Using VO Standards in Science Applications.

Abstract

This document describes how the various current and evolving VO standards have been used in VO applications and the concerns and issues that have been raised with regards to VO standards during software development.

This document reflects the experiences of the members of the IVOA Applications Interest Group and is largely based upon the demonstrations and conversation during the AIG session during the Cambridge Interoperability session (5/25) and subsequent discussion in the groups forum.

Overview of Standards and Released Applications

Much of the early work of the Virtual Observatory has been the development of standard formats and protocols through which VO applications communicate. This note discusses how the applications that have currently been built or are being built have been able to use these standards. The applications considered here are primarily those that a VO end-user, e.g., an astronomer or perhaps a member of the public would interact with. The views expressed here try to reflect those of the developers of such applications.

There are several VO standards that have been developed to the level of being used in VO applications. Since many of the standards discussed here arose for the national efforts or preceded the establishment of formal IVOA procedures for standards, few of the VO standards discussed here has gone through the formal IVOA process. Standards currently being used in applications include

- *VOTable* is a standard XML format for tabular information. Version 1.0 of VOTable was described in a document published after considerable discussion in the IVOA. Version 1.1 is currently being discussed in the formal IVOA process and should be a recommendation soon.

- *Uniform Column/Content Descriptors* provide a standard vocabulary for describing the content of data, particularly the columns of VOTables. Version 1.0 reflects the development of UCDs prior to the VO. A version 1.1 is being developed within the formal IVOA process.

- The *Simple Image Access protocol* was proposed by the NVO team and SIAP servers have been developed by many IVOA members. Version 1.0 is defined by an NVO proposal and version 1.1 is currently being within the VO framework.

- The *SkyNode* protocol has been going through the standard process in the VOQL group and basically provides mechanisms for distributed queries with special handling of spatial joins. The SkyNode protocol includes a standard for specifying a query, *VOQL*, which we do not list separately since as far as I know, there is no use of VOQL outside of SkyNodes (but this may change in the future). SkyNode has been developed within the IVOA standards framework and is now at version 0.84.

- While the *Cone Search* protocol does not seem to be considered a formal VO standard it has been implemented by many data providers. It is unclear if there will be any successor protocol or if the functionality will be subsumed into VOQL or SkyNode.

- The *Resource Metadata* standard describes the metadata that should be associated with a given VO resource. It provides some standardization of the vocabulary to be used and organization of the metadata. The instantiation of the resource metadata is less clearly defined. The resource metadata standard has gone through the formal IVOA process.

Some additional protocols have are under development and are influencing software development even though the definition of the standards is still in considerable flux:

- The *VOService* standard has begun to define the characteristics and capabilities that an atomic VO capability should have.

- A number of *registry* services have begun to operate and considerable work is underway to provide a standard way to access the registry. Standards for the coordination of the registry contents are also being developed but these do not directly affect registry client software.

Some of the areas where VO efforts have not yet had substantial impact on released applications include data models and grid applications. Some grid applications have been built in the NVO and Astrogrid but have not yet been released for public use.

Table 1 describes the tools demonstrated in the May 2004 interoperability meetings and suggests the range of applications that are now available to the scientific community. These tools demonstrate that using just the existing standards quite powerful capabilities can be presented to the community. These standards provide substantial leverage for accessing observational catalogs and archives. Users can find information from distributed resources and view and edit them in a common environment. Most of the query protocols are strongly reliant on positional queries, but there is some support for searches on other criteria. Table 2 (in the attached Excel spreadsheet) gives a rough guide to the use of VO standards within services.

There are few detailed analysis capabilities (spectral fitting, background subtraction, …) that use VO tools and no non-prototype tools for accessing theory data.

Service	Description
Data Inventory Service	Data discovery tool that enables users to find relevant datasets and send information to appropriate analysis tools
VOPlot	A standalone program or integrated package for plotting data in VOTable format
AVO demonstration	A set of capabilities linked to the Aladin tool that enabled the extraction and discovery of high-redshift, obscured QSOs.
TOPCAT	An editing and display program that provides capabilities for handling tables in a variety of formats including VOTables.
Aladin	A user tool to query, display and analyze spatial information from both catalogs and images
Common Execution Environment	An environment for controlling the execution of s series of related VO tasks.
GAVO cross-matcher	A Web tool for merging catalog data and classifying the resulting objects. The GAVO team is also working on other VO tools.
SkyQuery	A tool for query distributed databases using efficient spatial searches.
VO Spectrum and Filter service	A database that users can search for spectra meeting various criteria. Users can also publish new spectra of filter curves.

Table 1. Some VO Applications

Issues and Concerns

In using VO standards, developers have raised a number of concerns that may need to be considered by the VO community.

Quality assurance and reference implementations

One issue that affects the usability of data that is nominally conformant to an existing standard is the quality of the implementation. VOTables may be subtly incorrect, UCDs may be used in ways that were not intended, or SIA services may be announced without mandatory parts of the protocol being supported. While it is up to the providers of these data or services to ensure correctness, it is beneficial when there are test suites that can be used. In some cases other VO services are testbeds: can they use the results generated. This is fine when the discovery is made by the developer of the new dataset or

application, but when users discover the problems it can weaken their interest in VO tools in general.

Reference implementations of VO tools and protocols may be essential to their adoptions by the community. While the first VO service protocols were relatively simple, new protocols and formats are increasingly complex. The proposed Simple Spectral Access Protocol has substantially options than the Simple Image Access Protocol. The SkyNode protocol is orders of magnitude more complex the cone search. The grouping structures of VOTable 1.1 add substantial complexity to the definition of VOTables. This trend is unlikely to abate. However the software resources available in the community to incorporate VO capabilities are unlikely to increase substantially. To enable a broad user community of data providers and users to access the easily adaptable implementations of VO protocols are needed. The Applications Interest Group can play a role in disseminating information about such 'reference' implementations.

We do not recommend 'standard' implementations. Users should always be free to build their own implementations of any protocol so long as it is conformant with the underlying descriptions. However 'reference' implementations will likely be used by many groups who do not have the resources to independently develop resources. The goal of reference implementations should be to provide clear and adaptable implementations of the formats or protocols.

The two issues of quality assurance and reference implementations are not independent. Comparison of one's own results with a response from a presumed correct implementation is a standard way to test software.

New areas for standardization

There may be a few areas where additional standards are needed beyond the areas where there is currently activity. One important area is the characterization of applications in some standardized way. While the VOService protocol addresses this at the very highest level, additional approaches are needed if the VO is ever to address how it can incorporate the very large legacy of existing software applications. This might start with looking at a few commonly used standard tools that are going to be used and reused throughout the VO, e.g., name resolver and other conversion utilities.

The relevance of many of the existing protocols for theory applications is unclear. The strong reliance on positional selections is likely not useful for theory. The theory interest group may wish to suggest areas of standardization that help the integration of their datasets within the VO. It is likely the case that treating theory as a single area will be far to simple. The development of format standards (e.g., VOTable and UCDs) has been more immediately useful within the theory community.

Prototypes and standards

A consensus of the AIG is that the prototyping of standards within 'real' applications is essential from the earliest possible stages of standards development. Not only do such prototypes help in the working groups in which the standard is being defined, it makes it much clearer to those outside the working group what the scope and impact of the standard is likely to be.

There was considerable discussion of the role of the AIG with respect to use case scenarios. It was recognized that these primarily arise from the science working groups, but may be augmented by suggestions from the AIG. Regardless of their origin, the AIG should serve to review use cases with regard to how well they map to the VO standards and especially to understand the gaps that may show us where we need to evolve or extend our standards.

The need to ensure maximal portability of software was reiterated, not so much for copying software from one site to another as to manage it over time.

While this AIG does not have specific deliverables to the VO, the following major activities are scheduled over the next few months. These are in addition to the standard announcement and assistance activities of this group.

Evaluation of use cases

Much of the development of the VO standards has flowed from use case scenarios generated by the VO science teams. These use cases are analyzed to guide the efforts of the various standards working groups. There the use cases are broken down and the working groups assess how the standards within their purview might help address the use case. Since no working group has responsibility for the use case as a whole, and since the science working group is primarily concerned with the science outcomes rather than the details of how the a service might be implemented, there seems to be no overarching analysis how the use cases might expose gaps or possible conflicts in the developing VO infrastructure.

The AIG may be an appropriate venue for such an analysis although precisely how formalized such a process should or could be is unclear.