Astronomy and Computing 11 (2015) 190–209

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

# The Virtual Astronomical Observatory: Re-engineering access to astronomical data

R.J. Hanisch [a,b,1], G.B. Berriman [c], T.J.W. Lazio [d,*], S. Emery Bunn [e], J. Evans [f], T.A. McGlynn [g], R. Plante [h]

[a] *Virtual Astronomical Observatory, 1600 14th Street NW, Suite 730, Washington, DC 20036, USA*
[b] *Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[c] *Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA*
[d] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA*
[e] *Center for Advanced Computing Research, California Institute of Technology, Pasadena, CA 91125, USA*
[f] *Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, MA 02138, USA*
[g] *High Energy Astrophysics Science Archive Research Center, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA*
[h] *National Center for Supercomputing Applications, University of Illinois, Urbana, IL 61801, USA*

ARTICLE INFO

*Article history:*
Received 31 October 2014
Received in revised form
11 March 2015
Accepted 16 March 2015
Available online 23 March 2015

*Keywords:*
Catalogs
Surveys
Virtual observatory tools
Data discovery
Data access
Applications

ABSTRACT

The US Virtual Astronomical Observatory was a software infrastructure and development project designed both to begin the establishment of an operational Virtual Observatory (VO) and to provide the US coordination with the international VO effort. The concept of the VO is to provide the means by which an astronomer is able to discover, access, and process data seamlessly, regardless of its physical location. This paper describes the origins of the VAO, including the predecessor efforts within the US National Virtual Observatory, and summarizes its main accomplishments. These accomplishments include the development of both scripting toolkits that allow scientists to incorporate VO data directly into their reduction and analysis environments and high-level science applications for data discovery, integration, analysis, and catalog cross-comparison. Working with the international community, and based on the experience from the software development, the VAO was a major contributor to international standards within the International Virtual Observatory Alliance. The VAO also demonstrated how an operational virtual observatory could be deployed, providing a robust operational environment in which VO services worldwide were routinely checked for aliveness and compliance with international standards. Finally, the VAO engaged in community outreach, developing a comprehensive web site with on-line tutorials, announcements, links to both US and internationally developed tools and services, and exhibits and hands-on training at annual meetings of the American Astronomical Society and through summer schools and community days. All digital products of the VAO Project, including software, documentation, and tutorials, are stored in a repository for community access. The enduring legacy of the VAO is an increasing expectation that new telescopes and facilities incorporate VO capabilities during the design of their data management systems.



## 1. Introduction

### 1.1. Beginnings

The formal Virtual Observatory (VO) program in the United States began with the 2000 Decadal Survey of the National Academy of Science, in which a National Virtual Observatory (NVO) was identified as the top priority small initiative (McKee et al., 2001).

* Corresponding author.
*E-mail addresses:* robert.hanisch@nist.gov (R.J. Hanisch), gbb@ipac.caltech.edu (G.B. Berriman), joseph.lazio@jpl.nasa.gov (T.J.W. Lazio), janet@cfa.harvard.edu (J. Evans), thomas.a.mcglynn@nasa.gov (T.A. McGlynn), rplante@illinois.edu (R. Plante).
[1] Current address: Office of Data & Informatics, Material Measurement Laboratory, National Institute of Standards & Technology, Gaithersburg, MD 20899, USA.

http://dx.doi.org/10.1016/j.ascom.2015.03.007

*The NVO is the committee's top-priority small initiative. NVO involves the integration of all major astronomical data archives into a digital database stored on a network of computers, the provision of advanced data exploration services for the astronomical community, and the development of data standards and tools for data mining. The committee recommends coordinated support from both NASA and the NSF, since NVO will serve both the space- and ground-based science communities.*

The NVO project and parallel projects in Europe and the UK were formulated through a series of meetings, beginning with "Virtual Observatories of the Future" (Brunner et al., 2001), held at the California Institute of Technology in 2000 June.

At the 2002 conference, "Toward an International Virtual Observatory" (Quinn and Górski, 2004), held in Garching, Germany, the International Virtual Observatory Alliance[2] (IVOA) was formed with the NVO, the Astrophysical Virtual Observatory (AVO, ESO), and AstroGrid (UK) as founding partners. R. Hanisch, the then-NVO Project Manager, was the first chair of the IVOA Executive Committee. In the subsequent decade, the IVOA has grown to have 21 member national projects.

The IVOA patterned itself on the World-Wide Web Consortium[3] (W3C) and adopted its process for the development of standards (Working Drafts → Proposed Recommendations → Recommendations) with the actual standards documents developed by a set of working groups. (See Section 3.1 for more details.) A Virtual Observatory Working Group was established under Commission 5 of the International Astronomical Union (IAU) in order to give IVOA Recommendations official status within the IAU, but this process has not been used in practice since there was already global acceptance of IVOA standards.

The NVO project focused on standards and infrastructure development, working closely in the context of the IVOA, and implemented a number of prototype science applications to demonstrate the utility of the underlying VO standards. NVO also ran an active program of engagement with the astronomical community through annual summer schools of one-week duration, exhibits at American Astronomical Society meetings, and the production of a major reference book, *The National Virtual Observatory: Tools and Techniques for Astronomical Research* (Graham et al., 2007). In a demonstration of this book's value, it was translated into Mandarin by members of the VO-China project.

The NVO project was funded by the National Science Foundation's Information Technology Research program, starting in 2001, and included organizations in astronomy and computer science. Its funding came to a planned close in 2008, after demonstrating the technology framework for supporting a VO.

### 1.2. Program

In 2010, the successor to the NVO, the Virtual Astronomical Observatory (VAO), was begun to sustain and evolve those technologies successfully demonstrated by the NVO as part of an operating virtual observatory. While there were numerous management and logistical barriers to the establishment of the VAO, the National Science Foundation (NSF) and National Aeronautics and Space Administration (NASA) agreed to fund the project jointly, with NSF support directed through the VAO, Limited Liability Company, and NASA support provided directly to the participating NASA data centers.

The VAO, LLC, was created as a 50–50 collaboration between the Association of Universities for Research in Astronomy (AURA) and the Associated Universities, Inc. (AUI), with an independent Board of Directors. This management structure was chosen deliberately so that the VAO would be perceived as belonging to the research community and have dedicated oversight. Executive authority within the VAO was provided by the Director, who worked with a Program Manager, Project Scientist, and Project Technologist. In order to provide advice on priorities for research tools, a Science Council was established. Within the VAO, a Program Council consisting of senior management representatives from each VAO member organization was also established. The Program Council worked with the VAO management to map Science Council priorities onto available resources and expertise, and thus to develop the annual program plan. Work packages for all organizations, whether funded by NSF or NASA, were agreed with the Director and Program Manager. The program plan covered all work at all organizations regardless of the source of funding.

Table 1 shows the VAO program history and funding. As a result of two major reviews, NSF and NASA redefined program priorities and reduced the overall budget from an original plan of \$27.5M (\$20M NSF + \$7.5M NASA) to \$16.5M (\$11M NSF + \$5.5M NASA). In addition to simple reductions in funding, these reviews were often accompanied by recommended changes in the direction of the project, and, ultimately, the project duration was reduced by seven months. Consequently, some activities that were started or intended to be started were reduced in scope or stopped early to respond to the combination of lower funding and recommended changes in direction. A specific example of this change in direction and cessation of activities was the Time Series Search Tool (Section 2.4), which was unable to be brought to the desired level of maturity.

### 1.3. Major accomplishments

The accomplishments of the NVO and VAO are extensive and will be described in further detail in the following sections of this paper. At a summary level, however, we note the following accomplishments:

- Major contributor to IVOA standards. Appendix B contains a list of IVOA standards to which NVO/VAO staff contributed. The list includes standards recommended by the IVOA Executive Committee and those submitted to the Executive Committee for recommendation.
- Leadership within the IVOA, within the executive, Working Groups, and Interest Groups.
- High-level science applications for data discovery, integration, analysis, and catalog cross-comparison.
- Scripting toolkits that allow scientists to incorporate VO data directly into their reduction and analysis environments.
- A robust operational environment in which VO services worldwide are routinely checked for aliveness and compliance with IVOA standards.
- Community engagement through AAS meetings, summer schools (NVO), and community days (VAO).
- Comprehensive web site with on-line tutorials, announcements, links to both US and internationally developed tools and services.
- Take up of VO standards and infrastructure within essentially every major data center and survey project in the United States, with approximately 1M VO-based data requests per month and some 2000 unique users.
- Prudent fiscal management, with overall management expenses kept below 15% and the project completed with an unspent balance of funds of less than 1% (for an \$11M [lifetime] budget over 4 years).

[2] http://www.ivoa.net/.
[3] http://www.w3.org/.

**Table 1**
VAO funding and review history.

| | | |
|---|---|---|
| 2010 Apr | NSF Cooperative Agreement issued | \$2M NSF + \$1.5M NASA (FY10)<br>\$4M NSF + \$1.5M NASA (FY11) |
| 2010 Aug | PEP v1.0 | |
| 2010 Oct | PEP v1.1 | |
| 2011 Apr | PEP and review | \$2M NSF + \$1M NASA (FY12) |
| 2012 Feb | PEP v2.0 | |
| 2012 Mar | PEP v2.1, v2.2 | |
| 2012 May | PEP v2.3 | |
| 2012 Jul | PEP and review | |
| 2012 Sep | Decision to terminate VAO, effective 2014 September | \$2M NSF + \$1M NASA (FY13)<br>\$1M NSF + \$0.5M NASA (FY14) |
| | Total funding | \$11M NSF + \$5.5M NASA |

PEP refers to the Project Execution Plan, an annual deliverable to the funding agencies. NSF's funding vehicle was a Cooperative Agreement (CA) with the VAO, LLC.

## 2. Science applications

The VAO developed three science applications (Data Discovery Tool, Iris Interoperable SED Access and Analysis tool, and the Catalog Cross Comparison Service) and one prototype application (Time Series Search tool), all described in more detail below. There were also various community-led efforts, that while not formal VAO projects, built upon VO standards and often involved VAO personnel in other capacities. These are also summarized below.

The motivations for developing these science applications were two-fold. First, before a standard is adopted as an IVOA Recommendation, it is expected that the Working Draft have *two* reference implementations. The objective is to ensure that the intentions of standards actually can be met in practice. In developing these science applications, the VAO provided feedback to the larger IVOA community on various aspects of IVOA standards. Second, these science applications were developed in concert with the research community, providing additional or new capabilities for addressing a variety of astronomical research questions. In the spirit that the VO is intended to enable data discovery and access for all astronomers, the applications do not serve any one observatory, wavelength, or type of user, but were intended for use by astronomers with multi-wavelength data from possibly a variety of telescopes that span the electromagnetic spectrum.

As part of a larger goal of developing an environment or "ecosystem" in which astronomical software can interact seamlessly and other tools can be contributed by the community, the development path for these science applications often included making them interoperable with other VO tools. In so doing, the VAO also provided feedback to the IVOA on the approaches toward interoperability. As a consequence of developing these applications, a number of libraries or services were developed that enable other developers to add functionality to the applications. Two examples are the SEDLIB (SED I/O library) and NED/SED service developed for Iris. Finally, by way of encouraging contributions, several collaborations (e.g., ASI Science Data Center (ASDC) archive plug-in for Iris) were fostered during VAO science applications development.

### 2.1. Data Discovery Tool

The Data Discovery Tool (DDT) is a web application for discovering all resources about an astrophysical object or a region of the sky (Section 3.1). Using protocols defined by the IVOA, the DDT searches those widely distributed resources that are found in the VO Registry and presents the results in a single unified Web page. In the spirit of the VAO being a working astrophysical observatory, the DDT was designed to serve as the initial steps toward a "portal", a means of discovering and accessing multi-wavelength data.

Many of the most popular US archives and catalog holdings are available for searches in the DDT, including the *Hubble Space Telescope*, *Chandra* X-ray Observatory, the Mikulski Archive for Space Telescopes (MAST), the High Energy Astrophysics Science Archive Research Center (HEASARC), Sloan Digital Sky Survey (SDSS), *Spitzer* Space Telescope, and the Two Micron All Sky Survey (2MASS), to name a few. A powerful filtering mechanism allows the user to quickly narrow the initial results to a short list of likely applicable data. Guidance on choosing appropriate data sets is provided by a variety of integrated displays, including an interactive data table, basic histogram and scatter plots, and an all-sky browser/visualizer with observation and catalog overlays (Fig. 1).

The DDT was developed incrementally with the first release of the application in 2011 June. Development continued over the next two years with five incremental releases that added features and addressed any deficiencies. Web-based user documentation and training videos were developed and updated for each release.

The DDT project utilized DataScope (GSFC/NVO, McGlynn, 2007) and Astroview (STScI) and shared synergy with the MAST archive development project at STScI. IVOA standards feedback was substantial. Experience from the DDT project was used to advocate for enhanced registry metadata, table access protocol improvements, and enhanced data access protocols to ensure support for bulk queries. Staff involved in DDT development also helped to write the IVOA standard on HEALPix Multi-Order Coverage maps (Boch et al., 2014) for describing sky coverage.

### 2.2. Interoperable SED access and analysis tool, Iris

Iris is a downloadable Graphical User Interface application that enables astronomers to build and analyze wide-band spectral energy distributions (SEDs, Doe et al., 2012; Laurino et al., 2014a,b). SED data may be loaded into Iris from a file on the user's local disk, from a remote URL, or directly from the NASA Extragalactic Database (NED) for analysis via the NED/SED Service. A plug-in component enables users to extend the functions of Iris. Iris utilized Sherpa (Freeman et al., 2001; Doe et al., 2006) and Specview (Busko, 2002) as the components that performed fitting and visualization in the application. Communication between Specview and Sherpa is managed by a Simple Application Messaging Protocol (SAMP) connection (Taylor et al., 2012a,b). Data can also be read into Iris and can be written out via the SAMP interface (Laurino et al., 2012). A separable library for SED data input/output (SEDLib) is also included and available independently from Iris (Fig. 2).

Iris was first released in 2011 October. Three incremental releases and one bug fix release followed. Iris is supported on sev-

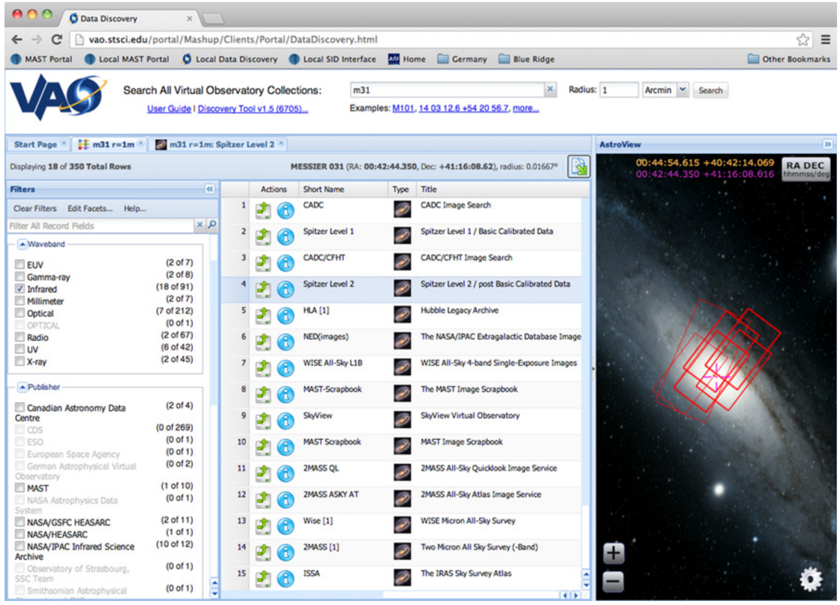


**Fig. 1.** Appearance of the Data Discovery Tool (DDT) after a search for M31 with a radius of 1′ showing the filters (left panel) that can be applied to the search results (center panel), and the AstroView component with field-of-view overlays representing the available data sets.

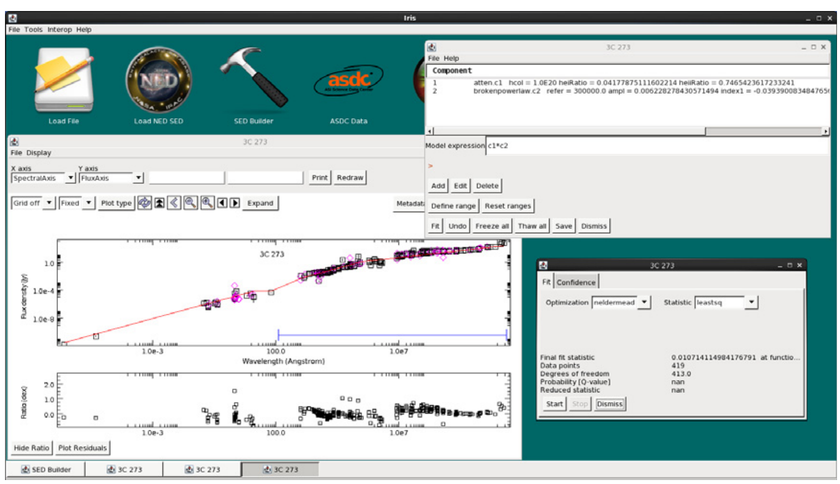


**Fig. 2.** VAO SED access and analysis tool Iris in operation. The Iris desktop holds the interactive windows for SED data review and analysis. Shown is a panel displaying the SED of 3C 273 with a model fit (red curve) and two panels from which the user can describe the model to fit an SED and control the fitting.

eral versions of the Mac OS X and Linux. Web-based documentation and user training videos are also provided. Iris was featured on the Astrobetter blog in 2013 September.[4]

There were two by-products of the Iris project—the NED/SED service and the SEDLib. There were collaborations with several groups including the ASI Science Data Center (ASDC) and CDS (Strasbourg). The collaborations led to Iris desktop plug-in services to access the respective SED data holdings (Laurino et al., 2013). The project provided feedback to the IVOA on the SAMP protocol, allowing for inclusion of a full SED into a single file extension, to TOPCAT (Taylor, 2005, 2011) for better support for SED plots, and inspired work toward a Virtual Observatory Data Model Language (VODML) by lead Iris developer O. Laurino.

4 http://www.astrobetter.com/release-iris-2-0-sed-analysis-tool/.

### 2.3. Scalable Cross-Comparison Service

The Scalable Cross Comparison (SCC) Service performs fast positional cross-matches between an input table of up to 1 million sources and common astronomical source catalogs for a user-specified match radius. The service returns a list of cross-identifications to the user. The output is a composite table consisting of records from the first table, joined to all the matching records in the second table, and the angular distance and position angles of the matches (Fig. 3).

The first release of the Scalable Cross Comparison Service was in 2012 January and was supported with three upgrades over the next 1.5 years. The indexing schemes that support large catalog cross-matching were provided by the Infrared Processing and Analysis Center (IPAC) and later adapted to the Wide-field Infrared Survey Explorer (WISE) and *Spitzer* projects.

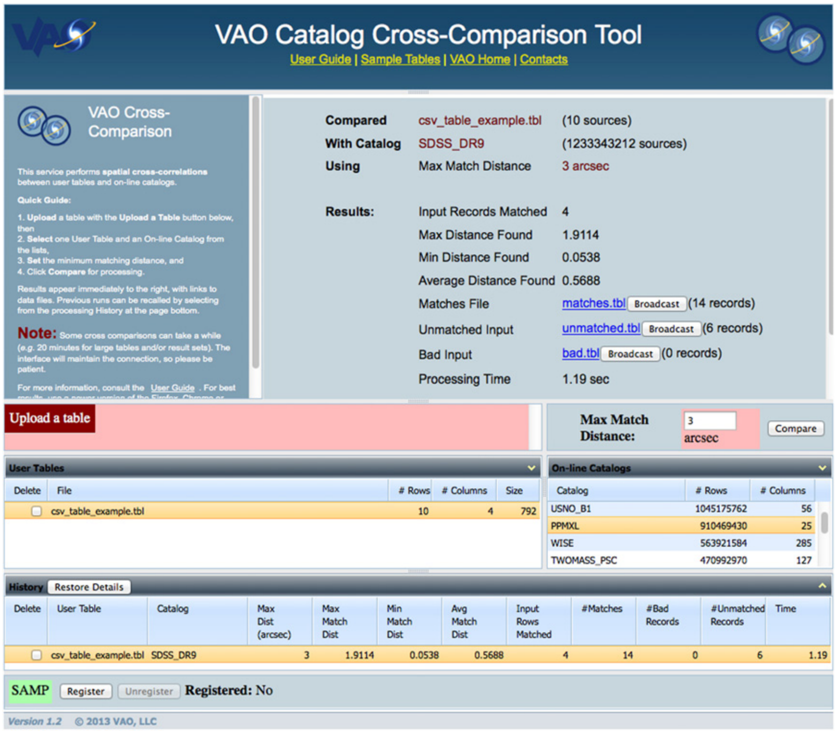


**Fig. 3.** The Scalable Cross Comparison Service. Shown are results for a user-uploaded table cross-matched with the Sloan Digital Sky Survey Data Release 9 catalog.

### 2.4. Time Series Search Tool

The Time Series Search Tool finds and retrieves time series data from three major archives and analyzes them with the NASA Exoplanet Archive periodogram application. The application was a prototype developed to demonstrate that the IVOA standards of the time-series protocol and data model met the needs of such a tool. The development of the Time Series tool ended after the first VAO re-plan.

### 2.5. Lessons learned

The VAO science applications group was distributed across multiple institutions, and Evans et al. (2012) described the management strategy this group. A key element of developing successful applications amongst a distributed group is managing unknowns. As the VAO science application lead might be unaware of the entire set of tasks assigned to an individual outside of the VAO efforts, coordinating task assignments and making organizational material and schedules easily available was important.

The VAO implemented a relatively lightweight process, tracked in a Wiki-based environment, in order to focus the distributed team on the requirements, design, and implementation of the applications. In addition to the developers themselves, a science stakeholder was assigned to each application and was key to bringing the view of the user to the development process. The stakeholder provided requirements, developed science use cases, handled technical questions, advised on development priorities, and performed unit tests. The use cases drove development and provided an opportunity to assess priorities and make course corrections. Internal product deliveries provided a test and assessment loop, and incremental releases (rather than one big software release) ensured that development was progressing as expected. A team lead managed priorities, schedule, and communication within the group.

Frequent communication was essential to ensuring that issues were resolved quickly and the team was working toward a common vision. The distance gap of distributed teams needs to be managed diligently. The VAO Wiki provided easy-to-access project information so that a team member could resume work quickly if he or she were sidetracked due to external project responsibilities. This process enabled the group of developers working on a project, at a distributed set of institutions and working on a part-time basis, to perform their tasks and collaborate efficiently (Evans et al., 2012).

### 2.6. Community developments

During the course of the VAO, there were diverse, community-led efforts to develop VO software. (In some cases, these efforts started during the NVO era, but continued into the VAO project.) Often these involved VAO personnel, either in the role of "consultants" or who were engaged through their work on other projects.

Examples of such community-led software efforts include VOEvent, a protocol for notifications or "alerts" from and between observatories (White et al., 2006); Montage, a user-controlled tool for generating science-quality image mosaics (Berriman et al., 2003); and seleste,[5] a tool designed to provide uniform access to distributed VO databases.

## 3. Standards and infrastructure

The core of the VAO program was the development of software to support the IVOA standards for discovery and access to distributed data. Key components of the VAO infrastructure include the resource registry (the collection of metadata describing on-line data collections and services), the data access layer

[5] http://cda.cfa.harvard.edu/seleste/.

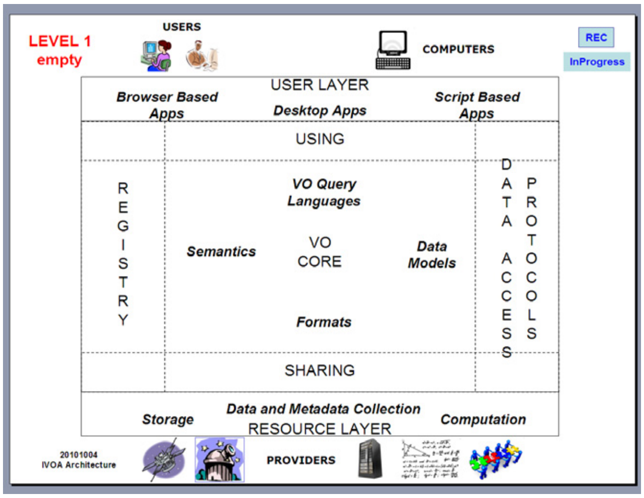


**Fig. 4.** Virtual Observatory architecture (Arviset and Gaudet, 2010). Users appear at the top of the figure and data providers and computational resources are at the bottom, connected by the VO bridge. The VO bridge itself comprises the registry of data providers and data services, the data access protocols for discovering and retrieving data, and the core infrastructure of query languages, data models, data formats, and semantic definitions.

protocols (images, spectra, tables, databases) and their validation tools, a distributed authentication service ("single sign-on"), and applications programming interfaces either built-in to existing software packages or available stand-alone that allow researchers to develop their own VO-enabled scripts. Much of the VAO infrastructure is now incorporated into the data services of major data centers using VAO-provided software libraries.

### 3.1. The VAO infrastructure in context

Fig. 4 shows the VO architecture (Arviset and Gaudet, 2010). In this diagram, the VO infrastructure serves as a bridge between data providers and users, and that bridge is supported by standards. On the provider side, data is connected into the infrastructure through standard services that present that data in terms of standard data models. On the other side, users are connected to the infrastructure via generic tools that understand the VO standards. Tools are no longer tied to a single archive, but rather can talk to any and all archives that speak the common VO language.

Providing the ability to discover and access data of interest is a significant motivation for the structure of the VO architecture. Fig. 5 illustrates the discovery framework. Registries represent the first step for data discovery in this framework. A registry is a database containing descriptions of data collections and services available in the VO (Demleitner et al., 2014). Conceptually a VO registry is similar to a "name server" for domain name service (DNS) on the Internet (Mockapetris, 1987).

There is no single master or central registry; however, there are registries called full searchable registries that aim to have descriptions of all the data collections, archives, and service providers known to the VO from around the world. This type of registry can populate itself through a process known as harvesting; it starts by contacting a special "boot-strapping" registry (run by the VAO) called the Registry of Registries that will return to it all of the other known registries in the VO ecosystem. In order for a new registry to enter the VO, it must be registered with the Registry of Registries.

Most of the registries within the VO are publishing registries. A registry of this type is typically run by a data center that uses it to advertise the data collections and services that it offers to the VO. The full searchable registry contacts each of

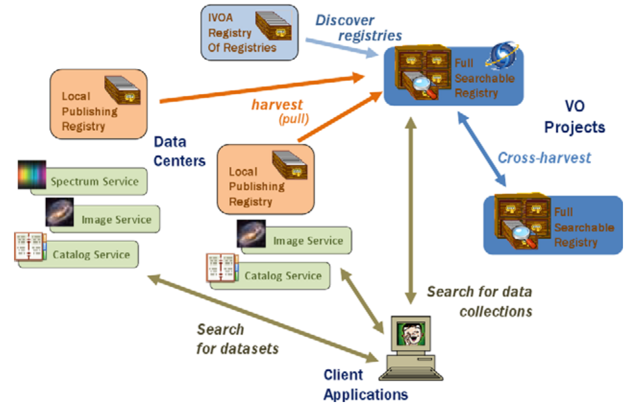


**Fig. 5.** Data discovery in the Virtual Observatory.

the publishing registries and pulls descriptions of all the data collections and services provided by the data center. At this point the full searchable registry is populated with descriptions of all of the resources known to the VO. Periodically it will re-query the other registries to obtain any new resources or other changes since the last harvest.

With an up-to-date full searchable registry available to it, a client application (e.g., Section 2) can discover any data known to the VO. It starts by asking the registry for a list of collections and services from each of the data centers that might have data relevant to the user's science question. Most of the services will be standard data access services for finding and downloading images, spectra, or catalog information from a particular archive or collection. The application can then send a query to all of the matching services to get back lists of available data sets. By browsing the returned metadata for these data sets, the user can choose which data sets to download.

### 3.2. VAO and the IVOA

Much of the work the VAO conducted in advancing standards was through engagement with the IVOA. The role of the IVOA is two-fold: first, to coordinate the efforts of all of the VO projects around the world, and second, to serve as a standards body for establishing VO interoperability.

From the IVOA's beginnings, the NVO and VAO were leaders in shaping the VO's global architecture and the standards that enable it, reflecting the significant data holdings of US institutions. NVO/VAO staff members served as chairs or vice-chairs of key IVOA working groups (Appendix C). The impact of this leadership is also seen in the standard documents; most of the IVOA recommendations across all of the areas of the VO have featured NVO/VAO team members either as first authors, secondary lead authors, editors, or major contributors (Appendix B).

The VAO produced many of the key reference implementations—software that demonstrates a standard in action and proves its viability. During the NVO era, there was a vigorous international debate regarding the character of the VO Registry and whether it should be relatively "coarse-grained" or "fine-grained", in terms of the amount of detail stored in the VO Registry. (See below.) The NVO created the first implementations of registries with several different architectures. The VAO was instrumental in demonstrating data access services through software packages like DALServer (Section 3.4.1) and TAPServer (Section 3.4.2).

The NVO/VAO led the IVOA in the development of service validators. A validator is an application that checks whether another service is compliant with VO standards. A validator performs this check by sending a series of queries to a VO service and examining the response to assess whether it follows all of the rules and recommendations described in the standard. The NVO developed the first validators in the IVOA to assist data providers, allowing them to check their data access services and fix any problems before publishing them to the VO. These NVO validators quickly became critical pieces of VO infrastructure and were continued by the VAO (Section 3.4.3); and other projects joined in to contribute validators for other service standards.

In other areas, though, the VAO benefited from international developments. Not only did the VAO benefit from technical comments, there were multiple occasions in which the VAO could produce a library or tool ultimately more rapidly because some of the initial development had been done by international partners (e.g., the development of the single sign-on capability, initially developed by Astrogrid).

### 3.3. The Registry

As described above, a VO registry is a database containing descriptions of data collections, archives, services, and other resources, and it represents the first step in data discovery. The NVO/VAO established itself as an early leader in the area of registries. In addition to creating some of the first registries, the VAO operated the Registry of Registries (RofR) on behalf of the IVOA. The RofR allows searchable registries to bootstrap their collection of resource descriptions.

The NVO and its IVOA partners developed several different types of registries with several different implementations. There was considerable debate over the registry design, and whether it should be "fine-grained" or "coarse-grained". A fine-grained registry contains detailed metadata about the datasets available at a VO resource (for example, it might contain the right ascension and declination of all observed positions in an archive). A coarse-grained registry would only contain information about the general sky coverage of an archive. The advantage of a fine-grained registry is that one need not query distributed resources explicitly to determine if they have data of interest, whereas with a coarse-grain registry data discovery is a two-step process. The problem with a fine-grained registry, however, is that many data collections are dynamic, so that any metadata cache has to be updated continuously. Also, the structure of a fine-grained registry will necessarily be much more complicated, and harvesting of metadata between fine-grained registries could easily become inefficient. Despite the efficiencies for search and discovery offered by fine-grained registries, the VO currently operates with coarse-grained registries. The VAO consolidated support around the coarse-grained, full searchable registry service at the Space Telescope Science Institute. There are ongoing efforts to build a fine-grained registry for mostly static data collections.

#### 3.3.1. VAO Directory Service

As part of the VAO's production registry, a Web-browser-based front end called the Directory Service[6] was provided. This tool is particularly useful for discovering collections and services related to a topic. By entering keywords into the search input box, the tool will return a list of resources whose description contains those keywords (Fig. 6).

#### 3.3.2. Registry upgrades

Over the last two years of the VAO project, an updating program was conducted to overhaul the underlying registry database and update it to support the latest IVOA registry metadata standards. This overhaul was also necessary to support a new standard for searching registries. This new standard leverages an existing IVOA standard for querying complex databases called the Table Access Protocol (TAP), for which client software already exists. (The TAP standard did not exist when the first registry search interfaces were standardized.) Completing this upgrade was critical to maintaining the registry in the eventual post-VAO era.

A final effort conducted in the VAO project was to complete registry curation activities aimed at improving the descriptive content of the registry. In particular, a specific approach was implemented to registering resources intended to make registry searches more effective and their results less confusing. This approach has recently been accepted as a best practice by the IVOA Registry Working Group. The VAO curation work started with an inventory of existing resources by publisher, followed by developing a set of recommendations for improving the resource descriptions that brings them into line with the best practice. The new registry resource publishing tool (described below) will be instrumental in communicating these recommendations to the publisher.

#### 3.3.3. Publishing registries and the resource publishing tool

A publishing registry is the vehicle for making a resource available to the VO. In particular, it can create new descriptions of resources and share them with the rest of the VO through the harvesting process. A data center, which may curate a number of data collections and offer a variety of services to access them, may operate their own publishing registry. Because such a registry does not need to serve end users directly, operating one is much simpler than running a searchable registry. During the NVO project, the VORegistry-in-a-Box product was developed that provides a simple but compliant publishing registry implementation through which a data center can maintain its own resource descriptions in-house. This product is still in production use within the VO (including by the Registry of Registries), and the VAO continued its support.

A searchable registry can also support the publishing function, which the VAO Registry at STScI does. In particular, it maintains resources descriptions on behalf of data providers who only have a few resources to share, relieving them from having to run their own publishing registry. In order to enable this feature, the VAO created the Resource Publishing Tool, a browser-based application that allows a data provider to create and share resource descriptions through the VAO Registry. It features a guided interface that steps a data provider through the process of describing a resource, prompting for metadata along the way. The tool also can check for the validity of values as they are entered, alerting the user of any problems. Draft descriptions can be saved for updating and publishing later, and already-published resource entries can be updated with this tool. Various techniques are used to minimize the amount of typing required to create a useful resource description. While the VAO Registry will share records created through this tool, the descriptions are considered "owned" by the user. Thus, to control access, the publishing tool uses the VAO Single Sign-On Services (described below).

### 3.4. Data access

Standard services that allow users to find and access to data from an archive are part of the VO architecture known as the data access layer (DAL; Fig. 4). In the VO architecture, there is a standard service for each type of dataset; e.g., the Simple Image Access protocol (SIAP, Tody and Plante, 2009) enables discovery and downloading of images from an archive, and the Simple Spectral Access (SSAP, Tody et al., 2012) protocol enables access to spectra. In this section, we describe the four different "toolkits" or new protocols that the VOA developed for improved data access within the VO.

6 http://vao.stsci.edu/directory.

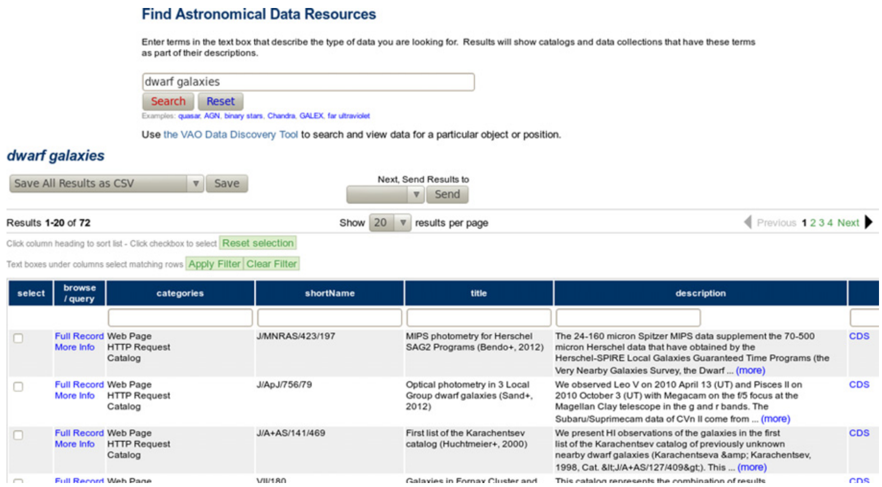


**Fig. 6.** Results from a search query submitted through the VAO Directory Service. The directory service results allow one to browse the descriptions, filter results, download matching descriptions in VOTable format, and when the resource is a standard data access service, even send it a sky position-based query.

### 3.4.1. DALServer

In order to help data providers share their data collections through standard VO services, the VAO created the DALServer Toolkit, a Java-based software package. When first developed as part of the NVO project, it served as a platform for developing reference implementations of standard VO services (like SIA and SSA) that demonstrated features of the standards. At about the same time, both Astrogrid and ESO were developing data access toolkits, and some of this development fed into the VAO concept.

During the VAO's final year, specific efforts were made to enhance the toolkit for use directly by data providers; this effort was considered "productization", as it focused on making the toolkit easier to use. The focus was on a simple class of use cases in which a small data provider had a simple catalog or a simple collection of images or spectra that they wished to share. By just editing configuration files and running a few scripts, the provider could deploy fully compliant VO services with no programming required. For more complicated situations, such as for a data center that might already operate custom data access services through their own data management system, they could use the underlying DALServer Library application programming interface (API) to adapt the VO services to their local infrastructure.

The first production release of the DALServer provided support for the four "simple" standards for data access recommended by the IVOA: namely, Simple Cone Search (SCS, for simple position-based querying of object and observation catalogs, Williams et al., 2008), Simple Image Access Protocol (SIAP, for finding images), Simple Spectral Access Protocol (SSAP, for finding spectra), and Simple Line Access Protocol (SLAP, for finding rest frequencies for spectral line emissions, Salgado et al., 2010). Toward the end of the VAO project, DALServer was extended to operate on multidimensional data sets (Section 3.4.4).

### 3.4.2. TAPServer

The Table Access Protocol (TAP) is an IVOA standard for querying complex catalogs that may be made up of several tables (e.g., the 2MASS catalog). When a TAP service is connected to a catalog, users can create complex, SQL-like queries that can join metadata from several tables. Such queries are critical for mining very large catalogs. Not surprisingly given its power and flexibility, a TAP service is one of the more complex IVOA standards to implement. To make deploying a TAP service easier, the VAO created the TAPServer toolkit.

Like DALServer, TAPServer is configuration file driven. That is, with no programming required one can wrap the toolkit around a collection of tables in a database and deploy it as a service accessible to the VO. Because of the VAO close-out schedule, only a limited amount of development could be completed, and there was no effort toward the "productization" of TAPServer. However, the code is included in the VAO Repository and available for community use. Some post-VAO targeted deployments are planned. For example, it will be deployed at the National Center for Supercomputing Applications (NCSA) to expose the Dark Energy Survey Source Catalog. In turn, DES scientists will be able to analyze the catalog using the seleste[7] TAP client, a tool that allows users to form complex queries with little or no knowledge of SQL.

### 3.4.3. Service validators

During the VAO project, service validators originally developed during the NVO project were continued and expanded. These validators have a web browser interface that allows a data center to enter a service access URL and test the services compliance with the appropriate standards; the result is a listing of errors, warnings, and recommendations for improving the service. These validators share a common Java-based toolkit platform called DALValidate. They also support a programmatic interface that allowed VAO Operations to automatically test VO services. (The VAO Operations team also engages other validators developed outside of the VAO.) Supported validators include those for Simple Cone Search, Simple Image Access, Publishing Registries, and VOResource records. The DALValidate software is available through the VAO Repository.

### 3.4.4. Image cube access

An emerging suite of telescopes is or soon will be generating multidimensional data (often termed "image cubes"). The most general data set, produced by an instrument measuring photons, would be $[I(\alpha, \delta, \nu, t), Q(\alpha, \delta, \nu, t), U(\alpha, \delta, \nu, t), V(\alpha, \delta, \nu, t)]$, where we have described the polarization properties by the Stokes parameters $(I, Q, U, V)$ and each polarization can be a function of position on the sky $(\alpha, \delta)$, frequency $\nu$ (or equivalently wavelength $\lambda$ or energy $E$), and time $t$. Radio interferometers have naturally produced such multidimensional data sets for some time, and the commissioning of the Jansky Very Large Array (JVLA) and the Atacama Large Millimeter/submillimeter Array (ALMA) is making such data sets much more common. X-ray telescopes have, for some time, been generating data that can be considered to be

7 http://cda.cfa.harvard.edu/seleste/.

extremely sparse image cubes. The introduction of integral field units (IFUs) both for ground-based telescopes and eventually for the *James Webb Space Telescope* is making these data more common at visible and infrared wavelengths.

Providing discovery of and access to multidimensional data was taken up as a key project upon endorsement of the VAO Board. The VAO also helped stimulate interest in multidimensional data within the IVOA. In the IVOA, it was recognized that although SIAP could support image cubes in a limited way, it lacked some of the metadata support and data access mechanisms needed to support the cubes being produced or soon to be produced. From the VAO perspective, not only would discovery and access to multidimensional data advance a new capability in the VO, it might also to engage the radio astronomy community more in VO activities.

The VAO produced an early prototype service that demonstrated a number of the key capabilities needed in a new standard for image cube discovery and access. This demonstration was instrumental for mapping out the strategy for an SIAP Version 2 (Dowler et al., 2014). In particular, the necessary standardization was broken down into three independent components: (1) the Image Data Model defines the semantic labels used to describe image cubes; (2) these labels are used by the SIAPV2 standard to annotate image search results; and (3) the Access Data standard defines how one can request cutouts or other transformations of image cubes.

While active in the development of the standards within the IVOA, the VAO continued prototyping access to image cube data. To ensure that the standards served the needs of real providers of image cubes, we established a collaboration with the National Radio Astronomy Observatory (NRAO). Our joint goals were first to create a real functional image cube access service based on the emerging SIAV2 draft serving real data from NRAO instruments, and second, to provide a useful architectural design along with software to support active archive operations. In this collaboration, NRAO provided the VAO project with requirements and use cases. NRAO desired an image service that could simultaneously provide data to both internal and external clients. One key client is the Common Astronomy Software Applications (CASA, Jaeger, 2008; CASA Consortium, 2011). Viewer that needs to request small visualizable parts of a larger cube. In return, the VAO project provided NRAO with general purpose software to both deliver data over the network to clients. The DALServer product (Section 3.4.1) was extended to provide server-side support for SIAV2, and VOClient (described below) was extended to support the client. In the spring of 2014, NRAO, using VAO-provided software, successfully demonstrated a service that provides access to image cube data, including image cutouts. This service allows their archive and CASA Viewer developers to test against a functional service.

### 3.5. *Data sharing*

A common interest among astronomers is making their data available to their colleagues. Data sharing can be essential part of a project in which team members are in different institutions, or it can be for legacy reasons to enable data re-purposing (using the data for studies not originally envisioned when the data were acquired) or for ensuring replication. While there are institutional data centers, both in the US and internationally, there are also so-called "long-tail" data, the many small collections of data products that are typically associated with published papers. Such data products tend to be highly processed by individual astronomers and are not typically available from traditional observatory or project archives.

The second key project conducted during the close-out plan was more exploratory in its full scope (though it supported an important end-user application). The VAO sought to understand how these products could be published to the VO in a low-effort way; in order to enable such access, the focus was on integrating data sharing and publishing into the overall scholarly publishing process which starts even before the first draft of a paper. Two products were developed.

#### 3.5.1. *SciDrive*

SciDrive (Mishin et al., 2014) is a Dropbox-like cloud storage application intended for use in scientific research. It was inspired by the SDSS MyDB (O'Mullane et al., 2004) and AstroGrid MySpace (Davenhall et al., 2004) developments, and it is based primarily on the OpenStack[8] software (in particular the OpenStack Swift component for object storage). It can be accessed from a web browser in which the user is presented with a view of a personal hierarchical directory space where one may save files by dragging-and-dropping file icons into the web page interface (Fig. 7). Also available is a desktop client that can (like Dropbox) monitor a local directory and automatically upload files that are moved into it. As many researchers already do with Dropbox, SciDrive can be a simple platform for sharing data within a research group; it provides a secure means to share read–write access to a collection within a restricted group or to send one-off permissions (read or read/write) to individuals. One difference from commercial storage providers is SciDrive's ability to scale to larger collections than with the typical free versions of storage.

SciDrive supports the VOSpace 2.0 interface, the IVOA standard for managing third party data transfers (Graham et al., 2014). This capability allows a user to seamlessly move files between different SciDrive instances (or other VO-compatible storage systems) located around the network. This feature is important for new VO capabilities in which web-based tools allow users to save application outputs to their personal space in the cloud. These outputs could be reloaded later into the tool for further analysis (e.g., as a "favorite" starting point) or loaded by other tools for synthesis with other data and analysis. A current example of this is the use of SciDrive with the Sloan Digital Sky Survey (SDSS) CasJobs: a SciDrive user can configure a directory to automatically detect uploaded table files and load them into the SkyServer database so that it can be correlated with the SDSS catalog.

The CasJobs connection highlights another unique feature of SciDrive: it supports plugins that enable special handling of certain types of data. This application is therefore a possible platform for publishing data by individual scientists and research groups. There has been experimentation with plugins that automatically extract the metadata from files that is needed to expose data to the VO. With such a feature, a research group could use SciDrive to organize a collection of data for publication. When the collection is ready for release a simple press of a button would expose the data publicly; the metadata would be automatically loaded into a database and the collection would be made available through standard IVOA services (e.g., using DALServer).

Completing this vision to a working implementation was beyond the scope of the VAO Project; nevertheless, operations and development at Johns Hopkins University (JHU) of the SciDrive platform continues (with NSF support from the Data Intensive Building Blocks program). Furthermore, VAO partners JHU and the National Center for Supercomputing Applications (NCSA) are collaborating on the emerging, community-driven initiative called the National Data Services (NDS) Consortium, which aims to address data publishing across all research fields. As the publishing scenario described above is much like one being discussed in the NDS community, we expect the development of SciDrive as a publishing platform to continue beyond the VAO project.

8 http://www.openstack.org/.

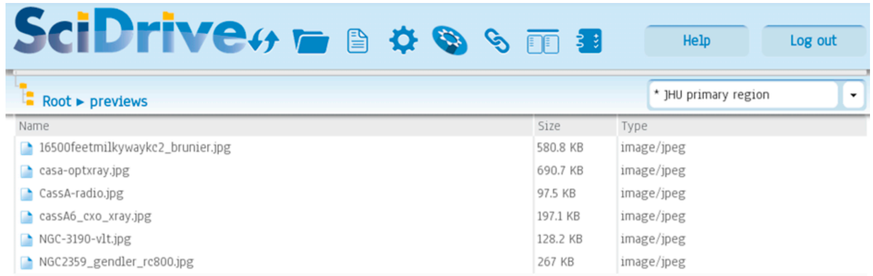


**Fig. 7.** The SciDrive file browser, viewed via a web browser.

#### 3.5.2. *Single sign-on services*

In order to restrict access to the user's personal space, SciDrive uses the VAO Login Services (Plante et al., 2012) for authentication. These services were created so that VO users could have a single login to connect any VO-compatible service or portal even when they are managed by different organizations. More than the simple convenience of a single login, a federated login system allows a user to access their proprietary data from one data center using analysis tools from another data center. A participating organization can choose to support VAO logins either as its primary identity or as an augmentation of its local authentication system.

Inspired by initial developments by Astrogrid for the single sign-on capability in an astronomical context, the VAO federated login is built on the OpenID standard[9] that is in broad use across the Internet. Associated with it are all the usual services that help users manage a login: the ability to reset forgotten passwords, edit the user profile, etc. The VAO Login service also leverages an OpenID feature for sharing user information with a portal in a privacy-conscious way; this can make registering users with a portal faster and simpler. One less common feature that is important for VO applications is the ability for transparently delivering X.509 certificates to the portal. This allows a portal to access private data at another site on the user's behalf. While the service requires the user's permission to do this, it is worth noting that the user never handles the certificates directly.

The development of the VAO Login service resulted in two release software products. First, VAOSSO provides the user identity server that powers the VAO services. This software can be configured either to run as a mirror of the VAO service (for high availability) or as a completely independent service. Second, VAOLogin is a toolkit that helps portal developers add support for VAO Logins.

Current applications using the VAO Login Services include SciDrive, the VAO Registry's Resource Publishing Tool, and the VAO Notification Service. The National Optical Astronomy Observatory (NOAO) Data Archive, which currently supports the predecessor NVO Login Service, is migrating to use of the VAO Login Services to augment their own local authentication system.

### 3.6. *Virtual astronomy on the desktop*

A key initiative of the VAO Standards and Infrastructure program was to make VO capabilities more available from a user's local machine. Not only was the goal to make VO capabilities integrated into both new and existing desktop applications, the VAO Project sought to deliver that power directly to scientists through custom scripts that they can create to conduct their research.

Because of its growing popularity as a scripting language for scientific research, Python[10] was a major focus of our scripting support, following upon the example set by AstroGrid's python package. Further, we enabled all VO-enhanced applications and scripts running on the desktop to work together using the Simple Application Messaging Protocol (SAMP, Taylor et al., 2012a), the IVOA standard that allows desktop and Web applications to exchange data.

#### 3.6.1. *VO-Enhanced Image Reduction and Analysis Facility (IRAF)*

The first VAO product supporting VO on the desktop was a VO-enhanced version of IRAF (Tody, 1986, 1993; National Optical Astronomy Observatories, 1999) developed by M. Fitzpatrick (NOAO). This included some general IRAF infrastructure enhancements including the ability to load data from arbitrary URLs as well as support for loading data in VOTable format. With these two capabilities, a suite of tasks was added to take advantage of VO services; these included an object name resolver, the ability to search the registry to find archives and services, the ability to search individual archives or catalogs, and the ability to download discovered data products. SAMP support was also added so that IRAF could send data to other non-IRAF tools running on the desktop; for example, images could be sent to Aladin (Bonnarel et al., 2000) and catalogs to TOPCAT for visualization.

#### 3.6.2. *VOClient*

This downloadable product provides direct access to VO services outside of a Web browser. The first VOClient release featured a suite of command-line tools that enables interactive use from UNIX/Linux shell; they can also be used to create customized shell scripts. The capabilities provided by these tools include discovering archives and catalogs via the VAO registry, searching individual archives for images and spectra, downloading discovered data across multiple archives, searching catalogs by position, resolving object names to sky positions, and sending data to other desktop tools (via SAMP).

The second release of the VOClient package focused more on the underlying set of core C libraries. These libraries can be used directly to add VO capabilities to C and C++ applications (as was done for the NRAO CASA Viewer). These libraries are intended to be the basis for bindings to other languages, such as Python and Perl.[11] The Python bindings in particular were a focus of the second release (which featured a common API with PyVO, described below). Finally, the second release featured a task framework that enables easy integration of legacy software, making it callable from Python.

9 http://openid.net/.

10 https://www.python.org/.

11 https://www.perl.org/.

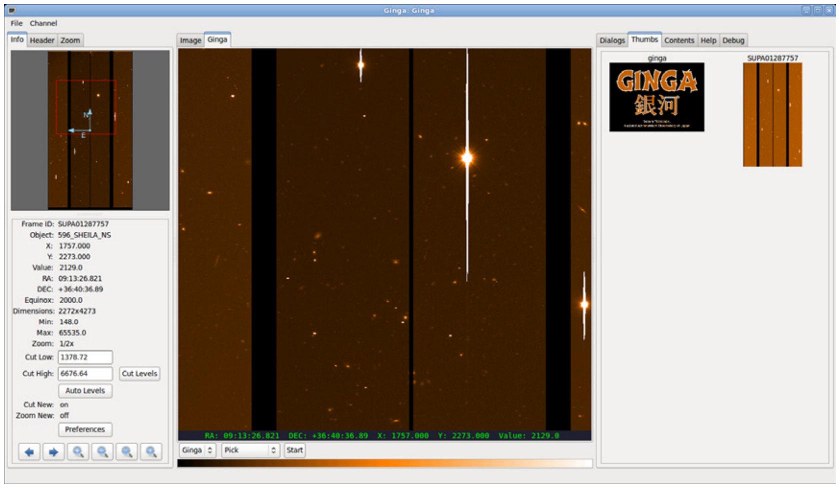


**Fig. 8.** Ginga Image Browser from the Subaru Telescope showing a VO plugin powered by PyVO. The rightmost panel represents a plugin that allows users to download images and catalogs from the VO for display and overlay in the viewer.

### 3.6.3. PyVO

This downloadable product represented a parallel effort to support Python with a slightly different focus. Through our community engagement, we found that many Python users prefer to use a pure Python implementation of a VO library, which PyVO provides, as opposed to a mixture of Python and Unix system commands. As for VOClient the audience is two-fold, the first being developers who want to integrate VO capabilities into their own Python applications. As an example, Fig. 8 shows the Ginga image browser, developed for the Subaru Telescope, to preview observatory images (Jeschke et al., 2013). Downloading of images and catalogs was an additional functionality added to the Ginga image browser using the PyVO python module.

PyVO was also aimed at the growing community of research astronomers using Python to create custom scripts to carry out their research and analysis. In fact, PyVO is built on top of the widely used Astropy package (Astropy Collaboration et al., 2013), an integrated set of astronomically-oriented modules. This allows users to discover and download data and process and analyze it with the robust capabilities of Astropy. This combination is an important key to doing VO science at a large scale, as it becomes very easy to apply common processing to a vast array of data either from a single survey or from distributed collection. It also becomes possible to continuously monitor the evolving holdings of an archive or the VO in general as new data sets are added.

The first evaluation version of PyVO was released in 2013. As this release date was close to the end of the VAO Project, we wanted to ensure further use and development of PyVO beyond the Project's end. Accordingly, we explicitly employed a strategy to build a community around the PyVO package. First, GitHub[12] was used to provide a web-based code repository for future community contributions. This approach has enabled important contributions from users outside of the VAO Project; as of this writing, there are 22 issue submissions from seven external users and seven code submissions from four external users. The other part of the strategy was to establish a strong tie to the Astropy community, which is quite large and active. (In fact, this tie is responsible for much of the external participation via GitHub.) To this end, we applied for and were given status as an Astropy "affiliate package". This connection also allows PyVO to become a proving ground for migrating addition VO capabilities into Astropy.

## 4. Operations

The VAO operations effort addressed two primary goals. The first was to enable science use of the VO, in the sense of being an "operational observatory", with a focus on the VAO-developed interfaces but not exclusively. Tools must work, should work consistently, and when problems arise they must be swiftly resolved. The second goal was to enable the services needed internally for the activities of the VAO itself. VAO personnel needed reliable access to the tools needed for software design and access, user support, testing, configuration management, bug tracking, and so forth.

The VAO provided a number of science services and tools directly to the scientific community (Sections 2 and 3): its home web site, a data portal and cross-correlation tool, the Iris SED tool, downloadable VO libraries for use by clients and servers, and cloud storage and secure access protocols. Internal services included the VAO infrastructure: the JIRA ticket system, a Jenkins testing service, SVN code repository, a YouTube channel,[13] a blog, and mailing lists. The VAO also supported the IVOA Web site and document repository; these were transferred to international partners in Italy and India. The VAO software repository[14] was established to ensure that VAO-developed resources are available indefinitely.

VAO services are supported by member institutions of the VAO with significant resources hosted at each of our sites: the Smithsonian Astrophysical Observatory, JHU, MAST, HEASARC, NRAO, NOAO, Caltech, and IPAC (IRSA and NED). Most recently the software repository has used free Google cloud-based services. Elements are distributed across the country and the Internet.

Supporting such a distributed system posed (and will pose) special operational concerns. Especially for its science users, the VAO worked to ensure that elements were seen as a coherent whole: science tools need to be available at a common location, forms should have consistent look-and-feel, and everything should be clearly visible through a consistent web presence even when the web sites are on various servers.

All elements were continuously monitored and a responsible party identified for each so that issues could be rapidly and decisively addressed. The operations staff met frequently (in

12 https://github.com/.

13 https://www.youtube.com/user/usvaoTV.

14 https://sites.google.com/site/usvirtualobservatory/.

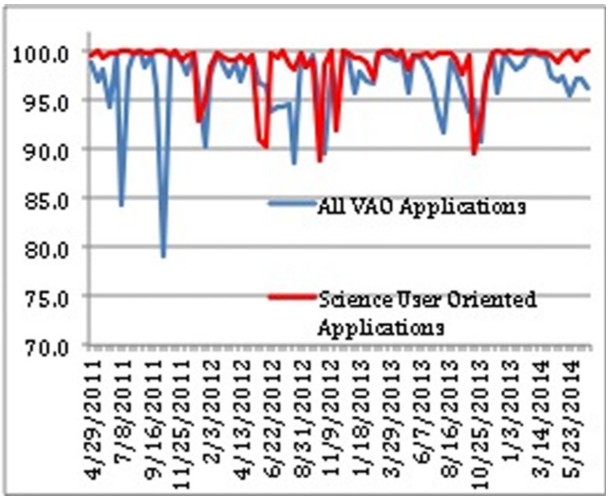


**Fig. 9.** Operational status for all VAO services since from spring 2011 through early summer 2014 in each biweekly period.

weekly telecons) and operational issues were rapidly escalated using an internal issue tracking software to whatever level was needed to ensure that they received the needed visibility.

## 4.1. Service monitoring

All VAO services were monitored hourly and a database of all tests was continuously updated. Each service was tested to ensure not only that the service was operational, but also that it responded sensibly to some simple request. When services failed a test, they were retested 15 minutes later. If the second test also failed, a message was automatically sent to the responsible parties and to the VAO operations monitor.

A web site was available giving the current status of all operational services, and the VAO home site reflected the operations status of VAO science services so that users were immediately informed if there was an issue. Statistics were collected in and reported in biweekly periods.

Fig. 9 shows the operational status for all VAO services from spring 2011 through early summer 2014 in each bi-weekly period. The blue line shows that some of the internal VAO services – not seen directly by our science users – have had significant downtime recently. This mostly reflects in our testing and validation tools. More critically, the red line indicates only one significant lapse, in 2013 October, for the science-oriented services since early 2013. This was directly due to the shutdown of US federal services that affected NASA sites.

## 4.2. Monitoring and validation of VO data providers

Since the effective operation of the VAO from the perspective of science users required that VAO data providers' services were available, in addition to testing the aliveness of VAO services, the VAO also monitored whether data services external to the VAO were working. Every site that published data through the VO was tested each hour. Not all published services were tested; rather a representative service from each of class of services at a site was tested. All tests were recorded and the current status of all VO sites could be seen at the VO monitoring web site. When a problem was detected, the VAO operations monitor contacted the responsible party and noted the problem. In many cases the VAO assisted such sites in rapidly bringing their services back on-line.

Occasionally a VO data-providing site is abandoned. When sites were not responsive after two months, the VAO monitoring service

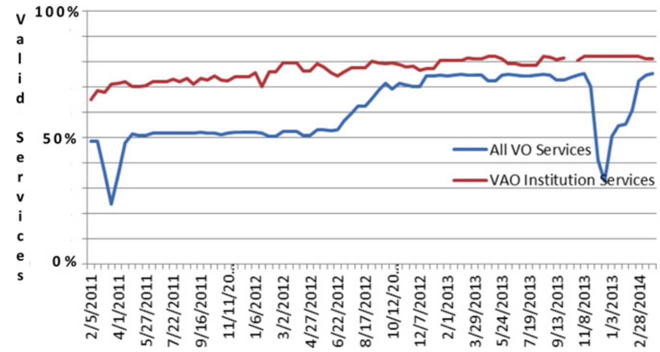


**Fig. 10.** Variation with time of the fraction of VO services that completely passed validation.

deprecated them in the VAO registry so that users would no longer see them in typical queries.

Each week approximately 5–10 service interruption issues were handled. In addition to testing whether services were available, the VAO also validated every published VO service using the catalog/table, image, spectral, or registry service validators. Each day approximately 300 services were validated and all validation issues were recorded in a database. This means that all published services were validated roughly once per month. Periodically, a summary report describing the VAO validation issues was prepared for each site, in order to provide concrete recommendations for resolution of validation issues.

A service that does not pass full validation can still provide valuable information, but obtaining more complete agreement with the IVOA standard ensures that tools work more robustly.

Fig. 10 shows the fraction of VO services that completely passed validation. The blue line shows all VO data providers, while the red line shows the services associated with institutions that were part of the VAO. In both cases there was a steady rise in compliance over the past several years. Two major drops in the overall compliance reflect bugs introduced at one of the major VO data providers outside the VAO. Seeing these declines our operations monitor worked with the provider, identifying specific services that were affected after initial bug fixes did not completely rectify the problem, and helped in their recovery.

## 4.3. Post-VAO operations

The disposition of VAO-developed services and resources is discussed in detail in other sections of this paper. Most science-oriented services will continue to be maintained by the existing institutions. The state of the internal VO services, mailing lists, documentation, blogs, and such will be maintained in the software repository. Critical infrastructure services, the web site, registry, and monitoring tools will be maintained as part of a coordinated NASA follow-on effort. This will also include at least some coordination of NASA VO operations efforts. Our experience has shown that the VO, a broadly distributed system, greatly benefits from clear and comprehensive mechanisms to identify and resolve operational issues. While the NASA follow-on effort may provide some minimal capabilities, it requires a broader national and international visibility. This is not currently something that is handled by the IVOA.

# 5. Community engagement and user support

During the course of the VAO, effort was undertaken to ensure that products and services delivered were robust and usable by research scientists and to reach out to the broader astronomical community. The outreach efforts aimed to expose VO products


**Fig. 11.** (*Left*) VAO web site home page. (*Right*) Science Tools and Services area within the VAO web site. Tools and services developed by the VAO appear at the top of the document, VO tools and services provided by the community appear as well.

and services to potential users, to assist in the take-up of those products and services, and to gather feedback in order to assure the maximum utility of the VO for astronomical research. This section describes the full scope of the efforts.

### 5.1. Web site

Fig. 11 shows the VAO web site, with an intended audience of professional astronomers and software developers. The web site was designed both to serve as an entry portal to the VAO and to provide a means for astronomers to find information about the VO—of the more than 3 million results of a search for "virtual observatory" with Google, the VAO web site is one of the top hits.

From the perspective of the end user, the web site had two key areas. The first was "Science Tools and Services". This web document provided access to the web services or software developed by the VAO. Further, as the project began to mature, community provided tools or services began to be developed, and links to those tool or services were added.

The second area of interest for end users was "Support and Community". Analogous to the "knowledge base" that might be provided by a commercial software provider, this area was designed to help users find answers to their questions, contact other users, or submit bug reports (Fig. 12).

### 5.2. Product testing

At the beginning of the VAO, quality control and testing activities were under the purview of User Support. The motivation for this structure was that User Support could serve as a proxy for the end user and ensure that the products and services could be used in a research setting. For most testing activities, the User Support role was to act as the coordinator of the activities and as reviewers. In addition, User Support took the lead for performing User Acceptance Testing (UAT), which was used, along with other tests and quality control reports, to prepare software release readiness reviews.

### 5.3. Documentation

User Support staff wrote or completed user documentation in order to help research scientists have a better understanding of VAO services and applications and how to use them. Documentation packages included deployment instructions, general descriptions, tutorials, cookbooks, and similar documents. The User Support staff and product developers also collaborated to produce video tutorials, which were then made available through a YouTube channel. All software documentation produced is available in the VAO Repository and the video tutorials remain available through YouTube.[15]

### 5.4. Scientific collaborations

During the course of the project, the VAO supported the scientific or technical work of multiple individuals or collaborations. The objectives of explicitly supporting such scientific collaborations was two-fold. First, we aimed to provide examples of the VO infrastructure and capabilities being used for astronomical research. Second, the interactions with the teams were anticipated to provide feedback to the development teams for improvements to the VO infrastructure and tools. The requests for support resulted both from *ad hoc* proposals to the VAO and from a formal call for proposals that the VAO issued in 2012. The following is a summary of the projects and work supported.

- "Real-Time Analysis of Radio Continuum Images and Time Series for ASKAP" (PI: T. Murphy). This proposal requested assistance in describing multi-dimensional radio wavelength data and publishing it to the VO. Interaction with this team was used as a key use case in developing the VAO Standards and Infrastructure effort toward multi-dimensional data and in interactions with the IVOA.
- "Integration of AAVSO Data Archives into the Virtual Astronomical Observatory" (PI: M. Templeton). This proposal requested assistance in publishing data from the American Association of Variable Star Observers into the VO. The VAO provided assistance to the AAVSO, and the data are now available.
- "Cosmic Assembly/Near-infrared Deep Extragalactic Legacy Survey (CANDELS)" (PIs: S. Faber and H. Ferguson). The VAO supported the CANDELS program by distributing supernovae detections with the VOEvent network and providing access to CANDELS images through standard VO image access protocols. CANDELS supported the VAO program by providing guidance on requirements for SED building and analysis tools.

15 http://www.youtube.com/user/usvaoTV.

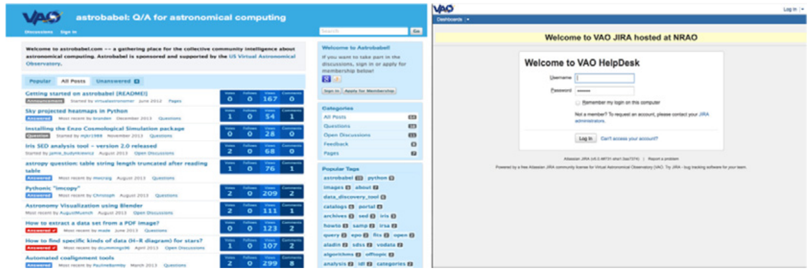


**Fig. 12.** (Left) VAO Forum at astrobabel.com, where users could post questions and interact with other users. (Right) VAO Help Desk.

**Fig. 13.** Collage of images from the VAO Booth at the 221st American Astronomical Society Meeting, Long Beach, CA (2013 January). Also shown is one of the VAO-related posters (Kinne et al., 2013).

- "Brown Dwarf Candidate Identification Through Cross-Matching" (PI: S. Metchev). The VAO supported a project that continued a search for extremely red L- and T-type brown dwarfs that had begun during the NVO. It involved cross-comparing the 2MASS and SDSS catalogs to identify candidates that were followed-up with spectroscopy at the Infrared Telescope Facility, Mauna Kea. The project identified the two reddest known L dwarfs, nine probable binaries, six of which were new and eight of which likely harbor T dwarf secondary stars, and derived an estimate of the space density of T dwarfs (Geißler et al., 2011).

In addition to these scientific collaborations, a scientifically motivated sub-award was issued to produce a cross-matched multi-wavelength catalog of more than 1M objects within a 10° radius of the SMC was produced ("A Catalog of Spectral Energy Distributions of Stars in the Small Magellanic Cloud", PI: B. Madore). The catalog is in the VAO Repository, and it has been incorporated into NED with value-added content.

### 5.5. Booths and exhibits at American Astronomical Society meetings

American Astronomical Society (AAS) meetings, principally those occurring during the winter, are one of the focal points for the US (and international) astronomical community. During the course of the project, the VAO had exhibit booths at AAS meetings (Fig. 13). The use of an exhibit booth built on experience gained from NASA Archives and National observatories, for which it was found that substantial fractions of the community could be engaged at low cost. As an illustration of the value of an AAS meeting, people stopping at the exhibit were offered the opportunity to sign up for the VAO mailing list. At each AAS meeting, the size of the VAO mailing list increased by approximately 20%.

### 5.6. VAO Community Days

VAO Community Days were a series of presentations and hands-on activities designed to take the VAO to the community, demonstrate capabilities, develop and encourage new users, and obtain feedback on VO tools and services (Fig. 14). Community Days were typically structured with a morning session led by VAO team members, with the option of an afternoon session for attendees to ask more detailed questions to VAO team members or to bring in their research questions to assess how VO tools and services could assist them. Community Days were aimed initially at locations where there were a large number of astronomers with the goal of making it easy for many to attend. Table 2 lists the VO Community Days that were held. Two VAO Community Days (at the University of Washington and Cornell University) were being planned when the VAO was directed to discontinue them in preparation for its close-out activities.

In addition to the VAO Community Days organized by the VAO, VAO Team Members also participated in similar activities organized by international organizations, including in Italy, Brazil, and Chile.


**Fig. 14.** (Left) Example of an announcement flyer for a VO Community Day. (Right) Scene from a VAO Community Day, in which a VAO team member is demonstrating how a VO tool could be used. Both of these examples are from the Community Day held at the University of Michigan.

**Table 2**
VAO Community Day locations and dates.

| | |
|---|---|
| Center for Astrophysics, Cambridge, MA | 2011-11-30 |
| Caltech, Pasadena, CA | 2011-12-09 |
| U. Arizona, Tucson, AZ | 2012-03-13 |
| U. Michigan, Ann Arbor, MI | 2012-11-14 |
| STScI, Baltimore, MD | 2012-11-27 |

### *5.7. Summer schools*

During the VAO, VAO Team Members participated in summer schools organized by other institutions, often presenting lectures or developing demonstrations. The NVO project hosted four Summer Schools between 2004 and 2008. During these week-long intensive sessions, over 160 participants worked with experienced VO users and software specialists to become familiar with how to discover, access, visualize, and analyze data, and how to use the data publication and high performance computing capabilities of the VO. Those attending were introduced to VO tools and utilities and use them to accomplish a variety of research goals including data mining, multiwavelength research, and time domain astronomy. In the second half of the session small teams created their own VO-enabled data analysis applications. Students were asked to work on team-based projects using VO protocols and software in service of astronomical science. At the end of the school when the projects were presented, Summer School faculty granted awards to the five best projects. Winning projects received financial support to attend and present their work at forthcoming winter AAS meetings.

One NVO Summer School led to the production of the book *The National Virtual Observatory: Tools and Techniques for Astronomical Research* (Graham et al., 2007), which contained the lectures and tutorials from that school. The volume also included a complete set of software libraries and worked examples to guide the astronomer/software developer through the process of developing VO-enabled programs in a variety of programming languages and scripting environments. Several chapters describe research results obtained by participants in the NVO Summer Schools using VO tools and technologies.

## 6. Long-term curation of VAO assets

### *6.1. The VAO repository*

The VAO is making available all its digital assets – including code, documentation, data-bases, reports – through a single Google Services repository, chosen because it is free of charge, stable, and openly accessible.[16] Its existence was announced through venues such as the AAS Newsletter, the IVOA Newsletter, and astronomy blogs and social forums. The code repository will contain all builds of the VAO software components, and all the information needed to build and use them. This content includes build instructions, release history, system requirements, license information, test results, documentation, user guides, and tutorials. The material has a common organization and look-and-feel. Currently, the repository contains builds of the science application codes, the VAO single sign-on and login codes, and the monitoring and validation software. In addition, the repository mirrors snapshots of all software that has been committed to the VAO SVN development repository, via automated weekly up-dates.

The VAO chose not to have a software licensing policy as there will be no organization to enforce it after close-out. The software is therefore released as public domain software, while duly honoring institutional licensing policies and licensing restrictions implied by the licensing of dependent third-party software. Thus, the Iris SED builder developed at SAO is released with an Apache 2.0 license and the cross-comparison code developed at Caltech/IPAC is released with a BSD 3-clause license.

All completed documentation has been posted to the repository,[17] including software documentation, project reports, and outreach material. All project presentations and papers are also available.[18] The VAO YouTube channel, blog, Facebook page and Twitter feed will remain live.

### *6.2. Transition of the VAO infrastructure to the NASA archives*

In response to a Call for Proposals issued by NASA in 2013 August, the NASA archives at STScI (MAST), IPAC (NED, IRSA, NASA Exoplanet Archives) and HEASARC submitted a proposal to sustain the core infrastructure components of the VAO within their "in-guide" budgets, beginning FY 2015 (2014 October 1). That proposal was accepted, and the NASA Archives began their activities to sustain the core VO infrastructure elements. A Project Scientist at HEASARC will coordinate VO activities between archives and report to NASA on VO-related activities.

16 https://sites.google.com/site/usvirtualobservatory/.
17 https://sites.google.com/site/usvirtualobservatory/home/documents.
18 https://sites.google.com/site/usvirtualobservatory/home/documents/publications-presentations.

## 7. The VAO Legacy

The impact of the US VO programs on the international VO can be seen in a number of ways:

- Significant contributions to at least 35 IVOA standards and documents, from the first basic standards and services (VOTable, Simple Cone Search) to sophisticated data models and advanced data access protocols (Table Access Protocol, ObsCore, SIAP Version 2, ...).
- Leadership of numerous IVOA Working Groups and Interest Groups, as well as leadership at the IVOA Executive level.
- A rich infrastructure for data discovery and access, with wide deployment and implementation at major data centers in the US.
- A robust operational environment in which distributed services are routinely validated against IVOA standards.
- A system of resource registries that enables discover of data and data services through the world.
- Exemplar science applications for data discovery, spectral energy distribution construction and analysis, and catalog cross-comparison.
- Desktop scripting tools including a native Python implementation.
- Cloud-based data storage for collaborative research and simple data sharing with the research community.
- Creation of a "data scientist" position at the American Astronomical Society whose responsibilities include "to help process and manage the increasing volume of digital data and to integrate it within the Virtual Observatory".
- A repository of all VAO products: software, documentation, tutorials, videos, news-letters, ....
- An increasing expectation that new telescopes and facilities incorporate VO capabilities during the design of their data management systems (e.g., Mahabal et al., 2012; Graham et al., 2012; Juric et al., 2013; Anderson et al., 2013; Seaman et al., 2014).

However, it is more difficult to measure impact quantitatively. Since the VAO was mostly about the deployment of software tools and infrastructure services, it can be challenging to attribute data accesses to the VAO as opposed to the underlying data services. Web applications are primarily entry points to VO services; scripting environments are needed for bulk processing. In the astronomy community at least, and probably in many other disciplines, new software can take many years to penetrate the community, and even then, there is not a strong culture of software citation. For example, we find that although some 22,000 peer-reviewed papers mention the VLA radio telescope, only 68 formally acknowledge the use of AIPS and only 59 acknowledge use of CASA, the two dominant reduction and analysis packages for radio interferometry data. Remarkably (or perhaps not, given the situation for software citation) of over 13,000 peer-reviewed publications in astronomy and astrophysics published in 2013, only 4% acknowledge use of the ADS (M. Kurtz 2014, private communication) and the ADS is probably the most widely-used software system in the field. Thus, counting acknowledgments to VAO or VO tools is unlikely to reflect accurately on community take-up.

On the other hand, VAO usage logs indicate close to one million VO-based data accesses per month at US data providers, and with ∼100 organizations who have published some 10,000 VO-compliant data services worldwide. VAO usage logs also show some 2000 distinct users of VAO services in the past three months (April–June 2014). The ADS lists over 2500 papers (about half of these peer-reviewed) citing "virtual observatory" in some context, and these papers are read as often and cited as often as other types of papers. Of course, without reading each and every paper one cannot be sure of the level of contamination in this sample (a paper saying "our observatory has photometry measurements of virtually thousands of stars" would count as a hit). A list of ∼100 papers that make explicit use of VO tools and services are listed at http://www.usvao.org/support-community/vo-related-publications/.

The VO concept has been adopted in numerous other fields, particular in space science (with seven VxOs within NASA), plus the Virtual Solar Observatory (NASA, NSO), Planetary Science Virtual Observatory (Europe), and the Deep Carbon Virtual Observatory (Rensselaer Polytechnic Institute). The VO concept was recently endorsed by a panel of neuroscientists convened by the Kavli Foundation and General Electric as a means for improving access and interoperability to the vast data sets being collected in the European Brain Project and US Brain Initiative. VAO and IVOA participants are now playing leading roles in the international Research Data Alliance and the newly formed US National Data Services Consortium.

### 7.1. Lessons learned

In looking back over the VAO project and its NVO predecessor, a number of "lessons learned" is apparent.

- Successful infrastructure is largely invisible and unappreciated. Developing metrics for measuring the success of software infrastructure is a difficult question that reaches across all scientific disciplines. The topic was, for example, discussed in detail at the 2015 NSF Software Infrastructure for Sustained Innovation (SI2) Principal Investigators meeting.[19] We urge scientists and funding agencies to investigate it collectively and develop guidelines for measuring the impact of software infrastructure. More attention should have been paid to explaining the VO infrastructure to the user community and the funding agencies, and we recommend that similar projects make such explanations a priority even in the earliest phases of development.
- Deployment of a distributed infrastructure takes considerable time. Community consensus and buy-in require early and ongoing participation. The VAO team inherited the solutions and approach of the its technology-driven predecessor, the NVO, primarily because the both projects had a common core staff. Consequently, the VAO was slow to engage the user community and deliver services that have value to astronomers in their day-to-day work. The approach eventually used, of bringing the VAO to astronomers through integration into widely used tools in consultation with the community, led to the successful delivery and take-up of the PyVO and VOClient toolkits. Nevertheless, and an earlier start would have led to more advanced and richly-featured services.
- It is important to do marketing to the research/user community, and to manage expectations. Promising too much is as bad or worse than delivering too little. Early promises for VO capabilities were overly ambitious and led to significant skepticism. This ambitious program was also the primary reason why the VAO was in the position of making a substantial number of deliveries in the final three months of the project, precisely when staff are moving on new projects at their home institutions. As a result, some deliveries were snapshots of the code rather the full featured and well documented deliveries. Thus, in addition to managing expectations, we recommend scheduling the majority of deliveries in the earlier phases of a project. Placing all the software in a central public repository ensures that all the code, whether full deliveries or snapshots, is available to the community for further development.

[19] http://cococubed.asu.edu/si2pimeeting2015/index.html.

- The absence of a dedicated test team, led to by a dedicated test engineer, that would be available to support development and execution of test plans across the VAO increased the overhead on managing and organizing testing. This overhead arose because test teams were assembled on-the-fly from available staff, and test plans were consequently begun late in the development phase. We recommend establishing an independent test team at the start of a project, who coordinate with developers throughout the development lifecycle.
- An essential element of the VO is the Registry, within which data providers indicate what services they provide. Initially, an approach of having an easy registration process was adopted, with the consequence that some of the services registered were either of low quality or poorly maintained. It is difficult to achieve the correct balance between easy registration to encourage a substantial Registry and substantial initial quality control that results in a Registry not containing expected services.
- A distributed project has both advantages and disadvantages:

  *Advantage* Access to a diversity of skills and different environments for validating technical approaches and implementations.

  *Disadvantage* Coordination of efforts takes time; staff members have competing priorities as most were not working on VAO full-time.

  For VAO the advantages outweighed the disadvantages, though there were certainly inefficiencies resulting from the distributed nature of the development work. These inefficiencies were minimized by having staff at only two or three organizations responsible for deliveries. For example, staff from SAO, STScI, and IPAC/NED developed Iris. Where appropriate, one organization was responsible for a component. HEASARC managed the operational monitoring system, for example, and NOAO managed the User Support system.
- Setting up an independent management entity such as the VAO, LLC, is a non-trivial effort, though in the VAO case it proved to be worthwhile and effective. Having a dedicated Board of Directors to provide focused advice was a great asset.
- Top-down imposition of standards is likely to fail. Attempts to turn the OpenSkyQuery protocol into an IVOA standard, for example, did not succeed because one group proposed the standard and suggested that everyone else just adopt it.
- Coordination at the international level is essential, but takes time and effort. It can be difficult to reach consensus, or even know if consensus has been reached, owing to different cultures and communications styles.
- Explicit definition of data models is important, even in cases where they seem obvious. Constructing data models after-the-fact leads to having to redefine protocols.
- Metadata collection and curation are essential and ongoing tasks, but complex, and represent a considerable investment. Across the entire VO, resources were never adequate to do a proper job of curation.

## 8. Conclusions

The NVO and VAO, working with international partners, have established the key infrastructure for data discovery, access, and interoperability in astronomy and this infrastructure extends world-wide by virtue of collaboration with the IVOA. This infrastructure is both widely adopted and heavily used, although because of the nature of infrastructure people are often unaware that they are using the VO. The IVOA has also developed a rich body of standards – 45 in all – in the remarkably short period of 12 years, and the international VO efforts remain strong. Through the transfer of VAO assets to NASA, with open source software and documentation, the VAO legacy will be preserved and, we hope, enhanced. The VAO legacy will also be protected through the establishment of the US Virtual Observatory Alliance under the AAS.

## Acknowledgments

The VAO program would not have been possible without the financial support of the National Science Foundation (AST-0834235) and NASA (NNX13AC07G to STScI/MAST), and it was supported NASA/HEASARC. Funding at IPAC has been provided by a grant from the National Aeronautics & Space Administration (NASA) to the Jet Propulsion Laboratory, operated by the California Institute of Technology under contract to NASA. We appreciate the wise guidance of the Board of Directors of the VAO, LLC, and the VAO Science Council, and we are grateful for feedback from the astronomical community that helped us improve our science tools and infrastructure. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Foremost, we acknowledge the dedication, commitment, and excellence of the VAO project team. We brought together the best of the best, from nine different organizations, and through pursuit of common goals created a data management infrastructure that has brought about a sea change in how we manage and share data in astronomy and that has become a model for data management in many other disciplines. We express our gratitude D. De Young (deceased, 2011 December) for his astute guidance throughout the NVO Project and in the initial phases of the VAO.

This research has made use of NASA's Astrophysics Data System Bibliographic Services.

## Appendix A. VAO institutions

The VAO was operated as a limited liability company, funded by the National Science Foundation with coordinated funding provided by the National Aeronautics and Space Administration. Table A.3 lists the institutions engaged in the scientific and technical development work of the VAO; business management was provided by the Associated Universities, Inc. (AUI).

## Appendix B. IVOA standards

This appendix lists International Virtual Observatory Alliance standards and recommendations for which VAO Team Members were identified either as authors or editors. Standards and recommendations are listed in reverse chronological order of adoption.

- "VOTable Format Definition," Version 1.3, IVOA Recommendation, 20 September 2013 (F. Ochsenbein, R. Williams, C. Davenhall, M. Demleitner, D. Durand, P. Fernique, D. Giaretta, R. Hanisch, T. McGlynn, A. Szalay, M. Taylor, A. Wicenec).
- "Data Access Layer Interface," Version 1.0, IVOA Recommendation, 29 November 2013 (P. Dowler, M. Demleitner, M. Taylor, D. Tody).
- "IVOA Registry Relational Schema," Version 1.0, IVOA Proposed Recommendation, 27 February 2014 (M. Demleitner, P. Harrison, M. Molinaro, G. Greene, T. Dower, M. Perdikeas).
- "MOC—HEALPix Multi-Order Coverage map," Version 1.0, IVOA Proposed Recommendation, 10 March 2014 (T. Boch, T. Donaldson, D. Durand, P. Fernique, W. O'Mullane, M. Reinecke, M. Taylor).
- "Simple Application Messaging Protocol," Version 1.3 IVOA Recommendation, 11 April 2012 (M. Taylor, T. Boch, M. Fitzpatrick, A. Allan, J. Fay, L. Paioro, J. Taylor, D. Tody).

**Table A.3**
Participating VAO institutions.

| NSF | NASA |
|---|---|
| California Institute of Technology (Caltech) | High Energy Astrophysics Science Archive Research Center (HEASARC) |
| Johns Hopkins University (JHU) | Infrared Processing and Analysis Center, California Institute of Technology (IPAC) |
| National Center for Supercomputing Applications (NCSA) | Jet Propulsion Laboratory, California Institute of Technology (JPL) |
| National Optical Astronomy Observatory (NOAO) | Space Telescope Science Institute (STScI) |
| National Radio Astronomy Observatory (NRAO) | |
| Smithsonian Astrophysical Observatory (SAO) | |

Institutions are listed according to which agency provided the significant funding for VAO work.

**Table C.4**
VAO Leadership within the IVOA.

| | Position | Individual | Term |
|---|---|---|---|
| Executive committee | Chair | R. Hanisch | 2002 June–2003 July |
| | Deputy Chair | D. De Young | 2006 August–2007 August |
| | Chair | D. De Young | 2007 August–2008 October |
| | Secretary | J. Evans | 2013 September–2014 September |
| Technical Working Group | Chair | R. Williams | 2002 June–2006 July |
| Technical Coordination Group | Chair | R. Williams | 2006 July–2008 May |
| | Deputy Chair | M. Graham | 2012 May–2014 September |
| Inter-operability Conference Program Organizing Committee | Member | R. Hanisch | 2003 March–2007 May |
| | Member | M. Graham | 2012 May–2014 September |
| Standards and Process Subcommittee | Member | R. Hanisch | 2007 September–2010 September |
| Document Coordinator | | S. Emery Bunn | 2010 July–2014 September |
| Applications Working Group | Chair | Tom McGlynn | 2008 July–2011 July |
| | Vice Chair | Tom Donaldson | 2014 May–2014 September |
| Data Access Layer Working Group | Chair | Doug Tody | 2003 June–2007 May |
| | Vice Chair | Mike Fitzpatrick | 2010 May–2013 May |
| Data Models Working Group | Chair | Jonathan McDowell | 2003 June |
| | Vice Chair | Omar Laurino | 2011 May–2014 May |
| | Vice Chair | Omar Laurino | 2014 May–2014 September |
| Grid and Web Services Working Group | Chair | Matthew Graham | 2006 December–2007 May |
| | Chair | Matthew Graham | 2007 May–2011 May |
| Registry Working Group | Chair | Ray Plante | 2006 September–2009 September |
| | Chair | Ray Plante | 2009 November–2010 November |
| | Vice Chair | Gretchen Greene | 2009 November–2010 November |
| | Chair | Gretchen Greene | 2011 January–2014 May |
| Standards and Processes Working Group | Chair | Bob Hanisch | 2003 June–2006 May |
| Uniform Content Descriptors Working Group | Chair | Roy Williams | 2003 June–2005 January |
| VO Event Working Group | Chair | Roy Williams | 2005 January–2008 January |
| | Chair | Rob Seaman | 2006 December–2008 May |
| | Chair | Rob Seaman | 2008 May–2011 May |
| | Vice Chair | Roy Williams | 2010 October–2011 October |
| | Chair | Matthew Graham | 2011 October–2012 October |
| Applications Interest Group | Chair | Tom McGlynn | 2004 January–2005 July |
| Data Curation and Preservation Interest Group | Chair | Bob Hanisch | 2007 May–2010 May |
| | Chair | Alberto Accomazzi | 2010 May–2014 May |
| Knowledge Discovery in Databases Interest Group | Chair | George Djorgovski | 2012 October–2014 September |
| Time Domain Interest Group | Chair | Matthew Graham | 2012 October–2013 May |
| | Vice Chair | Mike Fitzpatrick | 2013 May–2014 September |

The term of the Chair of the Executive Committee was increased to 18 months beginning in 2007 August.
In 2005 July, the Technical Working Group was reformulated as the Technical Coordination Group.
The Chair and Vice Chair of the Data Models Working Group were both granted one year extensions in 2014 May.
The Standards and Processes Working Group was deactivated in 2005 May.
The Uniform Content Descriptors Working Group was renamed to the Semantics Working Group in 2005 October.
The VO Event Working Group was converted to the Time Domain Interest Group in 2012 October.
The Applications Interest Group was converted to the Applications Working Group in 2007 January.

- "Simple Line Access Protocol," Version 1.0, IVOA Recommendation, 09 December 2010 (J. Salgado, P. Osuna, M. Guainazzi, I. Barbarisi, M.-L. Dubernet, D. Tody).
- "Simple Spectral Access Protocol," Version 1.1, IVOA Recommendation, 10 February 2012 (D. Tody, M. Dolensky, J. McDowell, F. Bonnarel, T. Budavari, I. Busko, A. Micol, P. Osuna, J. Salgado, P. Skoda, R. Thompson, F. Valdes, and the Data Access Layer working group).
- "Table Access Protocol," Version 1.0, IVOA Recommendation, 27 March 2010 (P. Dowler, G. Rixon, D. Tody).
- "TAPRegExt: a VOResource Schema Extension for Describing TAP Services," Version 1.0, IVOA Recommendation, 27 August 2012 (M. Demleitner, P. Dowler, R. Plante, G. Rixon, M. Taylor).
- "IVOA Spectral Data Model," Version 2.0, IVOA Proposed Recommendation, 09 March 2014 (J. McDowell, D. Tody, T. Budavari, M. Dolensky, I. Kamp, K. McCusker, P. Protopapas, A. Rots, R. Thompson, F. Valdes, P. Skoda, B. Rino, S. Derriere, J. Salgado, O. Laurino, and the IVOA Data Access Layer and Data Model Working Groups).
- "Observation Data Model Core Components and its Implementation in the Table Access Protocol," Version 1.0, IVOA

Recommendation, 28 October 2011 (M. Louys, F. Bonnarel, D. Schade, P. Dowler, A. Micol, D. Durand, D. Tody, L. Michel, J. Salgado, I. Chilingarian, B. Rino, J. de Dios Santander, P. Skoda).

- "VOSpace Specification," Version 2.0, IVOA Recommendation, 29 March 2013 (M. Graham, D. Morris, G. Rixon, P. Dowler, A. Schaaff, D. Tody).
- "IVOA Credential Delegation Protocol," Version 1.0, IVOA Recommendation, 18 February 2010 (M. Graham, R. Plante, G. Rixon, G. Taffoni).
- "Web Services Basic Profile," Version 1.0, IVOA Recommendation, 16 December 2010 (A. Schaaff, M. Graham).
- "StandardsRegExt: a VOResource Schema Extension for Describing IVOA Standards," Version 1.0, IVOA Recommendation, 08 May 2012 (P. Harrison, D. Burke, R. Plante, G. Rixon, D. Morris, and the IVOA Registry Working Group).
- "Describing Simple Data Access Services," Version 1.0, IVOA Recommendation, 25 November 2013 (R. Plante, J. Delago, P. Harrison, D. Tody, and the IVOA Registry Working Group).
- "VODataService: a VOResource Schema Extension for Describing Collections and Services," Version 1.1, IVOA Recommendation, 02 December 2010 (R. Plante, A. Stébé, K. Benson, P. Dowler, M. Graham, G. Greene, P. Harrison, G. Lemson, T. Linde, G. Rixon).
- "IVOA Registry Relational Schema," Version 1.0, IVOA Proposed Recommendation, 27 February 2014 (M. Demleitner, P. Harrison, M. Molinaro, G. Greene, T. Dower, M. Perdikeas).
- "IVOA Document Standards," Version 1.2, IVOA Recommendation, 13 April 2010 (R.J. Hanisch, C. Arviset, F. Genova, B. Rino).
- "Sky Event Reporting Metadata," Version 2.0, IVOA Recommendation, 11 July 2011 (R. Seaman, R. Williams, A. Allan, S. Barthelmy, J. Bloom, J. Brewer, R. Denny, M. Fitzpatrick, M. Graham, N. Gray, F. Hessman, S. Marka, A. Rots, T. Vestrand, P. Wozniak).
- "IVOA Support Interfaces," Version 1.0, IVOA Recommendation, 31 May 2011 (Grid and Web Services Working Group, M. Graham, G. Rixon).

## Appendix C. International Virtual Observatory Alliance leadership

Table C.4 lists VAO Team members who served in various leadership positions within the IVOA.

## References


Anderson, K.R., Rosolowsky, E., Dowler, P., 2013. CyberSKA radio imaging metadata and VO compliance engineering. In: Friedel, D.N. (Ed.), Astronomical Data Analysis Software and Systems XXII. In: Astronomical Society of the Pacific Conference Series, vol. 475. p. 231.

Arviset, C., Gaudet, S., The IVOA Technical Coordination Group, November 2010. IVOA Architecture, Version 1.0. International Virtual Observatory Alliance. URL: http://www.ivoa.net/documents/Notes/IVOAArchitecture/.

Astropy Collaboration, Robitaille, T.P., Tollerud, E.J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A.M., Kerzendorf, W.E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M.M., Nair, P.H., Unther, H.M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J.E.H., Singer, L., Fox, R., Weaver, B.A., Zabalza, V., Edwards, Z.I., Azalee Bostroem, K., Burke, D.J., Casey, A.R., Crawford, S.M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P.L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M., Streicher, O., 2013. Astropy: A community Python package for astronomy. A&A 558, A33.

Berriman, G.B., Good, J.C., Curkendall, D.W., Jacob, J.C., Katz, D.S., Prince, T.A., Williams, R., 2003. Montage: an on-demand image mosaic service for the NVO. In: Payne, H.E., Jedrzejewski, R.I., Hook, R.N. (Eds.), Astronomical Data Analysis Software and Systems XII. In: Astronomical Society of the Pacific Conference Series, vol. 295. p. 343.

Boch, T., Donaldson, T., Durand, D., Fernique, P., O'Mullane, W., Reinecke, M., Taylor, M., June 2014. MOC—HEALPix Multi-Order Coverage map, Version 1.0. International Virtual Observatory Alliance, iVOA Recommendation. URL: http://www.ivoa.net/twiki/bin/view/IVOA/IvoaApplications.

Bonnarel, F., Fernique, P., Bienaymé, O., Egret, D., Genova, F., Louys, M., Ochsenbein, F., Wenger, M., Bartlett, J.G., 2000. The ALADIN interactive sky atlas. A reference tool for identification of astronomical sources. A&AS 143, 33–40.

Brunner, R.J., Djorgovski, S.G., Szalay, A.S. (Eds.), 2001. Virtual Observatories of the Future. In: Astronomical Society of the Pacific Conference Series, vol. 225.

Busko, I., 2002. Specview: a Java tool for spectral visualization and model fitting of multi-instrument data. In: Starck, J.-L., Murtagh, F.D. (Eds.), Astronomical Data Analysis II. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 4847. pp. 410–418.

CASA Consortium, 2011. CASA: Common Astronomy Software Applications. Astrophysics Source Code Library.

Davenhall, A.C., Qin, C.L., Noddle, K.T., Walton, N.A., 2004. The astroGrid MySpace system. In: Ochsenbein, F., Allen, M.G., Egret, D. (Eds.), Astronomical Data Analysis Software and Systems (ADASS) XIII. In: Astronomical Society of the Pacific Conference Series, vol. 314. p. 330.

Demleitner, M., Greene, G., Le Sidaner, P., Plante, R.L., 2014. The virtual observatory registry. Astron. Comput. 7, 101–107. doi:10.1016/j.ascom.2014.07.001.

Doe, S., Bonaventura, N., Busko, I., D'Abrusco, R., Cresitello-Dittmar, M., Ebert, R., Evans, J., Laurino, O., McDowell, J., Pevunova, O., Refsdal, B., 2012. Iris: the VAO SED application. In: Ballester, P., Egret, D., Lorente, N.P.F. (Eds.), Astronomical Data Analysis Software and Systems XXI. In: Astronomical Society of the Pacific Conference Series, vol. 461. p. 893.

Doe, S., Nguyen, D., Stawarz, C., Siemiginowska, A., Burke, D., Evans, I., Evans, J., McDowell, J., Refsdal, B., Houck, J., Nowak, M., 2006. Sherpa: goals and design for chandra and beyond. In: Gabriel, C., Arviset, C., Ponz, D., Enrique, S. (Eds.), Astronomical Data Analysis Software and Systems XV. In: Astronomical Society of the Pacific Conference Series, vol. 351. p. 77.

Dowler, P., Tody, D., Bonnarel, F., Plante, R., July 2014. Simple Image Access Version 2.0. International Virtual Observatory Alliance. URL: http://www.ivoa.net/documents/SIA/.

Evans, J.D., Plante, R.L., Boneventura, N., Busko, I., Cresitello-Dittmar, M., D'Abrusco, R., Doe, S., Ebert, R., Laurino, O., Pevunova, O., Refsdal, B., Thomas, B., 2012. Managing distributed software development in the Virtual Astronomical Observatory. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 8449. p. 84490I.

Freeman, P., Doe, S., Siemiginowska, A., 2001. Sherpa: a mission-independent data analysis application. In: Starck, J.-L., Murtagh, F.D. (Eds.), Astronomical Data Analysis. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 4477. pp. 76–87.

Geißler, K., Metchev, S., Kirkpatrick, J.D., Berriman, G.B., Looper, D., 2011. A Cross-match of 2MASS and SDSS. II. Peculiar L Dwarfs, Unresolved Binaries, and the Space Density of T Dwarf Secondaries. Astrophys. J. 732, 56.

Graham, M.J., Djorgovski, S.G., Donalek, C., Drake, A.J., Mahabal, A.A., Plante, R.L., Kantor, J., Good, J.C., 2012. Connecting the time domain community with the Virtual Astronomical Observatory. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 8448. p. 84480P.

Graham, M.J., Fitzpatrick, M.J., McGlynn, T.A. (Eds.), 2007. The National Virtual Observatory: Tools and Techniques for Astronomical Research. In: Astronomical Society of the Pacific Conference Series, vol. 382.

Graham, M., Morris, D., Rixon, G., Dowler, P., Schaaff, A., Tody, D., Major, B., 2014. VOSpace Service Specification, Version 2.1. International Virtual Observatory Alliance. URL: http://www.ivoa.net/documents/VOSpace/.

Jaeger, S., 2008. The Common Astronomy Software Application (CASA). In: Argyle, R.W., Bunclark, P.S., Lewis, J.R. (Eds.), Astronomical Data Analysis Software and Systems XVII. In: Astronomical Society of the Pacific Conference Series, vol. 394. p. 623.

Jeschke, E., Inagaki, T., Kackley, R., 2013. Introducing the ginga FITS viewer and toolkit. In: Friedel, D.N. (Ed.), Astronomical Data Analysis Software and Systems XXII. In: Astronomical Society of the Pacific Conference Series, vol. 475. p. 319.

Juric, M., Kantor, J., Axelrod, T.S., Dubois-Felsmann, G.P., Becla, J., Lim, K., LSST Collaboration, , LSST Science Collaborations, , 2013. LSST data products: enabling LSST science. In: American Astronomical Society Meeting Abstracts, vol. 221. p. #247.01.

Kinne, R.C., Templeton, M.R., Henden, A.A., Zografou, P., Harbo, P., Evans, J., Rots, A.H., LAZIO, J., 2013. Distributing variable star data to the virtual observatory. In: American Astronomical Society Meeting Abstracts, vol. 221. p. #240.37.

Laurino, O., Budynkiewicz, J., Busko, I., Cresitello-Dittmar, M., D'Abrusco, R., Doe, S., Evans, J., Pevunova, O., 2014a. Iris: constructing and analyzing spectral energy distributions with the virtual observatory. In: Manset, N., Forshay, P. (Eds.), Astronomical Society of the Pacific Conference Series, 485. p. 19.

Laurino, O., Budynkiewicz, J., D'Abrusco, R., Bonaventura, N., Busko, I., Cresitello-Dittmar, M., Doe, S.M., Ebert, R., Evans, J.D., Norris, P., Pevunova, O., Refsdal, B., Thomas, B., Thompson, R., 2014b. Iris: an extensible application for building and analyzing spectral energy distributions. Astron. Comput.

Laurino, O., Busko, I., Cresitello-Dittmar, M., D'Abrusco, R., Doe, S., Evans, J., Pevunova, O., 2013. Extending iris: the VAO SED analysis tool. In: Friedel, D.N. (Ed.), Astronomical Data Analysis Software and Systems XXII. In: Astronomical Society of the Pacific Conference Series, vol. 475. p. 295.

Laurino, O., D'Abrusco, R., Cresitello-Dittmar, M., Evans, J., McDowell, J., 2012. SedImporter: a tool and an extensible framework for constructing interoperable spectral energy distribution data files. In: Ballester, P., Egret, D., Lorente, N.P.F. (Eds.), Astronomical Data Analysis Software and Systems XXI. In: Astronomical Society of the Pacific Conference Series, vol. 461. p. 391.

Mahabal, A., Kembhavi, A., Williams, R., 2011. Astronomy with cutting-edge ICT: from transients in the sky to data over the continents (India-US). In: Proceedings of the Asia-Pacific Advanced Network, vol. 32. p. 143. doi:10.7125/APAN.32.18.

McGlynn, T.A., 2007. Web-based tools—using the virtual observatory datascope tool. In: Graham, M.J., Fitzpatrick, M.J., McGlynn, T.A. (Eds.), Astronomical Society of the Pacific Conference Series, vol. 382. p. 51. (Chapter 6).

McKee, C.F., Taylor Jr., J.H., Hollenbach, D.J., Boroson, T., Freedman, W., Jewitt, D.C., Kahn, S.M., Moran Jr., J.M., Nelson, J.E., 2001. Astronomy and Astrophysics in the New Millennium. National Academy Press.

Mishin, D., Medvedev, D., Szalay, A.S., Plante, R., Graham, M., 2014. Data Sharing and Publication Using the SciDrive Service. In: Manset, N., Forshay, P. (Eds.), Astronomical Data Analysis Software and Systems XXIII. In: Astronomical Society of the Pacific Conference Series, vol. 485. p. 465.

Mockapetris, P., November 1987. RFC 1034: Domain Names—Concepts and Facilities. The Internet Engineering Task Force. URL: http://tools.ietf.org/html/rfc1034.

National Optical Astronomy Observatories, 1999. IRAF: Image Reduction and Analysis Facility. Astrophysics Source Code Library.

O'Mullane, W., Gray, J., Li, N., Budavári, T., Nieto-Santisteban, M.A., Szalay, A.S., 2004. Batch query system with interactive local storage for SDSS and the VO. In: Ochsenbein, F., Allen, M.G., Egret, D. (Eds.), Astronomical Data Analysis Software and Systems (ADASS) XIII. In: Astronomical Society of the Pacific Conference Series, vol. 314. p. 372.

Plante, R., Yekkirala, V., Baker, W., 2012. Enabling OpenID Authentication for VO-integrated Portals. In: Ballester, P., Egret, D., Lorente, N.P.F. (Eds.), Astronomical Data Analysis Software and Systems XXI. In: Astronomical Society of the Pacific Conference Series, vol. 461. p. 423.

Quinn, P.J., Górski, K.M. (Eds.), 2004. Toward an International Virtual Observatory.

Salgado, J., Osuna, P., Guainazzi, M., Barbarisi, I., Dubernet, M.-L., Tody, D., December 2010. Simple Line Access Protocol Version 1.0. International Virtual Observatory Alliance. URL: http://www.ivoa.net/documents/SLAP/.

Seaman, R.L., Vestrand, W.T., Hessman, F.V., 2014. Reengineering observatory operations for the time domain. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 9149. p. 914906.

Taylor, M.B., 2005. TOPCAT & STIL: Starlink Table/VOTable Processing Software. In: Shopbell, P., Britton, M., Ebert, R. (Eds.), Astronomical Data Analysis Software and Systems XIV. In: Astronomical Society of the Pacific Conference Series, vol. 347. p. 29.

Taylor, M., 2011. TOPCAT: Tool for OPerations on Catalogues And Tables. Astrophysics Source Code Library.

Taylor, M.B., Boch, T., Fay, J., Fitzpatrick, M., Paioro, L., 2012b. SAMP: application messaging for desktop and Web applications. In: Ballester, P., Egret, D., Lorente, N.P.F. (Eds.), Astronomical Data Analysis Software and Systems XXI. In: Astronomical Society of the Pacific Conference Series, vol. 461. p. 279.

Taylor, M., Boch, T., Fitzpatrick, M., Allan, A., Fay, J., Paioro, L., Taylor, J., Tody, D., April 2012a. IVOA recommendation: simple application messaging protocol, Version 1.3. International Virtual Observatory Alliance. URL: http://www.ivoa.net/twiki/bin/view/IVOA/IvoaApplications.

Tody, D., 1986. The IRAF Data Reduction and Analysis System. In: Crawford, D.L. (Ed.), Instrumentation in astronomy VI. In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 627. p. 733.

Tody, D., 1993. IRAF in the Nineties. In: Hanisch, R.J., Brissenden, R.J.V., Barnes, J. (Eds.), Astronomical Data Analysis Software and Systems II. In: Astronomical Society of the Pacific Conference Series, vol. 52. p. 173.

Tody, D., Dolensky, M., McDowell, J., Bonnarel, F., Budavari, T., Busko, I., Micol, A., Osuna, P., Salgado, J., Skoda, P., Thompson, R., Valdes, F., Data Access Layer Working Group, February 2012. Simple Spectral Access Protocol Version 1.1. International Virtual Observatory Alliance. URL: http://www.ivoa.net/documents/SSA/.

Tody, D., Plante, R., November 2009. Simple Image Access Specification Version 1.0. International Virtual Observatory Alliance. URL: http://www.ivoa.net/documents/SIA/20091116/.

White, R.R., Allan, A., Barthelmy, S., Bloom, J., Graham, M., Hessman, F.V., Marka, S., Rots, A., Scholberg, K., Seaman, R., Stoughton, C., Vestrand, W.T., Williams, R., Wozniak, P.R., 2006. Astronomical network event and observation notification. Astron. Nachr. 327, 775.

Williams, R., Hanisch, R., Szalay, A., Plante, R., February 2008. Simple Cone Search Version 1.03. International Virtual Observatory Alliance. URL: http://www.ivoa.net/documents/latest/ConeSearch.html.