

Proposal for a Quantity-based Data Model in the Virtual Observatory

Brian Thomas and Edward Shaya¹

Code 630.1, Goddard Space Flight Center/NASA, Greenbelt, MD 20771

Abstract. We propose the beginnings of a data model for the Virtual Observatory (VO) built up from simple “quantity” objects. In this paper we present how an object-oriented, domain (or namespace)-scoped simple quantity may be used to describe astronomical data. Our model is designed around the requirements that it be searchable and serve as a transport mechanism for all types of VO data and meta-data. In this paper we describe this model in terms of an OWL ontology and UML diagrams. An XML schema is available online.

1. Introduction: an object- and domain-oriented approach

The VO community is currently attempting to formulate a data model which might be shared across all data repositories and used to facilitate the query, exchange and fusion of astronomical data. This fundamental requirement must also be coupled with the need for this data model to be able to replicate the structure and content of the data at any participating VO data repository. One possible avenue is to utilize an object-oriented methodology to reuse important concepts and perhaps allow for a mechanism whereby a computer may decompose advanced child concepts into more digestible parent concept parts. As Plante (2002) has argued this kind of approach, if done within a single domain, will result in an increasing problem both for initial development and long-term maintenance of the VO data model meta-data. If we however sever the connection of the data model to the astronomical meta-data we will allow for much greater progress on the problem of developing the data model and its maintenance.

What is needed by the VO community is a new data model which contains the ability to be extended in an object-oriented fashion, has domains which may allow for separate communities within the VO to apply their expertise without stepping on other areas of work, and has the ability to be used as an exchange format across the Internet.

Recently Plante (2003) has argued for a quantity-based data model, where a quantity is used to model all of the data interactions. We have taken this idea to develop our proposal paper based on our own work with XDF¹ and guided

¹Astronomy Department, University of Maryland

¹<http://xml.gsfc.nasa.gov/XDF>

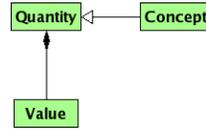


Figure 1. UML diagram of relationship between a concept, quantity and its value

by the work of many others (refs above plus: McDowell 2002ab; Williams 2002; UCD²) and larger astronomy communities (FITS³) In this paper we propose the beginnings of such a data model and describe how adopting an object-, domain-, and quantity-oriented approach may serve to promote a sharable, searchable standard amongst the various data repositories of the VO.

2. Theoretical basis : definition of quantities

We start by introducing the "Quantity" as the object that associates a value with a concept. The concept may be any meaningful term or idea which the VO community wishes to use (for example "X-ray star", "visual flux", "CCD camera type", "index", etc.) and the value indicates the amount or degree of the concept. The value can be stated as a number, a word ("full", "partial", "high", etc.), or a symbol.

To serve as a general framework for manipulating concepts and their values we create an entity or object which may serve as the package for all of these tuples. This entity, the "quantity" is inherited (passes on all of its properties) to all concepts (figure 1) and we thus tie the meaning to the class of the object.

The value is itself a one-dimensional array, which may have one or more "datum", e.g. the value V may be defined:

$$V = d_1, d_2, d_3, \dots d_n \quad (1)$$

where n is the number of datum in V and each d_i are the individual datum that may be either scalars, vectors or other quantities (tuples). In the last case the child quantity serves as data but we may also use child quantities to specify meta-data. To do so we insert the child quantity directly within the parent quantity with the designated relationship "metaData" (figure 2).

As the datum in V must describe scientific information (at least most of the time) each datum must have associated with it some description of its accuracy (errors) and scientific units. Machine understandability requires that each datum should also be described by some type of a data format. In our model, we consider that all datum in V for a given Q are "homogeneous" meaning that every datum of a given Q has the same units and data format. Accuracy of scientific data can, and often does, vary on a datum by datum basis. Thus, we

²<http://vizier.u-strasbg.fr/doc/UCD.htm>

³<http://fits.gsfc.nasa.gov>



Figure 2. Quantities use other quantities to capture meta-data

infer the existence of an array of accuracy values which is of the same size as the array of V to which it refers.

Further description of the basis of our model, including how the model may hold vectors and arbitrary tuples (tables) may be found in Thomas & Shaya (2003).

3. Practical application - development of domain-based ontologies

In practical application our data model must be realizable in software, thus we have described our model using the OWL⁴ ontology language (figure 3). We choose OWL because it may be used to derive both the XML schema and UML diagrams (see Thomas & Shaya 2003) which will serve as the basis for generating both code and XML instance documents. Additionally, OWL, and ontologies in general, can be used to formally define the relationships between classes as well as their instances. Of course, placed on this footing we may now talk of quantities as classes in the object-oriented sense, which may be "extended" to create concepts that will be queried for and traded between participants in the VO. As a starting point, some important concepts that the community might choose to extend the quantities into in the VO domain include the space-time schema of Rots (2003) and the UCD descriptors.

The development of ontologies that show the relationship between all of the concepts in the VO will be an important task and we believe that the quantity-based data model can form the basis for all the other VO ontologies. We need not, nor is it desirable, create a single all-encompassing ontology. It is better to have a spare, shared group ontology, working in tandem with richer, but localized ontologies. Mapping between localized ontologies can serve to bridge differences in defined concepts. Under this formulation then, each ontology should belong to its own namespace or domain.

4. Discussion

With this model, we have managed to meet all of our set requirements and design goals. The model is sparse, with few classes, yet powerful enough to encompass many types of data that will exist in the VO. By casting the model in terms of an ontology, we may use it to support sophisticated concept-based searches. Furthermore, we have managed to cast this model in terms of an XML representation, which is important for transport of information across the Internet as well as used in designing catalogs for searching holdings at data repositories. Additional issues are discussed in Thomas & Shaya (2003).

⁴<http://www.w3.org/TR/owl-features>

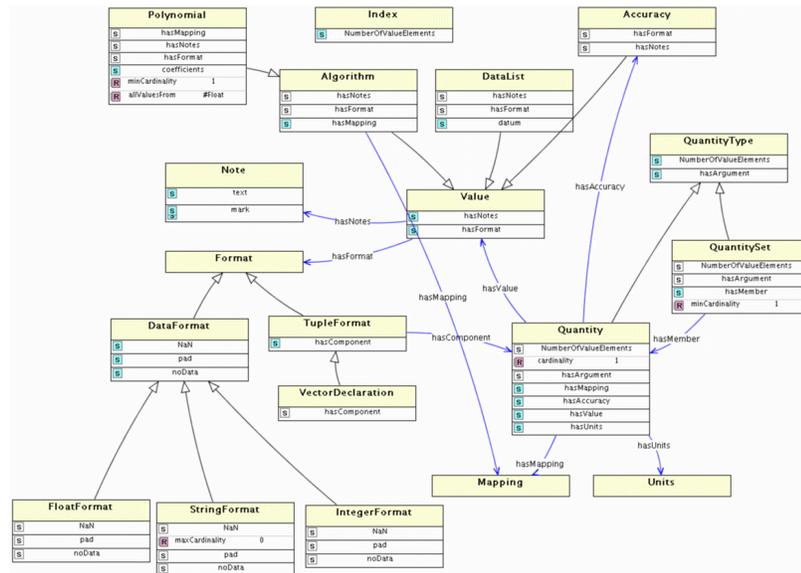


Figure 3. OWL diagram of quantity

5. Summary

We have produced a rudimentary data model based on theoretical considerations and mapped them into an ontological space. In addition, we have given thought to how this data model may be utilized in terms of software, and how the community would approach designing concepts for search and interchange. Software resources related to this work, including an XML schema, an XML example that wraps a FITS file and simple Java code may be found at the following URL: <http://nvo.gsfc.nasa.gov/QuantityDataModel>

Acknowledgment: Support for this work was provided in part by NSF through Cooperative Agreement AST0122449 to the Johns Hopkins University.

References

- McDowell, J., et al. 2002a, “Data Models for the VO: overview”, <http://bill.cacr.caltech.edu/cfdocs/usvo-pubs/files/vodm003.ps>
- McDowell, J., et al. 2002b, “Data Models for the VO: Metadata objects for the VO”, <http://bill.cacr.caltech.edu/cfdocs/usvo-pubs/files/vodm003.ps>
- Plante, R. 2002, “A Scalable Metadata Framework for the Virtual Observatory”, <http://bill.cacr.caltech.edu/cfdocs/usvo-pubs/files/fw-draft2.pdf>
- Plante, R. 2003, NVO Cambridge workshop presentation.
- Rots, A. 2003, “Space-Time Coordinate Specification for VO Metadata”
- Thomas, B. and Shaya, E. 2003, astro-ph/0312604
- Williams, R., et al. 2002, “VOTable: A Proposed XML Format for Astronomical Tables”