ADASS XIII
ASP Conference Series, Vol. XXX, 2004
F. Ochsenbein and M. Allen eds.

Open SkyQuery – VO Compliant Dynamic Federation of Astronomical Archives

Tamás Budavári[1], Alex Szalay[1], Tanu Malik[1], Ani Thakar[1], William O'Mullane[1], Roy Williams[2], Jim Gray[3], Bob Mann[4], Naoki Yasuda[5]

Abstract. We discuss the redesign of the SkyQuery architecture, originally built as a simple proof of concept for dynamic federation of astronomical archives. In keeping with the Virtual Observatory philosophy of hierarchical services, the design of Open SkyQuery is based upon higher level services extending the basic functionality of the current VO standard, the ConeSearch. Open SkyQuery implements the VO specifications for data access, retrieval and spatial join. Data are published via Web Services called SkyNodes providing a rich functionality including footprint coverage. SkyNodes are discovered through the VO registry. We propose to have at least two levels of SkyNode compliance (Basic and Full). We will also provide templates for publishing data into a SkyNode.

1. Motivation

With the advent of large CCD detectors, the way astronomy is done changes rapidly. Because of the exponential growth in the size and speed of the silicon chips, new surveys are expected to have significantly higher data rates. These survey projects become both the authors and publishers of their data (Szalay et al. 2002). In this exponential world, only 10% of all astronomical information is available in central archives at any given time. In order to have access to all up-to-date observations, we need to find a way to federate geographically separated astronomical archives.

Current sky surveys such as SDSS, 2MASS, DPOSS have proven that discoveries are always made at the boundaries when going deeper or using more colors. By covering different wavelength ranges, surveys can very well complement one another if one finds a way to combine them. In the past, crossmatching of many catalogs was prohibitively complex and expensive. The Virtual Observatory and Open SkyQuery in particular are to make it simple and affordable.

[1]Dept. of Physics & Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA

[2]CACR, California Institute of Technology, Pasadena, CA, 91125, USA

[3]Microsoft Bay Area Research Center, San Francisco, CA 94105, USA

[4]Institute for Astronomy, University of Edinburgh, Edinburgh, EH9 3HJ, UK

[5]National Astronomical Observatory of Japan, Tokyo 181-8588, Japan

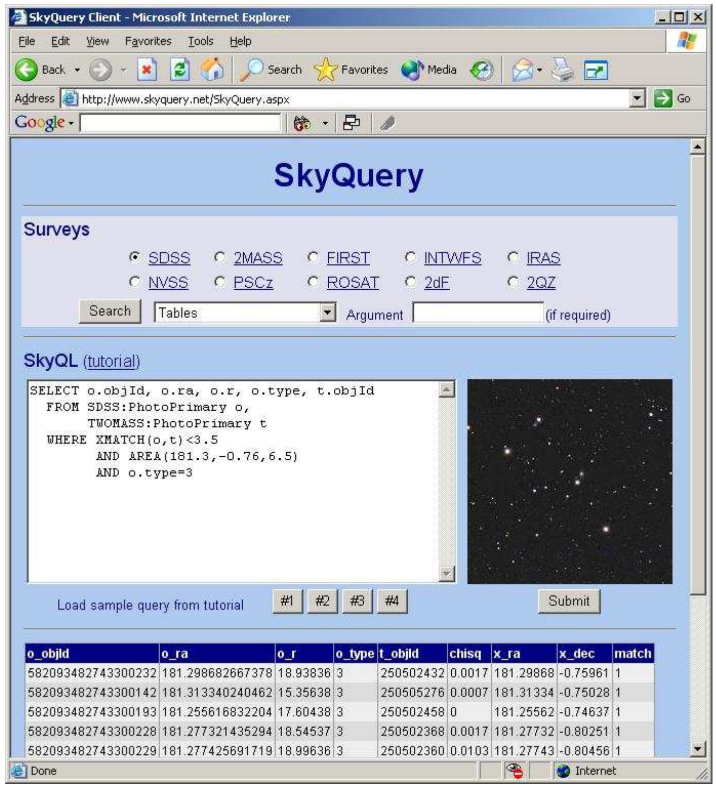

Figure 1. The SkyQuery web site currently provides access to 10 catalogs, altogether close to one terabyte (1TB) online astronomical data.

2. The Prototype SkyQuery

The prototype SkyQuery was built last year in six weeks as a feasibility study (Budavári et al. 2002, Malik et al. 2002). It used a hierarchy of XML Web Services to implement a distributed query system that provided seamless access to SDSS, 2MASS and FIRST data. Since the launch of the SkyQuery web site, many other catalogs have become available (Purger et al. 2003) including the Isaac Newton Telescope's Wide Field Survey (INTWFS), IRAS, NVSS, 2dF, PSCz, 2QZ and Rosat, see Figure 1.

3. Building on Virtual Observatory Standards

The SkyQuery architecture is being redesigned to utilize the recently emerging VO standards such as the VOTable, the Astronomical Data Query Language (ADQL; Yasuda et al. 2003), the VO Query Language (VOQL) and the VO Registry services (Greene et al. 2003). The data are going to be published by the SkyNodes that implement XML Web Services to extend the basic functionality of the ConeSearch.

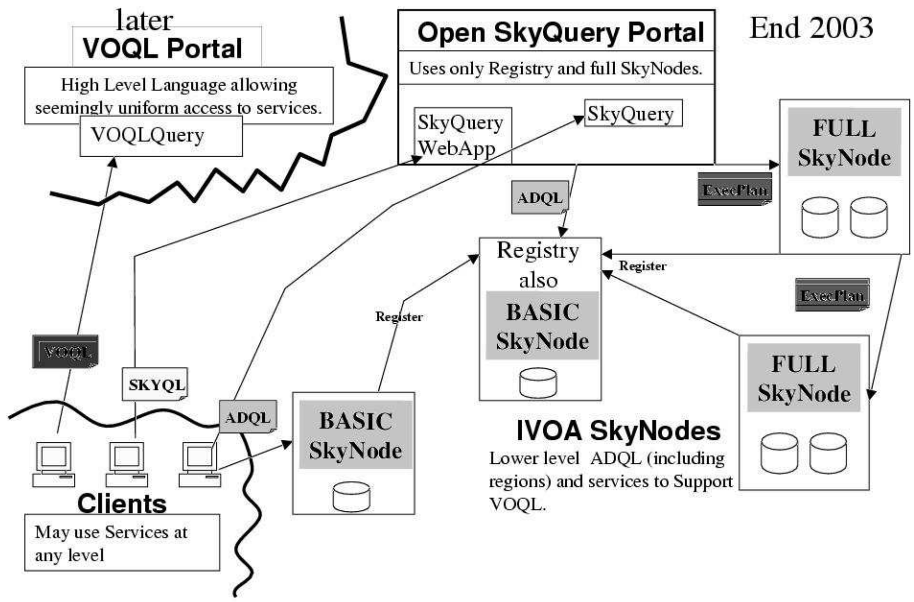

Figure 2. Open SkyQuery architecture: Basic and Full IVOA SkyNodes are discovered dynamically in the VO registry by the SkyQuery portal.

4. Open SkyNode

We propose to have three layers of VOQL building on top of one another, see also in Yasuda et al. (2003):

- **VOQL1–ADQL:** ADQL and VOTable to query a single node
- **VOQL2–SkyQL:** SQL-like query language and federation system
- **VOQL3–Sky???:** a future query language

The SkyNode is essentially the implementation that provides the necessary services, i.e. automatic crossmatching and participation in federated queries called SkyQL. They may publish only a small amount of data, i.e. a single FITS file or an entire survey such as SDSS. We distinguish between at least two levels in the implementation of the SkyNode: **Basic SkyNodes** confirm to the Layer-1 specifications. They know how to execute ADQL requests to query their own data and return the results in VOTable format. **Full SkyNodes** support all methods required to be part of a federated query (Layer-2). Advanced versions will also implement footprint services in order to work out dynamically their intersection when used in the same SkyQL query. For large surveys, it is essentially a must to implement a sky indexing scheme as well, such as the Hierarchical Triangular Mesh (HTM; Kunszt et al. 2001) for quick lookup and spatial joins.

Figure 2 illustrates the Open SkyQuery architecture and shows the relations between the components.

5. SkyQuery Strategy

In order to ensure fast response, one needs to optimize the query plan. Our simulations show that the simple sequential execution proves to be optimal because today the wire speed is the limiting factor. One needs to arrange the SkyNodes in *ascending* order of the number of matching records so that the *least* amount of data is transferred. This simplifies the logic of the portal significantly. However, the SkyNodes are designed to deal with more complicated query plans, so that the system may be enhanced easily later on. Another possible enhancement might be the asynchronous data flow (O'Mullane et al. 2003).

6. Concluding Remarks

The emerging Virtual Observatory infrastructure makes it possible to develop a new generation of astronomical tools. These online tools promise to be easy to use and to open new dimensions for scientists.

Open SkyQuery is just one of the first steps. Its catalog services will enable us to analyze geographically separated IVOA archives, a.k.a. SkyNodes, as if they were part of the same dataset. As of today, the VO building blocks are already in place to make Open SkyQuery a reality.

Acknowledgments. SkyQuery is supported by NSF Awards 0122449 and 9980044, and NASA AISRP awards NAG5-10742 (2001) and NAG5-12092 (2002).

Links

`http://www.skyquery.net/`
`http://skyservice.pha.jhu.edu/develop/vo/adql/`
`http://www.ivoa.net/twiki/bin/view/IVOA/IvoaVOQL`

References

Budavári, T., et al. 2002, in APS Conf. Ser., Vol. 295, Astronomical Data Analysis Software and Systems XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook (San Francisco: ASP), [O10.1]

Greene, G., et al. 2003, this volume, [P3.8]

Kunszt, P. Z., Szalay, A. S., and Thakar, A. R. 2001, Mining the Sky: Proc. of the MPA/ESO/MPE workshop, Garching, A.J.Banday, S. Zaroubi, M. Bartelmann (ed.), (Springer-Verlag Berlin Heidelberg), 631.

Malik, T., et al. 2002, CIDR '03, p.17, *'SkyQuery: A WebService Approach to Federate Databases'*

O'Mullane, W., et al. 2003, this volume, [O4.4]

Purger, N., et al. 2003, this volume, [P2.26]

Szalay, A. S., et al. 2002, Proc. of SPIE, 4846, *'Web Services for the VO'*

Yasuda, N., et al. 2003, this volume, [P3.10]