ASTRONOMICAL VIRTUAL OBSERVATORIES THROUGH INTERNATIONAL COLLABORATION

Masatoshi Ohishi

Astronomy Data Center, National Astronomical Observatory of Japan, 2-21-1, Osawa, Mitaka, Tokyo 181-8588 Japan
Email: masatoshi.ohishi@nao.ac.jp

ABSTRACT

Astronomical Virtual Observatories (VOs) are emerging research environment for astronomy, and 16 countries and a region have funded to develop their VOs based on international standard protocols for interoperability. The 16 funded VO projects have established the International Virtual Observatory Alliance (http://www.ivoa.net/) to develop the standard interoperable interfaces such as registry (meta data), data access, query languages, output format (VOTable), data model, application interface, and so on. The IVOA members have constructed each VO environment through the IVOA interfaces. National Astronomical Observatory of Japan (NAOJ) started its VO project (Japanese Virtual Observatory - JVO) in 2002, and developed its VO system. We have succeeded to interoperate the latest JVO system with other VOs in the USA and Europe since December 2004. Observed data by the Subaru telescope, satellite data taken by the JAXA/ISAS, etc. are connected to the JVO system. Successful interoperation of the JVO system with other VOs means that astronomers in the world will be able to utilize top-level data obtained by these telescopes from anywhere in the world at anytime. System design of the JVO system, experiences during our development including problems of current standard protocols defined in the IVOA, and proposals to resolve these problems in the near future are described.

Keywords: Astronomy, Virtual Observatory, Standardized Interface, International Collaboration, Database, GRID

1 INTRODUCTION

The National Astronomical Observatory of Japan (NAOJ) operates the Subaru telescope in Hawaii and the large millimeter-wave telescopes in Nobeyama. All the observed data are digitally archived and are accessible via internet. The radio telescopes of Nobeyama produce about 1 TBytes per year, and the Subaru telescope outputs about 20 TBytes per year. NAOJ has joined the international project, ALMA, and it is foreseen that more than a few PBytes of data per year will be produced for the astronomy community.

Because astronomical objects radiate electromagnetic waves in wide frequency range, it has been recognized that multi-wavelength analyses are essential to understand physical and chemical behavior of galaxies, stars, planets and so on. The idea of the Virtual Observatory (VO) has recently appeared to resolve such situation, and the system has been developed in 15 countries and one region around the world. These individual VO projects have established an alliance, the International Virtual Observatory Alliance (IVOA; http://www.ivoa.net/), to define standardized protocols for their interoperations.

Japanese Virtual Observatory (JVO) is designed to seamlessly access to multi-wavelength, federated databases and data analyses systems for astronomers through high speed network facility. The basic concept and a new query language to access to the distributed databases, JVO Query Language, are described in Ohishi et al. (2006).

Since JVO is a member of the IVOA, and have adopted many recommendations in implementing its portal system. Thus the JVO portal will provide access to more than 1,900 astronomical resources worldwide, as of January 2009. Further JVO has been collaborating with the Japanese national research Grid initiative (NAREGI; http://www.naregi.org/index_e.html) toward utilizing distributed data analysis environment. This paper describes the current status of the JVO portal system and an experiment to federate several servers in NAOJ and KEK (Institute for High Energy Accelerator Research Organization, located in Tsukuba, Japan) by means of the NAREGI middleware.

2 VIRTUAL OBSERVATORY IN ASTRONOMY

IVOA was formed in June 2002 with a mission to "*facilitate the international coordination and collaboration necessary for the development and deployment of the tools, systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory.*" The IVOA now comprises 16 VO projects from Armenia, Australia, Canada, China, Europe, France, Germany, Hungary, India, Italy, Japan, Korea, Russia, Spain, the United Kingdom, and the United States of America. Since the resources are distributed in many countries of the world, it is crucial to develop standard interfaces for interoperation among individual "national" virtual observatories, which, then, construct an International Virtual Astronomical Observatory.

The work of the IVOA focuses on the development of standard interfaces to federate distributed astronomical resources in the member countries. Figure 1 shows a schematic diagram of the architecture of the virtual observatories in astronomy, taken from Williams et al. (2004). It would be easily understood that IVOA has been developing the middle layer of the three-tiered architecture, that interconnects the physical resources (databases, disks, storages, CPUs, etc.) and the VO users who access to the VOs through portals, VO tools, etc.

The Registry Services play very important roles in the astronomical virtual observatories since they are used to discover wanted resources as well as in publishing resources. IVOA has followed the format on meta-data defined in the Open Archival Initiative of Protocol for Metadata Harvesting (OAI-PMH) to describe and exchange meta-data on astronomical resources. The Data Services provide VO users with accesses to the wanted astronomical data (images, spectra, catalogues, time-series data, simulation data, etc.). The Data Services consist of Simple Image Access Protocol (SIAP) for image data, Simple Spectrum Access Protocol (SSAP) for spectrum and time-series data, and Astronomical Data Query Language (ADQL) to access table data (catalogue data). SIAP and SSAP are based on the HTTP protocol, while ADQL is an extension to the SQL. And the Compute Services provide access to astronomical data analysis tools that are used to derive physical parameters from the retrieved data (position of a celestial object, size, brightness of the object, and simulation services, etc.).

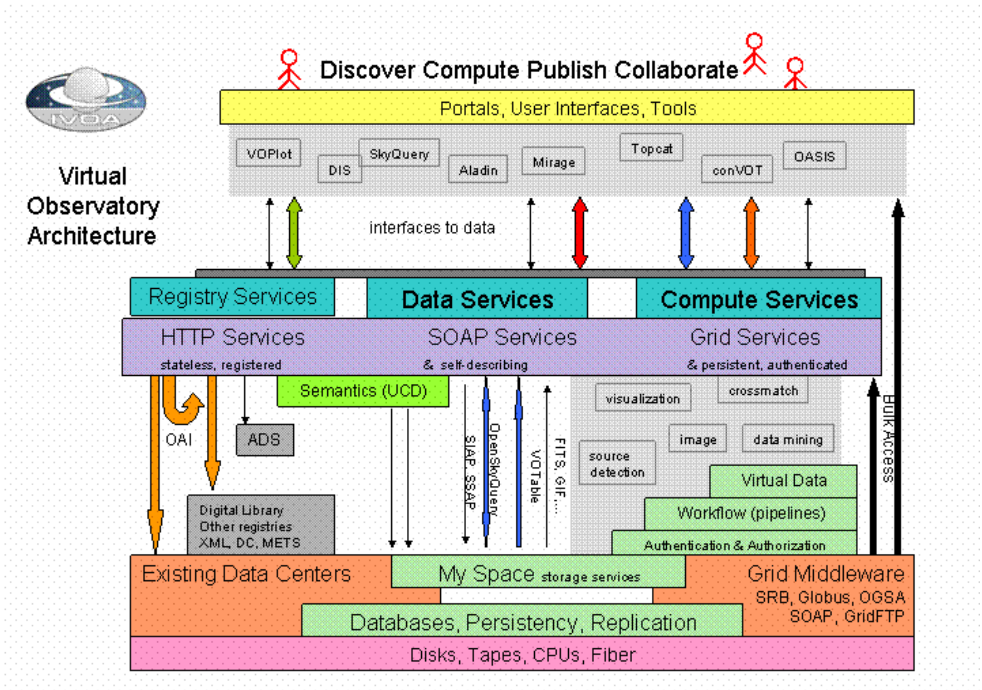

Figure 1. Architecture of Virtual Observatories in Astronomy to Federate Distributed Resources (Williams et al., 2004).

2.1 Architecture of the JVO Portal System

Schematic diagram of the JVO portal system is shown in Figure 2. Its fundamental design is similar to our previous prototype systems (Tanaka et al., 2005; Ohishi et al., 2006) and the one defined in the IVOA (Williams et al., 2004). In this system we have adopted the Web Services, standardized protocols for the VOs (SIAP for images, SSAP for spectra, and ADQL for catalogs), and introduction of resource metadata exchange mechanism based on the OAI-PMH.

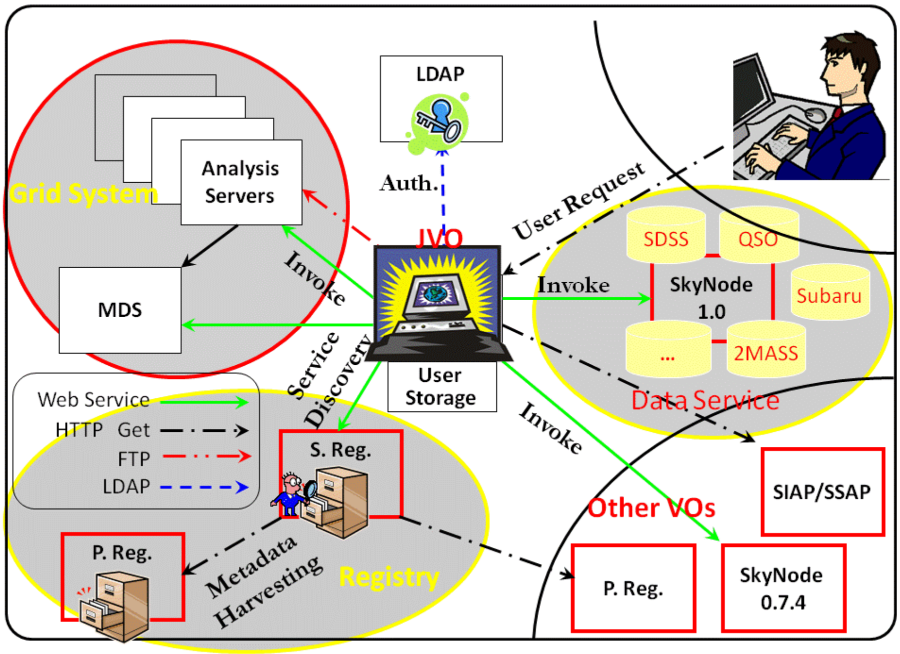

Figure 2. Architecture of the JVO Portal System.

Queries issued by a user are sent from the user terminal to the JVO portal, and are parsed into single queries to appropriate servers that are found by consulting the registry. The parsed querying processes are executed by the scheduler, being passed to individual servers through appropriate protocol. The query results are sent back with the VOTable format that has been defined by the IVOA by means of XML.

Major database servers connected to this prototype are Suprime-CAM (highly sensitive wide-field camera) from Subaru telescope, ASCA (X-ray satellite) database operated by the Institute of Space and Aeronautical Science, Japan Aerospace Exploration Agency (JAXA /ISAS), and others. We succeeded to interoperate the JVO prototype with VOs in the North America (NVO, VO-Canada) and Europe (European Southern Observatory, European Space Astronomy Centre, AstroGrid in the UK, Strasbourg Data Centre (CDS)) in December 2004, and more than 1,900 astronomical resources are accessible as of January 2009.

The protocols to access images, spectra and catalogs have been developed separately. Therefore it is necessary to prepare separate query interface to access to each resource. Astronomical researches may sometimes utilize images, spectra and catalogs simultaneously, and it is desirable to use an integrated query interface to make such queries easier. (Shirasaki et al., 2005) proposed such an integrated query language by using the “virtual column” concept. We implemented the integrated query language interface to the interoperable JVO prototype, and succeeded to retrieve images/spectra/catalogs though a single interface. Further details of the JVO system is described in Shirasaki et al. (2009), and actual scientific results obtained by using virtual observatories in astronomy are summarized by Tanaka et al. (2009).

Figure 3 shows an access statistics to the JVO portal system, as of January 2009. The figure has two statistics: downloaded data size per month in kBytes (in histogram) and total requested pages per month (in a graph). Note that the ordinates are the logarithmic scales. When we opened the Subaru Suprime-CAM on-demand mosaicking service, the downloaded data size has dramatically increased. This suggests that there has been a high potential to use the Subaru data in many countries.

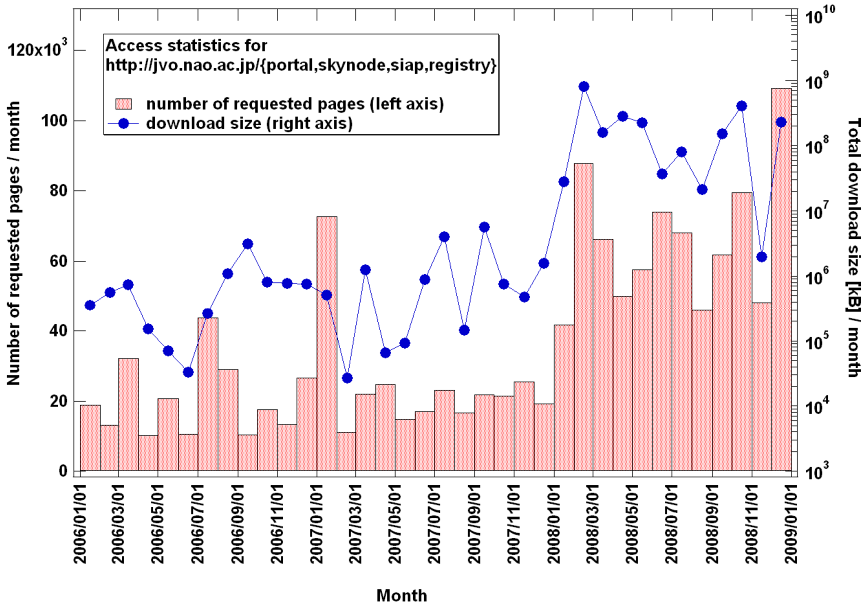

Figure 3. Access Statistics of the JVO Portal System, as of January 2009.

2.2 Virtual Integration of Multiple Catalogues

Since we started experimental operations of the JVO system, we noticed that several databases, such as Subaru databases, Sloan Digital Sky Survey (SDSS), the Two Micron All Sky Survey (2MASS), and so on, are frequently accessed. Since the total records in these databases count up to about 10^9, this meant that the accesses to these databases could be the bottle-neck of the JVO performance. Therefore we developed an integrated, quick "*Multiple Catalog Search*" system through designing a common metadata table which consists of position, wavelength (frequency) and intensity (flux) extracted from individual databases together with pointers (links) for further detailed information (Tanaka et al., 2008).

We also utilized the Table-partitioning by means of the HTM (Hierarchical Triangular Mesh) with 32,768 upper levels. After implementation we made performance measurements, and found that our method is more efficient than the PostgreSQL by a factor of 150 at maximum (Tanaka et al., 2008) (see Figure 4).

Search radius	Result objects	Elaspted time (sec)			# of HTM conditions	
arcmin	#	Postgre SQL	Our method	ratio	Postgre SQL	Our method
1	2	6.460	0.042	154	32	32
10	165	3.807	0.030	127	16	16
60	6697	6.468	0.107	60	32	32
100	26720	2.016	0.307	6.6	4	16
180	57246	9.044	0.709	12.8	48	72

Figure 4. Results of Performance Measurement and Comparison with the PostgreSQL.

2.3 NAREGI Federation between NAOJ and KEK

NAREGI, the National Research Grid Initiative, was created in 2003 by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan. From 2006 through 2007, the research and the development were continued under the "Science Grid NAREGI" Program of the "*Development and Application of Advanced High-performance Supercomputer project*" being promoted by MEXT. NAREGI aimed to make fundamental building blocks in the Cyber Science Infrastructure (CSI), which has been operated by the National Institute of Informatics (NII) in Japan, and its goal is to provide a large-scale computing environment for widely-distributed, advanced research and education (the Science Grid). The NAREGI Grid Middleware Ver. 1.0 has been released in May 2008. NII endeavors to build the Grid infrastructure by continuing software maintenance and user support services.

Prior to the official release of the NAREGI Grid Middleware Ver. 1.0, we were invited to use its β version to evaluate its functionalities and usages. Thus we constructed a test bed together with the KEK, which is shown in Figure 5. Because past NAREGI test beds were constructed among the NAREGI project member institutes through the VPN connections, we intended not to use the VPN connection since the Grid middleware would be used, in most cases, without the VPN connections. The test bed consists of several Linux servers in the NAOJ site and the KEK site. The servers in the NAOJ site were the quad core machines, and a few NAREGI nodes were constructed on a single CPU by means of the VMware environment. On the KEK site each node was constructed on separate machines. Both sites were connected through the firewall to simulate the real use of the NAREGI middleware.

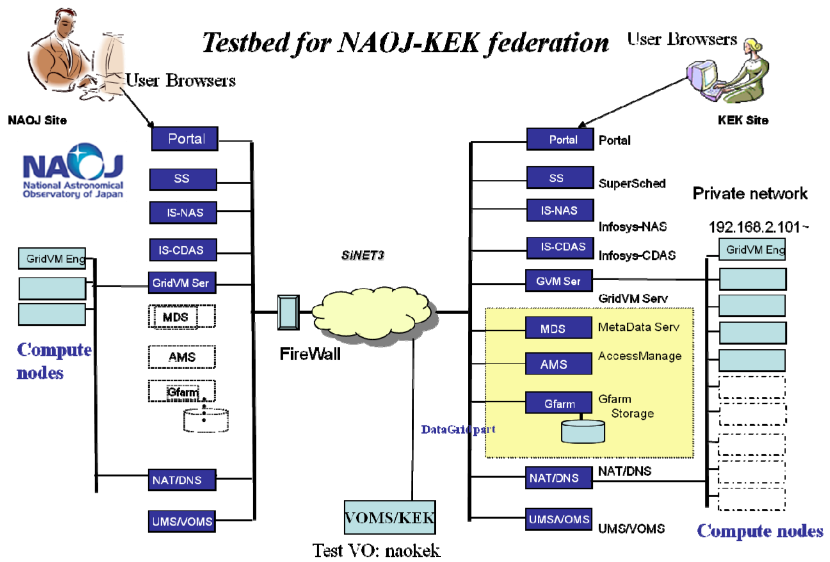

Figure 5. Configuration of the Federation Experiment.

On the computing nodes in NAOJ and KEK we installed some astronomy libraries to perform data reduction of the Subaru Suprime-CAM images. A single image from the Suprime-CAM has the images size of 160 MBytes, and 17 image files were used for a single data reduction. These image files were transferred by means of the GridFTP command from the NAOJ site to the KEK site, and the data reduction jobs were submitted from the NAOJ site to the KEK site through the NAREGI super-scheduler. A single data reduction took almost 10 hours because there are many CPU-heavy steps, such as flat, correction of image distortion, astrometry (measurement of star positions), mosaicking, and coadding all the images). The resultant images were, then, transferred from the KEK site to the NAOJ site by using the GridFTP command, and visualization software on the NAOJ site was used to show the final images. We experienced several problems during the installation processes, the data reduction process went quite well. For example we were able to detect more than 50,000 objects in a single mosaicked image (see Figure 6).

The problems we found during our federation experiment were reported to the NAREGI project, which were already fixed before the official release of the NAREGI Grid Middleware Ver. 1.0.

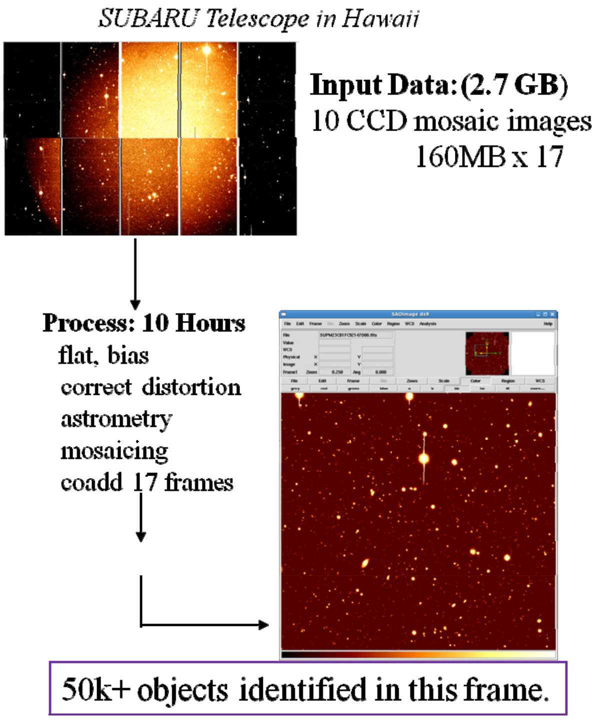

Figure 6. Data Processing Flow of the Federation Experiment.

3 SUMMARY

We have opened the JVO system to the astronomy community in March 2008, meaning that it was established to access data servers via VO interfaces. Thus it is necessary for astronomical researches to analyze obtained data on the VO environment. We have been designing a work flow language to access remotely located analysis engines. The work flow language will enable users not only data queries but data analyses through a single user interface (Tanaka et al., 2006).

Then we plan to implement Single-Sign-On mechanism for secure user access, and other standardized interfaces by the IVOA, and to disseminate standard VO interfaces to astronomy communities not only in Japan but in (East) Asian countries to further promote international collaborations. Such activity is crucial for the ALMA era that we expect to produce more than a few PBytes per year.

4 ACKNOWLEDGEMENTS

This work was supported by the Core-to-Core Program of the Japan Society for the Promotion of Science (JSPS), the Grant-in-Aid for Scientific Research in Priority Area (KAKENHI, 18049074 and 19024070) carried out by the Ministry of Education, Culture, Sports, Science and Technology of Japan, and the CSI project promotion program by the National Institute of Informatics.

5 REFERENCES

Ohishi, M., et al., (2006), in ASP Conf. Ser. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, & E. Solano (San Francisco: ASP), 375

Shirasaki, Y., et al., (2005), in ASP Conf. Ser. 347, ADASS XIV, ed. P. Shopbell, M. Britton, and R. Ebert (San Francisco: ASP), 262

Shirasaki, Y., Tanaka, M., Ohishi, M., Mizumoto, Y., Kawanomoto, S., Honda, S., et al. (2009), Application of Information Technologies on Astronomy: Japanese Virtual Observatory (JVO) Portal. *Proceedings of International Symposium: Fifty Years after IGY*, Tsukuba, Japan

Tanaka, M., et al., (2005), in ASP Conf. Ser. 347, ADASS XIV, ed. P. Shopbell, M. Britton, and R. Ebert (San Francisco: ASP), 267

Tanaka, M., et al., (2006), in ASP Conf. Ser. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, & E. Solano (San Francisco: ASP), 460

Tanaka, M. et al. (2008), in ASP Conf. Ser. 394, ADASS XVII, ed. J. Lewis, R. Argyle, P. Bunclark, D. Evans, & E. Gonzalez-Solares (San Francisco: ASP), 261

Tanaka, M. Shirasaki, Y., Ohishi, M., Mizumoto, Y., Kawanomoto, S., Yasuda, N., & Honda, S. (2009), Astronomical Research using Virtual Observatories, *Proceedings of International Symposium: Fifty Years after IGY*, Tsukuba, Japan

Williams, R., Hanisch, R., Linde, A., McDowell, J., Moore, R., Ochsenbein, F., Ohishi, M., Rixon, G., Szalay, A., & Tody, D. (2004) Virtual Observatory Architecture Overview: Retrieved from the World Wide Web, March 6, 2010: http://www.ivoa.net/Documents/Notes/IVOArch/IVOArch-20040615.html