

Evaluation of Interoperability Tools

AVO Work Area 2 Report, 2004
WP 2.4

Mark G. Allen, Centre de Données astronomiques, Strasbourg (allen@astro.u-strasbg.fr)

1. Introduction

This report evaluates the interoperability of the tools and services used for the 2003 and 2004 science case demonstrations of the Astrophysical Virtual Observatory ([AVO](#)) project.

Interoperability broadly refers to the goal of having software tools, on-line services and astronomical data all working together in a seamless, readily compatible manner.

Clearly there are different levels of interoperability. Take the example of a table of Astronomical data, at a very basic level, tools and services may work with tables from different sources by simply using a common table format for inputs and outputs. To identify columns of a tables containing similar information requires a higher level of interoperability where the column metadata needs to be interpreted. Further, combining columns in a scientifically meaningful way, for example merging tables of photometric measurements to construct Spectral Energy Distributions, will impose many more requirements to achieve 'Scientific Interoperability'.

Scientific interoperability is a high goal for automatic VO systems because often the final steps of combining scientific data will require customized procedures which only the scientist can perform. While current developments in describing data with "Data Models", and Ontologies for expressing knowledge and relationships, are addressing some of these issues, it is important that VO systems continuously provide the framework in which the scientist can do science.

The AVO has done this by using a science driven approach. AVO has developed prototype tools for use in real science cases at annual scientific demonstrations. The exacting nature of scientific investigation has pushed the interoperability requirements to a high level, and has allowed us to identify many aspects of interoperability that need to be addressed in order to build a VO.

This report includes evaluations of all the VO tools used in the science demonstrations, with an emphasis on interoperability issues. We make a distinction between the interoperability and functionality of tools. Interoperability being the way tools communicate and pass information, and functionality being the set of tasks the tool can perform. Many additional functionalities could be implemented in the tools, but as prototypes the developments have been focused on the demo science cases rather than enabling fully generalized functions.

A previous report [Doing Science with VO tools? \(Mann & Allen 2003\)](#) highlighted aspects of VO interoperability by applying existing VO tools to a specific scientific problem. Many of the minor bugs encountered in that exercise have since been fixed as the tools have matured from the first prototypes, and also a number of the desired VO functionalities described in that report have been realized in the tool-sets.

2. Tools and Services Used

The AVO demonstrations utilized a wide range of tools and services developed by the AVO team, and by external partners and parties. Extensive use was made of existing tools, in particular those from the CDS, with the main prototype software component based on the CDS Aladin suite. Below we list all the tools and services, with a brief description of their usage in the science demonstrations, the standards/protocols used and any other interoperability aspects.

2.1 CDS 'Aladin for AVO'

The CDS Aladin interface was used as the main component of the AVO prototype. This is because it already contained many of the desired features for a VO portal including the ability to access and visualize data and catalogs from remote servers. The features of Aladin that were of particular importance for the science demonstrations included;

- Access to the Vizier catalogue server
- Selection of catalogs (or sky-region bounded subsets)
- Access to image servers for browsing and selecting image data
- Overlaying multiple catalogues on multiple images
- Filtering of catalogs
- Manipulation of catalogue data (computing new columns etc.)
- Cross-matching catalogues
- Launching other applications
- Talk-back between Aladin visualization and external applications

Aladin provides interoperability by integrating access to multiple services via a single interface. This is done via the GLU registry of services, which (among other functions) provides a means for listing the parameters and calling sequence for each service in a standardized way. This is similar to the ideas developed for IVOA registries, here implemented for services callable via URL.

Basic interoperability of images and catalogues in Aladin is enabled by use of FITS format for images, and use of a set of table formats for catalogues including VOTable, and a basic ascii format. Essentially Aladin has evolved to be compatible with the information formats provided by the services.

In addition to the GLU system, other interoperability enhancements were made to the Aladin for AVO prototype. These include a mechanism for linking external applications into Aladin (ExtApp), a set of data access mechanisms, and a hierarchical 'data tree' visual representation for image and spectra metadata.

Aladin ExtApp

In order to work together with other VO tools such as VOPlot a method for linking an external application was implemented. In the case of VOPlot, catalogs (or selected points from a catalog) loaded into Aladin may be directly plotted in VOPlot via exchange of a VOTable representation of the catalog points. Also, a callback mechanism was put in place to

allow VOPlot to then control the display of the points in Aladin, for example, to show only subsets of points filtered by VOPlot.

Aladin Data Access Methods

Aladin has been updated for compatibility with the Simple Image Access (SIA) and Simple Spectral Access (SSA) protocols. SSA servers for ISO and XMM data were explicitly linked into Aladin, making it easy to query these services for a named object, or coordinate location (region) on the sky. The SSA servers are easily made available via Aladin by inserting them into the GLU registry. Similarly, spectra available via the Simple Spectral Access protocol can be accessed by Aladin, an example of this is the set of VLT spectra of CDFS objects used in the science demonstrations.

Hierarchical Data Tree

A Data Tree function was developed to help handle and visualize data accessed via SIA, SSA, and indeed any image available via a URL. This function allows the data description to be represented in a hierarchical fashion, and greatly facilitates browsing and selection of the relevant data. For example, it uses the astrometry metadata to automatically generate the field of view outlines of the images overlaid on the sky. For spectra where position angle metadata is available, the non-dispersion direction of the slit can be projected on the image.

2.2 CDS Services: Vizier, Aladin Image Server, and SIMBAD

The CDS Services: The Aladin image server, Vizier, and Simbad are the major services which provide the catalog and image data used by AVO demonstrations.

The Vizier catalogue server is a very sophisticated service that provides a number of access mechanisms for selecting catalogs, or subsets of catalogs defined by constraints on the columns. All the catalogues are described by metadata in the [Readme standard](#) devised by the CDS. Vizier also has a interactive web form interface that allows selection of catalogs by many different approaches including references, author name, wavelength regime, and subject keywords. With its multiple access mechanisms, and output possibilities in all formats, Vizier is a leading example of an interoperable service.

The Aladin image server holds large collections of publicly available images. It includes various DSS version, 2MASS and SLOAN. For the AVO demo, smaller data sets were included to support the science cases. These include the GOODS version 1.0 ACS imaging data and the ESO Imaging Survey (EIS) data. data included in this image server automatically becomes interoperable with other imaging data available via Aladin. The image cutout server function of the Aladin image server makes this service extremely useful for obtaining manageable size imeg cutouts from large datasets. Including more and more small data sets in this server is not scalable because of the maintenance involved. We foresee that a deployable image server of this kind would greatly enhance access to many more datasets from providers who wish to participate in the VO.

The Vizier interface within Aladin provides access to the Vizier service through the GLU registry. Vizier has also implemented access via the IVOA registry protocols, so that IVOA registries can obtain the metadata of tables in Vizier. This represents a significant input into the requirements for developing these registries.

2.3 VOPlot

VOPlot is a plotting facility for VOTables developed by the VO-India project. In collaboration with VO-India, VOPlot has been included as plug-in to the Aladin tool. This has been accomplished with the Extapp mechanism, and includes a number of callback methods to enhance the interactive connection between the two tools. The benefit gained from being directly connected is simply ease of use, allowing catalogue data to be used immediately in VOPlot without going through the loop of saving local versions of the catalogues. While originally thought of as a quick-look tool, VOPlot is also rapidly developing into a sophisticated plotter.

2.4 Topcat

Topcat is a table spreadsheet application developed by the Starlink project. Topcat is the visual interface to the powerful STIL libraries for manipulating tables of all formats, including ALL flavors of the VOTable format. While not linked into the Aladin software (it could be done via ExtApp) this tool was used extensively in the preparation of the science demonstrations because of its flexible handling of VOTable data, and its capabilities for converting between different formats (in particular FITS, ascii and VOTable as used here).

2.5 Specview

Specview is a tool for viewing and analysing spectra, developed by STScI. It has been linked to the aladin tool via the ExtApp method. It allows Aladin to provide the locations of spectra via the SSA protocol, and have them automatically loaded into Specview. Specview has a set of native formats, including various FITS and ascii formats. Greater compatibility with VO formats (VOTable) is being sought for the Jan 2005 demo.

2.6 VOSpec

VOSpec is another spectra tool developed by P. Osuna (ESA). This tool is designed for VO purposes from the beginning and has simple access method for displaying spectra available via SSA. This tool has tackled the issue of characterizing the units used on the spectra flux and wavelength scales. The dimensional equation and scale equation method proposed is proving to be very simple and practical. Further enhancements are being implemented for Jan 2005, and this tool will probably lead the developments in terms of interoperability of VO compatible spectra.

2.7 SExtractor

A service for running SExtractor remotely was developed for the 2003 demo. The service was linked to Aladin via an early version of the ExtApp method. This service highlighted the VO concept of remote computing, whereby the an image location plus a set of SExtractor parameters were sent to the remote service, and the resulting VOTable catalog of measured sources was returned to the Aladin tool. This first implementation was fragile, and was limited in terms of using different SExtractor modes.

3. Interoperability issues

3.1 Scientific interoperability

The tools developed and used in the science demonstrations have provided a framework in which an astronomer can integrate diverse information from various distributed sources with a fairly high level of technical interoperability. To bring this information together scientifically, requires significant input from the astronomer, and understanding of how the system works. This is because while the system provides a framework for expressing units and semantic description of values (UCDs), the system does not yet 'know' how to use this information together. So, unit conversions, conversions between photometric systems, and other calculations such as converting redshifts to distances must all be input by the astronomer. Complicated and customized processes will probably always need to be done with an astronomer's expert knowledge. However, basic unit conversion, and astronomy awareness within 'the VO' will significantly enhance the system, and astronomers generally expect that the VO systems will address this. To enable scientific interoperability requires that there are standard ways of expressing the photometric systems, and units, and the relationships between different data components. This falls within the mandate of the Data Modeling effort being done by IVOA. The development of tools which can handle these higher level astronomy functions are desperately waiting for the maturation of data model descriptions. Specific scientific interoperability issues encountered during the AVO science demonstrations are listed below.

Lack of photometry metadata standards

A large amount of catalog data with magnitude measurements do not have clear links to information describing the photometric systems. This makes conversion of these values to accurate fluxes tedious and sometimes impossible. While many subtleties exist in performing these transformations, a system for photometric metadata, and basic methods for handling it would be of enormous benefit for making photometric data from different telescopes interoperable.

PSF information missing

In order to scientifically combine images, for example to put a set of multi-wavelength data onto a common pixel grid for synthetic aperture photometry, requires good knowledge of point spread function. When dealing with a large set of diverse data it is important that the PSF be described in the image metadata.

Loss of metadata in non-VO tools

Much of the benefit of VO formats such as VOTable, is the direct inclusion of metadata. This metadata can however be lost when the data is exported to a non-VO tool, as necessary for various processing steps in our science demonstration scenarios. In particular the tables generated as part of the Obscured Type II AGN case needed to be processed in ways not (yet) possible in VO tools. This causes the metadata to be stripped of in order to input ascii (or other) formats. Re-attaching the metadata afterward, and creating the new metadata for added parts is rather tedious.

Extending the standards

The AVO prototype tools were largely built on existing software which already used conventions for doing some VO like things. Also, developing the prototype pushed the standards beyond their definitions in order to make things work. Both of these mean that the prototype has requirements on data and services beyond the IVOA standards. This situation makes us well placed to input into the IVOA process, but it also causes various inconsistencies between tools.

For example, the RA and DEC columns computed by Vizier (on all Vizier tables) are hidden in the Aladin tool, and it is sometimes not clear which coordinates are being used in the overlay, or in the coordinates automatically passed to tools like VOPlot. Also, the Aladin image server goes beyond the SIA standard because of the need of a hierarchical description of data which is not (yet) possible in SIA.

3.2 Interoperability Limitations of Current tools

Here we briefly list some of the interoperability limitations of the current tools.

Cross matching of tables has been implanted in both a local process, and a remote calculation manner. These are limited in the sizes of tables they can handle because of memory limitations - on the local machine, and on the VOTable parser of the remote service.

The CDS services, Vizier and SIMBAD are naturally very interoperable. These services were designed to work via dedicated powerful interfaces, allowing for local saving of results. The heavy use of these services in the science demonstrations has pushed the requirements for interoperability between these services via the Aladin /AVO prototype. To address this, more and more of the functionalities of these services have been implemented in the Vizier and SIMBAD interfaces within the prototype. This brings up an important issue for interoperability between tools. It seems to be very cumbersome to keep pushing interfaces of separate services into a single interface, and it would be preferable to define less intertwined interoperability interfaces that allows launching of individual services that maintains their original identity.

A similar issue occurs with the VO tools called by the AVO prototype. For VO tools such as VOPlot, Specview, VOSpec and the X-match tool, we adopted a plug-in style connections. This was a convenient way to implement basic interoperability between these components. This however will not scale to a large (greater than say 20) number of plugins due to the maintenance of these plugins. Again, this suggests we need a cleaner way of connecting interoperable tools.

Both these issues really relate to the need for a VO portal. This central component is often described as the main launching point, and control panel for performing tasks with VO tools. The AVO prototype allows this kind of thing, but since it was not designed with the full VO scope in mind, it will not scale to a full VO portal. The lessons learned from the prototype will be extremely valuable when designing VO portals. It is clear that this will be a challenging development, and with the current fast growth of VO tools and services, there is an immediate need to start designing VO portals.

3.3. Interoperability with AstroGrid infrastructure

The upcoming January 2005 final AVO demo aims to show a high level of interoperability between existing VO tools and the AstroGrid infrastructure. This infrastructure is designed to facilitate distributed computing, using remote compute servers controlled by a portal. These workflows, are to link together various VO services and store intermediate results in a distributed storage system called MySpace. These functions will rely on interoperability standards for accessing registries, for querying services and databases, and for directing results to various VO tools. Given the rapid development of these capabilities for the Jan 2005 demo, discussion of the interoperability issues is deferred until after the demo.

4. Summary

The fact that publication quality science is being done with VO tools shows that they have already made a significant first step toward the goal of an operational Virtual Observatory. The 'first science' for the AVO was enabled by prototype tools that provide access to catalog and data archive servers, and capabilities for smart browsing and selection of data. The tools allow flexible use of catalogs, image and spectral data, plus a high level of connectivity between VO tools for cross-matching, plotting and making measurements on data. Moreover, the current prototype tools have been very good for demonstrating the potential of doing science in the "VO way"

Moving beyond the demonstration level will however require significant development in the way that VO tools are connected together, and we need the VO portal.