JVOの研究開発:全体進捗
データベース国際連携の成功

国立天文台　大石雅寿

masatoshi.ohishi@nao.ac.jp

JVO collaborators

Project Scientists

NAOJ
- Mizumoto
- Oe
- Shirasaki
- Tanaka
- Honda
- Kawanomoto

ICRR
- Yasuda

Ochanomizu U.
- Masunaga

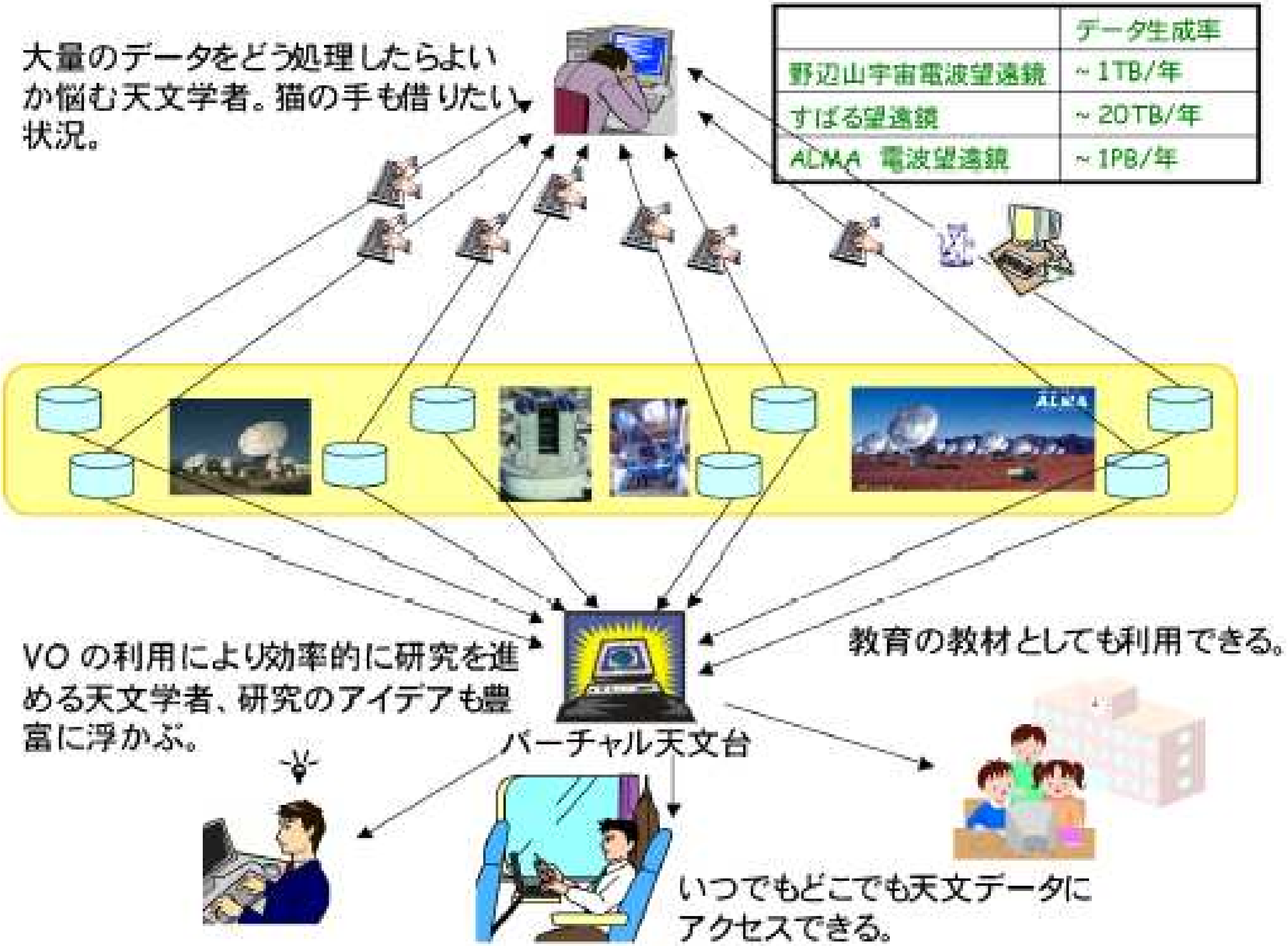
System Engineers

Fujitsu Ltd.
- Monzen
- Kawarai
- Ishihara
- Yamazaki

SEC Ltd.
- Morita
- Nakamoto
- Kobayashi
- Yoshida

Supporter

NII
- Miura



	データ生成率
野辺山宇宙電波望遠鏡	~1TB/年
すばる望遠鏡	~20TB/年
ALMA 電波望遠鏡	~1PB/年

What is the Virtual Observatory… and what it is not…

The VO is:

- A set of international standards to share complex data
- A modular set of tools to work with distributed data
- A simple environment to publish data to
- An essential part of the research astronomer's toolkit
- A catalyst for world-wide access to astronomical archives
- A vehicle for education and public outreach

The VO is not:

- A replacement for building new telescopes and instruments
- A centralized repository for data
- A data quality enforcement organization

VO Projects in the world

- 15 countries and a region
- International Virtual Observatory Alliance (IVOA) Standards to interoperate VOs
- Japan – Language to access federated DB

http://www.ivoa.net/

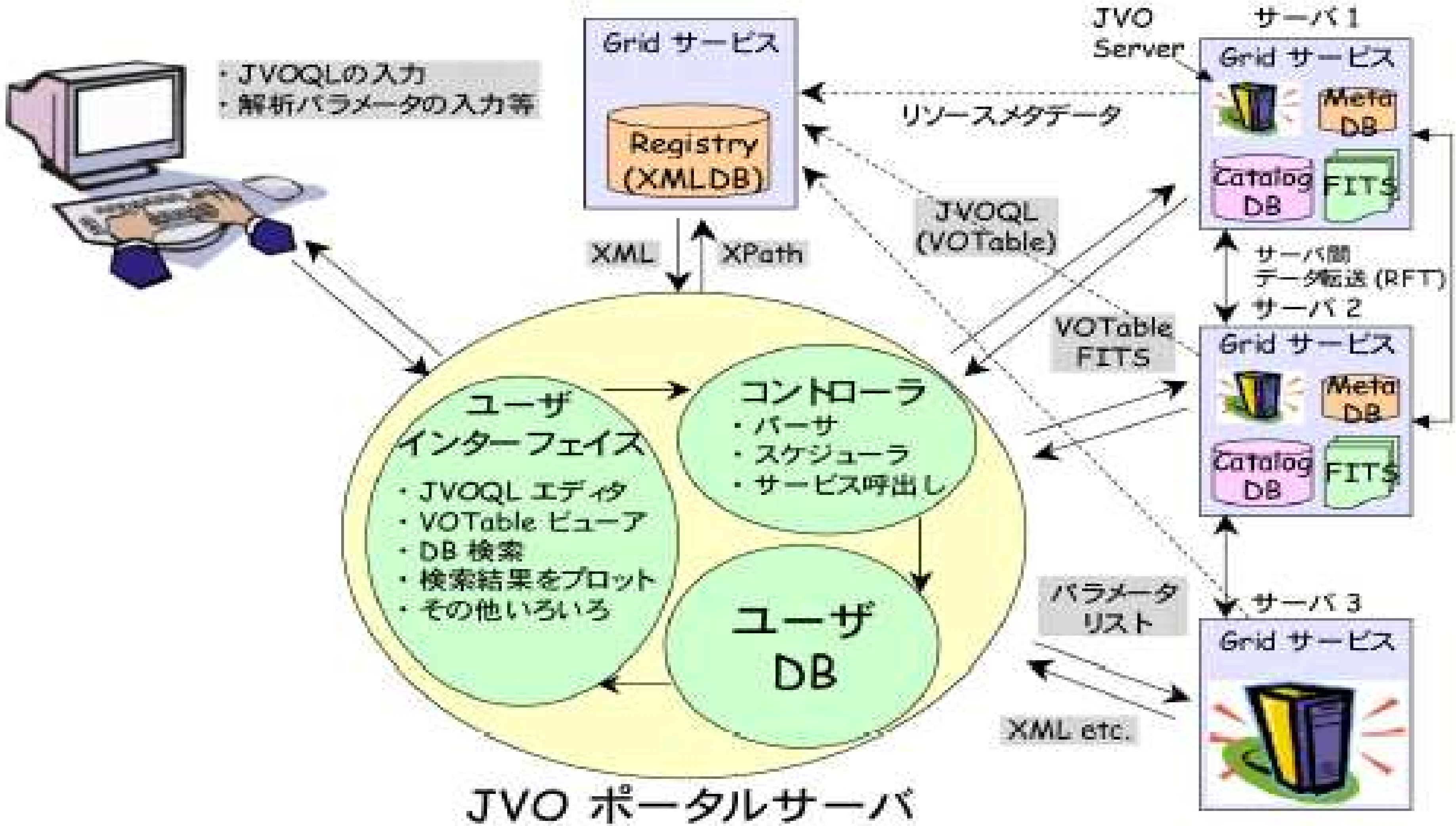
Standardization in IVOA

- Query language to distributed DBs (VOQL)
- Meta-data：contents, protocol to interchange
  based on OAI-PMH
- Protocols to retrieve images, spectra, and so on
  SkyNode, SIAP, SSAP, STC, etc.
- Unified attribute names in DBs
  UCD (Unified Contents Descriptions)
- Output Format：VOTable (XML)
  incorporates FITS
- etc

International Endorsements

- IAU XXVth GA Res. (2003 Jul.)
- OECD Rec. (‘04 Aug)
  - place archives that may be accessible via internet
  - provide adequate funding as long-term issues

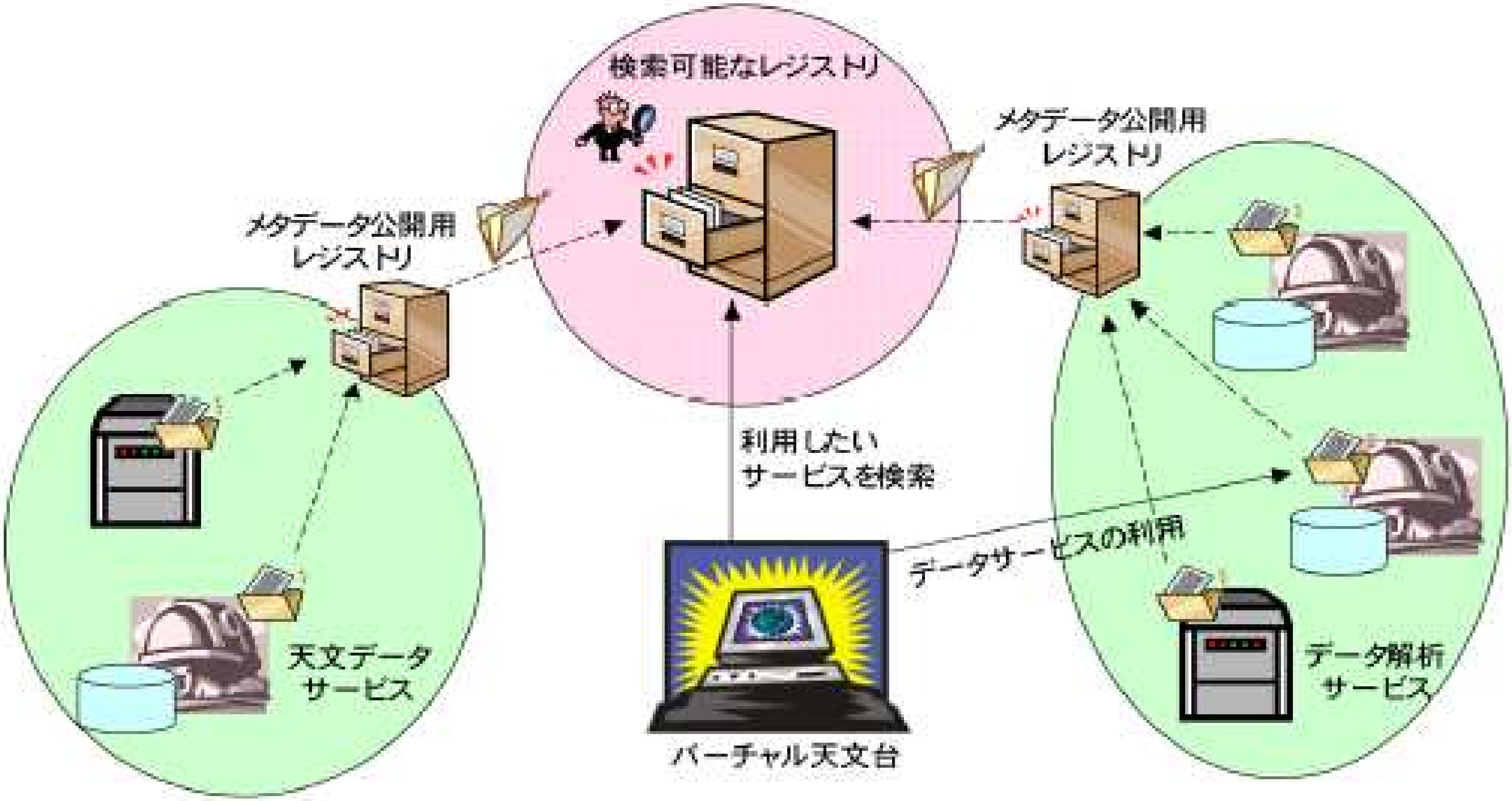
Schematic diagram of JVO



DBs available under JVO

- Subaru SupCAM (partial)
- SXDS
- SMOKA (catalog)
- SDSS
- 2MASS
- JAXA/ISAS – ASCA
- More to come

Exchange of Meta Data:OAI-PMH



Successful Interoperations

- Accesses from / to other VOs in Europe and US have been available since Dec 2004
- 85 resources are accessible as of today

70		More Info	HyperLeda FITS Archive Simple Spectrum Data Access	SIAP	URL	
71		More Info	Sloan Digital Sky Survey DR3- Filter z	SIAP	URL	
72		More Info	Sloan Digital Sky Survey DR3- Filter g	SIAP	URL	
73		More Info	Sloan Digital Sky Survey DR3- Filter i	SIAP	URL	
74		More Info	2MASS All-Sky Quicklook Image Service	SIAP	URL	
75		More Info	INES: The IUE Newly Extracted Spectra	SIAP	URL	
76		More Info	ASCA SIA Service	SIAP	URL	
77		More Info	MAST Image Scrapbook	SIAP	URL	
78		More Info	JVO Publishing Registry	Registry	URL	
79		More Info	NCSA Radio Astronomy Imaging Registry	Registry	URL	
80		More Info	Minnesota Automated Plate Scanner	Registry	URL	
81		More Info	CADC Registry	Registry	URL	
82		More Info	Astrogrid Full Registry	Registry	URL	

JVO is seen from the UK VO

AstroGrid Registry | AstroGrid

Server
Home
Admin
Investigate
IVORN
Lookup
Browse
Query
Keyword
Query
Register
Enter
Resource

Registry Browser

Version: 0.9

Find IVORNs including: jvo [List]

Browse for another version 0.9 [List]

Title	Type	AuthorityID	ResourceKey	Up
JVO Publishing Registry	vg:Registry	o	publishingregistry	200 16
JVO Publishing Registry	vg:Registry	jvo	publishingregistry	200 21
the Subaru/XMM-Newton Deep Survey (SXDS) SkyNode Service	sn:OpenSkyNode	jvo/skynode	sxds	200 20
Subaru/XMM-Newton Deep Survey 01	jsn:OpenSkyNodeJ	jvo/skynodej	sxds	200 20
JVO	vr:Organisation	jvo	jvo	200 18
the Subaru/XMM-Newton Deep Survey (SXDS) SIA Service	sia:SimpleImageAccess	jvo/siap	sxds	200 20
JVO Authority	vg:Authority	jvo	null?!	200 21

Access time to the US VO

表 1　カニ星雲データへの検索時間

Table 1　Time to query data of the Crab Nebula

波長	サーベイ名	サーバ	時間 (秒)
X 線	Chandra	cda.harvard.edu	1.715
赤外線	2MASS	mercury.cacr.caltech.edu	3.536
電波	VLA	adil.ncsa.uiuc.edu	7.115

Analysis Tools

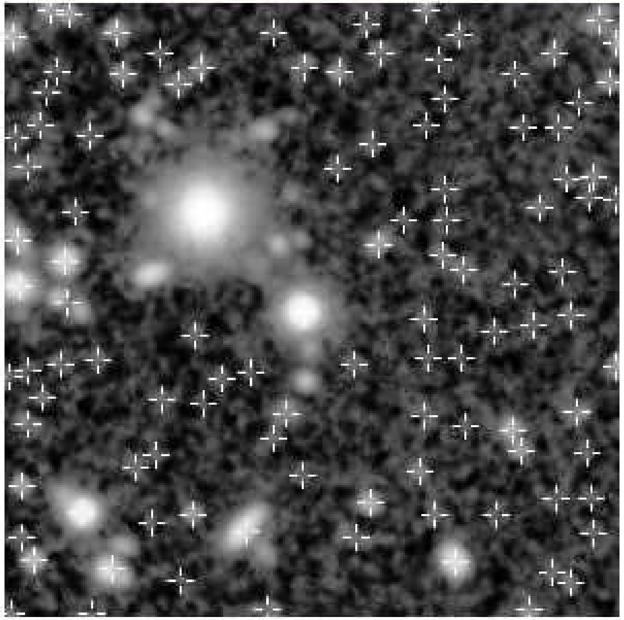
- Sextractor – extract source parameters
  generate personal catalogs
- HyperZ – derive photometric Z
- Aladin – Image viewer
- VOPlot – Plot VOTables
- SpecView – SED generator
- More to be added
  - Legacy softwares, Data mining, personal DBs, etc.

QSO searches

Image Viewer

Status | Registry | Search | Result | Database | QSO Search | Image Viewer | Logout

Name	Origin	Scale	Contrast
fits0	http://gridq.dc.nao.a	hist	min = 0.0 max = 65529.0 auto = true



Scale :	hist eq.
Contrast :	auto min = 0.0 max = 65529.0

Submit

VOTable :	votable0

Overplot

User ID	User Name	Group	Last Login
ohishi	Masatoshi Ohishi	jvo	Tue Mar 29 20:13:07 JST 2005

Total memory = 266403kB Used momory = 162745kB (61%)

Science results (EU)

- Padovani et al. (2004)) demonstrates that VO tools are mature enough to produce cutting-edge science results by exploiting astronomical data beyond classical identification limits ($R \lesssim 25$)

A&A 424, 545 559 (2004)
DOI: 10.1051/0004-6361:20041153

Astronomy & Astrophysics

Discovery of optically faint obscured quasars with Virtual Observatory tools

P. Padovani[1], M. G. Allen[2], P. Rosati[3], and N. A. Walton[4]

[1] ST-ECF, European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany
e-mail: Paolo.Padovani@eso.org
[2] Centre de Données astronomiques de Strasbourg (UMR 7550), 11 rue de l'Université, 67000 Strasbourg, France
e-mail: allen@astro.u strasbg.fr
[3] European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching bei München, Germany
e-mail: Piero.Rosati@eso.org
[4] Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
e-mail: naw@ast.cam.ac.uk

Received 23 April 2004 / Accepted 2 June 2004

Abstract. We use Virtual Observatory (VO) tools to identify optically faint, obscured (i.e., type 2) active galactic nuclei (AGN) in the two Great Observatories Origins Deep Survey (GOODS) fields. By employing publicly available X-ray and optical data and catalogues we discover 68 type 2 AGN candidates. The X-ray powers of these sources are estimated by using a previously known correlation between X-ray luminosity and X-ray-to-optical flux ratio. Thirty-one of our candidates have high estimated powers ($L_x > 10^{44}$ erg/s) and therefore qualify as optically obscured quasars, the so-called "QSO 2". Based on the derived X-ray powers, our candidates are likely to be at relatively high redshifts, $z \sim 3$, with the QSO 2 at $z \sim 4$. By going ~3 mag fainter than previously known type 2 AGN in the two GOODS fields we are sampling a region of redshift – power space which was previously unreachable with classical methods. Our method brings to 40 the number of QSO 2 in the GOODS fields, an improvement of a factor ~ 4 when compared to the only 9 such sources previously known. We derive a QSO 2 surface density down to 10^{-15} erg cm^{-2} s^{-1} in the 0.5–8 keV band of $\gtrsim 330$ deg^{-2}, ~30% of which is made up of previously known sources. This is larger than current estimates and some predictions and suggests that the surface density of QSO 2 at faint flux limits has been underestimated. This work demonstrates that VO tools are mature enough to produce cutting-edge science results by exploiting astronomical data beyond "classical" identification limits ($R \lesssim 25$) with interoperable tools for statistical identification of sources using multiwavelength information.

Key words. astronomical data bases: miscellaneous – methods: statistical – galaxies: quasars: general – X-rays: galaxies

1. Introduction

The unified model for active galactic nuclei (AGN) is largely accepted (e.g., Urry & Padovani 1995; see also the very recent results by Jaffe et al. 2004). The apparent disparate properties and nomenclature of active galaxies can be explained by the physics of black hole, accretion disk, jet, and obscuring torus convolved with the geometry of the viewing angle. Type 1 sources are those in which we have an unimpeded view of the central regions and therefore exhibit the straight physics of AGN with no absorption. Type 2 objects arise when the view is obscured by the torus. While many examples of local, and therefore relatively low-power, type 2 AGN are known (the Seyfert 2s), it has been debated if their high-power counterparts, that is optically obscured, radio-quiet type 2 QSO, exist. Indeed, until very recently, very few, if any, examples of this class were known. Apart from their importance for AGN models, type 2 sources are expected to make a significant fraction of the X-ray background (see, e.g., Comastri et al. 2001) and are therefore also cosmologically very relevant. These sources are heavily reddened and therefore fall through the "standard" (optical) methods of quasar selection. The hard X-rays, however, are thought to be able to penetrate the torus. Type 2 QSO, therefore, should have narrow, if any, permitted lines (and might look like normal galaxies in the optical/UV band), powerful hard X-ray emission, and, in some cases, a high equivalent width Fe K line (e.g., Norman et al. 2002).

In this paper we use Virtual Observatory (VO) tools to identify 68 type 2 AGN candidates in the two Great Observatories Origins Deep Survey (GOODS) fields (Giavalisco et al. 2004a), ~1/2 of which qualify as QSO 2 candidates. Based on the properties of already known sources, we expect the large majority of these to be obscured quasars whose identification is only possible through their X-ray emission.

Science results (US)

- McGlynn et al. (2004) classified all unidentified ROSAT WGACAT objects using VO data access methods to cross-correlate multi-wavelength catalogs
  - Technique applied to find candidate X-ray binaries and now to SDSS photometric catalog
- More than 400 papers related to “virtual observatory” in ADS

THE ASTROPHYSICAL JOURNAL, 616:000–000, 2004 December 1

AUTOMATED CLASSIFICATION OF *ROSAT* SOURCES USING HETEROGENEOUS MULTIWAVELENGTH SOURCE CATALOGS

T. A. McGlynn,[1] A. A. Suchkov,[2] E. L. Winter,[1,3] R. J. Hanisch,[2] R. L. White,[2] F. Ochsenbein,[4] S. Derriere,[4] W. Voges,[5] M. F. Corcoran,[1,6] S. A. Drake,[1,6] and M. Donahue[7]
Received 2004 February 19; accepted 2004 August 5

ABSTRACT

We describe an online system for automated classification of X-ray sources, ClassX, and we present preliminary results of classification of the three major catalogs of *ROSAT* sources, *ROSAT* All-Sky Survey (RASS) Bright Source Catalog, RASS Faint Source Catalog, and WGACAT, into six class categories: stars, white dwarfs, X-ray binaries, galaxies, active galactic nuclei, and clusters of galaxies. ClassX is based on a machine-learning technology. It represents a system of classifiers, each classifier consisting of a considerable number of oblique decision trees. These trees are built as the classifier is “trained” to recognize various classes of objects using a training sample of sources of known object types. Each source is characterized by a preselected set of parameters, or attributes; the same set is then used as the classifier conducts classification of sources of unknown identity. The ClassX pipeline features an automatic search for X-ray source counterparts among heterogeneous data sets in online data archives using Virtual Observatory protocols; it retrieves from those archives all the attributes required by the selected classifier and inputs them to the classifier. The user input to ClassX is typically a file with target coordinates, optionally complemented with target IDs. The output contains the class name, attributes, and class probabilities for all classified targets. We discuss ways to characterize and assess the classifier quality and performance, and we present the respective validation procedures. On the basis of both internal validation and external verification, we conclude that the ClassX classifiers yield reasonable and reliable classifications for *ROSAT* sources and have the potential to broaden class representation significantly for rare object types.

Subject headings: methods: statistical — surveys — X-rays: binaries — X-rays: general — X-rays: stars

1. INTRODUCTION

The classification of cosmic sources into physically distinct classes is a key element of research in all domains of astrophysics. Traditionally, this has involved painstaking manual analysis of detailed, homogeneous sets of observations. More recently, automated classifier tools have been used to help in the classification of objects from huge but still largely homogeneous surveys. Examples include analysis of the First (Odewahn 1995) and Second (Weir et al. 1995) Digital Sky Surveys and the Sloan Digital Sky Survey (SDSS; Stoughton et al. 2002). In this paper, we discuss how we can go beyond using single large surveys and combine information from multiple heterogeneous databases to classify astronomical sources. Using dynamic cross-correlations of electronically available data sets, the ClassX team has developed a series of classifiers that rapidly sort X-ray sources into classes. These facilities are now available to the community at the ClassX Web site.[8] Classification is distinct from correlation and identification with objects at other wavelengths. Our classification tools can use the nonexistence of counterparts at other wavelengths or use ensembles of potential counterparts to establish limits to parameters.

Our initial work has concentrated on the more than 100,000 unclassified sources detected by the *ROSAT* observatory[9] from 1990 to 1999. These high-energy sources are particularly rich in interesting objects: QSOs and other active galactic nuclei (AGNs), clusters of galaxies, young stars, and multiple systems containing white dwarf (WD), neutron star, or black hole companions. The *ROSAT* samples have been used in prior investigations (e.g., Rutledge et al. 2000; Zhang & Zhao 2003), but still only about 10% of the sources observed by *ROSAT* have a reliable classification. In most cases this identification rests on cross-correlation between the *ROSAT* object and tables of classified sources. In some cases detailed follow-up observations have been performed on a source-by-source basis. This is extraordinarily expensive in both telescope time and the time of astronomers analyzing these data. Direct comparison of *ROSAT* sources with massive optical catalogs (e.g., Rutledge et al. 2000; or the similar efforts for *XMM-Newton* data; cf. Watson et al. 2003, Yuan et al. 2003, and Lamar et al. 2003) enables the cross-identification of *ROSAT* sources, but unless the class of the counterpart is known, this does not determine the type of the source. However, flux information from multiple catalogs allows us to try to classify sources with more information than is available from the X-ray observations alone.

Our approach differs from most previous efforts at multispectral classification in several basic ways. First it does not specifically constrain the information that is used to distinguish our output categories. Other authors have looked at the X-ray to optical ratios (Maccacaro et al. 1988) or X-ray/optical/radio

[1] NASA Goddard Space Flight Center, Greenbelt, MD 20771.
[2] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218.
[3] Science Systems and Applications, Inc., Lanham, MD 20706.
[4] Centre de Données Astronomiques de Strasbourg, Observatoire Astronomique, UMR 7550, 11 rue de l'Université, F-67000 Strasbourg, France.
[5] Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse, D-85740 Garching, Germany.
[6] Universities Space Research Association, 7501 Forbes Boulevard, Seabrook, MD 20706.
[7] Michigan State University, East Lansing, MI 48834.
[8] See http://heasarc.gsfc.nasa.gov/classx.
[9] See http://wave.xray.mpe.mpg.de/*ROSAT*.

Future Prospects

Start to design / implement operational system

- Upgrade user-interfaces
- Add spectrum data
- Add analysis engines by Java-wrapping
- Download to / Upload from user machines
- Science corresponding to use cases !!
- Design workflows
- ….

Acknowledgement

- 日本学術振興会 先端研究拠点事業
- 科研費　特定領域研究「情報学」
- 国立天文台
  - 研究費も含めて