*Astronomical Data Analysis Software and Systems XIII*
*ASP Conference Series, Vol. 314, 2004*
*F. Ochsenbein, M. Allen, and D. Egret, eds.*

# A Unified Domain Model for Astronomy

Gerard Lemson

*Max-Planck-Institut für extraterrestrische Physik, Garching, Germany*

Patrick Dowler

*National Research Council Canada, Victoria, BC, Canada*

A. J. Banday

*Max-Planck-Institut für Astrophysik, Garching, Germany*

**Abstract.** We propose a framework for constructing a unified, conceptual domain model for astronomy. We believe such a model to be an essential ingredient for a future Virtual Observatory (VO). We also give a high level, skeleton proposal for this VO domain model. We indicate where details must be filled in for more specialized models. We describe one detailed base level model, a component model for Quantity, which is a generalisation of previous informal proposals. Our domain model puts it in a larger context that includes such concepts as measurement, error, and units.

## 1. Introduction

Standard methodologies (see Fowler 1997, Booch 1994, Halpin 2001) identify various phases in the software development process. The first phase analyses the *universe of discourse* (UoD), that is the world that we are interested in talking about in the context of a particular project (Halpin 2001). The goal is to come to a comprehensive *domain model* containing all the relevant concepts and their interrelations. There are reasons why such a model is an essential ingredient for the development of a Virtual Observatory (VO). It will provide a common grammar and vocabulary for expressing the varied data products existing in distributed astronomical databases. It will define the set of concepts that a user of the VO can use in queries to these archives. Without such a common, "Esperanto" data model it will not be possible to achieve true interoperability between different archives. In different terms, we believe the conceptual model arising from the standard analysis phase to be equivalent to an ontology.

## 2. Modelling Domain

We define the UoD for the VO as "the work that astronomers, astrophysicists and support scientists do and the results they have obtained". Our motivation



© Copyright 2004 Astronomical Society of the Pacific. All rights reserved.

for this choice is that we believe that users of the VO are ultimately interested in the results of the work done by other astronomers. Users are not "just" interested in getting access to images, simulation results or other physical results of astronomical research, stored in some astronomical archive, but will want to know what is actually represented by these results, how they were obtained, what experiments were executed and how. The latter is what we mean by the term "work".

When we say that we believe VO users *will* be interested in the experiments that produced the results, we mean that they *should* be interested in them. One of the main tasks of the VO is to enable other astronomers to do rigorous science with the results and services that are made available through it by their colleagues. It is obvious that results can only be interpreted through knowledge of the process that produced the results – the "provenance". We believe the VO has both the chance and duty to formalize the concepts underlying this provenance by including them explicitly in the modelling effort.

## 3. Modelling Concepts

Here we list some of the core concepts we believe need to be modelled explicitly to give a proper description of the world of astronomy, the UoD of the VO. In the full model (`http://www.g-vo.org/materials/UDM-Poster.pdf`), these have been translated into UML classes and worked out in more detail.

- **experiment**
  By experiment, we mean that work that leads to results that are of interest to users of the VO.
- **result**
  The data resulting from astronomical work that users of the VO are assumed to be interested in.
- **protocol**
  It should be possible to retrieve *what* was actually done, *why* and *how*. This we call the protocol of the experiment. Separating this concept from experiment has an obvious use when a certain protocol is followed in multiple experiments, eg. using an analysis tool such as SExtractor.
- **objective/observable**
  The goal of the experiment, as described in the protocol. The objective of an experiment is to assign values to variables.
- **variable**
  One can assign values to variables by measuring them or calculating them. Thus, the variables are the things which are to be or which have been determined by an experiment.
- **measurement**
  A result consists of values that have been assigned to variables. A measurement is a special kind of "value assignment", namely the common one corresponding to a *numerical* value.
- **value (quantity, category)**
  This is the smallest atomic unit in the result hierarchy. The concept value corresponds to a value assigned to a property. SI calls this a "value of a physical quantity" (NIST 1999).

- **uncertainty/confidence/error**
In the scientific process of assigning values to properties by measurements, errors will be made. These may be statistical, systematic, and possibly correlated. However they are caused, to interpret measurement values it is imperative that they be accompanied by the confidence/uncertainty associated with them.
- **phenomenon**
This concept corresponds to SI's "quantity in the general sense", or to Fowler's PhenomenonType (Fowler 1997). It is the generalization of "a property that a body or substance can have" (NIST 1999). Phenomena are the ultimate things scientists are trying to determine.
- **property**
This is a phenomenon assigned to a particular *thing*, as in "a property of" that thing. For example, *colour* is a phenomenon and *the colour of that galaxy* is a property of that galaxy.
- **subject**
Subject is the concept that unifies the *thing* mentioned above and in the SI definition for "quantity in the general sense". It includes other concepts like body, substance, region, etc. It is really an anchor concept that corresponds to a collection of properties, since a particular *thing* is defined by properties that we can measure, know, or be interested in.
- **unit**
Unit defines the "measuring rod" that is required to interpret numerical values (quantities) assigned to numeric properties. SI defines it as a particular property of a particular, identified subject. All values of the same phenomenon can be expressed by giving the ratio of that property to the property of the identified subject.
- **reference system**
Just as a unit specifies the measuring rod for interpreting numerical values assigned to a property, so the reference system can be thought to specify the zero point. Examples include a coordinate system for positions on the sky, and a magnitude system for expressing fluxes.
- **standard**
To enable interoperability, or more simply, for us to be able to understand each other's (meta-)data, it is not sufficient to provide a model (the grammar of the "Esperanto" discussed above), we also need to agree on a number of standard instances – the "vocabulary".
- **physical artifacts (database, file)**
Data will be stored in some physical datastore which may be a filesystem or a more formal database system. We need to be able to identify and locate these containers.

## 4. Model Detail: Measurement and Quantity

Modelling quantitative values with units and errors has received considerable attention in the IVOA data modelling working group. In Figure 1 we propose a model for this area. The main difference between this and other proposals (see `http://www.ivoa.net/twiki/bin/view/IVOA/IVOADMQuantityWP` for IVOA Quan-

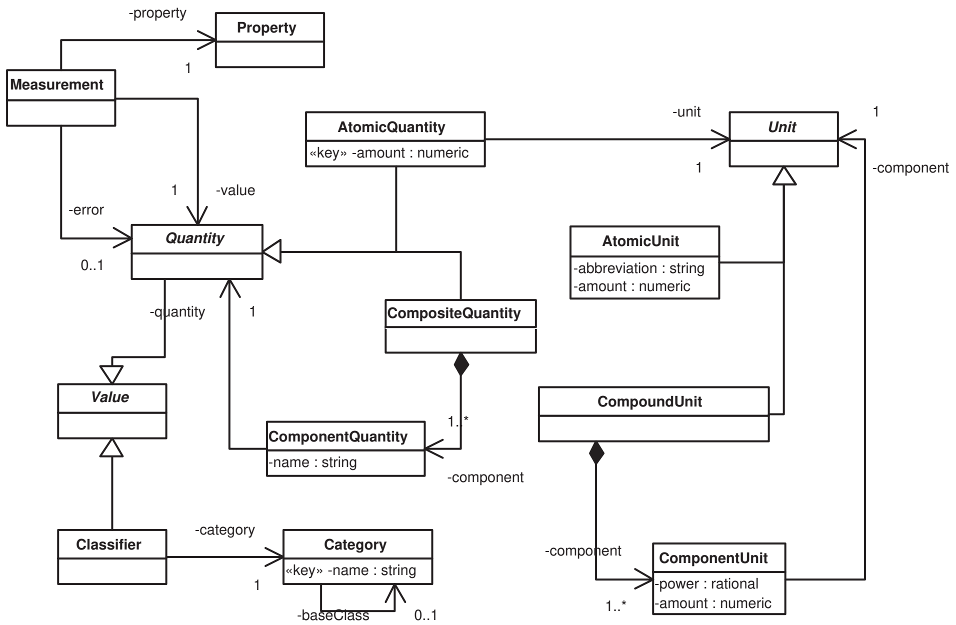


Figure 1. A detail of the domain model, dealing with measurements and values.

tity data modeling resources ) is that we believe that the core concept is the *measurement*, which we define as the act of assigning a value plus error to a property, not the quantity itself. We further generalize the Quantity concept to values and classifiers. To see how this model fits within the large scale framework we refer the reader to the full diagram (`http://www.g-vo.org/materials/UDM-Poster.pdf`).

**Acknowledgments.** This work was undertaken as part of a Canadian Virtual Observatory (`http://services.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/cvo/`, CVO) and German Astrophysical Virtual Observatory (`http://www.g-vo.org`, GAVO) collaborative research project. CVO is sponsored by the National Research Council (NRC) and the Canadian Space Agency (CSA). GAVO is sponsored by the German Federal Ministry for Education and Research (BMBF).

## References

Grady Booch 1994, Object-oriented Analysis and Design, 2nd edition, Addison-Wesley

Martin Fowler 1997, Analysis Patterns, Addison-Wesley

NIST 1999, SI Specification,
`http://physics.nist.gov/cuu/Units/introduction.html`

Terry Halpin 2001, Information Modelling and Relational Databases, Morgan Kaufmann