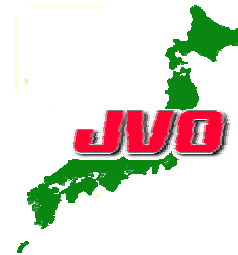


# 天文学への情報学の導入 ~ ~ Japanese Virtual Observatory プロジェクト

Masatoshi Ohishi (NAOJ)  
masatoshi.ohishi@nao.ac.jp



Mar. 22, 2005

電腦Davis ワークショップ

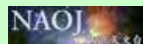
1

## JVO collaborators



### Project Scientists

NAOJ



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

ICRR



- Yasuda

Ochanomizu U.

- Masunaga



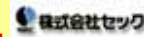
### System Engineers

Fujitsu Ltd.



- Monzen
- Kawarai
- Ishihara
- Yamazaki

SEC Ltd.



- Morita
- Nakamoto
- Kobayashi
- Yoshida

### Supporter

NII

- Miura



電腦Davis ワークショップ

2

# Nature of Astronomers



- Much higher sensitivity  
more distant objects
- New frequencies / wavelength  
unknown aspects
- Wider areas  
statistical studies
- **WE NEED MORE DATA !!**

Mar. 22, 2005

電脳Davis ワークショップ

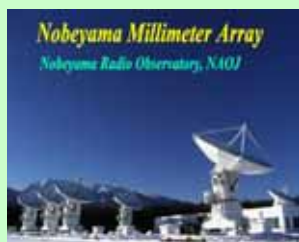
3

## Data Resources in NAOJ

- **Subaru** 8.2m Optical-Infrared Telescope
- **Kiso** 105cm Schmidt Camera
- **Okayama** 188cm Optical Telescope
- **Nobeyama 45m** Radio Telescope
- **Nobeyama Millimeter Array**
- **Nobeyama Radioheliograph**
- **VSOP**
- **VERA**
- **ALMA**



Nobeyama 45m



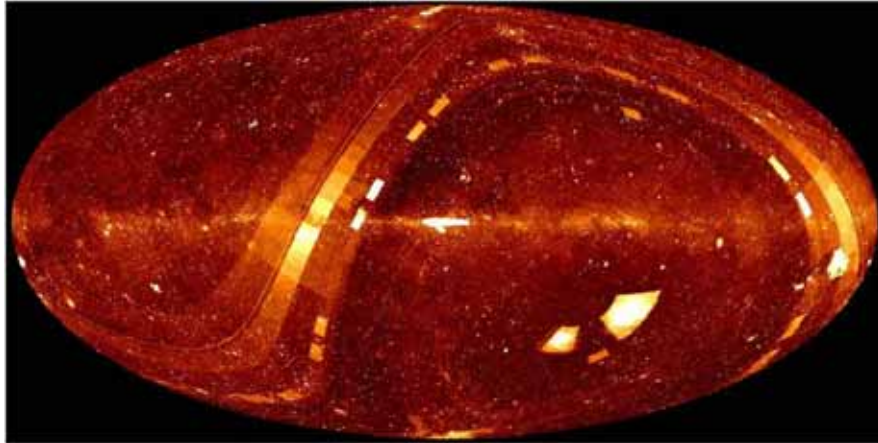
Nobeyama Millimeter Array  
Nobeyama Radio Observatory, NAOJ



Subaru



Catalogues, published tables  
Whole sky view shows heterogeneity



Mar. 22, 2005

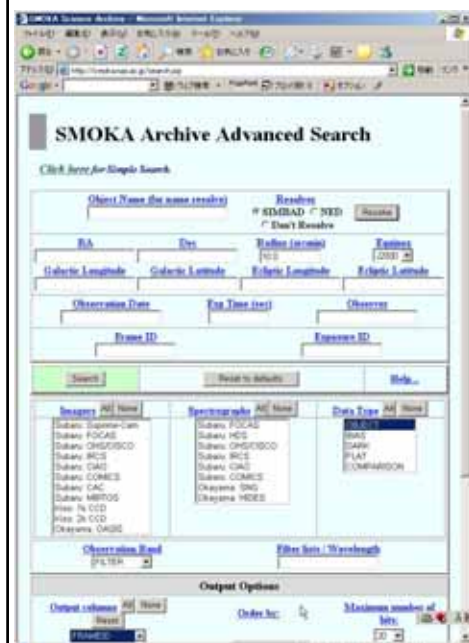
電腦Davis ワークショップ

5

*Simple Spectrum Access*

*Image / Spectrum / Catalog interoperability*

# SMOKA Archive in NAOJ



<http://smoka.nao.ac.jp/>

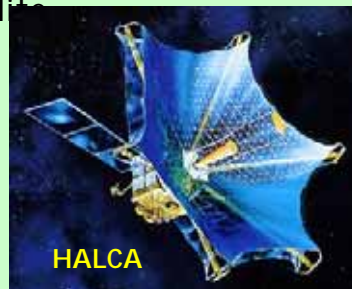
Public science archive of the

- **Subaru Telescope**,
- **188cm telescope** at Okayama Astrophysical Observatory,
- **105cm Schmidt telescope** at Kiso Observatory / University of Tokyo.

Reduced data of Subaru Suprime-Cam is now available.

## Data Resources in JAXA/ISAS

- **ASCA** X-ray astronomy satellite
- **YOHKO** solar physics satellite
- **Ginga** X-ray astronomy satellite
- **HALCA** VLBI satellite
- **Geotail** geomagnetosphere satellite
- **Akebono** aurora observation satellite
- **ASTRO-F** Infrared satellite
- **ASTRO-E2** X-ray satellite
- **SOLAR-B**

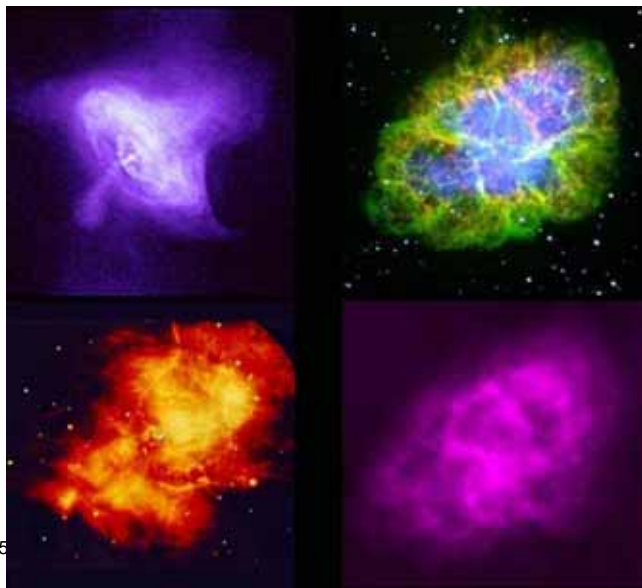




# DARTS



## Multi-wavelength astronomy



Mar. 22, 2005

10

# Data Productivity



- Nobeyama Radio : ~1TB/yr
- Subaru@Hawaii : ~20TB/yr
- ASTRO-F(will be launched)  
: several 100 GB in total
- ALMA(planned) : ~PB/yr

Flood of excellent data (survey data)  
Digitized & Archived

Mar. 22, 2005

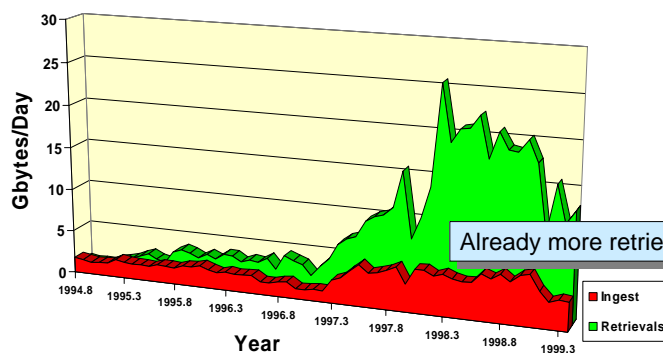
電腦Davis ワークショップ

11

# Astronomy based on Archives



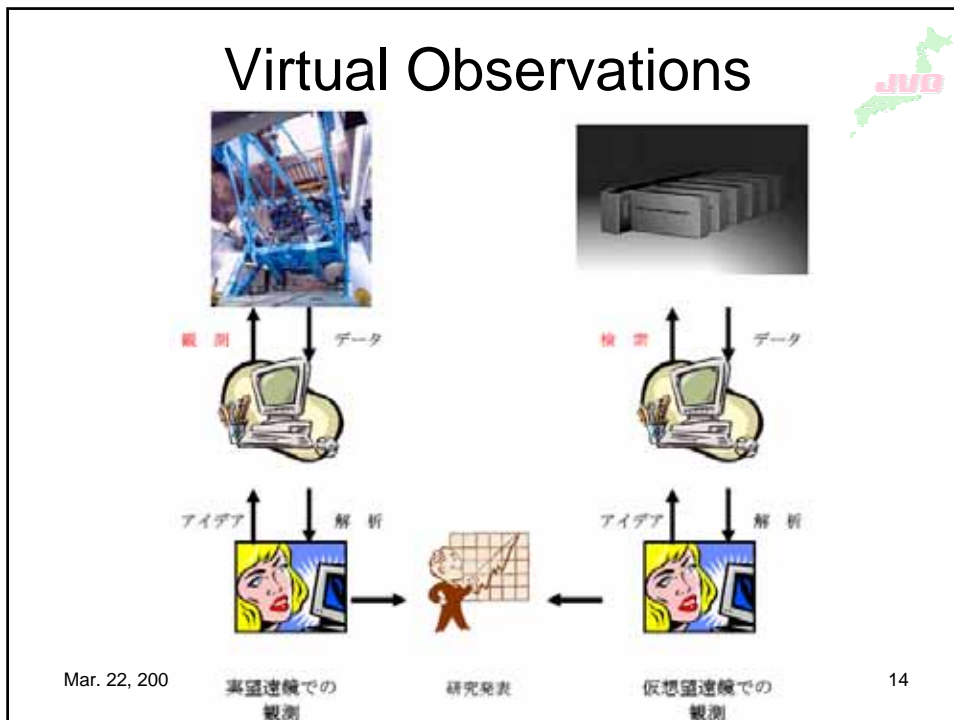
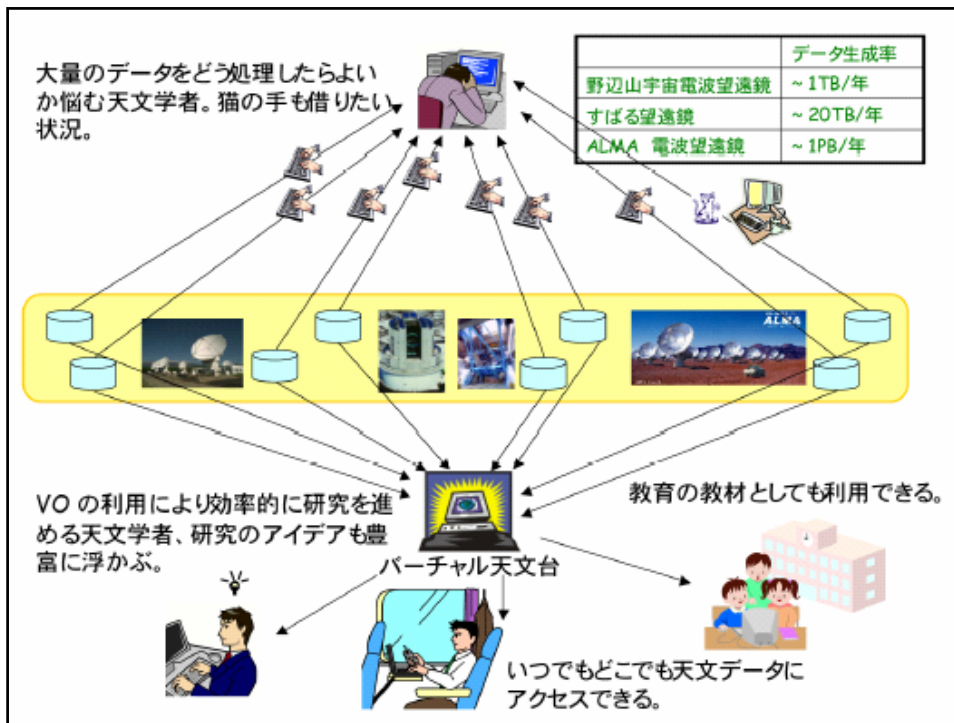
- large fraction of astro-papers based on archives
- HST archive use growing faster than archive



Mar. 22, 2005

電腦Davis ワークショップ

graphics from  
US NVO  
project 12



# 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

Mar. 22, 2005

電脳Davis ワークショップ

15

# VO Projects in the world



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



Mar. 22, 2005

電脳



# Standardization in IVOA

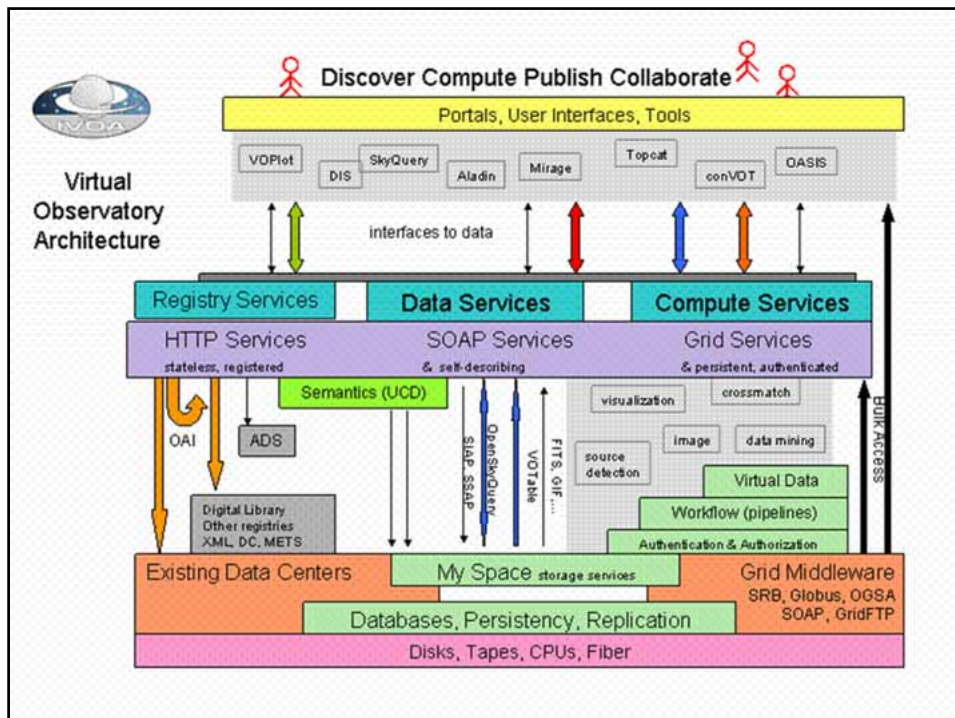


- Query language to distributed DBs (VOQL)
- Meta-data : contents, protocol to interchange based on OAI-PMH
- Protocols to retrieve images, spectra, and so on SkyNode, SIAP, SSAP, STC, etc.
- Unified attribute names in DBs UCD (Unified Contents Descriptions)
- Output Format : VOTable (XML) incorporates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

Mar. 22, 2005

電腦Davis ワークショップ

17



## IAU XXVth GA Res. (2003 Jul.)



### Recommends

1. that, **data obtained at major astronomical facilities should**, after a reasonable proprietary period in which they are available only to observers or other designated users of the facility, **be placed in an archive where they may be accessed via the internet** by all research astronomers. As far as possible, the data should be accompanied by appropriate metadata and other information or tools to make them scientifically valuable,
2. that, **such data should not be subject to intellectual property rights**. The form in which data are made available, and the subsequent processing of such data, may be appropriately protected by copyright laws, but the fair usage (including educational purposes) of the archive data themselves should not be subject to restrictions,
3. that, **funding agencies provide encouragement and support to enable data produced by astronomical research** that they fund to be deposited, after some proprietary period as defined above, in recognized data archives which provide unrestricted access to **these data**.

Mar. 22, 2005

電腦Davis ワークショップ

19

## OECD Rec. ('04 Aug)



### Recommendations:

New projects and facilities must take the data management, storage, maintenance, and dissemination into account at the earliest planning stages, consulting potential users in the process. Agencies and governments should consider **adopting the IAU resolutions** as the basis for progress in this field.

Agencies should recognise that **this is an important long-term issue and should coordinate plans, provide adequate funding on a long-term basis, and support development and maintenance of the needed infrastructure**.

Agencies should encourage broadening of existing VO collaboration into a fully **representative global activity**.

Mar. 22, 2005

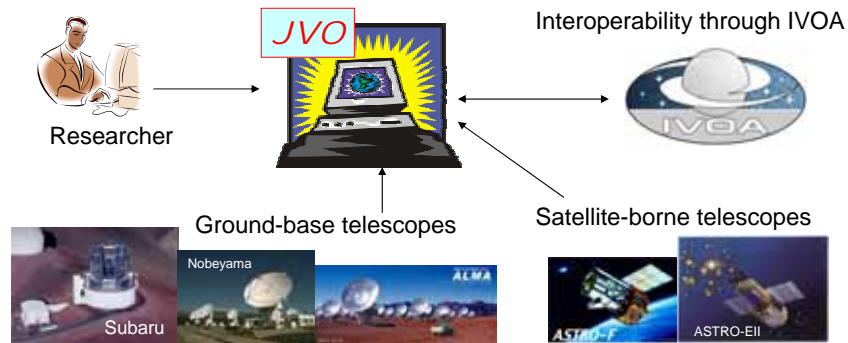
電腦Davis ワークショップ

20



# JVO : Japanese Virtual Observatory

- Purpose:
  - Easy access to federated Astronomical databases
  - Interoperability through IVOA



Mar. 22, 2005

電腦Davis ワークショップ

21

# Development history

- JVO Project start April 2002
- Prototype 1 finish March 2003
- Prototype 2 finish March 2004
- Prototype 3 partially finished

Mar. 22, 2005

電腦Davis ワークショップ

22

# Development Strategy



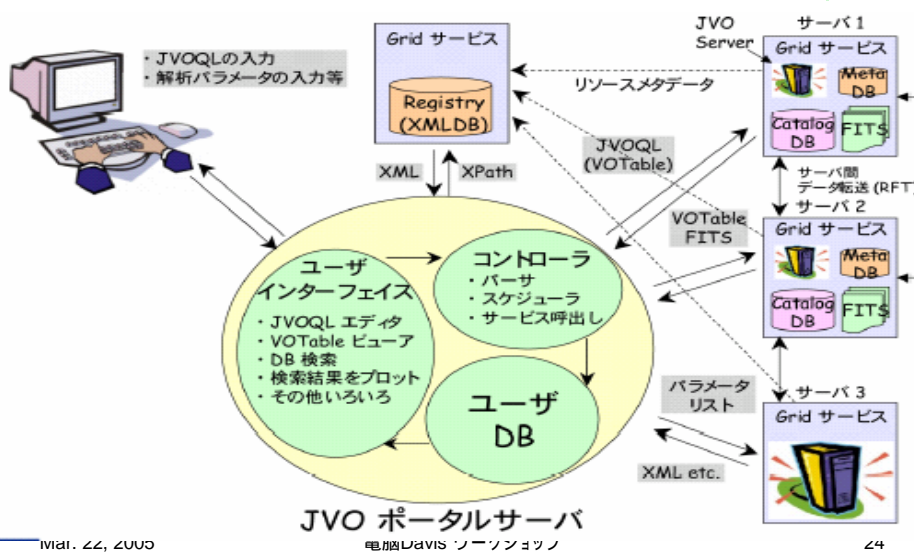
- “Top-down” approach
  - set “Science Use Cases”
  - study and design “Overall System”
    - How to federate distributed computers?
  - build “Prototype System”
    - with minimal capabilities
    - to achieve use cases
- “Build-and-scrap” prototype

Mar. 22, 2005

電脳Davis ワークショップ

23

# Schematic diagram of JVO



Mar. 22, 2005

電脳Davis ワークショップ

24

## Design toward Prototype 3

- Support **IVOA Standard Protocols**
  - SIAP, SSAP, SkyNode
  - implement ADQL
- improve registry, employing **OAI-PMH** architecture
- flexible workflow architecture
- introduce User management
  - LDAP
  - User Storage Area (support VOStore?)
- API to control JVO with SOAP

Mar. 22, 2005

電腦Davis ワークショップ

25

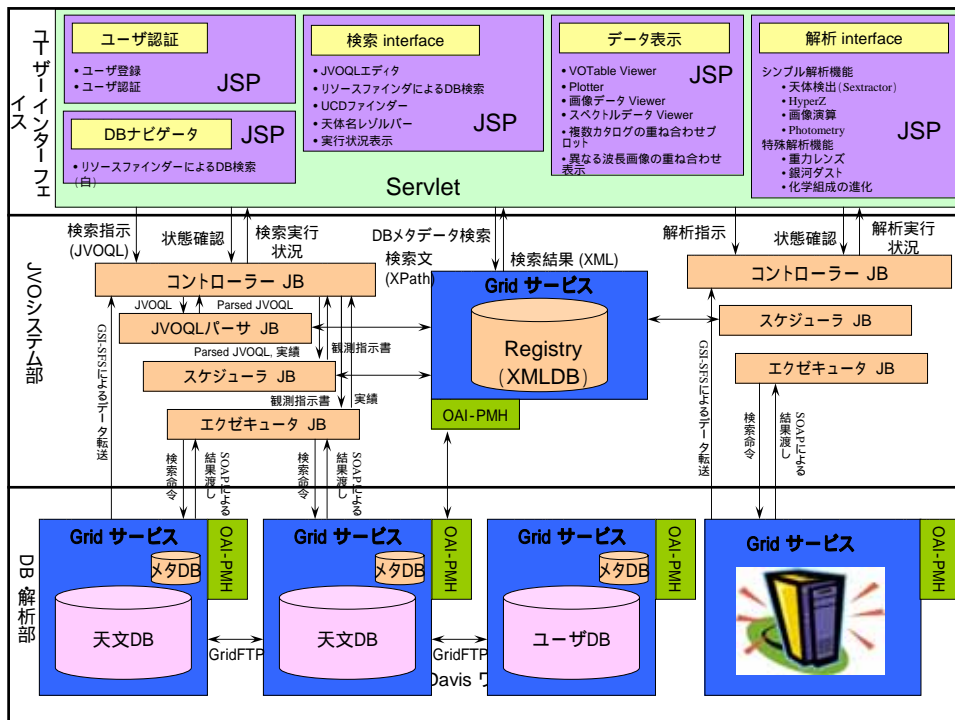
## Science targets in prototype 3

- Cosmic String Search through the gravitational lens effect.
- Environmental dependence of QSO evolution.
- Automatic classification of late type star.
- Search for very metal-poor stars.
- etc.

Mar. 22, 2005

電腦Davis ワークショップ

26

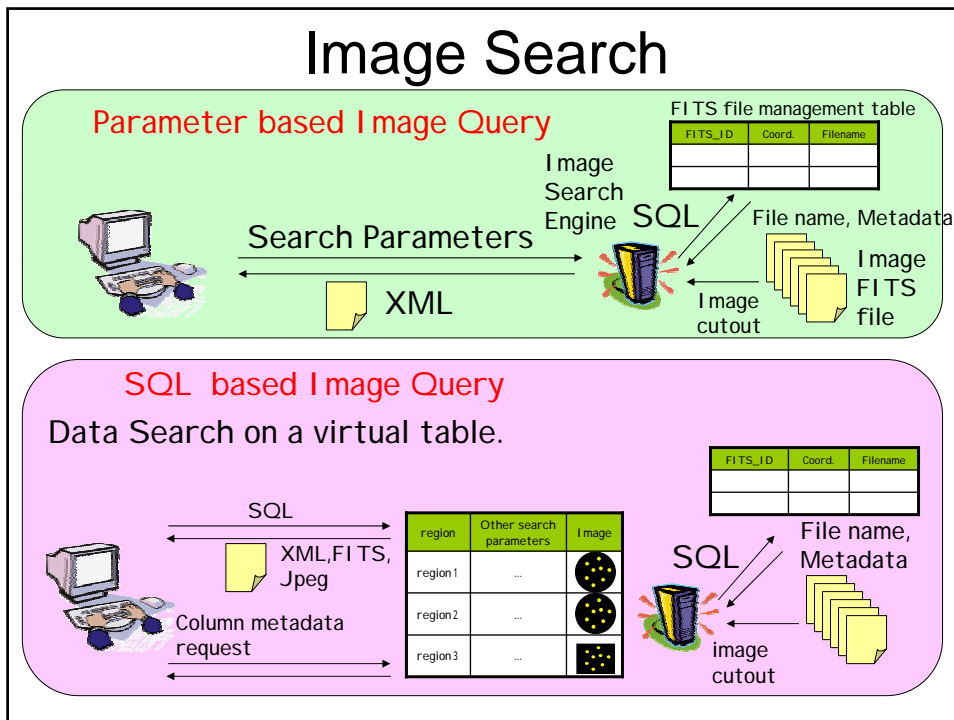
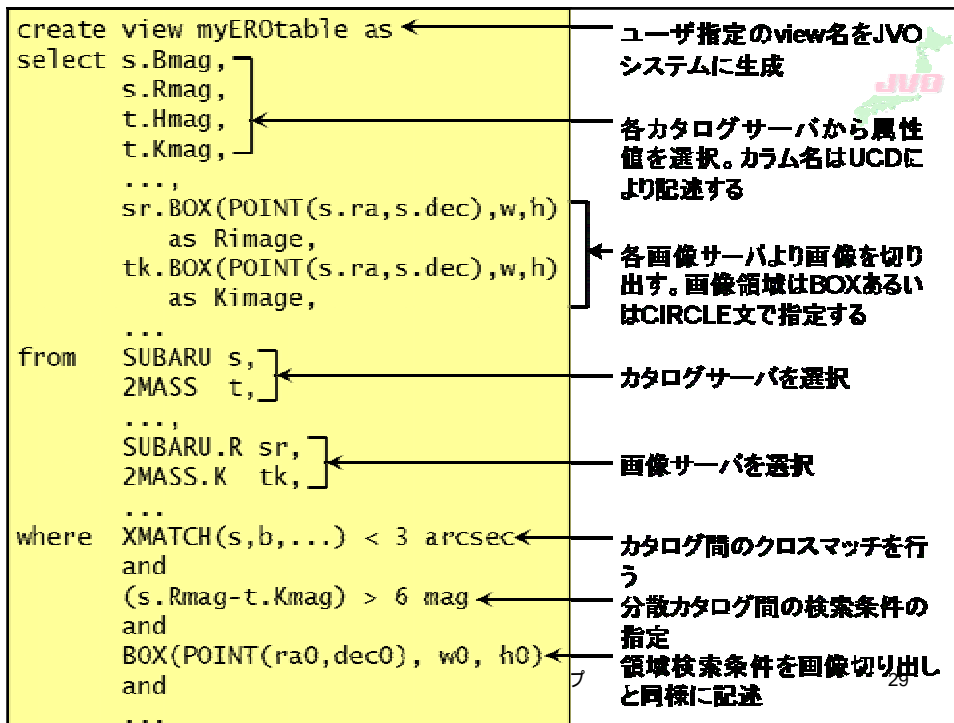


## JVO Query Language (JVOQL)

JVOQL is designed as a prototype of VO Query Language.

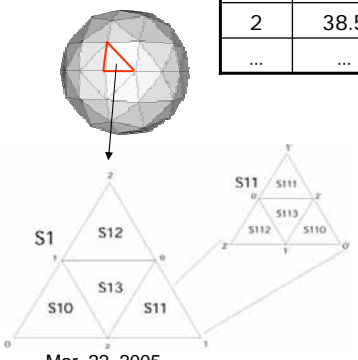
Characteristics of the JVOQL:

- **SQL** based Query Language
- Query Language for the **distributed astronomical DB**.
- Can search and retrieve **observational data** as well as catalog data
- Upward compatible with the **ADQL** and **SIAP** syntax.
- **Scalable** syntax structure. Very simple core syntax and extension syntax packages.



## Region Search using HTM index

Region search is a common search criterion for an astronomical database. For efficient search data should be properly indexed on the object coordinate.



Mar 22, 2005  
<http://www.sdss.jhu.edu/htm/>

id	ra	dec	mag
1	12.3	-23.4	18.4
2	38.5	+34.2	16.5
...	...	...	...

id	htm
1	16522516
2	16754765
...	...

Select ra, dec, mag  
 From Catalog  
 Where Point(ra,dec) within Box((20,+15), 1.0)

Select c.ra, c.dec, c.mag  
 From Catalog as c  
 Natural Left Join htmIndex as i  
 Where i.htm between 16522500 and 16522512  
 OR  
 i.htm between 16522500 and 16522512  
 ...

## Parsing JVOQL and Generating Workflow



- “JVOQL Parser” generates query for each host
- “Scheduler” generates:
  - count query job for host1
  - count query job for host2
- “Executer” executes jobs on remote hosts
- “Scheduler” generates based on the result of execution
  - query job for host1
  - xmatch job for host2
  - image query for host1 and host2
- “Executer” executes jobs on remote hosts

Mar. 22, 2005

電腦Davis ワークショップ

32



# Remote execution



- Proto1:
  - Globus toolkit ver 2
  - using `globus-job-run` command
  - 1 call = 30 sec
    - 1 query ~10 min!
- Proto2:
  - Globus toolkit ver 3
  - using Grid Service
  - 1 call = 2-3 sec
    - overhead time is only ~ 30 ms

Mar. 22, 2005

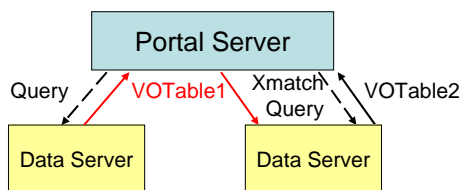
電腦Davis ワークショップ

33

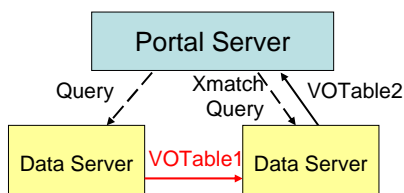
# Data Transfer



- Web/Grid Service
  - Query result is always returned to Portal server



- GridFTP
  - Query result can be directly transferred to XMatch server



Mar. 22, 2005

電腦Davis ワークショップ

34

## Data Transfer



- We tried:
  - GridFTP, RFT
  - GSI-SFS
- Learned:
  - SFS needs to be modified
    - A server cannot be a client.
  - RFT is promising in Globus environment
  - need support for HTTP and Web Services

Mar. 22, 2005

電腦Davis ワークショップ

35

## Meta Data



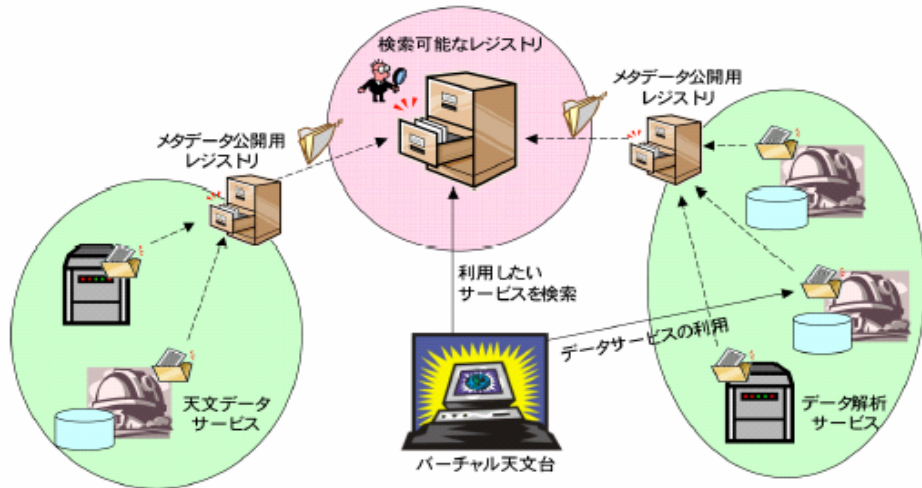
- Two Categories
  1. Locations of Data Server
    - Necessary info to search for appropriate Servers
  2. Contents of Data Server
    - Similar to FITS header
    - Instruments, Wavelengths, Dates, Areas of sky , ...
- Implementations
  - Prototype 1: UDDI
    - Locations only; impossible to query contents
  - Prototype 2: XMLDB
    - Based on IVOA Standards (XML format)
    - Queries are made through XMLDB (Karearea)

Mar. 22, 2005

電腦Davis ワークショップ

36

# Exchange of Meta Data: OAI-PMH



Mar. 22, 2005

电脑DAVIS ワークショップ

37

# Resource Metadata

- Identity metadata
- service metadata
- column metadata
- curation metadata
- content metadata

title	string	subject	string,list
short_name	string	description	string,free text
identifier	URI	source	string
publisher	string	reference_url	URL
publisher_id	URI	type	string,list
creator	string	coverage_spatial	string
creator_logo	URL	coverage_region_of_regard	float,decimal degrees
contributer	string	coverage_spectral	string,list
date	string	coverage_spectral_bandpass	string,list
version	string	coverage_spectral_central_wavelength	float
contact_name	string	coverage_spectral_minimum_wavelength	float
contact_email	e-mail address	coverage_spectral_maximum_wavelength	float
service_interface_url	URL	coverage_temporal_start_time	string
service_base_url	URL	coverage_temporal_stop_time	string
service_http_result	MIME type	coverage_depth	float
service_standard_uri	URI	coverage_depth_unit	string
service_standard_url	URL	coverage_object_density	float
service_msr	float,decimal degrees	coverage_object_count	int
ucd	string	coverage_sky_fraction	float
unit	string	resolution_spatial	float
datatype	string	resolution_spectral	float
width	int	resolution_temporal	float
precision	string	content_level	string,list
arraysize	string	facility	string,list
		instrument	string,list
		format	string,list
		right	string

	identity	curation	service	content	column
catalog					x
table					x
column					

Mar. 22, 2005

电脑DAVIS ワークショップ

38

# Access to Archives



Search

Results

The screenshot shows two browser windows. The left window, titled 'DB Search', displays a search interface with fields for 'User', 'Total memory', and 'Used memory'. It includes sections for 'Simple Search' and 'Advanced Search' with various filters like 'Wavelength' and 'Telescope'. The right window, titled 'Results', shows a 'Catalog List' with a table of search results. The table has columns for 'Title' and 'Description'. Below the table is the JVO logo and the text 'JAPANESE VIRTUAL OBSERVATORY'.

Title	Description
Hydra Digital Sky Survey	Hydra Digital Sky Survey (DSS)
Hydra SDSS Survey	The Hydra SDSS Survey Deep Survey (SDSS) is a wide area multi-wavelength survey of a ~1 square-degree region of the
Hydra Deep Field (band)	The Hydra Deep Field (band Catalog (TEXT)
Hydra Deep Field (band)	The Hydra Deep Field (band Catalog (TEXT)

Mar. 22, 2005

電腦Davis ワークショップ

39

# Automatic Generation of Queries

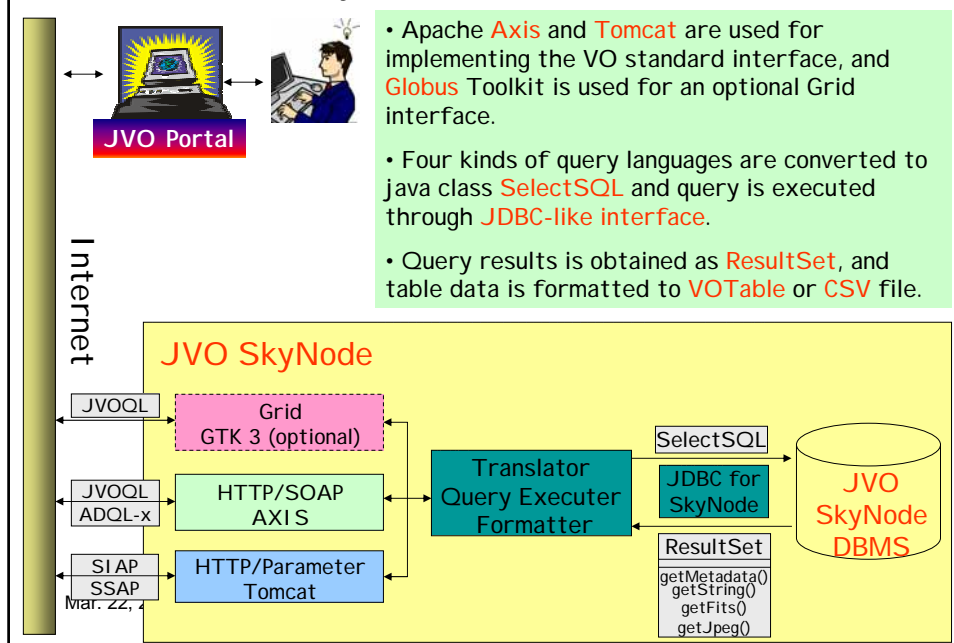


The screenshot shows the 'JVO Query Language Editor' interface. It includes a section for '0. SQL Editor' with a text area containing SQL code. Below this are sections for '1. Region Selection' and '2. Catalog Selection', each with input fields and buttons for defining search parameters.

Mar. 22, 2005

40

# JVO SkyNode Architecture



## Free software used in JVO Skynode

- **Java (J2SE)** : Generally used in the development.
- **JAXB 1.0**: used for generating Java class files from VO standard schema, ADQL, VOTable, VOResource etc...
- **JavaCC** : used for [parsing JVOQL](#) and constructing SelectSQL java object.
- **PostgreSQL** : Backend DBMS.
- **HTM library** : developed by JHU, used for [region search](#).
- **Apache AXIS, Tomcat** : Web service and servlet.
- **Globus Toolkit** : Grid service.
- etc...

## DBs available in JVO



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

Mar. 22, 2005

電脳Davis ワークショップ

43

## Analysis Tools



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

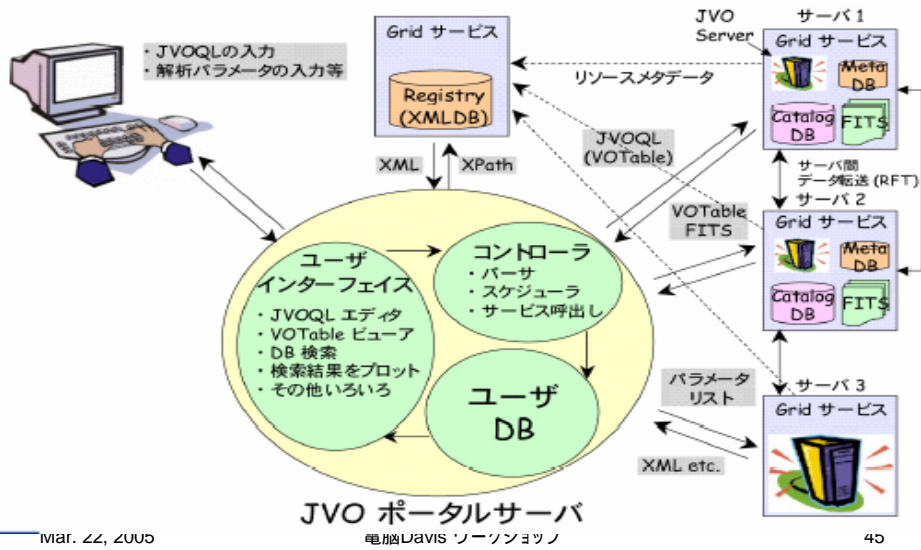
Mar. 22, 2005

電脳Davis ワークショップ

44



# Schematic diagram of JVO



# 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

# JVO is seen from the UK VO

**AstroGrid Registry**

**Registry Browser**

Version: 0.9  
Find IVORNs including:    
Browse for another version 0.9

Title	Type	AuthorityID	ResourceKey	Up
JVO Publishing Registry	vg:Registry	jvo	publishingregistry	20018
JVO Publishing Registry	vg:Registry	jvo	publishingregistry	20021
The Subaru/3MM-Newton Deep Survey (SDSS) SkyNode Service	sn:OpenSkyNode	jvo/skynode	snids	20020
The Subaru/3MM-Newton Deep Survey G1	sn:OpenSkyNode	jvo/skynode	snids	20020
JVO	vr:Organisation	jvo	jvo	20018
The Subaru/3MM-Newton Deep Survey (SDSS) SIA Service	sia:SimpleImageAccess	jvo/siaj	snids	20020
JVO Authority	vg:Authority	jvo	auth	20021

Mar. 22, 2005

47

**JVO Data Search**

[Home](#) | [Registry](#) | [Search](#) | [Login](#) | [Logout](#)

User ID	User Name	Group	Last Login
ohishi	Masatoshi Ohishi	jvo	Tue Jan 11 23:07:19 JST 2005

Total memory = 155742kB Used memory = 118748kB (76%)

**JVOQL**

Please select

**Search Region**

Target:   Coordinate:   
Size:  [deg]

Coordinates by VOTable:

Mar. 22, 20

48

# JVO Searchable Registry

[Status](#) | [Registry](#) | [Search](#) | [Result](#) | [Logout](#) |

User ID	User Name	Group	Last Login
ohishi	Masatoshi Ohishi	jvo	Tue Jan 11 23:07:19 JST 2005

Total memory = 155742kB Used memory = 113499kB (72%)

## Search

AND

## Registries

## Data Services

## Harvest

## Register Metadata

## Remove Metadata

## Get Service Information by Identifier



59	<input type="radio"/>	<a href="#">More Info</a>	HST Hubble HELIX Observations	SIAP	<a href="#">URL</a>		
60	<input type="radio"/>	<a href="#">More Info</a>	The ALADIN image server	SIAP	<a href="#">URL</a>		
61	<input type="radio"/>	<a href="#">More Info</a>	Digitized Sky Survey 1 - Red	SIAP	<a href="#">URL</a>		
62	<input type="radio"/>	<a href="#">More Info</a>	Infrared Space Observatory Simple Spectrum Data Access	SIAP	<a href="#">URL</a>		
63	<input type="radio"/>	<a href="#">More Info</a>	Hubble Space Telescope Faint Object Spectrograph	SIAP	<a href="#">URL</a>		
64	<input type="radio"/>	<a href="#">More Info</a>	Sloan Digital Sky Survey Simple Spectrum Data Access	SIAP	<a href="#">URL</a>		
65	<input type="radio"/>	<a href="#">More Info</a>	JVO Publishing Registry	Registry	<a href="#">URL</a>		
66	<input type="radio"/>	<a href="#">More Info</a>	NCSA Radio Astronomy Imaging Registry	Registry	<a href="#">URL</a>		
67	<input type="radio"/>	<a href="#">More Info</a>	Minnesota Automated Plate Scanner	Registry	<a href="#">URL</a>	unknown	

Select the checked service and go to the search page.





## JVO Data Search

[Status](#) | [Registry](#) | [Search](#) | [Result](#) | [Logout](#) |

User ID	User Name	Group	Last Login
ohishi	Masatoshi Ohishi	jvo	Tue Jan 11 23:07:19 JST 2005

Total memory = 155742kB Used memory = 117026kB (75%)

### JVOQL

```
SELECT *
FROM irsa.ipac.caltech.edu:ISSA
WHERE region = BOX( 83.633212,22.014460), 0.2, 0.2)
```

Mar. 22, 2005

電腦Davis ワークショップ

51



## JVO Data Search Status

[Status](#) | [Registry](#) | [Search](#) | [Result](#) | [Logout](#) |

User ID	User Name	Group	Last Login
ohishi	Masatoshi Ohishi	jvo	Tue Jan 11 23:07:19 JST 2005

Total memory = 155742kB Used memory = 124641kB (80%)


Obs. Name	Process ID	Server	Flag	Elapsed Time (sec)	Status
OBS_20050111233330	proc_0001	irsa.ipac.caltech.edu		3.341	OK <a href="#">VO Table</a> <a href="#">URLs</a>



Mar. 22, 2005

電腦Davis ワークショップ

52

Status | [Registry](#) | [Search](#) | [Result](#) | [Logout](#) | 

User ID	User Name	Group	Last Login
ohishi	Masatoshi Ohishi	jvo	Tue Jan 11 23:07:19 JST 2005

Total memory = 155742kB Used memory = 117013kB (75%)

Obs. Name : OBS\_20050111233330

Show Selected Images | Reset

check	download		POS_EQ_RA_MAIN	POS_EQ_DEC_MAIN		
<input type="checkbox"/>	<a href="#">Download</a>	1600	84.7414377	20.0273093	05h 38m 57.95s	+20d 01m 38.3s
<input type="checkbox"/>	<a href="#">Download</a>	1601	84.7414377	20.0273093	05h 38m 57.95s	+20d 01m 38.3s
<input type="checkbox"/>	<a href="#">Download</a>	1602	84.7414377	20.0273093	05h 38m 57.95s	+20d 01m 38.3s
<input type="checkbox"/>	<a href="#">Download</a>	1603	84.7414377	20.0273093	05h 38m 57.95s	+20d 01m 38.3s

Show Selected Images | Reset

Mar. 22, 2005 電脳Davis ワークショップ

## Cosmic Strings?

- Theoretical origin of the elementary particles which existed (?) at  $t \sim 10^{-35}$  s ,  $T \sim 10^{15}$  GeV
- Not a POINT, has its SIZE
- Width  $< 10^{-22}$ m, Length  $\sim$  size of the Universe.
- Mass of 10km string  $\sim$  the Earth

Mar. 22, 2005 電脳Davis ワークショップ 54

## Gravitational Lens

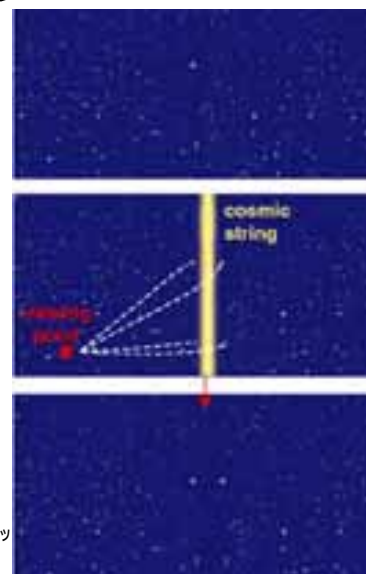
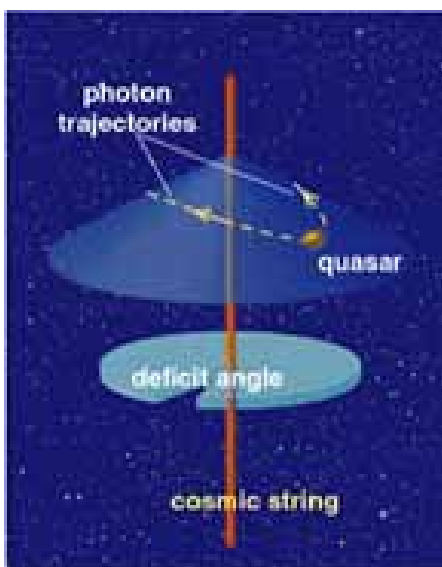


- A phenomenon that the space-time is distorted by a huge mass object, and multiple images of another object behind the massive object are observed around it.



Mar. 22, 200

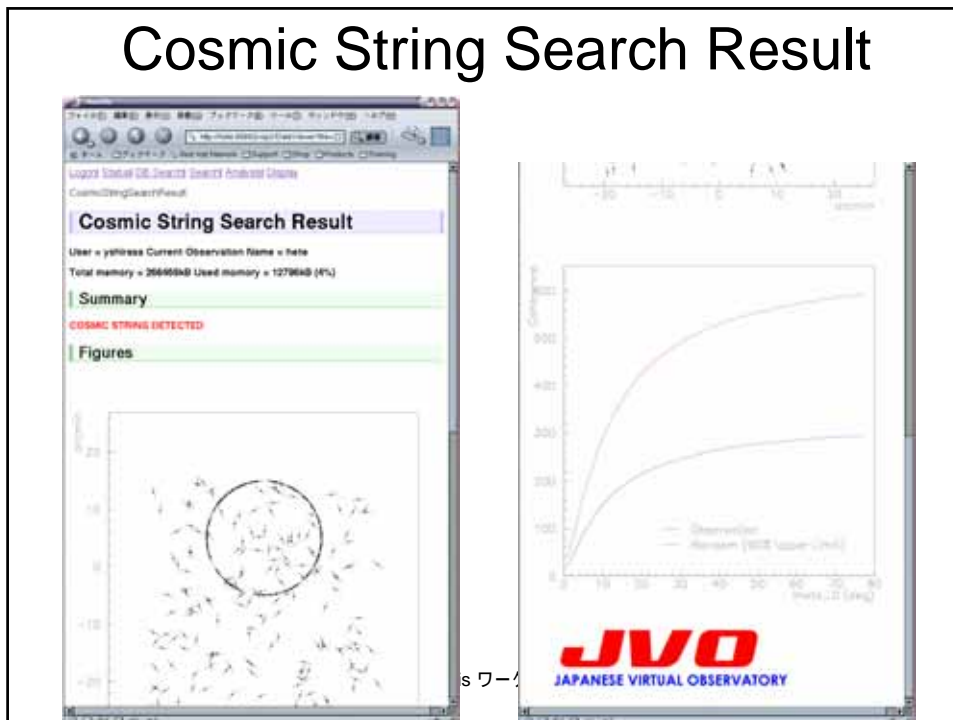
## Gravitational Lens by Cosmic Strings



クシヨツ



# Cosmic String Search Result



## Work flow for Gravitational lens search

1. Retrieve Subaru catalog data in a specified region.
2. Calculate brightness.
3. Define condition to select quasars.
4. Make a list of pair quasar objects.
5. Retrieve image data of the pair objects.
6. Narrow candidates by analyzing the image data.

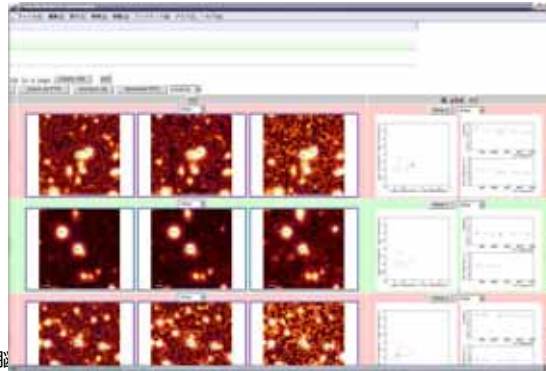
Mar. 22, 2005

電腦Davis ワークショップ

58

## Search for Gravitational Lenses produced (?) by Cosmic Strings

- SXDS data observed by Subaru
- Query results were obtained less than **5** min, displaying SEDs
- It has been proven that VO can accelerate researches.



Mar. 22, 2005

電腦

## QSO銀河探查のフロー

- QSO候補カタログに対応しそうな光で見える銀河の画像をすばる望遠鏡のDBから探す
- 得た各銀河のパラメータを読み取る
- 銀河の位置がQSOと一致するものをオーバープロットする

Mar. 22, 2005

電腦Davis ワークショップ

60

# 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

( $R < 25$ )

電脳Davis

Mar. 22, 2005

Astronomy Astrophysics

**Discovery of optically faint obscured quasars with Virtual Observatory tools**

F. Padovani<sup>1</sup>, M. G. Alford<sup>2</sup>, F. Beudef<sup>3</sup>, and N. A. Wilson<sup>4</sup>

<sup>1</sup> *IFGC European Southern Observatory, Karl Schwarzschild Str. 1, 85748 Garching bei München, Germany  
 e-mail: [Padovani@eso.org](mailto:Padovani@eso.org)*  
<sup>2</sup> *Centre de Cosmologie et d'astrophysique de l'Observatoire de Strasbourg, 21 rue de l'Université, 67081 Strasbourg, France  
 e-mail: [alford@uha.fr](mailto:alford@uha.fr)*  
<sup>3</sup> *European Southern Observatory, Karl Schwarzschild Str. 1, 85748 Garching bei München, Germany  
 e-mail: [Beudef@eso.org](mailto:Beudef@eso.org)*  
<sup>4</sup> *Centre de Cosmologie et d'astrophysique de l'Observatoire de Strasbourg, 21 rue de l'Université, 67081 Strasbourg, France  
 e-mail: [wilson@uha.fr](mailto:wilson@uha.fr)*

Received 11 April 2004 / accepted 1 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-Gaia-Observation (2GO) survey (2GO AGN). By exploiting publicly available X-ray and optical data and catalogues we discover 88 type-2 AGN candidates. Our X-ray search is complemented by a previously unattempted image co-alignment between X-ray templates and 30 optical filters sets. The use of our co-aligned high-resolution images ( $\sim 1''$  resolution) and multi-wavelength, optically obscured images, we select 2GO type-2 AGN candidates (we consider only those with a minimum high redshift:  $z > 0.5$  with the QSO 1 + 2 + 4 fit). We report 61 new high-redshift type-2 AGN in the two-Gaia-Observation (2GO) survey, a sample of 61 optically obscured, optically obscured, optically obscured type-2 AGN candidates. This list is a first step towards a complete sample of optically obscured, optically obscured type-2 AGN. We also report 27 type-2 AGN candidates in the 2GO survey. Our study shows that the use of Virtual Observatory (VO) tools is now possible to produce cutting-edge science results by exploiting astronomical data beyond classical identification limits.

**Key words:** cosmology – galaxies: active – methods: statistical – galaxies: general – X-ray galaxies

**1. INTRODUCTION**

The field of active galactic nuclei (AGN) is largely unexplored (e.g. Ueda & Padovani 2004), with only a few recent results by Ueda et al. (2004). The apparent disparate population and composition of active galaxies can be explained by the physical conditions, accretion disk, jet, and obscuring torus associated with the activity of the emitting galaxy. Type-2 sources are those in which we have an unobscured view of the central engine and the emission from the accretion disk of AGN with an absorption. Type-2 objects are those that are obscured by the torus. While many examples of local and moderate redshift type-2 AGN are known (the largest 30), it has been argued that high-redshift type-2 AGN are abundant, but optically obscured, redshifted type-2 AGN are likely to be very rare, very few, if any, examples of this class are known. Again from their importance for

AGN models, type-2 sources are expected to make a significant fraction of the X-ray background flux, e.g. Comastri et al. (2004) estimate that they contribute roughly 50% of the background. Their discovery is therefore a key step towards the “standard” spectral method of galaxy selection. The hard X-ray sources are thought to be able to penetrate the torus. Type-2 QSOs therefore, should have narrow, if any, permitted lines and require the use of narrow-line emission and broad, permitted hard X-ray emission, and, in some cases, a high redshift redshift ( $z > 0.5$ ) (e.g. Ueda et al. 2004).

In this paper we use Virtual Observatory (VO) tools to identify 88 type-2 AGN candidates in the two-Gaia-Observation (2GO) survey (2GO AGN). By exploiting publicly available X-ray and optical data and catalogues we discover 88 type-2 AGN candidates. This list is a first step towards a complete sample of optically obscured, optically obscured type-2 AGN. We also report 27 type-2 AGN candidates in the 2GO survey. Our study shows that the use of Virtual Observatory (VO) tools is now possible to produce cutting-edge science results by exploiting astronomical data beyond classical identification limits.

# 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

Mar. 22, 2005

電脳Davis

The Astrophysical Journal, 610:1269–1286, 2004 December 1  
 © 2004 The American Astronomical Society. All rights reserved. Printed in the USA.

AUTOMATED CLASSIFICATION OF ROSAT SOURCES USING HETEROGENEOUS MULTIWAVELENGTH SOURCE CATALOGS

T. A. McGlynn<sup>1</sup>, A. A. Scoville<sup>2</sup>, E. L. Wynn<sup>3</sup>, R. J. Hanna<sup>4</sup>, R. L. Wynn<sup>5</sup>, F. Ostriker<sup>6</sup>, S. Gezari<sup>7</sup>, W. York<sup>8</sup>, M. F. Corwin, Jr.<sup>9</sup>, A. Chak<sup>10</sup>, and M. Courteau<sup>11</sup>

Accepted 2004 February 10; accepted 2004 August 1

**ABSTRACT**

We describe an online system for automated classification of X-ray sources, CLASS-X, and present preliminary results of classification of the three major catalogues of ROSAT sources: ROSAT All-Sky Survey (RASS) Bright Source Catalogue, RASS First Source Catalogue, and WGACAT, into six clear categories: stars, white dwarfs, X-ray binaries, galaxies, active galactic nuclei, and clusters of galaxies. CLASS-X 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 physical type. Each source is characterized by a provided set of parameters, or attributes; for some set to them used as the classifier conducts classification of sources of unknown identity. The CLASS-X pipeline features an automatic search for X-ray source counterparts among heterogeneous data sets in widely data archives using Virtual Observatory protocols; it retrieves from those archives all the attributes required by the selected classifier and reports them to the classifier. The user input to CLASS-X 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 manual validation and external verification, we conclude that the CLASS-X classifier yields meaningful and reliable classifications for ROSAT sources and have the potential to broaden their 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, heterogeneous sets of observations. Many recently automated classifier tools have been used to help in the classification of objects from large but still largely homogeneous surveys. Examples include analysis of the First (Schlafly, 1997) and Second (Wu et al. 1993) Digital Sky Surveys and the Sloan Digital Sky Survey (SDSS; Schlafly 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 CLASS-X team has developed a series of classifiers that rapidly sort X-ray source sets into classes. These facilities are now available to the community at the CLASS-X Web site.<sup>1</sup> Classification is derived from correlation and identification with objects at other wavelengths. Our classification tools can use the nonredundant counterparts of other wavelengths or

we ensembles of potential counterparts to establish links to counterparts.

Our initial work has concentrated on the more than 100,000 unidentified sources listed by the ROSAT Observatory<sup>2</sup> 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 sources have been used in previous catalogs (e.g., Padovani 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 stems from correlations between the ROSAT optical tables of classified sources. In some cases detailed follow-up observations have been performed in a search for source links. This is extraordinarily expensive in both telescope time and the time of astronomer early-career staff. Direct comparison of ROSAT sources with massive optical catalogs (e.g., Padovani et al. 2000, or the entire 6th or 7th SDSS data set; Padovani et al. 2001; York et al. 2001), and Lauer et al. 2002) enables the cross-identification of ROSAT sources, but neither the class of the counterpart is known, nor does it determine the type of the source. However, this information from multiple catalogs allows us to try to classify sources with more information than is available from the X-ray observations alone.

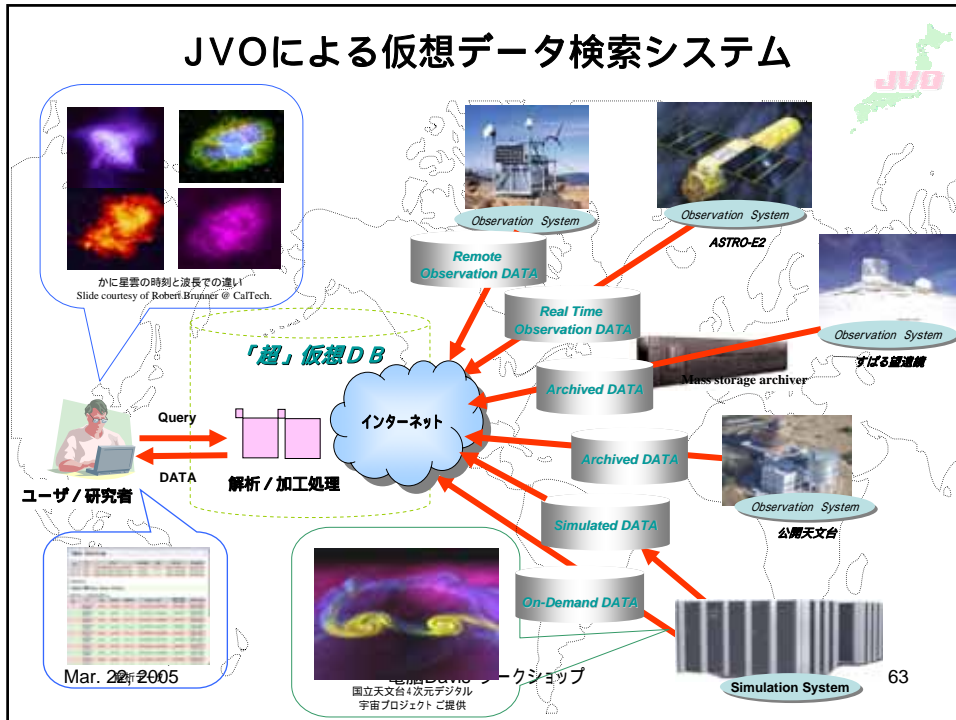
Our approach differs from most previous efforts at multi-object classification in several basic ways. First it does not specifically contain the information that is used to distinguish our output categories. Other authors have looked at the X-ray to optical ratios (Maccarini et al. 1988) or X-ray/optical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

Mar. 22, 2005

電脳Davis

2

# JVOによる仮想データ検索システム



# ALMA in Chile (sub-mm wave)



Mar. 22, 2005

電腦Davis ワークショップ

64



END

Mar. 22, 2005

電腦Davis ワークショップ

65