Merger of STC, Quantity and Characterization packages into a query-able VO catalog.

Brian Thomas (thomas@astro.umd.edu) and
Ed Shaya (edward.j.shaya.1@gsfc.nasa.gov)
Draft: April 1, 2005

Abstract

This document quickly describes a recent attempt we have made to bring together the serializations of the IVOA STC (Space-Time Coordinates), Quantity and Characterization data models into line with each other. As a test of this merger effort, and interesting work in its own right, we have tried to translate some existing catalog data into a data model which utilizes all of these new, now interacting, data models.

Merger of Existing Models

Currently the STC, Quantity and Characterization data models are separate, unrelated models (figure 1). The STC model contains 2 sub-components which include the "Coords" (coordinates) and "Region" models. Merger of these models would serve to provide a uniform structure throughout the schema (for example, utilization of Q within stc, characterization models provides uniform structure for location of errors, values, ucds, and meta-data locations) and reduces replication of effort. A further issue lies with the super-large STC datamodel family which incorporates model elements which really are outside of its pervue (in our opinion), namely definitions of units and accuracy (errors).

So, to be clear a reasonable overview of what has changed in the "merger" is the following:

1. Any part of the Characterization or STC data models which hold scientific data now utilize the Quantity data model.
2. The STC data model was split further into 5 parts: "stc" (the main coordinate system model), "coords", "region", "units" and "accuracy". Accuracy elements are largely inheriting from Quantity datamodel, as all errors are quantities.

3. Drop replicated Quantity machinery from stc, coords such as meta-data treatment.
4. Implementation of Characterization data model within STC and region
5. Quantity utilization of new units and STC packages instead of older place-holders.

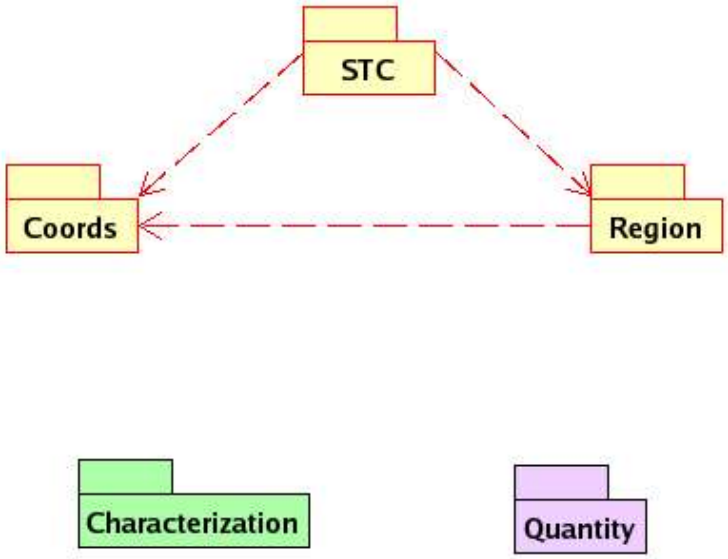

Figure 1. Existing VO Packages. Color indicates "family" of each data-model. The STC family contains 3 component packages: "STC" (main), "Coords" and "Region". STC datamodel accesses Coords and Region packages. Region package needs to re-use only elements of the Coords package.

An overview of the new family of packages, and how they access each other is shown in figure 2.

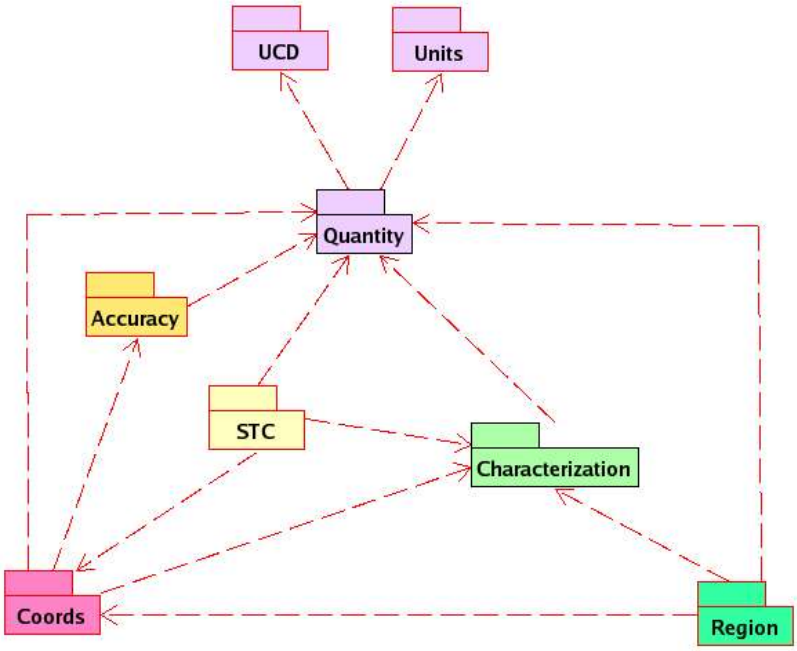

Figure 2. New package associations. A rather more complex set of associations is derived if re-utilization of data model components is a desired goal. Sub-component packages of the STC are differently colored to indicate their essentially separate nature from the old parent family.

VO Catalog Test case

From discussions with other members of the VO catalog working group it appears clear that the focus is on attempting to model 'source' catalogs (essentially 'source' is equivalent to any astronomical object such as a flare, star, galaxy, nebula, HII region, etc). As a first step, this means the catalog model is fairly simple, e.g. a parent class that is a *optimized collection of one or more homogeneous objects* and has some meta-data about the catalog itself. The emphasized 'optimized collection ..' part above means that we seek a serialization which compresses the information about our source objects in a non-lossy way, and in order to keep things simple, don't try to have collections of heterogeneous objects (for the time being).

Our analysis of the old ADC catalog archive holdings (http:adc.astro.umd.edu, largely identical to catalog holdings at CDS http://cdsweb.u-strasbg.fr) shows that a reasonable catalog model for sources is to allow the catalog to hold 'catalog objects' which inherit from the Quantity data model. Moreover, to indicate the types of meta-data structures that

a catalog might have we have also included information about history (Provenance) and any related references (Publications). There are ongoing attempts to define these later data models (especially within the Observation data model working group) but here we have elected to simply create some minimal place-holder data models until these more mature models may be merged in.

This simple catalog model is shown in figure 3. An expanded view of how the catalog data model package relates to the other VO data model packages is shown in figure 4.

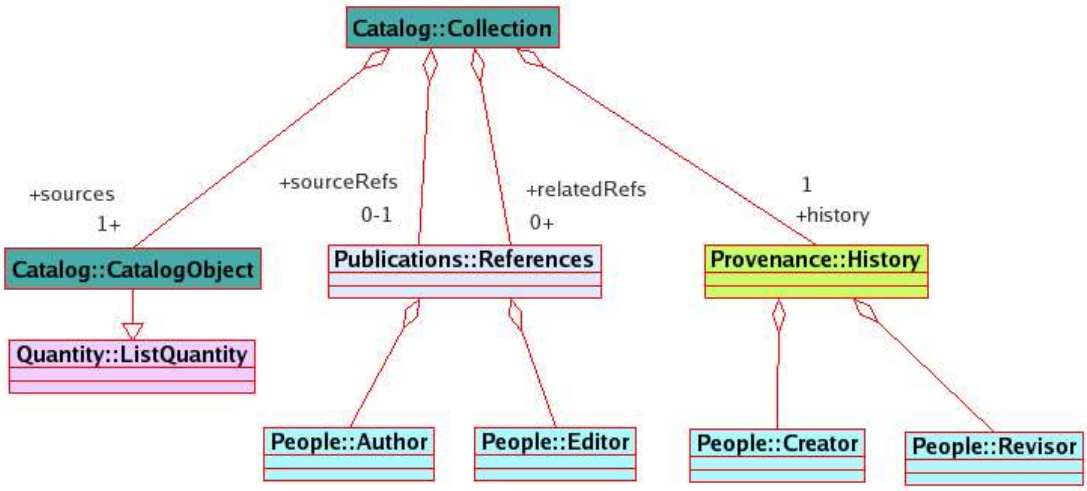

Figure 3. Class diagram of selected classes illustrating structure within the Catalog Collection.

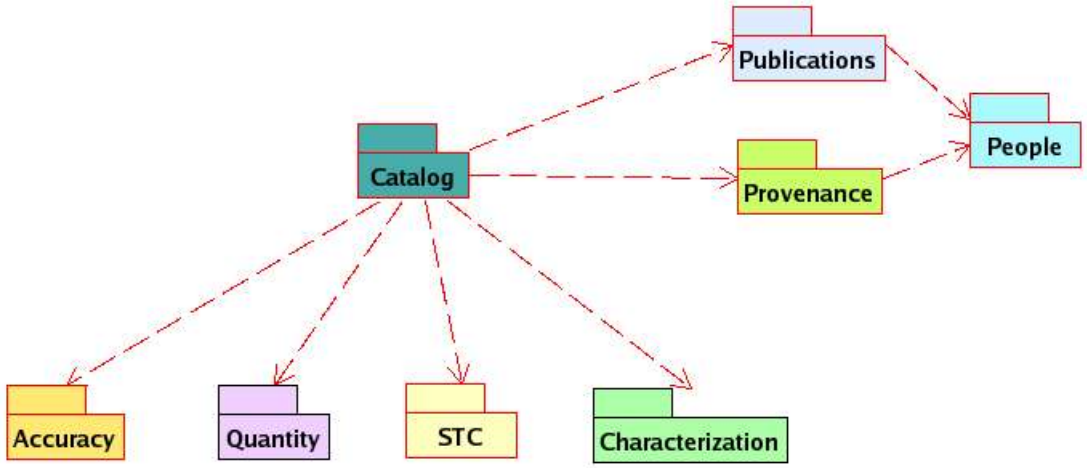

Figure 4. Showing the main associations of the Catalog package to other VO packages.

Some associations between packages suppressed for clarity within this diagram.

Serializations and Examples

As the VO is a networked information system, XML makes an excellent choice for a serialization of these data models. We have created the requisite schema, as well as a test serialization of the espenak catalog (ADC/CDS catalog 1005), and made them available at:

http://archive.astro.umd.edu/VOCatalogTest

Present Status of work

This is, of course, a work in progress and is presented as a possible future structure for component data models within the VO as well as a Catalog data model. We are seeking to collaborate and further refine the these package as well as the associations between the packages. Furthermore, we are working to integrate these schema into use with a Java based package, based on our earlier work with QML (Quantity Modeling Language) and QDM (Quantity Data Model), both which are available at

http://www.data-model.net