samples as well as variables which are the same for each sample; thus for a 1-dimensional slit spectrum, the frequency range extremes of the spectrum (independent variable) as well as the start/stop time of observation and the region of the sky covered by the slit aperture (constants for the observation) will be described by the coverage. The coverage may have multiple ranges for a given parameter - particularly useful in the case of the temporal axis, where an observation may consist of the co-addition of several widely separated time ranges. For two-dimensional parameters such as sky position, the coverage can be described by Regions (whose interface is described separately).

Sensitivity The most detailed level of description is called Sensitivity, a relative sensitivity function which goes beyond the on/off coverage description to a description of the relative cell-to-cell sensitivity in the data. This includes filter transmission curves, flat fields, sensitivity maps, etc.

However, in practice Support and Sensitivity are not used to describe the case in which there are a large number of small interruptions to the data. This arises in the temporal domain with detectors which have dead time between each sample, or in the spatial domain with pixels with gaps between them so that the active area does not completely fill the focal plane. In these cases analysis systems may handle the problem with a statistical correction, correcting the effective sensitivity by a Fill Factor (usually constant for an observation but sometimes varying with the coordinates). This Fill factor can actually be obtained by the ratio between the SamplingPrecision and the Sample Extent (see below). It can easily be described by a method. The sampling function (level 4, see below) could be another answer to that problem.

The final level of Characterisation in this sequence is Absolute Sensitivity, which includes the upper limit value of the Observable (e.g. limiting magnitude) at a given position, and the value corresponding to one detector count in cases where that concept applies.

The values of these Coverage (and Sensitivity) characterising one Observation may be derived from a number of factors, some of them described by other Characterisation features and others by Provenance details (for ex-

ample, the spectral sensitivity may have been derived from the Instrument and Filter in the Provenance). These links between various attributes in the model will be reflected in the data model implementation (using for example "attribute formulae").

2.5.2 Resolution and Sampling Precision

The concepts of resolution and sampling precision (or pixelization) are related. If the sampling precision period is small compared to the resolution, the knowledge of a single data value is limited by the resolution (oversampling). If the sampling precision period is coarse compared to the resolution, knowledge of a single data value is limited by the sampling (undersampling).

Resolution Ultimately resolution describes the continuous smearing of our knowledge about the data, or more precisely the probability that a photon (or other observable) which has one set of attributes is measured as having a different set of attributes. Mathematically, if the physical attributes (e.g. position, time, energy) of the photons are \mathbf{x} (e.g. $x_0 = \text{energy}$, $x_1 = \text{RA}$, $x_2 = \text{Dec}$, $x_3 = \text{time}$, etc.), and the measured attributes are \mathbf{y} (e.g. $y_1 = \text{spectral channel}$, $y_2, y_3 = \text{pixel position}$, $y_4 = \text{time bin}$) then given a flux of photons $S(\mathbf{x})$ the detected number of photons is

$$N(y_1, y_2, \dots) = N(\mathbf{y}) = \int \mathbf{S}(\mathbf{x})\mathbf{A}(\mathbf{x})\mathbf{R}(\mathbf{x}, \mathbf{y})d\mathbf{x}$$

where \mathbf{A} is the probability that a photon is detected at all (the quantum efficiency) and $R(x_1, x_2, \dots, y_1, y_2, \dots)$ is the smearing of measured values (PSF, line spread function, etc.).

In the most detailed case, the \mathbf{R} function may be specified as a function of the coordinates - for instance, a PSF which varies as a function of detector position and energy. The first level of simplification is to specify a single function which applies to the whole observation - e.g. a single PSF. This function may either be provided as a parameterized predefined function (e.g. gaussian) or as an array. The final level of simplification is to give a single number characterising the resolution, effectively implying a single-parameter default predefined function. Several versions of these simple resolution parameters may be supported; initially the standard deviation of a gaussian can be used to represent the resolution average value. The concept of Resolution Bounds gives the extreme values of this resolution parameter.

Sampling Sampling (or pixelization or precision or quantization) describes the truncation of data values as part of the data acquisition or data processing (e.g. via resampling). If the mapping of the data coordinates (the pixelized/truncated ones) to the coordinate axes is nonlinear, the sampling precision varies from sample to sample; as a simplification one can give the limits (bounds) in between it is changing; the last simplification level is the

definition of a ‘characteristic sampling precision’ for the whole observation. Beside the Sampling Period, the Sample Extent shows how far the sampling differs from the pure ”Dirac comb” case. It is useful to give it, because it could also affect the resolution in some cases. The ratio between the period and the extent is actually the ”filling factor” which could be provided by a method. In the same way, the Nyquist parameter - the ratio between the resolution FWHM and the Sampling period - will also be provided by a method.

In some use cases, it may be difficult to distinguish the Sampling function from the Coverage support. Considering a Time series where each individual sub-observation is only known by its start and stop time: the union of time intervals limited by these Start/Stop times is definitely the Time Support of this observation, defined as the set of instants for which there is data. But it can be seen obviously also as a sampling function giving the variable Period and Extent of the SamplingPrecision. This kind of ambiguity occurs each time the Sample Extent is much larger than the Resolution, and the Sampling Period is again pretty larger than the Extent itself. So in such a time series the ”Exposure Time” is given either by the Support ”length” corrected by the filling factor or by adding all the Sample Extents of the dataset.

The distinction between continuous smearing (resolution) and discrete quantization (sampling) often - but not always - reflects a physical distinction between the atmosphere/telescope optics combination and the discrete pixels and A/D signal conversion of a detector. More importantly for our model, it reflects aspects of the data which are handled differently in downstream data analysis.

The resolution can also be interpreted as the lowest detectable frequency on a given axis. Radio astronomers emphasize the need to give the ”largest” detectable spatial frequency as an important characterising item. For this version, this parameter could be given as the highest limit of the VISIBILITY axis Coverage Bounds, the lowest limit of it being actually the spatial resolution.

2.6 The layers of description

Organising the information into layers structure helps in structuring meta-data according to the various tasks they may be involved in. Data search and retrieval need the basic information such as pointing position, field of view, bandpass, date, for instance. On the contrary for a full data analysis or recalibration work , one will need to access to specific information such as the variation of the sampling within the field of view, the sensitivity or the signal to noise ratio, etc

The four levels described for Coverage have been generalised to the Resolution and SamplingPrecision concepts. The first level describes a reference value The second level provides the limits or Bounds. The third level indi-

cates the effective range or region covered on the axis. For Resolution and SamplingPrecision this level 3 will allow describing several disjoint ranges of local estimations).

The level 4 of this hierarchy encodes (represents) the variation in the sensitivity to the Observable(s) of one property. For example, level 4 of the Resolution property is the place to describe its variations along different axes: position (spatial), velocity, spectral,... **If available, in practice, such variations will often be packaged with the data, in such a way that level 4 could give a simple pointer to the appropriate information.** To implement encoding the variation of the property along different Characterisation axes, one could use different approaches: as an analytical function if available, or as a tabulated function, like a map for 2D signals or a lookup table for 1D observations like spectra or light curves. In fact the SensitivityMap level is a way to link the observed data to external calibration or ancillary data, with the same types of axes and dimensions as the observation itself. It should also contain documentation. These additional related data are sometimes stored and distributed with the data themselves in the same packaging. Many astronomers would like the ability to know if some data corrections or calibration data may be retrieved in addition to the observational data and if so, what their nature is. Advanced VO tools could use such variation metadata to recalibrate data on demand. In the first version of this model, a simple link to a SensitivityMap will be provided, with a documentation link as well. This allows to make software aware of its existence. In the version 2 of the model, further details and classification of various possible maps will be provided.

3 The Model

3.1 The role and structure of the Model

In order to describe the organisation of the metadata under the characterisation framework, place holders have been modeled, following our Properties/Axis/Levels perspective and using UML diagrams. This is a way to make explicit the main structures and their relationships. The model offers different views of the characterisation concepts. Fig 3 shows the relationships between the main concepts. The CharacterisationAxis box attached to each property class (for instance Resolution) is a template parameter in UML. It just mentions that there can be one Resolution class for each relevant axis. In other terms, it represents the axis along which the property is assessed. Fig 4 illustrates how the properties of the data are gathered under the Characterisation container class. To provide a complete characterisation of one data set, one should repeat one Characterisation tree for each CharacterisationAxis value.

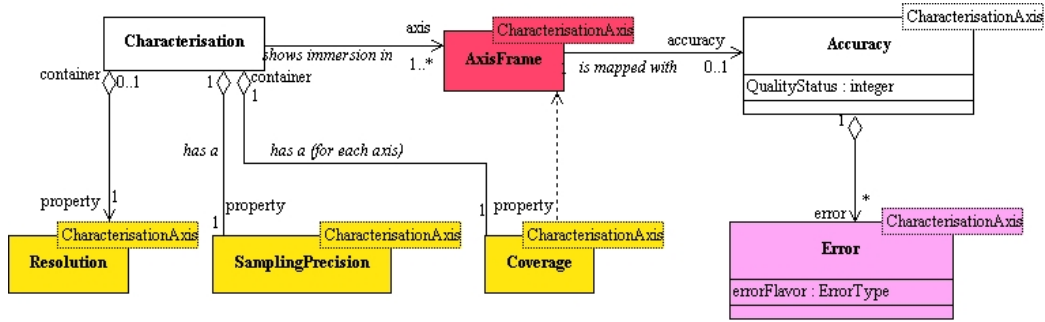


Figure 3: This class diagram emphasises the Property/Axis perspective. The Characterisation class is a container that gathers the properties for each axis. The axis is described by the AxisFrame class. The Accuracy class linked to it has pointers to Errors descriptions. The link 'shows immersion in' represents the list of all the relevant axes for one observation/dataset.

3.2 AxisFrame

The information related to this CharacterisationAxis parameter is described by a specific class: AxisFrame. It is related to the Frame concept in Quantity, containing the UCD, units, name, and a holder for the STC coordinate frame and observatory location. It also provides an Accuracy object gathering the errors on the axes.

Other elements in the AxisFrame class include the number of bins present on any axis, and flags to indicate the calibration status, independency and sampling properties of the axis:

- CalibrationStatus can be:
 - UNCALIBRATED: data on this axis are not in true physical units
 - CALIBRATED: data on this axis are fully calibrated
 - RELATIVE: the data on this axis are calibrated, but there is an unknown constant multiplicative or additive factor relating the data to the astronomy (e.g. flux in a spectrum when it is not known how much of the light from the target fell inside the slit).
 - NORMALIZED: this data were calibrated by dividing by another data set, and so are dimensionless.
- IndependentAxis: a flag which is TRUE for dataset axes or independent variables and FALSE for the Observable(s).
- Undersampling: a flag which is TRUE if the data are undersampled on this axis.
- RegularSampling: a flag which is TRUE if the mapping between the bins and the axis world coordinate is sufficiently close to being linear

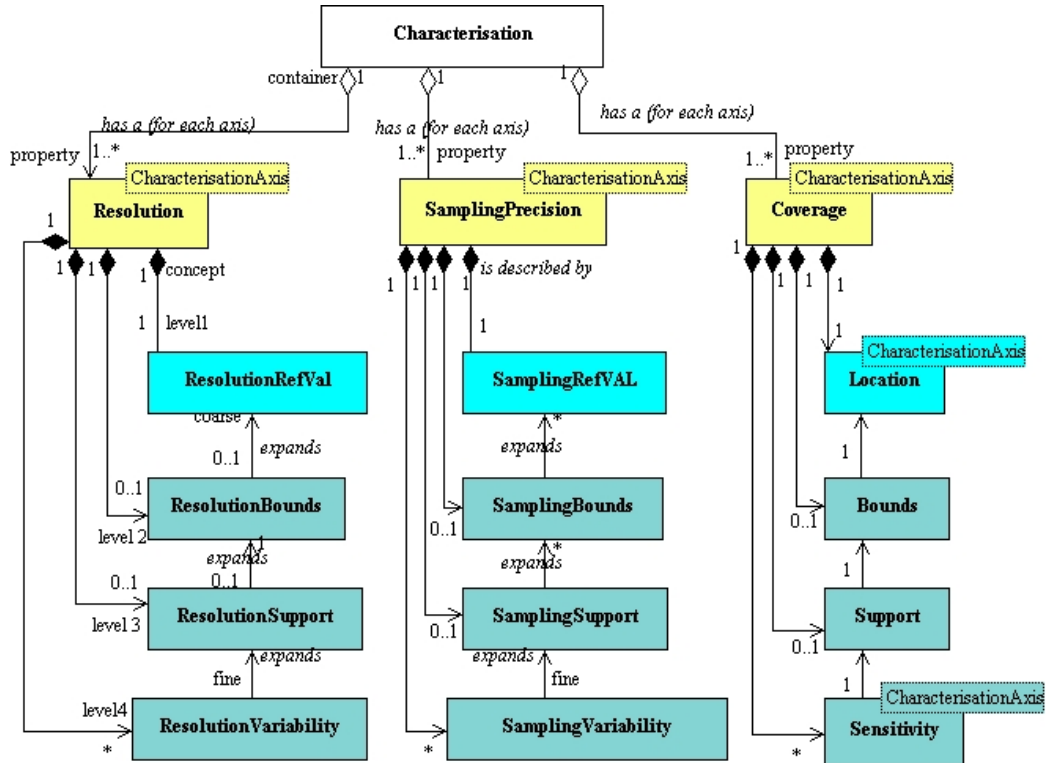


Figure 4: UML diagram: The layered structure of characterisation. This diagram synthesises the Property/Axis/Layer approach. The concepts are represented in yellow. The coarse description is designed by the blue boxes, while the grey ones represent the complementary metadata. The Bounds, Support and Sensitivity classes have nested levels of detail to add knowledge about the Coverage of an Observation. Symmetrically, Resolution and Sampling may also have the 4-level structure of description. The complete Characterisation for one observation is obtained by filling the tree for each relevant axis: spatial, spectral, temporal, etc.

that if you estimate coordinate values by using the lower bound and the pixel size the resulting error won't be important. For high precision use, data analysis will need to consult the actual WCS in the dataset.

If it happens that some deep level object, e.g. Sensitivity, needs to have its own AxisFrame object it can then be redefined at this place overriding the factorised top level AxisFrame object. It may be useful to redefine only the unit and CoordFrame at some levels (to change the spatial orientation for example).

3.3 Errors in Characterisation: the Accuracy class

All the items in Characterisation are assessments of physical properties and therefore may have errors. In addition, users may also need to have an assessment of the errors on the various coordinates and data values themselves.

First, let's consider the errors on Characterisation attribute values (e.g. the uncertainty in the spectral resolution). It's important to have such a description available, although one might concede that it's very unusual for data to be that well characterised. In the Quantity and STC data modes, Q:Quantity and STC:Coordinate are data types which contain uncertainty or "accuracy" values as well as the data values. Building up the Characterisation serialization on either of these data types, can automatically provide the option to add an error value to any of our attributes, without cluttering up the model by emphasizing this aspect.

Now, consider the uncertainties on the coordinate axes and on the data (observable) values. An Accuracy object is defined which is linked to the AxisFrame object, and gathers Error classes. Different types of errors may be hooked here: statistical and/or systematic, depending on the axis kind and its calibration status. The Error concept supports similar multiple levels of description to that used in Coverage: for each type of uncertainty, there can be a typical value, the bounds on the value, and the very detailed error value for each sampling element (e.g. pixel). Some modern observing programs now convey these error maps with the data in the same packaging. The need for an equivalent to the Support concept is not obvious here. This is valid for every Characterisation axis and can be considered as the error on the mapping from the pixel number to the world coordinate.

3.4 Properties and levels

Depending on the use case (see 1.1) the Characterisation DM is used for, all the levels are not necessarily filled. For data discovery, inspection, and comparison, the coverage part is needed till the level 3: Support gives the covered regions. The Resolution and Sampling information are probably well enough described with the two first levels: RefVal and Bounds.

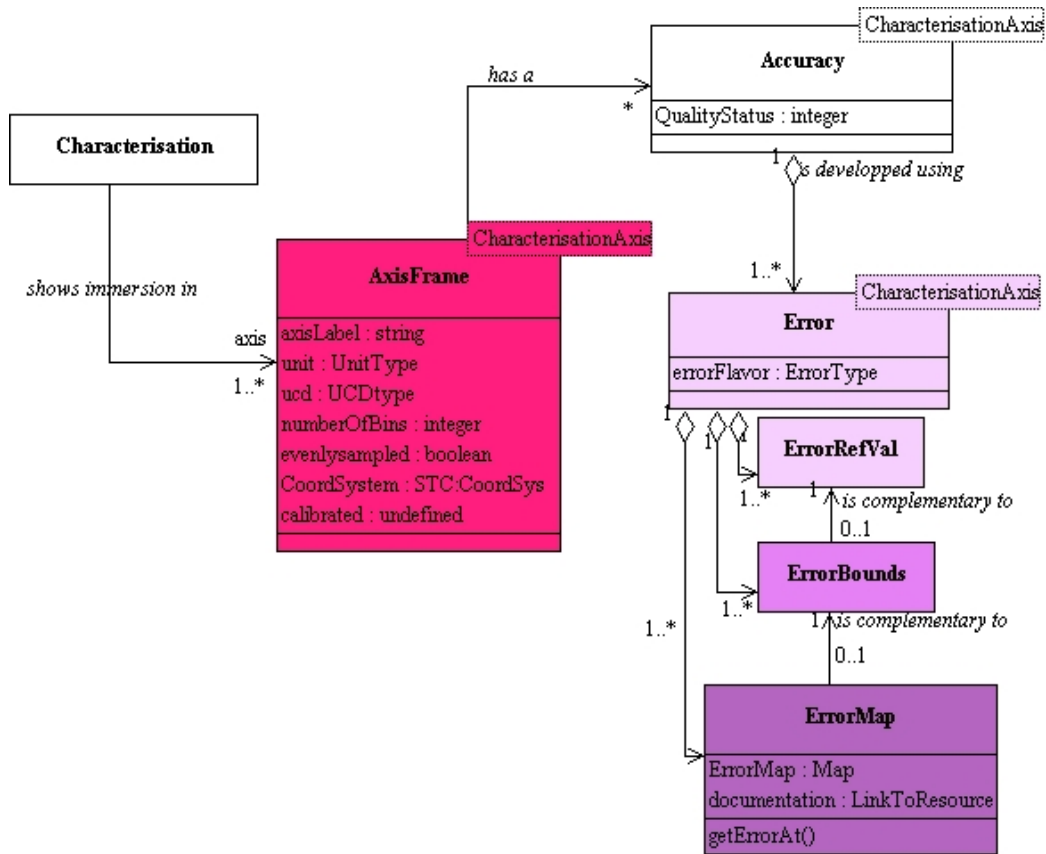


Figure 5: This class diagram illustrates the *AxisFrame* class and its relations with *Accuracy* and *Error* classes. *Accuracy* allows to gather different types of errors, like systematic or statistical. The quality of the mapping along the axis is encoded using a quality status word.

For the data processing use-case, then non uniformities of the PSF, for instance, or from the sampling step are relevant and can be available via the level 4. It is also necessary for data processing and interpretation.

Accommodating the various needs identified in the use cases, a layered structure as described in figure 4 is proposed, the top 3 levels containing the essential metadata for data retrieval, inspection and comparison and the following one providing extra details for advanced processing.

For complex observations obtained by combinations of several individual observations, a ComplexCharacterisation structure, that gathers simple characterisation objects, could be defined in a later version .

Dependencies on other parts of a more general "Observation" or "DataSet" data model could also be added later.

3.5 Navigation in the model: by axis or by properties?

The structure of the model is clearly hierarchical with the characterisation class as the root element. Whether browsing in the model is done by considering the properties first, and then their corresponding classes for each axis, or the contrary: axis description first, and then properties with the various levels, gives two kinds of navigation trees. They correspond to two possible serialisations of the model.

On one hand, according to the goal and design of VO services, one might want to see characterisation metadata organised in terms of properties instead of axes such as Spatial, Spectral, Temporal, Observable metadata. It could be useful for example for representation of data where axes are intertwined. On the other hand, factorising the Axis information for the multi layer description of one property is advantageous because it optimises the length of XML serialisations.

The inherent table structure as described in the example tables should be accessible both ways by property and by axis. This version of the UML model actually supports both possibilities, and can then be used to build two different XML schemas. **But in the next section the "Axis First" serialisation (and corresponding examples) *only* is presented. The "Property First" serialisation is delayed to a next version of this model.**

3.6 Implementing the DM with existing pieces of other models: Quantity and STC

Quantity tackles the problem of values representation that is dimension, coding, errors, units, UCDs, etc... Characterisation could make a fundamental use of the Quantity Frame class, as a template for its AxisFrame container. Any basic class such as Location , Support or Bounds, could also be implemented as a Quantity, but this would require another relationship between the Quantity datamodel and STC.

STC, the metadata scheme for Space-Time Coordinates (see <http://www.ivoa.net/Documents/PR/STC/STC-20050315.html>) encompasses the description of most of the Characterisation Axes defined, except the Observable one. Sensitivity is the only property not present in STC. This proposal is actually a general coordinate specification. A full STC structure for Characterisation could simply be reused. But this will prevent from having this multi-layer, multi-function scheme emphasized above, because the granularity of the top STC class does not match the nested structure. STC intermediate level objects are then more simple to use and may play the role of building blocks in the general scheme. In the model, the Location and Support classes may have direct links to the STC Coords and CoordArea metadata. This is illustrated by the UML diagram of Fig. 6.

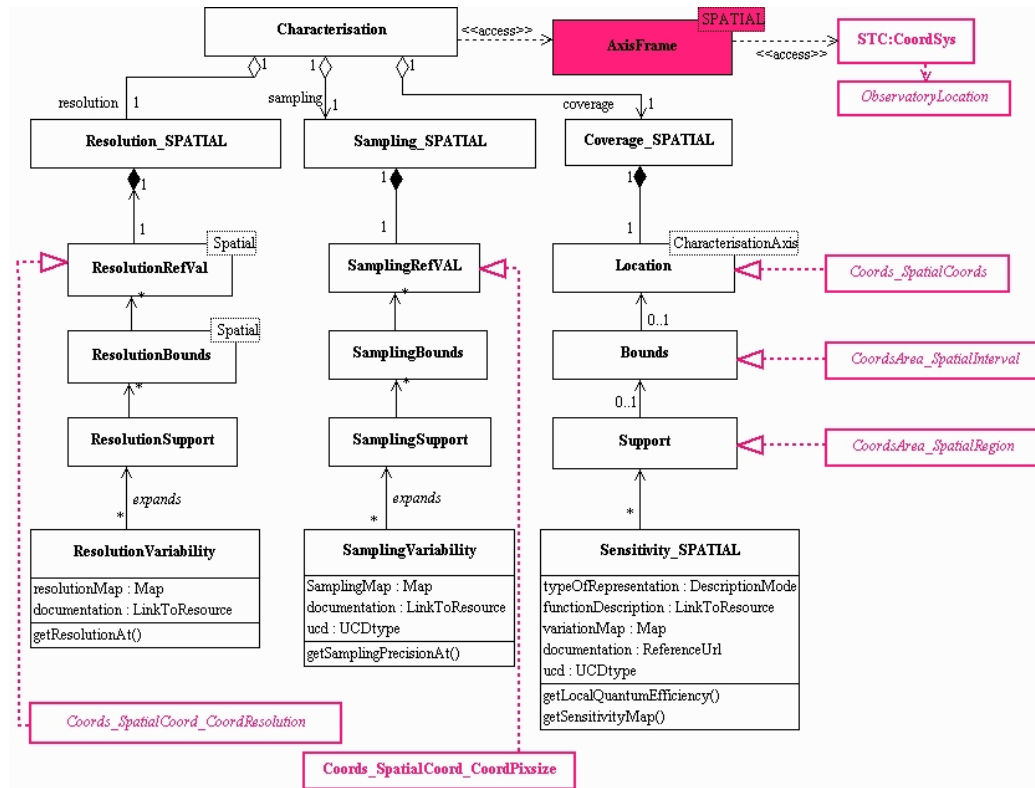


Figure 6: UML diagram: Expressing the spatial properties as a subtree of Characterisation . Here is an example of how STC components (in pink italics) may be used to implement the different levels of the Coverage description. The first level of Resolution and Sampling: RefVal also have STC counterparts.

With this a-priori, we can construct a Coverage object which consists of an arbitrary number of axes. Some of these axes will be the same as the axes of the main Observation Data, while others will represent phenomena that have been integrated over. For example, the simple 2D sky image has celestial coordinate axes, but has also been observed over a finite integration

time and wavelength band. The time and spectral axes are not present in the main data array, but their bounds - and even, for such things as color corrections, their sensitivity as a function of the coordinate within the bounds - may be represented in the Coverage.

The STC CoordSys object is needed as a reference to define the Coverage axes.

The Resolution and Pixel-Size concepts are represented in STC at a deep level inside the Coordinates class (together with the Name/Value/Error in the Coordinate object). This allows any coordinate to be properly interpreted in terms of resolution, which is necessary. But in the DM, we'd prefer to have it separate from Coords so that we could select metadata by resolution at the upper level of description. In this case, Coordinates could have a link to the Resolution Class. It is also a way to factorise the information and prevent redundancy or incoherence between coarse and detailed levels of description.

Since the space, time, spectral axes are particularly important for astronomy, one needs to verify that a complete and consistent space-time-spectral description is present. We recommend that implementations include a method to return an STC::AstroCoordSys object to provide this checking; an incomplete description will not be able to return one.

The STC definition also emphasizes the need to know the space-time coordinates of the Observatory (actually the aperture), potentially as a function of time. This will be the Observatory location for raw data, or the barycenter location for barycenter-corrected data, etc. In the future if the Characterisation data model is completed by a Provenance object (see Observation IVOA note) , it will definitely contain the actual Observatory Location.

4 XML Serialization

4.1 XML schema

4.1.1 Design of the schema

Due to the Hierarchical nature of the Model, the XML serialization of Characterisation is based here on a single tree. The root element called "Characterisation" is just aggregating a set of CharacterisationAxis elements³ for each of the axes. The AxisFrame element, (which could be derived from Quantity data model), is defined at the top level of CharacterisationAxis to help to immediately label the corresponding axis as "spatial", "time", etc.. In addition it includes one common CoordSys element as well as an ObservatoryLocation borrowed from STC XML schema. In simple cases Data handlers will probably reuse predefined elements included from an external STC library.

³These elements are containers gathering the result of dynamical grouping of properties for a given characterization axis

AxisFrames could also be present at lower levels, but will usually refer to the common one. Coverage implements different elements according to the four levels of description extensively described above.

STC substructures may be reused in the following way: Location implements STC:Coords, Bounds uses STC:Interval and Support STC:CoordArea. Resolution Refval can be implemented via STC:CResolution and the SampleExtent via CPixSize. This was represented using implementation links on the UML diagram in the Fig.6 for the spatial axis.

A full XML serialisation is provided, as an XML schema, for simple observations, at the following site:

<http://alinda.u-strasbg.fr/Model/Characterisation/schema/char.0.92.xsd>.

An XML instance document describing an IFU dataset characterisation is described in Appendix A and also browsable at

<http://alinda.u-strasbg.fr/Model/Characterisation/MPFS.xml>.

Three other implementations have been also considered:

- a more independent schema redefining internally only the specifically required STC and Quantity types. It has the drawback that it makes no reuse of other data models and could prevent from reusing related software too.
- a schema based on a STC-Quantity collaborative set of schema, which was able to manage any kind of STC and Characterisation type as a restriction on a very general Quantity type. The main drawback of this attempt was that it was not based on the current official version of the STC schemata.
- a schema where a full high level STC structure is defined together with the Characterisation types and where each STC element is referring to the appropriate Characterisation element - a variant referring in the other direction is also possible.

Full implementation of Characterisation software classes will probably benefit from a version of this schema based on Quantity and STC. Nevertheless, more compatibility between these two schemata is obviously needed before doing that.

4.1.2 Building blocks of the schemata

In order to illustrate how the XML schemata is derived from the UML Model, building blocks of the Schemata, corresponding to some main classes of the UML diagram are shown here.

The principle is to map the main classes in XML elements building up a hierarchy from the most englobing concept down to the more specific ones. Aggregated classes are easily translated as aggregated subelements. The

attributes of an UML class are also coded as sublevel elements.

This serialisation applies such rules and elaborated some specific techniques for the UML to XML translation in a way very similar to the work of Carlson. [1]. The examples shown here are 'handmade' translations of the UML model. The discussion about automated translation will be discussed in the version 2 of the Characterisation draft. The similarity and derivation process from UML to XML are expressed in the graphical views of the XML schema at Fig. 7 to 11.

4.1.3 Utypes generation: select one ordering strategy

One application of such a model is to provide a naming convention for every metadata considered within the model, in order to be able to identify one concept in various models or serialisations. The idea is that by navigating in the model following the logical links provided, it is possible to construct identifiers called Utypes that could be understandable by any VO tool aware of the model. To avoid multiplicity, the Utypes are build from the XML schema representation of the model which already enforces a hierarchical structure. For instance, the size of the sampling element in a 2D image along the spatial axis corresponds to

Characterisation.CharacterisationAxis.AxisFrame.AxisName.SamplingPrecision.SamplingRefval.

4.1.4 VOTable serialisation

A VOTable serialisation of the MPFS data set is shown in appendix B. Each CharacterisationAxis is seen as a table, where each property itself is seen as a Group of Fields. UML class attributes are serialised as FIELDS. Utypes are set on each Table, Group, and Field according to the following rule: In this example, a Utype is elaborated for each VOTable item in the serialisation. It is a string based on instance variable paths in object-oriented programming mode.

Other ways of deriving utypes from a valid Xpath to the equivalent XML element in the XML Characterisation schema have been studied. The main difference is that this option may use constrained element (or attribute) value in the Utype path. IVOA obviously has to define a single and robust rule to define this concept.

References

- [1] David Carlson. *Modeling XML Applications with UML*. Addison-Wesley, 2001.

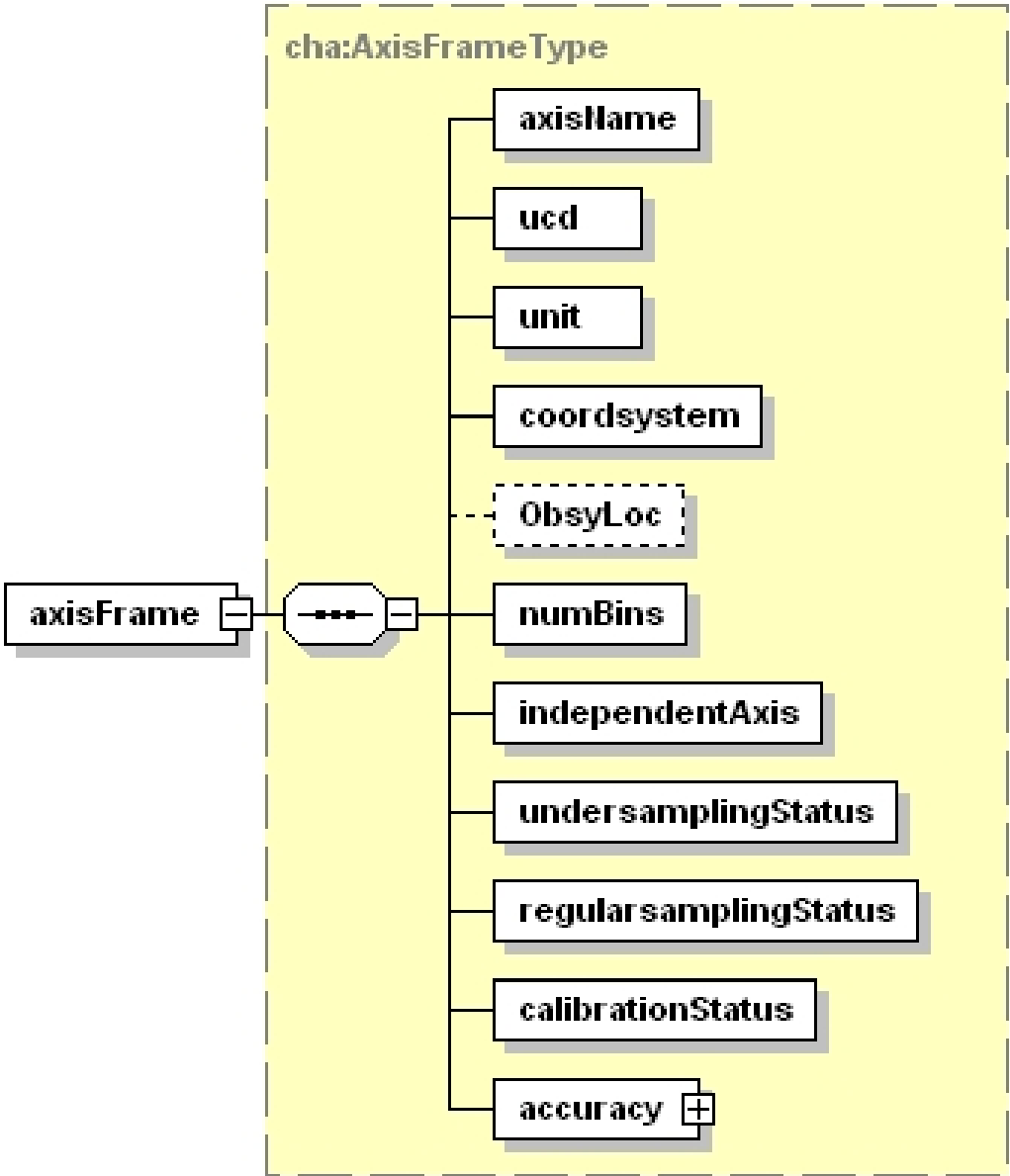


Figure 7: The `coordsystem` and `ObsyLoc` items are STC elements. Other elements are just copied from the UML class attributes.

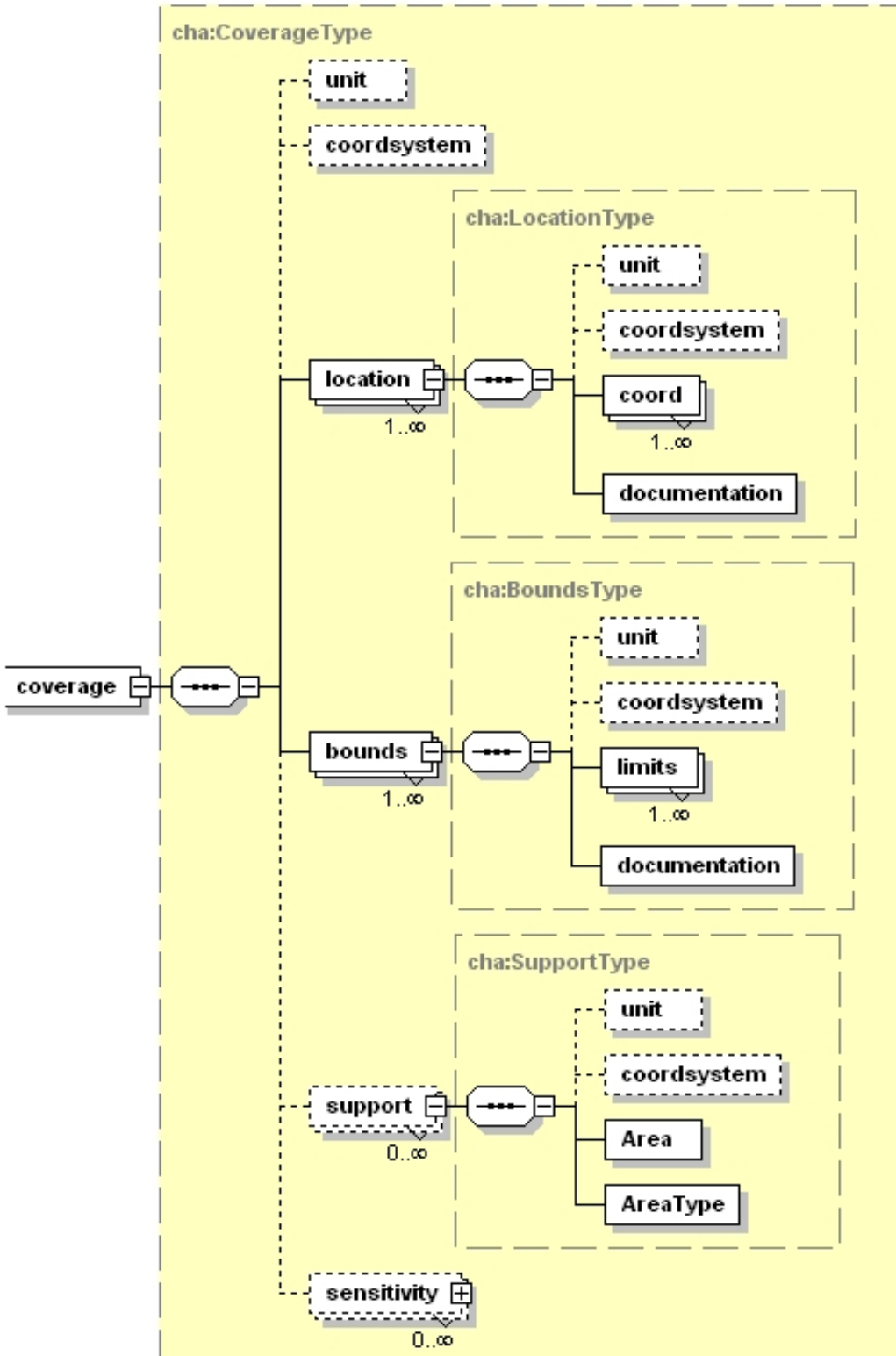


Figure 8: *The coordsystem and unit items can be factorised at the top of the Coverage structure, but may be redefined at each level when necessary.*

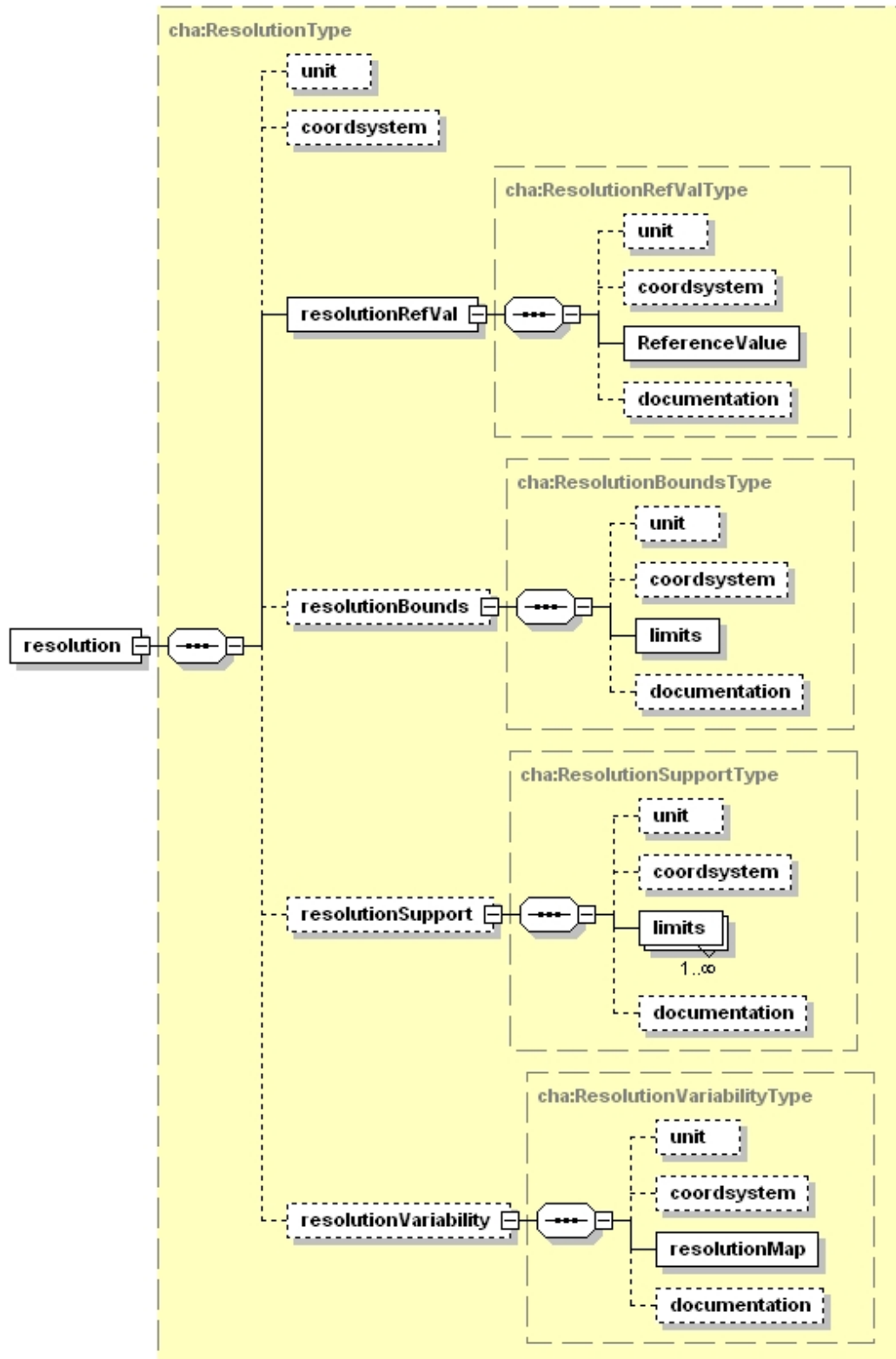


Figure 9: This graphical view was generated with XMLSPY from the resolution element of the schema. As designed in the UML class, the resolution item contains 4 possible subelements, with the RefVal and Bounds as mandatory elements.

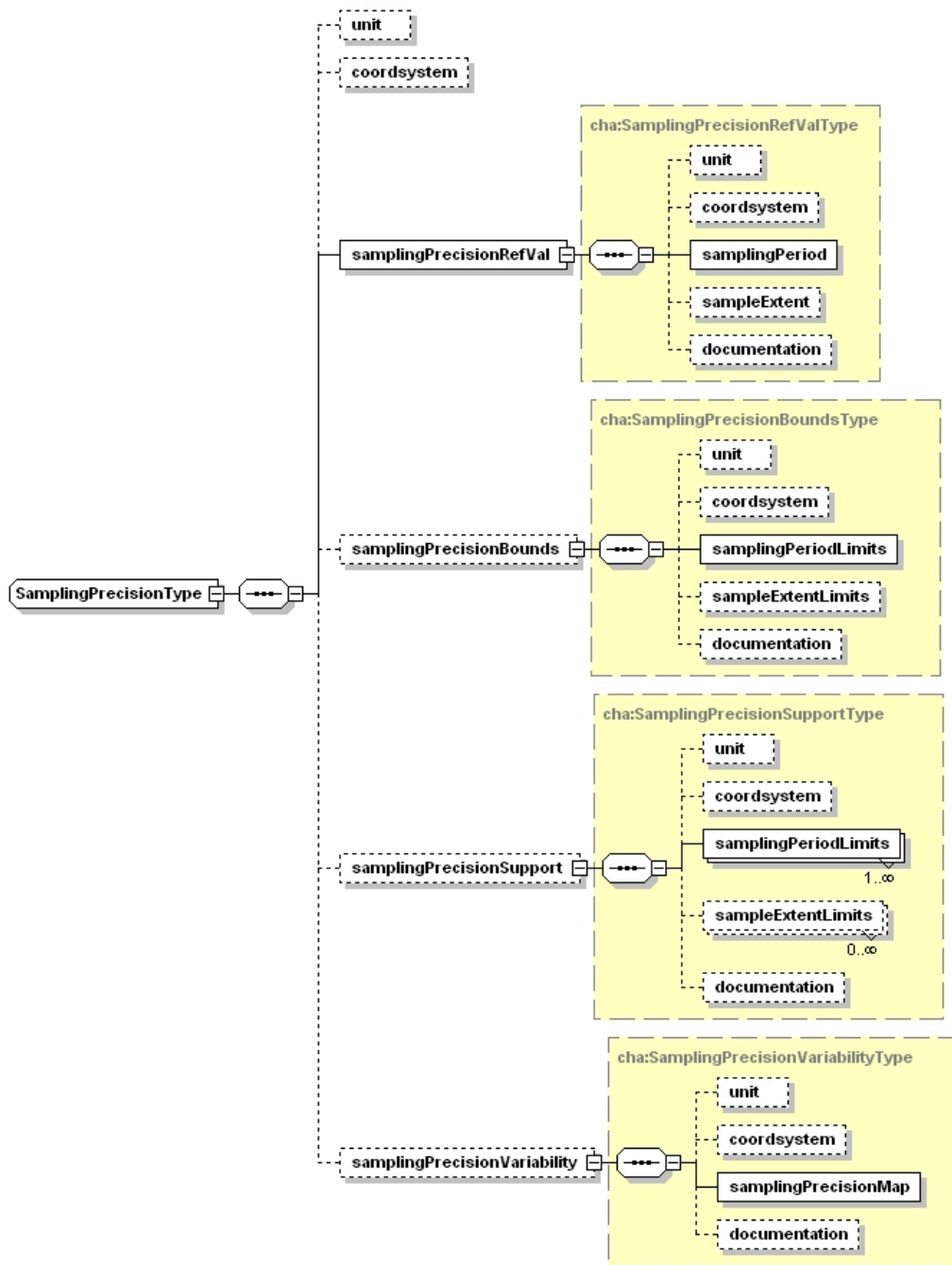


Figure 10: This graphical view was generated with XMLSPY from the `samplingPrecision` element of the schema. As designed in the UML class, the `samplingPrecision` item contains 4 possible subelements, with the `RefVal` and `Bounds` as mandatory elements.

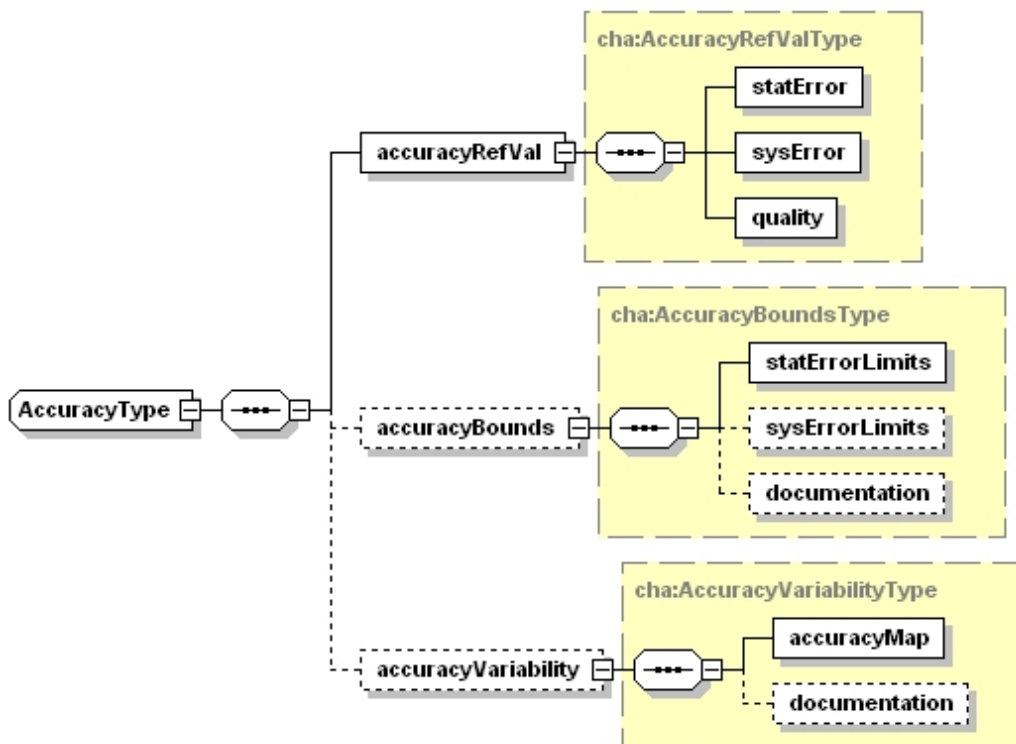


Figure 11: *This graphical view was generated with XMLSPY from the accuracy element of the schema.*

5 Appendix A : XML serialisation example

Here is an XML instance document representing the characterisation of an IFU data set, taken with the Russian MPFS instrument. It relies on the XML schema mentioned above. The xml file can be browsed easily at <http://alinda.u-strasbg.fr/Model/Characterisation/examples/MPFS.xml>

```
<?xml version="1.0" encoding="UTF-8"?> <!-- edited with XMLSpy v2005
rel. 3 U (http://www.altova.com) by bonnarel (CDS) --> <MPFS
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:stc="http://www.ivoa.net/xml/STC/v1.20"
xmlns:crd="http://www.ivoa.net/xml/STC/STCcoords/v1.20"
xmlns:cha="urn:vo-characterization"> <Characterization ID="MPFS">
  <characterizationAxis>
    <axisFrame>
      <axisName>spatial</axisName>
      <calibrationStatus>CALIBRATED</calibrationStatus>
      <ucd>pos</ucd>
      <unit>deg</unit>
      <coordsystem ID="FK5">
      </coordsystem>
      <ObsyLoc></ObsyLoc>
      <accuracy>
        <accuracyRefVal>
          <statErr>
            0.00055
          </statErr>
        </accuracyRefVal>
      </accuracy>
      <independentAxis>TRUE</independentAxis>
      <numBins>16 16</numBins>
      <undersamplingStatus>FALSE</undersamplingStatus>
      <regularsamplingStatus>TRUE</regularsamplingStatus>
    </axisFrame>
    <coverage>
      <location>
        <coord coordsystem_id="FK5">
          <crd:Position2D unit="deg">
            <crd:Name>RA,Dec</crd:Name>
            <crd:Value2>190.37379 11.366944 </crd:Value2>
          </crd:Position2D>
        </coord>
        <documentation>
        </documentation>
      </location>
      <bounds>
        <limits>
          <stc:LoLimit2Vec>190.37157 11.364722 </stc:LoLimit2Vec>
          <stc:HiLimit2Vec>190.37601 11.369167 </stc:HiLimit2Vec>
        </limits>
        <documentation>
        </documentation>
      </bounds>
    </coverage>
    <resolution>
      <unit> arcsec </unit>
      <resolutionRefVal>
        <ReferenceValue> 1.4 </ReferenceValue>
      </resolutionRefVal>
    </resolution>
    <samplingPrecision>
      <unit> arcsec </unit>
      <samplingPrecisionRefVal>
      <samplingPeriod> 1.0 </samplingPeriod>
      </samplingPrecisionRefVal>
    </samplingPrecision>
  </characterizationAxis>
  <characterizationAxis>
    <axisFrame>
      <axisName>time</axisName>
      <calibrationStatus>UNCALIBRATED</calibrationStatus>
      <ucd>time</ucd>
      <unit> s </unit>
      <coordsystem ID="UTC">
      </coordsystem>
      <ObsyLoc>
      </ObsyLoc>
      <accuracy>
      </accuracy>
      <independentAxis>TRUE</independentAxis>
      <numBins>1</numBins>
      <undersamplingStatus></undersamplingStatus>
      <regularsamplingStatus></regularsamplingStatus>
    </axisFrame>
    <coverage>
```

```

    <location>
      <coord coordsystem_id="UTC">
        <crd:Time unit="s">
          <crd:Name>Time</crd:Name>
          <crd:Value>2004/24/05 22:23:58 </crd:Value>
        </crd:Time>
      </coord>
      <documentation>
    </documentation>
    </location>
  </bounds>
  <limits>
    <stc:LoLimit> </stc:LoLimit>
    <stc:HiLimit> </stc:HiLimit>
  </limits>
  <documentation>
</documentation>
</bounds>
</coverage>
<resolution>
  <resolutionRefVal>
    <ReferenceValue> </ReferenceValue>
  </resolutionRefVal>
</resolution>
<samplingPrecision>
  <samplingPrecisionRefVal>
    <samplingPeriod> </samplingPeriod>
  </samplingPrecisionRefVal>
</samplingPrecision>
</characterizationAxis>
<characterizationAxis>
  <axisFrame>
    <axisName>spectral</axisName>
    <calibrationStatus>CALIBRATED</calibrationStatus>
    <ucd>em</ucd>
    <unit> um </unit>
    <coordsystem>
    </coordsystem>
    <ObsyLoc>
    </ObsyLoc>
    <accuracy>*>
      <accuracyRefVal>
      <statErr>
        0.0001
      </statErr>
      </accuracyRefVal>
    </accuracy>
    <independentAxis>TRUE</independentAxis>
    <numBins>2048</numBins>
    <undersamplingStatus>FALSE</undersamplingStatus>
    <regularsamplingStatus>FALSE</regularsamplingStatus>
  </axisFrame>
  <coverage>
    <location>
      <coord>
        <crd:Spectral unit="um">
          <crd:Name>Wavelength</crd:Name>
          <crd:Value>0.4858137 </crd:Value>
        </crd:Spectral>
      </coord>
      <documentation>
    </documentation>
    </location>
    <bounds>
      <limits>
        <stc:LoLimit>0.4140 </stc:LoLimit>
        <stc:HiLimit>0.56548382 </stc:HiLimit>
      </limits>
      <documentation>
    </documentation>
    </bounds>
  </coverage>
  <resolution>
    <unit> km/s </unit>
    <resolutionRefVal>
      <ReferenceValue>78.6162 </ReferenceValue>
    </resolutionRefVal>
    <resolutionBounds>
      <limits>
        <stc:LoLimit> 48.3233 </stc:LoLimit>
        <stc:HiLimit> 101.142 </stc:HiLimit>
      </limits>
    </resolutionBounds>
  </resolution>
  <samplingPrecision>
    <unit> km/s </unit>
    <samplingPrecisionRefVal>
      <samplingPeriod>40.0000</samplingPeriod>
    </samplingPrecisionRefVal>
    <samplingPrecisionBounds>
      <limits>
        <stc:LoLimit> 40.0000 </stc:LoLimit>

```

```

        <stc:HiLimit> 40.0000 </stc:HiLimit>
      </limits>
    </samplingPrecisionBounds>
  </samplingPrecision>
</characterizationAxis>
<characterizationAxis>
  <axisFrame>
    <axisName>flux</axisName>
    <calibrationStatus>UNCALIBRATED</calibrationStatus>
    <ucd>phot</ucd>
    <unit> </unit>
    <coordsystem>
    </coordsystem>
    <ObsyLoc>
    </ObsyLoc>
    <accuracy>
      <accuracyRefVal>
        <statErr>
          5.63e-17
        </statErr>
      </accuracyRefVal>
      <accuracyBounds>
        <statErrorLimits>
          5.80e-19 1.12e-16
        </statErrorLimits>
      </accuracyBounds>
    </accuracy>
    <independentAxis>FALSE</independentAxis>
    <numBins></numBins>
    <undersamplingStatus>FALSE</undersamplingStatus>
    <regularsamplingStatus>TRUE</regularsamplingStatus>
  </axisFrame>

  <coverage>
    <location>
      <coord>
        <Flux unit="">
          <crd:Name>Flux</crd:Name>
          <crd:Value>2.3519587e-17 </crd:Value>
        </Flux>
      </coord>
      <documentation>
      </documentation>
    </location>
    <bounds>
      <limits>
        <stc:LoLimit>-2.8933970e-15 </stc:LoLimit>
        <stc:HiLimit> 1.1838107e-14 </stc:HiLimit>
      </limits>
      <documentation>
      </documentation>
    </bounds>
  </coverage>
  <resolution>
    <resolutionRefVal>
      <ReferenceValue> </ReferenceValue>
    </resolutionRefVal>
  </resolution>
  <samplingPrecision>
    <samplingPrecisionRefVal>
      <samplingPeriod> </samplingPeriod>
    </samplingPrecisionRefVal>
  </samplingPrecision>
</characterizationAxis>
</Characterization>
</MPFS>

```

6 Appendix B : VOTable serialisation example

Here is another kind of serialisation, using the VOTable format and applying the Utype mechanism to map the various VOTable items to the Characterisation Data Model classes and attributes. See the xml file at <http://alinda.u-strasbg.fr/Model/Characterisation/examples/MPFSVot.xml>

```

<?xml version="1.0" encoding="UTF-8"?>
<VOTABLE version="1.1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://www.ivoa.net/xml/VOTable/v1.1">

```

```

<DESCRIPTION>
  Votable serialization of the MPFS.xml example of document
  characterizing an IFU
  F. Bonnarel Oct 5 2005
  based on Char WD of the same day edited by JCM,FB,IC,ML,AM,AR
</DESCRIPTION>

<RESOURCE utype="cha:characterization">
  <DESCRIPTION>
    This RESOURCE element is a container holding
    the full characterization of the IFU observation
  </DESCRIPTION>

<TABLE utype="cha:characterization.characterizationAxis">
  <DESCRIPTION> Spatial characterization </DESCRIPTION>

  <FIELD ID="Na" name="Name" datatype="char" arraysize="*"
    utype="cha:characterization.characterizationAxis.axisFrame.axisName" />
  <FIELD ID="Uc" name="Ucd" datatype="char" arraysize="*"
    utype="cha:characterization.characterizationAxis.axisFrame.ucd" />
  <FIELD ID="Ca" name="Calibration status" datatype="char" arraysize="*"
    utype="cha:characterization.characterizationAxis.axisFrame.calibrationStatus" >
    <VALUES>
      <OPTION value="CALIBRATED"/>
      <OPTION value="UNCALIBRATED"/>
      <OPTION value="RELATIVE"/>
    </VALUES>
  </FIELD>
  <FIELD ID="CooSys" name="Coordinate system" datatype="char" arraysize="*"
    utype="cha:characterization.characterizationAxis.axisFrame.coordsystem" />
  <FIELD ID="Ste" name=" Accuracy statistical error" datatype="double"
    utype="cha:characterization.characterizationAxis.axisFrame.accuracy.accuracyRefVal.statError"
    ucd="pos.eq;stat.error" />
  <FIELD ID="Sye" name=" Accuracy systematic error" datatype="double"
    utype="cha:characterization.characterizationAxis.axisFrame.accuracy.accracyRefVal.sysError"
    ucd="pos.eq;sys.error" />
  <FIELD ID="Ia" name="independent Axis flag" datatype="boolean"
    utype="cha:characterization.characterizationAxis.axisFrame.independentAxis" />
  <FIELD ID="Nb" name="number of Bins" datatype="int"
    utype="cha:characterization.characterizationAxis.axisFrame.numBins" />
  <FIELD ID="usSt" name=" unsampled Status" datatype="double"
    utype="cha:characterization.characterizationAxis.axisFrame.undersamplingStatus" />
  <FIELD ID="rsSt" name=" regular sampling Status" datatype="double"
    utype="cha:characterization/characterizationAxis/axisFrame/regularsamplingStatus" />
  <FIELD ID="RA" name="Right ascension" datatype="double"
    utype="cha:characterization.characterizationAxis.coverage.location.coord.stc:Position2D.Value2" />
  <FIELD ID="dec" name="Declination" datatype="double"
    utype="cha:characterization.characterizationAxis.coverage.location.coord.stc:Position2D/Value2"
    ucd="pos.eq" />
  <FIELD ID="LoLi" name="Spatial bounds low limit" datatype="double" arraysize="2"
    utype="cha:characterization/characterizationAxis/coverage/bounds/limits/stc:LoLimit2Vec"
    ucd="pos.eq"/>
  <FIELD ID="HiLi" name="Spatial bounds high limit" datatype="double" arraysize="2"
    utype="cha:characterization/characterizationAxis/coverage/bounds/limits/stc:HiLimit2Vec"
    ucd="pos.eq" />
  <FIELD ID="Res" name="Spatial Resolution" datatype="double"
    utype="cha:characterization.characterizationAxis.resolution.ResolutionRefVal.ReferenceValue" />
  <FIELD ID="Sam" name="Sampling Precision" datatype="double"
    utype="cha:characterization.characterizationAxis.samplingPrecision.SamplingPrecisionRefVal.SamplingPeriod" />
  <GROUP utype="cha:characterization.characterizationAxis.axisFrame">
    <FIELDref ref="Na" />
    <FIELDref ref="Uc" />
    <FIELDref ref="Ca" />
    <FIELDref ref="CooSys" />
    <FIELDref ref="Ste" />
    <FIELDref ref="Sye" />
    <FIELDref ref="Ia" />
    <FIELDref ref="Nb" />
    <FIELDref ref="usSt" />
    <FIELDref ref="rsSt" />
  </GROUP>
  <GROUP utype="cha:characterization.characterizationAxis.coverage">
    <GROUP utype="cha:characterization.characterizationAxis.coverage.location">
      <FIELDref ref="RA" />
      <FIELDref ref="dec" />
    </GROUP>
    <GROUP utype="cha:characterization.characterizationAxis.coverage.bounds">
      <FIELDref ref="LoLi" />
      <FIELDref ref="HiLi" />
    </GROUP>
  </GROUP>
  <GROUP utype="cha:characterization.characterizationAxis.resolution">
    <FIELDref ref="Res" />
  </GROUP>
  <GROUP utype="cha:characterization.characterizationAxis.samplingPrecision">
    <FIELDref ref="Sam" />
  </GROUP>
  <DATA><TABLEDATA>
<TR><TD>spatial</TD><TD>pos</TD><TD>CALIBRATED</TD><TD>FK5</TD><TD></TD></TR>

```

```

<TD></TD><TD></TD><TD>TRUE</TD><TD></TD><TD>FALSE</TD><TD>TRUE</TD>
<TD>190.37379</TD><TD>11.366944</TD><TD>190.37157 11.364722</TD>
<TD>190.37601 11.369167</TD><TD>1.4</TD><TD>1.0</TD></TR>
</TABLEDATA></DATA>
</TABLE>
<TABLE utype="cha:characterization.characterizationAxis">
<DESCRIPTION> Spatial characterization </DESCRIPTION>
<FIELD ID="TNa" name="Name" datatype="char" arraysize="*"
  utype="cha:characterization.characterizationAxis.axisFrame.axisName" />
<FIELD ID="TUC" name="Ucd" datatype="char" arraysize="*"
  utype="cha:characterization.characterizationAxis.axisFrame.ucd" />
<FIELD ID="TCa" name="Calibration status" datatype="char" arraysize="*"
  utype="cha:characterization.characterizationAxis.axisFrame.calibrationStatus" >
<VALUES>
<OPTION value="CALIBRATED"/>
<OPTION value="UNCALIBRATED"/>
<OPTION value="RELATIVE"/>
</VALUES>
</FIELD>
<FIELD ID="TCooSys" name="Coordinate system" datatype="char" arraysize="*" *
  utype="cha:characterization.characterizationAxis.axisFrame.coordsystem" />
<FIELD ID="TSte" name=" Accuracy statistical error" datatype="double"
  utype="cha:characterization.characterizationAxis.axisFrame.accuracy.accuracyRefVal.statError"
  ucd="pos.eq;stat.error" />
<FIELD ID="TSye" name=" Accuracy systematic error" datatype="double"
  utype="cha:characterization/characterizationAxis/axisFrame/accuracy/accuracyRefVal/sysError"
  ucd="pos.eq;sys.error" />
<FIELD ID="Tia" name="independent Axis flag" datatype="boolean"
  utype="cha:characterization.characterizationAxis.axisFrame.independentAxis" />
<FIELD ID="Tnb" name="number of Bins" datatype="int"
  utype="cha:characterization.characterizationAxis.axisFrame.numBins" />
<FIELD ID="TusSt" name=" unsampled Status" datatype="double"
  utype="cha:characterization.characterizationAxis.axisFrame.undersamplingStatus" />
<FIELD ID="TrsSt" name=" regular sampling Status" datatype="double"
  utype="cha:characterization.characterizationAxis.axisFrame.regularsamplingStatus" />
<FIELD ID="STi" name="observation time" datatype="double"
  utype="cha:characterization.characterizationAxis.coverage.location.coord.stc:Position.Value"
  ucd="time" />
<FIELD ID="TLi" name="Time bounds low limit" datatype="double" arraysize="2"
  utype="cha:characterization.characterizationAxis.coverage.bounds.limits.stc:LoLimit2Vec"
  ucd="pos.eq"/>
<FIELD ID="THiLi" name="Time bounds high limit" datatype="double" arraysize="2"
  utype="cha:characterization.characterizationAxis.coverage.bounds.limits.stc:HiLimit2Vec"
  ucd="pos.eq" />
<FIELD ID="Tres" name="Time Resolution" datatype="double"
  utype="cha:characterization.characterizationAxis.resolution.ResolutionRefVal.ReferenceValue" />
<FIELD ID="TSam" name="Sampling Precision" datatype="double"
  utype="cha:characterization.characterizationAxis.samplingPrecision.SamplingPrecisionRefVal.SamplingPeriod" />
<GROUP utype="cha:characterization.characterizationAxis.AxisFrame">
<FIELDref ref="TNa" />
<FIELDref ref="TUC" />
<FIELDref ref="TCa" />
<FIELDref ref="TCooSys" />
<FIELDref ref="TSte" />
<FIELDref ref="TSye" />
<FIELDref ref="Tia" />
<FIELDref ref="Tnb" />
<FIELDref ref="TusSt" />
<FIELDref ref="TrsSt" />
</GROUP>
<GROUP utype="cha:characterization.characterizationAxis.coverage">
<GROUP utype="cha:characterization.characterizationAxis.coverage.location">
<FIELDref ref="STi" />
</GROUP>
<GROUP utype="cha:characterization.characterizationAxis.coverage.bounds">
<FIELDref ref="TLi" />
<FIELDref ref="THiLi" />
</GROUP>
</GROUP>
<GROUP utype="cha:characterization.characterizationAxis.resolution">
<FIELDref ref="Tres" />
</GROUP>
<GROUP utype="cha:characterization.characterizationAxis.samplingPrecision">
<FIELDref ref="TSam" />
</GROUP>
</TABLE>
<DATA><TABLEDATA>
<TR><TD>temporal</TD><TD>time</TD><TD>UNCALIBRATED</TD><TD>UTC</TD><TD></TD>
<TD></TD><TD></TD><TD></TD><TD></TD><TD></TD><TD></TD><TD></TD>
<TD>2004/24/05 22:23:58</TD><TD></TD><TD></TD><TD></TD><TD></TD><TD></TD><TD></TD></TR>
</TABLEDATA></DATA>
</TABLE>
<TABLE utype="cha:characterization.characterizationAxis">
<DESCRIPTION> Spectral characterization </DESCRIPTION>
<FIELD ID="SNa" name="Name" datatype="char" arraysize="*"
  utype="cha:characterization.characterizationAxis.axisFrame.axisName" />
<FIELD ID="SUC" name="Ucd" datatype="char" arraysize="*"
  utype="cha:characterization.characterizationAxis.axisFrame.ucd" />
<FIELD ID="SCa" name="Calibration status" datatype="char" arraysize="*"
  utype="cha:characterization.characterizationAxis.axisFrame.calibrationStatus" >
<VALUES>
<OPTION value="CALIBRATED"/>

```

```

        <OPTION value="UNCALIBRATED"/>
        <OPTION value="RELATIVE"/>
    </VALUES>
</FIELD>
<FIELD ID="SCooSys" name="Coordinate system" datatype="char" arraysize="*"
    utype="cha:characterization.characterizationAxis.axisFrame.coordsystem" />
<FIELD ID="SSte" name="Accuracy statistical error" datatype="double"
    utype="cha:characterization.characterizationAxis.axisFrame.accuracy.accuracyRefVal.statError"
    ucd="pos.eq;stat.error" />
<FIELD ID="SSye" name="Accuracy systematic error" datatype="double"
    utype="cha:characterization.characterizationAxis.axisFrame.accuracy.accuracyRefVal.sysError"
    ucd="pos.eq;sys.error" />
<FIELD ID="Sia" name="independent Axis flag" datatype="boolean"
    utype="cha:characterization.characterizationAxis.axisFrame.independentAxis" />
<FIELD ID="Snb" name="number of Bins" datatype="int"
    utype="cha:characterization.characterizationAxis.axisFrame.numBins" />
<FIELD ID="SusSt" name="unsampled Status" datatype="double"
    utype="cha:characterization.characterizationAxis.axisFrame.undersamplingStatus" />
<FIELD ID="SrsSt" name="regular sampling Status" datatype="double"
    utype="cha:characterization.characterizationAxis.axisFrame.regularsamplingStatus" />
<FIELD ID="Wl" name="Mid wavelenght" datatype="double"
    utype="cha:characterization.characterizationAxis.coverage.location.coord.stc:Position.Value"
    ucd="em" />
<FIELD ID="SLoLi" name="Spectral band low limit" datatype="double" arraysize="1"
    utype="cha:characterization.characterizationAxis.coverage.bounds.limits.stc:LoLimit"
    ucd="em"/>
<FIELD ID="SHiLi" name="Spectral band high limit" datatype="double" arraysize="1"
    utype="cha:characterization.characterizationAxis.coverage.bounds.limits.stc:HiLimit"
    ucd="em" />
<FIELD ID="SRes" name="Spectral Resolution" datatype="double"
    utype="cha:characterization.characterizationAxis.resolution.ResolutionRefVal.ReferenceValue"
    unit="km/s"/>
<FIELD ID="SSam" name="Spectral sampling period" datatype="double"
    utype="cha:characterization.characterizationAxis.samplingPrecision.samplingPrecisionRefVal.samplingPeriod"
    unit="km/s"/>
<GROUP utype="cha:characterization.characterizationAxis.axisFrame">
    <FIELDref ref="SNa" />
    <FIELDref ref="SUc" />
    <FIELDref ref="SCa" />
    <FIELDref ref="SCooSys" />
    <FIELDref ref="SSte" />
    <FIELDref ref="SSye" />
    <FIELDref ref="Sia" />
    <FIELDref ref="Snb" />
    <FIELDref ref="SusSt" />
    <FIELDref ref="SrsSt" />
</GROUP>
<GROUP utype="cha:characterization.characterizationAxis.coverage">
<GROUP utype="cha:characterization.characterizationAxis.coverage.location">
    <FIELDref ref="Wl" />
</GROUP>
<GROUP utype="cha:characterization.characterizationAxis.coverage.bounds">
    <FIELDref ref="SLoLi" />
    <FIELDref ref="SHiLi" />
</GROUP>
</GROUP>
<GROUP utype="cha:characterization.characterizationAxis.resolution">
    <FIELDref ref="SRes" />
</GROUP>
<GROUP utype="cha:characterization.characterizationAxis.samplingPrecision">
    <FIELDref ref="SSam" />
</GROUP>
</GROUP>
<DATA><TABLEDATA>
<TR><TD>spectral</TD><TD>em.wl</TD><TD>CALIBRATED</TD><TD></TD><TD></TD></TR>
<TR><TD></TD><TD></TD><TD>TRUE</TD><TD></TD><TD>FALSE</TD><TD>FALSE</TD>
<TD>0.4858137</TD><TD>0.4140</TD><TD>0.5654382</TD><TD>78.6162</TD><TD>40.0</TD></TR>
</TABLEDATA></DATA>
</TABLE>
<TABLE utype="cha:characterization.characterizationAxis">
<DESCRIPTION> Flux characterization </DESCRIPTION>
<FIELD ID="FNa" name="Name" datatype="char" arraysize="*"
    utype="cha:characterization.characterizationAxis.axisFrame.axisName" />
<FIELD ID="FUc" name="Ucd" datatype="char" arraysize="*"
    utype="cha:characterization.characterizationAxis.axisFrame.ucd" />
<FIELD ID="FCa" name="Calibration status" datatype="char" arraysize="*"
    utype="cha:characterization.characterizationAxis.axisFrame.calibrationStatus" >
    <VALUES>
        <OPTION value="CALIBRATED"/>
        <OPTION value="UNCALIBRATED"/>
        <OPTION value="RELATIVE"/>
    </VALUES>
</FIELD>
<FIELD ID="FCooSys" name="Coordinate system" datatype="char" arraysize="*"
    utype="cha:characterization.characterizationAxis.axisFrame.coordsystem" />
<FIELD ID="FSte" name="Accuracy statistical error" datatype="double"
    utype="cha:characterization.characterizationAxis.axisFrame.accuracy.accuracyRefVal.statError"
    ucd="pos.eq;stat.error" />
<FIELD ID="FSye" name="Accuracy systematic error" datatype="double"
    utype="cha:characterization.characterizationAxis.axisFrame.accuracy.accuracyRefVal.sysError"
    ucd="pos.eq;sys.error" />
<FIELD ID="Fia" name="independent Axis flag" datatype="boolean"
    utype="cha:characterization.characterizationAxis.axisFrame.independentAxis" />

```

```

<FIELD ID="Fnb" name="number of Bins" datatype="int"
  utype="cha:characterization.characterizationAxis.axisFrame.numBins" />
<FIELD ID="FusSt" name=" unsampled Status" datatype="double"
  utype="cha:characterization.characterizationAxis.axisFrame.undersamplingStatus" />
<FIELD ID="FrsSt" name=" regular sampling Status" datatype="double"
  utype="cha:characterization.characterizationAxis.axisFrame.regularsamplingStatus" />
<FIELD ID="Fmid" name="Mid Flux" datatype="double"
  utype="cha:characterization.characterizationAxis.coverage.location.coord.stc:Position.Value"
  ucd="phot" />
<FIELD ID="FLoLi" name="Flux lowest detection" datatype="double" arraysize="1"
  utype="cha:characterization.characterizationAxis.coverage.bounds.limits.stc:LoLimit"
  ucd="phot"/>
<FIELD ID="FHiLi" name="Flux saturation limit" datatype="double" arraysize="1"
  utype="cha:characterization.characterizationAxis.coverage.bounds.limits.stc:HiLimit"
  ucd="phot" />
<FIELD ID="FRes" name="Flux Resolution" datatype="double"
  utype="cha:characterization.characterizationAxis.resolution.ResolutionRefVal.ReferenceValue" />
<FIELD ID="FSam" name="Flux digitazation step" datatype="double"
  utype="samplingPrecisionRefVal.samplingPeriod" />
<GROUP utype="cha:characterization.characterizationAxis.axisFrame">
  <FIELDref ref="FNa" />
  <FIELDref ref="FUc" />
  <FIELDref ref="FCa" />
  <FIELDref ref="FCooSys" />
  <FIELDref ref="FSte" />
  <FIELDref ref="Fia" />
  <FIELDref ref="Fnb" />
  <FIELDref ref="FSye" />
  <FIELDref ref="FusSt" />
  <FIELDref ref="FrsSt" />
</GROUP>
<GROUP utype="cha:characterization.characterizationAxis.coverage">
  <GROUP utype="cha:characterization.characterizationAxis.coverage.location">
    <FIELDref ref="Fmi" />
  </GROUP>
  <GROUP utype="cha:characterization.characterizationAxis.coverage.bounds">
    <FIELDref ref="FLoLi" />
    <FIELDref ref="FHiLi" />
  </GROUP>
  <GROUP utype="cha:characterization.characterizationAxis.resolution">
    <FIELDref ref="FRes" />
  </GROUP>
  <GROUP utype="cha:characterization.characterizationAxis.samplingPrecision">
    <FIELDref ref="FSam" />
  </GROUP>
</GROUP>
<DATA><TABLEDATA>
<TR><TD>flux</TD><TD>phot</TD><TD>CALIBRATED</TD><TD></TD><TD></TD><TD></TD></TR>
<TR><TD>FALSE</TD><TD></TD><TD>FALSE</TD><TD>TRUE</TD><TD>2.3519587e-17</TD>
<TD>-2.8933970e-15</TD><TD>1.1838107e-14</TD><TD></TD><TD></TD></TR>
</TABLEDATA></DATA>
</TABLE>
</RESOURCE>
</VOTABLE>

```