ASTRONOMICAL RESEARCH USING VIRTUAL OBSERVATORIES

M. Tanaka\textsuperscript{1}*, Y. Shirasaki\textsuperscript{1}, M. Ohishi\textsuperscript{1}, Y. Mizumoto\textsuperscript{1}, S. Kawanomoto\textsuperscript{1}, N. Yasuda\textsuperscript{2}, and S. Honda\textsuperscript{3}

\textsuperscript{1}National Astronomical Observatory of Japan
Email: masahiro.tanaka@nao.ac.jp
\textsuperscript{2}Institute for Cosmic Ray Research, The University of Tokyo
\textsuperscript{3}Gunma Astronomical Observatory

ABSTRACT

The Virtual Observatory (VO) for Astronomy is a framework that empowers astronomical research by providing standard methods to find, access, and utilize astronomical data archives distributed around the world. VO projects in the world have been strenuously developing VO software tools and/or portal systems. Interoperability among VO projects has been achieved with the VO standard protocols defined by the International Virtual Observatory Alliance (IVOA). As a result, VO technologies are now used in obtaining astronomical research results from a huge amount of data. We describe typical examples of astronomical research enabled by the astronomical VO, and describe how the VO technologies are used in the research.

Keywords: Astronomical Virtual Observatory, Database, Brown dwarf, AGN, Quasar, Galaxy

1 INTRODUCTION

Advanced observational instruments and homogeneous survey programs have produced various kinds of huge and high-quality data archives, which are now indispensable for studies in astronomy, especially for statistical studies using multi-wavelength data ranging from radio to gamma ray regions. Utilizing such huge data archives requires the aid of new information technologies. The Virtual Observatory (VO) for Astronomy (Ohishi et al. 2009) aims at providing the frameworks of standard methods to find, access, and utilize astronomical data archives distributed around the world. The VO framework, developed by 16 national and international projects, provides standard interfaces (Quinn et al. 2004). The interoperability among VO projects has been achieved with the VO standard protocols defined by the International Virtual Observatory Alliance (IVOA) (Hanisch 2006). IVOA have discussed and defined standards on data formats, data discovery, data access, and so on. VO projects in the world have been strenuously developing VO software tools and/or portal systems based on the IVOA standards. As a result, VO technologies have now been used in obtaining astronomical research results. In this paper, we introduce three example cases of astronomical research enabled by VO in the USA, Europe and Japan, and describe how the VO technologies are used in the astronomical research.

2 DISCOVERY OF BROWN DWARFS

The first example (Berriman et al., 2003) is an early demonstration project performed by NVO (National Virtual Observatory), the VO project in the USA and one of leading projects in the IVOA. They discovered a single new brown dwarf using the NVO prototype. In this section their demo scenario is described.

Brown dwarfs are substellar objects floating freely in space unlike planets and having too low in mass to sustain hydrogen-burning nuclear fusion reactions like normal stars. They are difficult to find in the sky, since they emit almost no optical light. Only 200 brown dwarfs are discovered until this study (Berriman et al., 2003). The NVO project searched for brown dwarfs in a huge amount of archived data obtained through survey observations. Since they are faint in the visual range of electromagnetic wave, it is necessary to use multiple archive data observed in different wavelengths.

2.1 Selection Criteria for Brown Dwarfs

In order to discover brown dwarfs, an observation in infrared is needed. Figure 1 is a schematic graph of Spectral Energy Distributions (SEDs) of thermal electromagnetic radiation from normal stars and brown dwarfs. It is noted that these SEDs represent thermal radiation only and are different from observed SEDs because real spectra from brown
dwarfs are affected by deep absorptions due to atmospheric atoms and molecules. This figure shows that normal stars are visible in the optical range, while brown dwarfs are faint in the optical, but the strongest emission is expected in the infrared range. This difference in SEDs is a major clue to identify brown dwarfs.

![Figure 1. Schematic plot of Spectral Energy Distributions of a brown dwarf and a normal star. The curves represent thermal radiation only. Real observed spectra exhibit absorption by atoms and molecules.](image)

2.2 Used Data

For this project two survey data are employed: 2MASS (Two Micron All-Sky Survey) and SDSS (Sloan Digital Sky Survey).

2MASS is a project to survey all the sky in infrared using two 1.3-m telescopes, one at Mt. Hopkins in Arizona, USA, and one at Cerro Tololo, Chile. The telescopes were able to observe the sky simultaneously at three infrared wavelengths: J (1.25 μm), H (1.65 μm), and Ks (2.17 μm). The NVO demo project used a catalog in 2MASS Second Incremental Data Release which covers about a half of the sky. The Point Source Catalog in the 2nd data release contains over 162 million objects. More recently, the 2MASS All-Sky Data Release (Skrutskie et al. 2006) covering almost all the sky has been available, including a Point Source Catalog containing over 470 million objects.

SDSS is a survey project to obtain multi-color, deep images covering more than a quarter of the sky and create 3-dimensional maps of the universe. The SDSS used a dedicated 2.5-meter telescope at Apache Point Observatory, New Mexico. The images are obtained in five bandpasses (u, g, r, i, and z) from ultraviolet to infrared. While the major target of SDSS is distant galaxies, the images also contain stars in our Galaxy. The NVO demo project used the catalog in SDSS Early Data Release in 2001 (Stoughton et al. 2002), covering 1 % of the sky and including 14 million astronomical objects. Among five bandpasses, they used the z band (0.9 μm).

The catalog data from these surveys are archived into a SkyNode server, a catalog server to accept an astronomical query language described in the next subsection.

2.3 VO Query Language

The Astronomical Data Query Language (ADQL, Yasuda et al. 2003) was proposed as a standard query language for the interoperability of the International Virtual Observatory. The primary use of ADQL is to express queries against an astronomical catalog, i.e., a tabular list of data on each celestial object. Thus ADQL is based on SQL. The fundamental data in astronomical catalogs are celestial coordinate and brightness of objects. Since there is no standard way to express a celestial coordinate in SQL, “Region” condition was added, as an extension to SQL, to specify a search area in the sky. Another extension in ADQL is “XMatch” (cross match) condition. XMatch is a function to join two or more catalogs and select matched data in the celestial coordinate as an identical celestial object. Since the coordinate involves positional uncertainty, the condition of XMatch needs a maximum distance between objects in the
two catalogs. XMatch is an advanced and non-mandatory function for ADQL, so it is not always accepted by data servers. An example of the query for brown dwarf search is shown below.

```
SELECT o.*, t.*
FROM SDSS o, TWOMASS t
WHERE Region('CIRCLE J2000 181.3 -0.76 6.5') AND
  XMATCH(o, t) < 3.5 arcsec AND ... cross match condition in VOQL
  (o.z - t.j_m) > 1 ... selection criteria for brown dwarfs
```

The last line is a condition to select brown dwarfs expressed in the conventional SQL language. In astronomy, the star brightness is often expressed in magnitude. o.z means a z-band apparent magnitude in the SDSS catalog, and t.j_m means a J-band apparent magnitude in the 2MASS catalog. Since the brightness of a celestial object depends on the distance to the object, apparent magnitude itself cannot be used to select brown dwarfs. On the other hand, the relative magnitude among different wavelengths can often be used as an intrinsic feature of the object. Since there is a difference in SED, between a normal star and a brown dwarf as is noted in Section 2.1, a condition to find brown dwarfs can be expressed as the query above.

### 2.4 Result

In the study of Berriman et al. (2003), the cross-match between 2MASS and SDSS covered an area of roughly 150 square degrees (about 0.4 % of the sky) and contains 326,020 matches. Filtering through a z-J color cut recovered the known brown dwarfs in that area; a T dwarf (SDSS 1346-0031) and a late-L dwarf (SDSS 1326-0038), while discovering three more brown dwarf candidates. Further spectroscopic observation identified one of these (2MASSI J0104075-005328) as an L5 dwarf. Thus Berriman et al. (2003) added a single new brown dwarf added to a list of ~200 known brown dwarfs.

### 3 DISCOVERY OF OBSCURED QUASARS

The second example is a study by Padovani et al. (2004). They searched for little-known type-2 quasars using VO tools developed mainly by AVO (Astronomical Virtual Observatory), current EURO VO.

Quasars (contraction of quasi-stellar object, QSO) were originally identified as point-like objects, similar to stars. However, they are so peculiar as to have quite high recession velocities, indicating that they are far in the distance and powerful in the electromagnetic radiation. Quasars are now considered to be a kind of objects which have an Active Galactic Nucleus (AGN) at their centers. Figure 2 shows a supposed structure of AGN. The radiation from AGN is believed to be a result of accretion of mass onto a supermassive black hole at the centre of the host galaxy. Astronomers have categorized AGNs into different types based on observational features. Now the unified model of AGN is accepted (e.g., Urry & Padovani 1995). The unified model explains that the apparent differences between different types arise simply because of their different orientations to the observer. Type 1 AGN objects are those in which we have an unimpeded view of the central regions and therefore exhibit the straight physics of AGN with no absorption. On the other hand type 2 AGN objects arise when the view is obscured by the torus.

![Figure 2. Sectional view of the supposed structure of AGN.](image-url)
Type 2 AGN is relatively difficult to detect since its center is obscured by the torus around the center. While many examples of local and relatively low-power type 2 AGN are known (Seyfert 2), very few of their high-power counterparts, that are optically obscured, radio-quiet type 2 QSO, were known until this study. In addition to their importance for AGN models, type 2 sources are expected to be an origin of the X-ray background and are therefore also cosmologically important.

Type 2 QSOs are heavily obscured in optical light and therefore not identified with the standard methods of quasar selection. The hard X-rays, however, are thought to be able to penetrate the torus. Type 2 QSOs, therefore, should be identified with powerful hard X-ray emission and X-ray line emission. Padovani et al. (2004) used Virtual Observatory (VO) tools to identify 68 type 2 AGN candidates in the two Great Observatories Origins Deep Survey (GOODS) fields (Giavalisco et al. 2004), a half of which were qualified as QSO 2 candidates.

3.1 AVO

This study was conducted using tools developed in the Astrophysical Virtual Observatory (AVO) project. The prototype consists of a suite of interoperable software, plus a set of conventions or standards for accessing remote data, and for launching remote calculations. The main component of the software is based on the CDS Aladin visualization interface (Bonnarel et al. 2000). This prototype VO portal allows efficient interactive manipulation of image and catalogue data, and provides access to remote data archives and image servers via the registry of services.

3.2 Method

To find type 2 QSO candidates, Padovani et al. (2004) used a relatively simple method based on the X-ray and optical fluxes. Two key physical properties that they used to identify type 2 AGN candidates are (1) obscured and (2) having high power to be classified as an AGN and not a starburst.

(1) Obscured

Optical data alone are not sufficient to find an AGN because the emission from an AGN could be diluted by the host galaxy at high redshift. However, AGNs reveal themselves by their hard X-ray emission and power. The X-ray has characteristics that the hard (high energy) X-ray photons are less absorbed by materials than the soft (low energy) X-ray photons. The Hardness Ratio (HR), a measure of the fraction of the hard photons relative to the soft photons, is used to identify an obscured AGN.

(2) High power in X-ray

Since AGNs or QSOs are known to radiate X-ray more powerfully than the normal galaxies, the X-ray power is employed to identify an AGN or a QSO. In general, when one wants to know the intrinsic brightness of a celestial object, he/she needs to know the distance to the object in addition to apparent brightness. The distance to a distant galaxy can be estimated from redshift (usually expressed with a parameter $z$), which is a result from recession velocity, by the Hubble’s law. The redshift can be obtained from spectroscopic observations. In this study, however, targets without spectroscopic data are important to discover new type 2 candidates. To estimate the X-ray power Padovani et al. (2004) used the correlation described by Fiore et al. (2003) between the $f(2-10 \text{ keV})/f(R)$ ratio and the hard X-ray power. The basis of this correlation is the fact that the $f(2-10 \text{ keV})/f(R)$ ratio is roughly equivalent to the ratio between the nuclear X-ray power and the R band luminosity of the host galaxy. Since the host galaxy R band luminosities (unlike the X-ray power) show only a modest amount of scatter, this flux ratio is a good indicator of the X-ray power.

3.3 Workflow and Result

The analysis workflow of Padovani et al. (2004) is described in this subsection. The workflow steps and the number of sources selected in each step are summarized in Table 1.
Table 1. Summary of the workflow to search for type 2 QSO and the number of selected sources in each step.

<table>
<thead>
<tr>
<th>Workflow</th>
<th>number of sources</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HDF-N</td>
</tr>
<tr>
<td>(1) X-ray catalog from Alexander et al. 2003</td>
<td>503</td>
</tr>
<tr>
<td>(2) Select with Hardness Ratio ≥ −0.2</td>
<td>190</td>
</tr>
<tr>
<td>(3) Cross-matching with Optical catalog</td>
<td>103</td>
</tr>
<tr>
<td>(4) Select unclassified type-2 AGN</td>
<td>47</td>
</tr>
<tr>
<td>(5) Qualified as type-2 QSO</td>
<td>16</td>
</tr>
</tbody>
</table>

(1) Data

Padovani et al. (2004) used the two fields of the Great Observatories Origins Deep Survey (GOODS) (Giavalisco et al. 2004) for this purpose. The GOODS is a joint observational program to survey the distant and faint universe across broad range of wavelengths, using space-based observatories including the Spitzer Space Telescope, Hubble, and Chandra, XMM-Newton, and ground observatories. The two GOODS fields are centered on the HDF-N (Hubble Deep Field North) and the CDF-S (Chandra Deep Field South). Padovani et al. (2004) used the X-ray catalogues produced by Alexander et al. (2003) using the Chandra telescope. These include 503 (HDF-N) and 326 (CDF-S) objects respectively, for a total of 829 sources. In the optical they used the publicly available GOODS Hubble Space Telescope (HST) Advanced Camera for Surveys (ACS) data.

(2) Selecting absorbed sources using the Hardness ratio

The Alexander et al. (2003) catalogues provide photon counts (proportional to the brightness in X-ray) in various X-ray bands. Padovani et al. (2004) defined the hardness ratio $HR = (H−S)/(H+S)$, where H is the hard X-ray counts (2.0-8.0 keV) and S is the soft X-ray counts (0.5-2.0 keV). They adopted the criterion by Szokoly et al. (2004) and identified those sources which have $HR \geq -0.2$ as absorbed sources. They found 294 (CDF-S: 104, HDF-N: 190) such absorbed sources which represent ~35% of the X-ray sources in the Alexander catalogues.

(3) Finding the optical counterparts

The optical counterparts to the X-ray sources were selected by cross-matching the absorbed X-ray sources with the GOODS ACS catalogues (29,599 sources in the CDF-S, 32,048 in the HDF-N). The GOODS catalogues contain sources that were detected in the z-band. To limit their sample to good matches, they used a criterion that the cross match distance be less than the combined optical and X-ray 3σ positional uncertainty for each individual match. Applying this distance/error <1 criterion they were able to limit the number of matches to 168 (CDF-S: 65, HDF-N: 103).

(4) Select unclassified sources

Previously classified sources and their spectroscopic redshifts were available from two catalogs: Szokoly et al. (2004) for the CDF-S and Barger et al. (2003) for the HDF-N. Cross-matching with these catalogs, 110 (CDF-S: 54, HDF-N: 56) sources out of 168 were found to be already classified in the two catalog. This leaves a total of 68 (CDF-S: 21, HDF-N: 47) unclassified candidates.

(5) Quality as type-2 QSO from the X-ray power

It is known that the normal galaxies, irrespective of their morphology, have X-ray powers that reach, at most, $L_X \leq 10^{42}$ erg/s (e.g., Forman et al. 1994). Therefore, any X-ray source with $HR \geq -0.2$ and $L_X > 10^{42}$ erg/s should be an obscured AGN. Furthermore, following Szokoly et al. (2004), any such source having $L_X > 10^{44}$ erg/s will be qualified as a type 2 QSO.

For the unclassified sources Padovani et al. (2004) estimated the X-ray power as follows; they first derived the $f(2-10$ keV)/$f(R)$ flux ratio, and then estimated the X-ray power from the correlation found by Fiore et al. (2003), namely $\log L_{10} = \log f(2-10$ keV)/$f(R) + 43.05$. From this estimation, 31 (CDF-S: 15, HDF-N: 16) candidates passed the criteria ($HR \geq -0.2$ and $L_X > 10^{44}$ erg/s) and qualified as type 2 QSO.
4 STUDY ON THE QSO ENVIRONMENT

The last example is a study by Shirasaki et al. (2006), on the evolution of number densities around QSOs (Quasi-Stellar Object, see the previous section) to investigate mechanism of formation and evolution of QSOs, and to understand the formation history of the large-scale structure of the Universe.

It has been suggested that QSOs and AGNs are associated with groups or clusters of galaxies. In astronomy, groups of galaxies are classified into small aggregates of galaxies that typically contain fewer than 30-50 galaxies in a diameter of 1-2 Mpc. On the other hand, clusters of galaxies are larger than groups, although there is no sharp boundary. Groups and clusters of galaxies are gravitationally-bound and form the large scale structure of the Universe. It is considered that member galaxies in clusters evolve by merging with each other. Thus clusters of distant galaxies are important for studying the evolution of galaxies and the large-scale structure of the Universe.

The correlation between QSOs and galaxies can be related with the formation and fuelling mechanism of a QSO, as well as the formation history of the large scale structure of the Universe. However, this kind of study had been limited to the local universe near our Galaxy (redshift z < 1.0) (e.g., Barr et al. 2003), yielding poor statistics (~tens of QSOs). Recent advanced observations have provided large samples of both QSOs and galaxies enough to allow statistically meaningful studies of QSO environments. However, the processing and analysis of such huge dataset is time consuming. We thus used the JVO system for the study on the QSO-galaxy correlation by exploring the Universe at high redshift, z > 1.0, to obtain reliable statistics (about hundreds or more).

4.1 JVO

Japanese Virtual Observatory (JVO; Shirasaki et al. 2009) is conducted by the National Astronomical Observatory of Japan (NAOJ). One of the major products is the JVO portal system. It is a web server to provide a user-friendly interface to discover, search, manipulate, and visualize astronomical data. This portal service provides a seamless access to the distributed astronomical databases. JVO also developed the data analysis services such as Subaru data reduction and general astronomical data analysis.

JVO also adopted the IVOA standards, such as Simple Image Access (SIA) and ADQL. We also constructed an OAI-PMH publishing registry, a web service based searchable registry, and VO data services based on SIA and SkyNode protocols. The JVO portal accepts JVO query language (JVOQL) which has similar syntax with the VO standard. JVOQL can describe a cross match query not only for the catalog services but also for image and spectrum services. We developed a SkyNode toolkit for building a VO compliant data service.

4.2 Distributed, Automated-Analysis System for Subaru Images

Shirasaki et al. (2006) utilized the JVO system to study environments of QSOs by combining an existing QSO catalog and image data obtained with SuprimeCam (Miyazaki et al. 2002), one of the instruments of Subaru telescope. In order to increase available dataset, raw data in pre-selected five fields are retrieved from the MASTARS (Takata et al. 2000) and SMOKA (Baba et al. 2001) services operated by NAOJ. Since the reduction and analysis of SuprimeCam data require huge computer resources, we developed a GRID computing system, composed of 29 CPUs in total. This system can be scaled-up easily.

The images are reduced with a standard analysis tool and registered as a SkyNode database. The fields are selected through cross-matching between the QSO database and SuprimeCam frame database. A workflow for this study is as follows:

- Select QSO coordinates from the QSO catalog database;
- Search multi-bands imaging data which covers the QSO regions;
- Select Subaru images at each QSO position with enough coverage around the QSO;
- Check if the selected image has at least three bands;
- Generate a source catalog around the QSO by processing the selected image through a SExtractor Web service;
- Calculate redshift probability P(z) by using the HyperZ;
- Select objects with P(z,QSO) > 0.5;
- Derive radial distribution (dead pix, bright objects); and
Calculate spatial QSO-Galaxy cross-correlation amplitude.

We succeeded to federate the catalog and image data services. By incorporating the workflow system, all the procedure will be done in an automatic way.

4.3 Tentative Result

At the time reported by Shirasaki et al. (2006), 100 QSOs/AGNs in 17 image fields were analyzed. Figures 3 and 4 are the result plots of the average radial profiles of number distributions around QSO/AGN. This initial result demonstrates that the JVO system is capable of a clustering analysis around distant QSOs at $z > 1$ where this kind of analysis had not been successful. Figure 3 exhibits statistically significant clustering of galaxies for $z = 0.5-1.5$, while Figure 4 shows that the statistics for $z = 1.5-4.0$ does not reach a significant level. However, a total of 700 QSOs/AGNs in 60 image fields were available for this analysis. Such a large sample of data allows statistical discussions for $z > 1$. The follow-up analysis has been already succeeded and a new result will be reported in the near future.

Figure 3. The average radial profiles of number distributions around QSO for $z = 0.5-1.5$

Figure 4. The average radial profiles of number distributions around QSO for $z = 1.5-4.0$

5 CONCLUDING REMARKS

We have briefly described an astronomical study using NVO, EuroVO, and JVO, respectively. The studies by NVO and EuroVO show that VO is useful for exploration studies such as discovering new objects. The study by JVO shows that VO is also useful for statistical studies. Once VO prevails, it will become a powerful tool for these kinds of
studies.

Figure 5 is a histogram of VO-related papers published in refereed journals, showing that VO-related papers are increasing year by year. This histogram is obtained from the query to Astrophysics Data System (ADS) with the key phrase “virtual observatory”. The total number of the articles in this histogram is 83 at the end of 2008. If non-refereed papers are counted, the total number exceeds 1000. Through a decade of development, VO is becoming a real tool for astronomy. This suggests that the similar situation would happen for other fields of science.

Figure 5. The histogram of VO-related published through refereed journals.

6 ACKNOWLEDGEMENTS

This work is supported by the JSPS Core-to-Core Program and Grant-in-aid for Information Science (15017289, 16016292, 18049074 and 19024070) carried out by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. YS is grateful for support under Grant-in-aid for Young Scientists (B) (17700085) carried out by the MEXT of Japan.

7 REFERENCES


