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

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

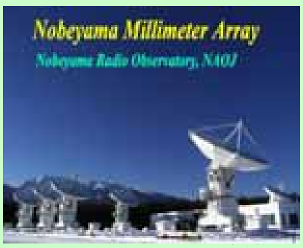
•Miura

Nature of Astronomers

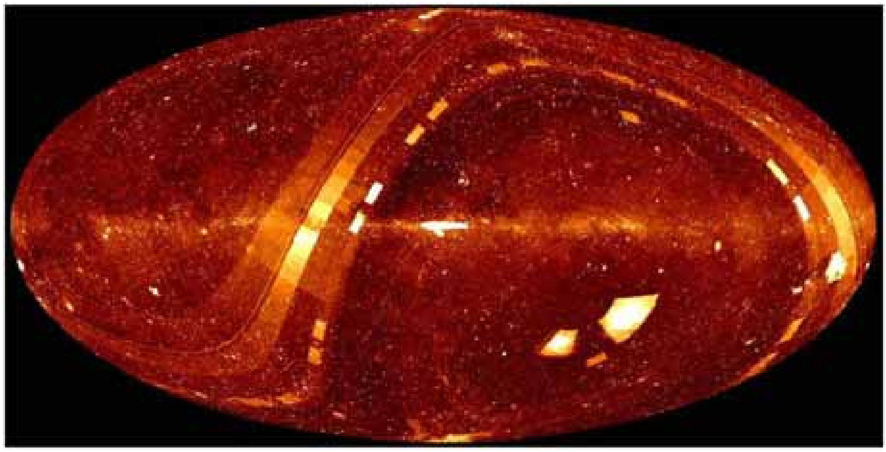
- Much higher sensitivity
  → more distant objects
- New frequencies / wavelength
  → unknown aspects
- Wider areas
  → statistical studies
- WE NEED MORE DATA !!

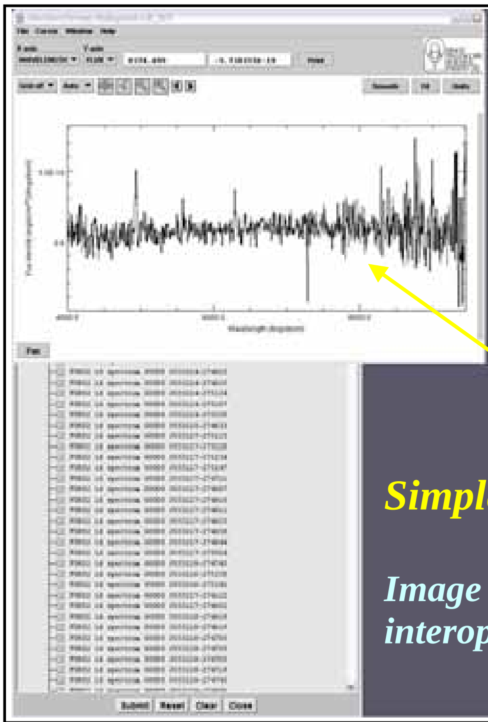
Data Resources in NAOJ

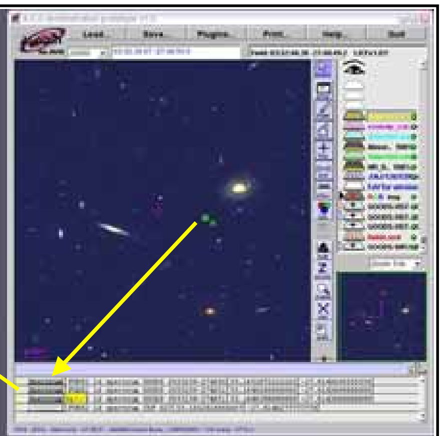
- Subaru 8.2m Optical-Infrared Telescope
- Kiso 105cm Schmidt Camera
- Okayama 188cm Optical Telescope
- Nobeyama 45m Radio Telescope
- Nobeyama Millimeter Array
- Nobeyama Radioheliograph
- VSOP
- VERA
- *ALMA*



Catalogues, published tables
Whole sky view shows heterogeneity



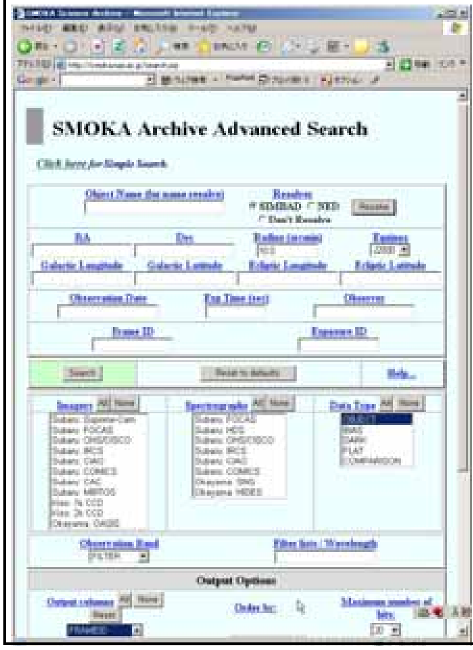




Simple Spectrum Access

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

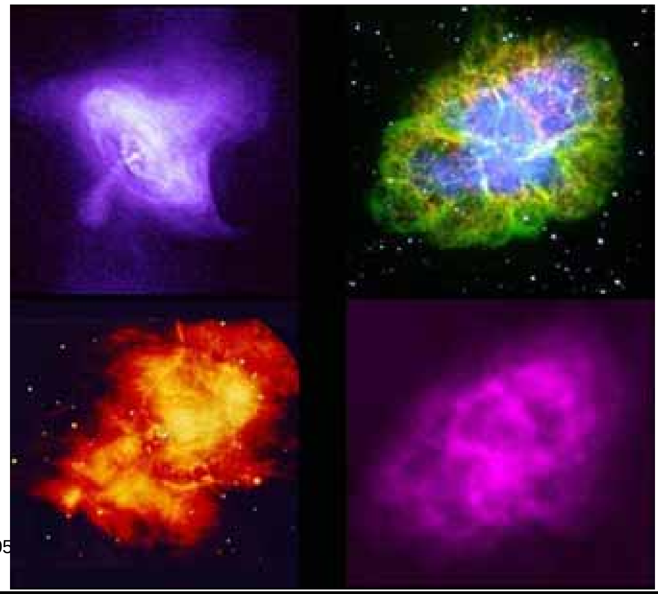
- ASCA X-ray astronomy satellite
- YOHKO solar physics satellite
- Ginga X-ray astronomy satellite
- HALCA VLBI satellite
- Geotail geomagnetosphere satellite
- Akebono aurora observation satelli
- *ASTRO-F Infrared satellite*
- *ASTRO-E2 X*
- *SOLAR-B*

DARTS

http://www.darts.isas.jaxa.jp

9

Multi-wavelength astronomy



Mar. 22, 2005

10

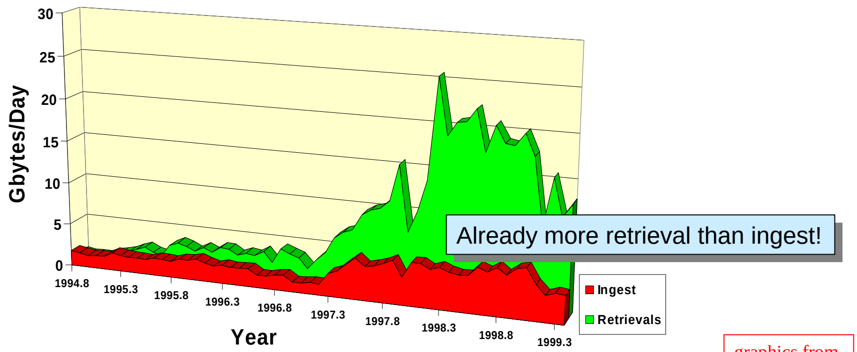
Data Productivity

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

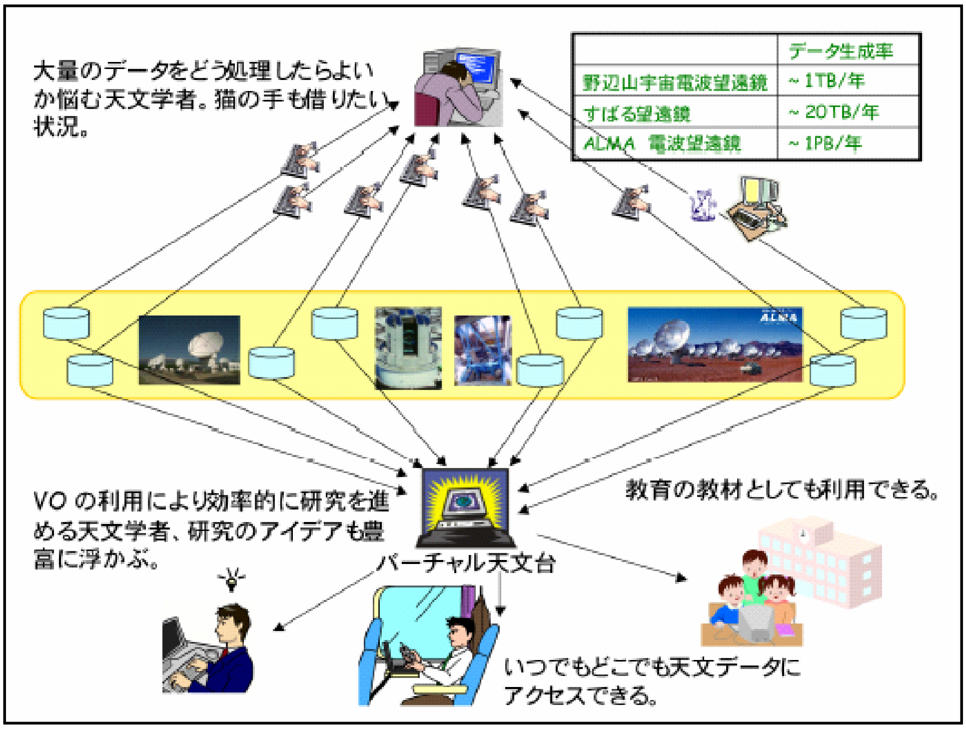
Flood of excellent data (survey data)
Digitized & Archived

Astronomy based on Archives

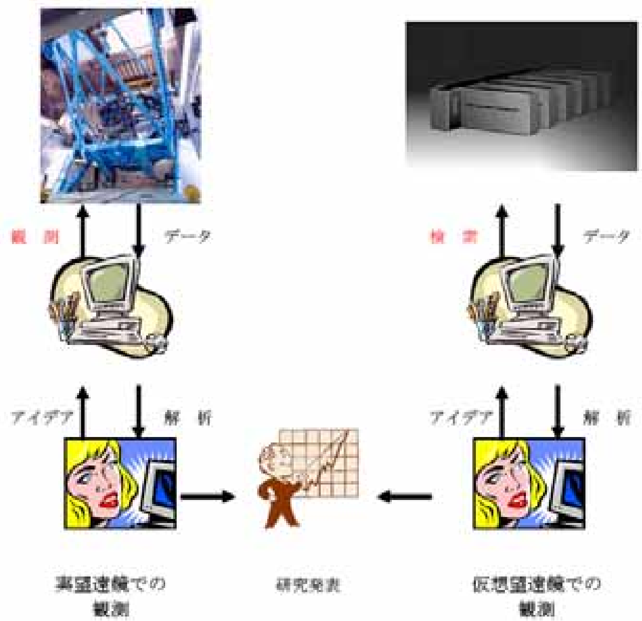
– large fraction of astro-papers based on archives
– HST archive use growing faster than archive



graphics from US NVO project



Virtual Observations



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:

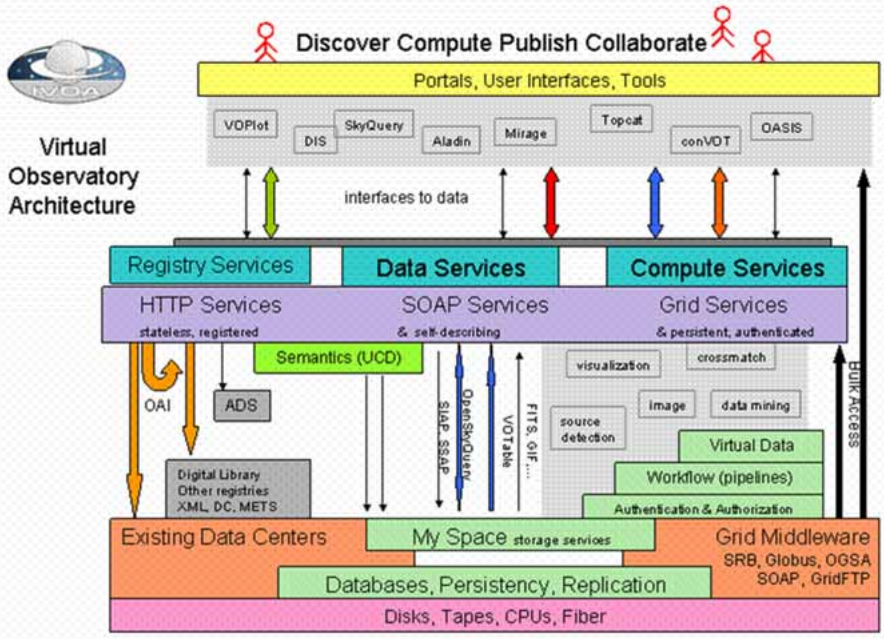
- A replacement for building new telescopes and instruments
- A centralized repository for data
- A data quality enforcement organization

VO Projects in the world

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

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



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.

OECD Rec. (‘04 Aug)

Recommendations:

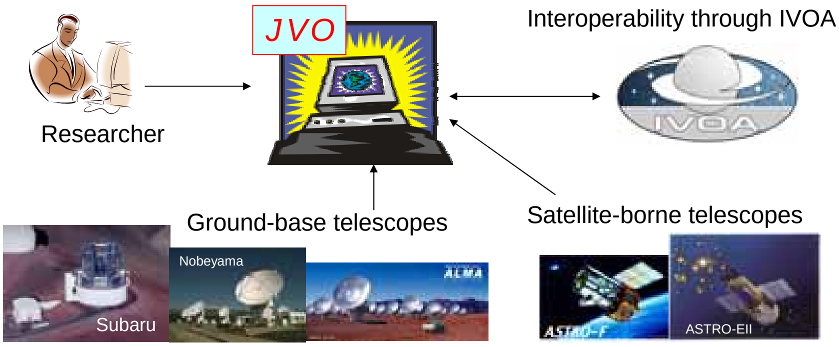
New projects and facilities must take the data management, storage, maintenance, and dissemination into account at the earliest planning stages, consulting potential users in the process. Agencies and governments should consider adopting the IAU resolutions as the basis for progress in this field.

Agencies should recognise that this is an important long-term issue and should co-ordinate 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.

JVO : Japanese Virtual Observatory

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



Mar. 22, 2005 電脳Davis ワークショップ 21

Development history

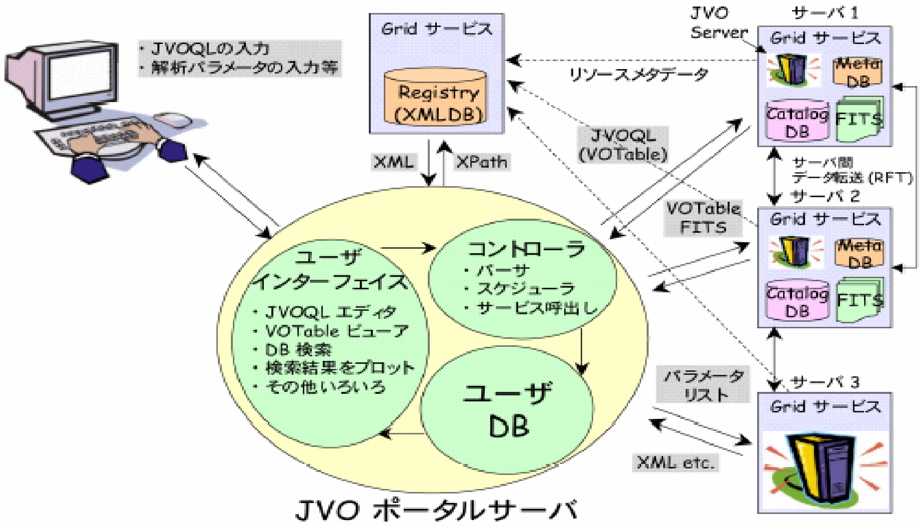
- JVO Project start — April 2002
- Prototype 1 finish — March 2003
- Prototype 2 finish — March 2004
- Prototype 3 partially finished

Mar. 22, 2005 電脳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

Schematic diagram of JVO



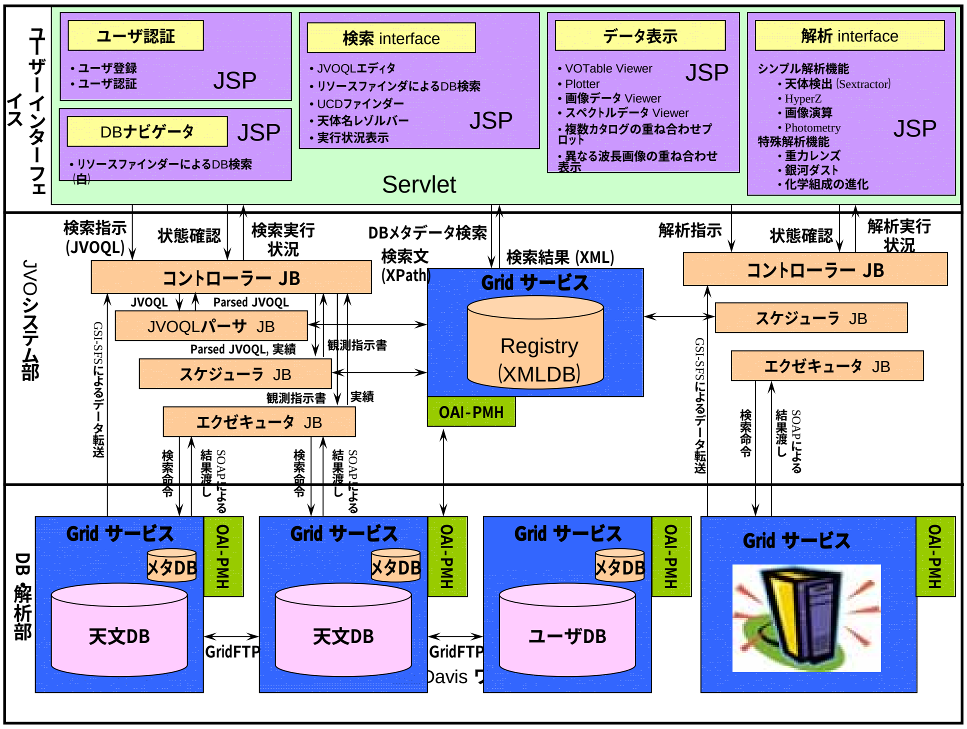
Design toward Prototype 3

- 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

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.



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.

Mar. 22, 2005 電脳Davis ワークショップ 28

```
create view myEROtable as
select s.Bmag,
       s.Rmag,
       t.Hmag,
       t.Kmag,
       ...,
       sr.BOX(POINT(s.ra,s.dec),w,h)
          as Rimage,
       tk.BOX(POINT(s.ra,s.dec),w,h)
          as Kimage,
       ...
from   SUBARU s,
       2MASS  t,
       ...,
       SUBARU.R sr,
       2MASS.K  tk,
       ...
where  XMATCH(s,b,...) < 3 arcsec
       and
       (s.Rmag-t.Kmag) > 6 mag
       and
       BOX(POINT(ra0,dec0), w0, h0)
       and
       ...
```

ユーザ指定のview名をJVOシステムに生成

各カタログサーバから属性値を選択。カラム名はUCDにより記述する

各画像サーバより画像を切り出す。画像領域はBOXあるいはCIRCLE文で指定する

カタログサーバを選択

画像サーバを選択

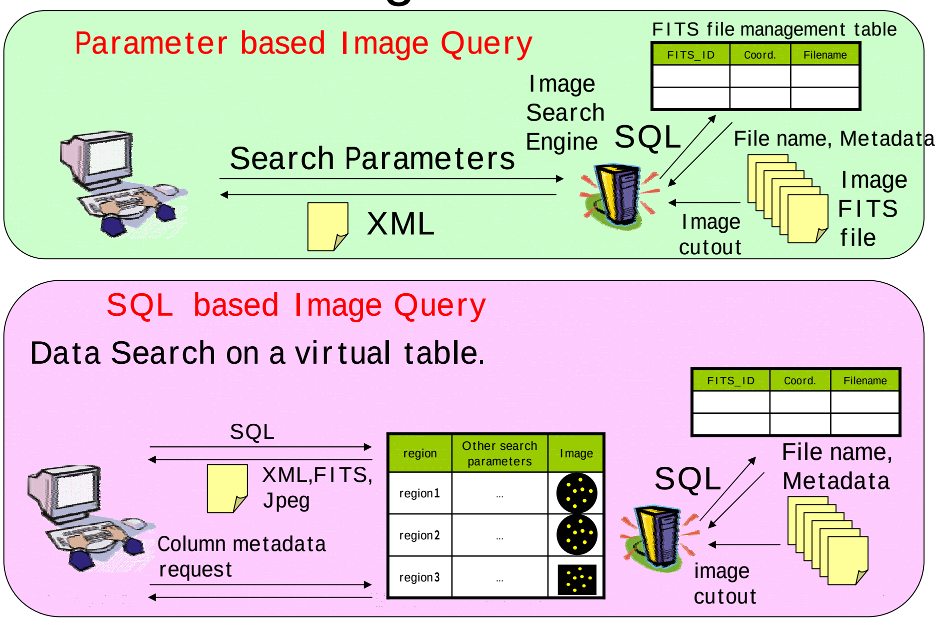
カタログ間のクロスマッチを行う

分散カタログ間の検索条件の指定

領域検索条件を画像切り出しと同様に記述

29

Image Search



Region Search using HTM index

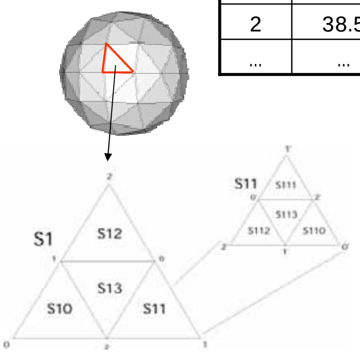
Region search is a common search criterion for an astronomical database. For efficient search data should be properly indexed on the object coordinate.

Catalog table

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

HTM Index table

id	htm
1	16522516
2	16754765
…	…



http://www.sdss.jhu.edu/htm/

```
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

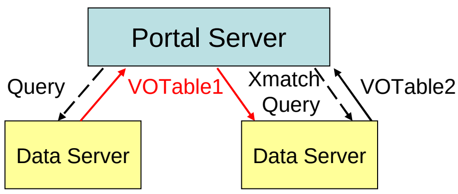
- “JVOQL Parser” generates query for each host
- “Scheduler” generates:
  - count query job for host1
  - count query job for host2
- “Executer” executes jobs on remote hosts
- “Scheduler” generates based on the result of execution
  - query job for host1
  - xmatch job for host2
  - image query for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

Remote execution

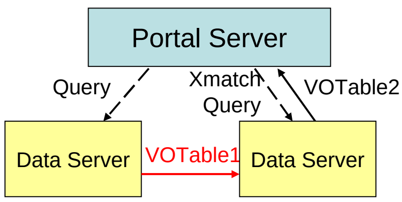
- Proto1:
  - Globus toolkit ver 2
  - using `globus-job-run` command
  - 1 call = 30 sec
    - 1 query ~10 min!
- Proto2:
  - Globus toolkit ver 3
  - using Grid Service
  - 1 call = 2-3 sec
    - overhead time is only ~ 30 ms

Data Transfer

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



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



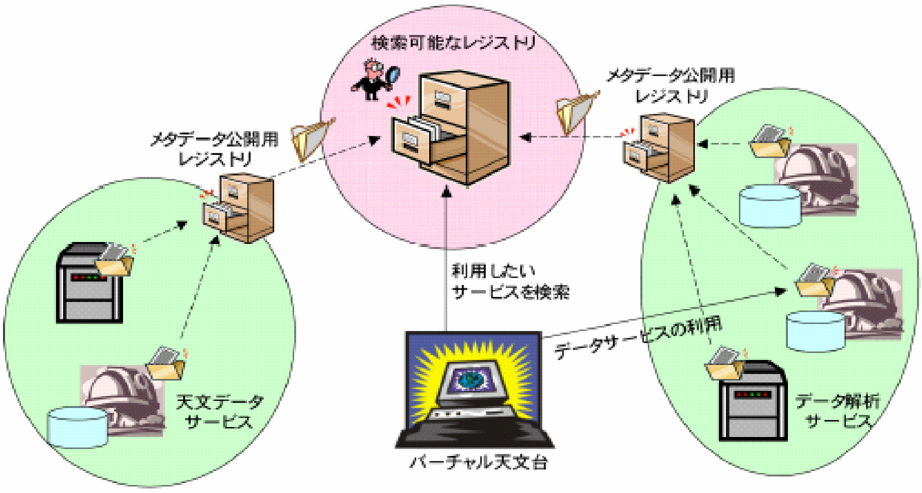
Data Transfer

- 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

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)

Exchange of Meta Data:OAI-PMH



Resource Metadata

Identity metadata | service metadata | column metadata
curation metadata | content metadata

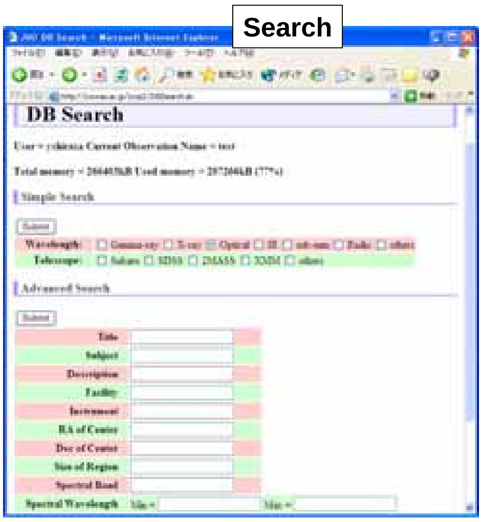
title	string
short_name	string
identifier	URI
publisher	string
publisher_id	URI
creator	string
creator_logo	URL
contributer	string
date	string
version	string
contact_name	string
contact_email	e-mail address
service_interface_url	URL
service_base_url	URL
service_http_result	MIME type
service_standard_uri	URI
service_standard_url	URL
service_msr	float,decimal degrees
ucd	string
unit	string
datatype	string
width	int
precision	string
arraysize	string

	identity	curation	service	content	column
catalog	○	○	○	○	×
table	○	○	○	○	×
column	○	○	○	○	○

subject	string,list
description	string,free text
source	string
reference_url	URL
type	string,list
coverage_spatial	string
coverage_region_of_regard	float,decimal degrees
coverage_spectral	string,list
coverage_spectral_bandpass	string,list
coverage_spectral_central_wavelength	float
coverage_spectral_minimum_wavelength	float
coverage_spectral_maximum_wavelength	float
coverage_temporal_start_time	string
coverage_temporal_stop_time	string
coverage_depth	float
coverage_depth_unit	string
coverage_object_density	float
coverage_object_count	int
coverage_sky_fraction	float
resolution_spatial	float
resolution_spectral	float
resolution_temporal	float
content_level	string,list
facility	string,list
instrument	string,list
format	string,list
right	string

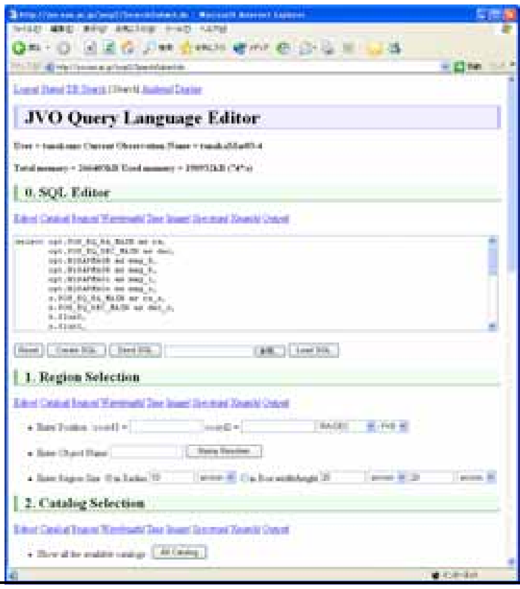
Access to Archives

Search

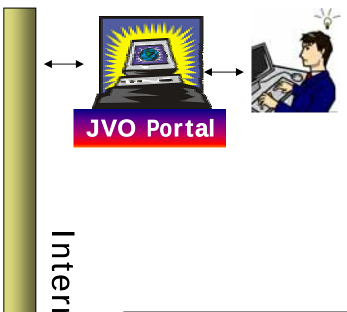


Results

Automatic Generation of Queries



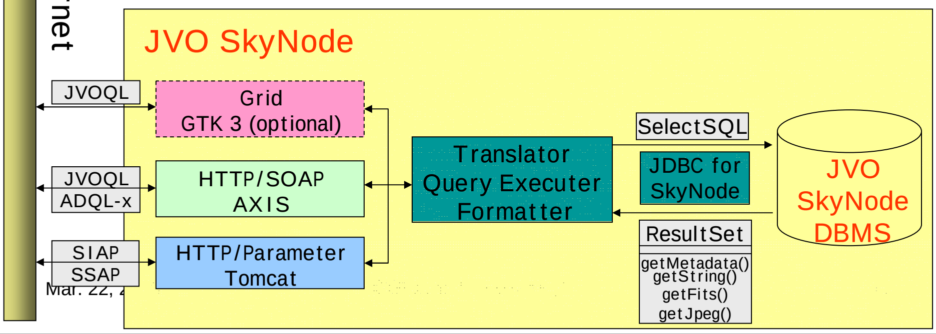
JVO SkyNode Architecture



• Apache Axis and Tomcat are used for implementing the VO standard interface, and Globus Toolkit is used for an optional Grid interface.

• Four kinds of query languages are converted to java class SelectSQL and query is executed through JDBC-like interface.

• Query results is obtained as ResultSet, and table data is formatted to VOTable or CSV file.



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...

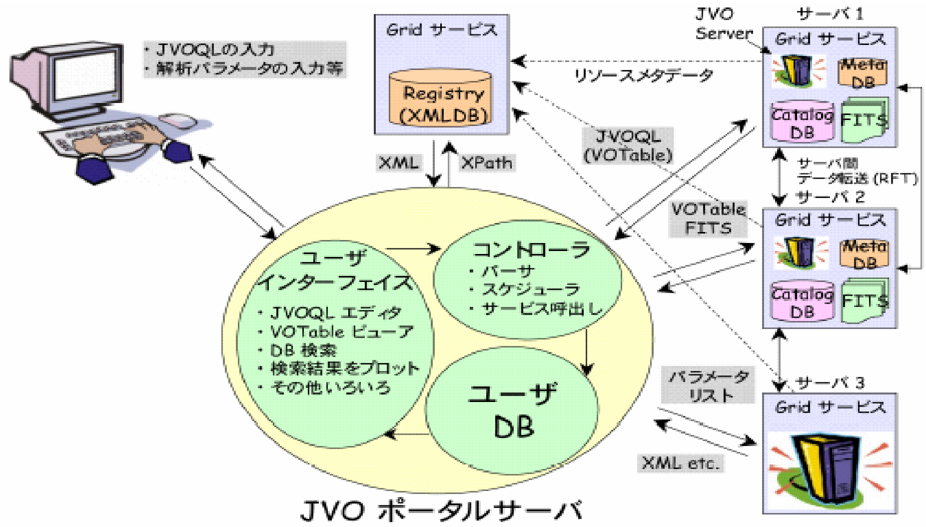
DBs available in JVO

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

Analysis Tools

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

Schematic diagram of JVO



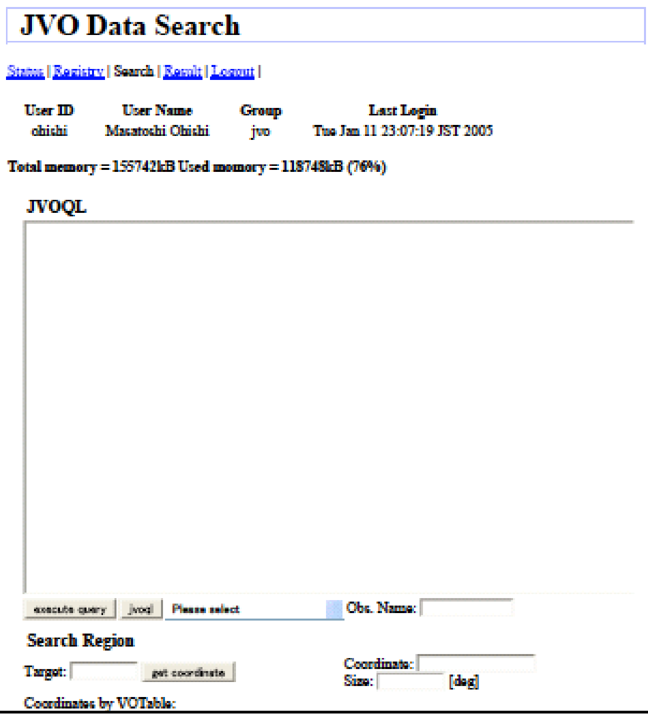
Access time to the US VO

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



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 momory = 113499kB (72%)

Search

[ ] AND [keyword search]

Registries	Data Services	Harvest	Register Metadata	Remove Metadata
registries	data services	harvest	register	remove

Get Service Infomation by Identifier

ivo://jvo/skynodej/sxds [getResource]

JVO JAPANESE VIRTUAL OBSERVATORY

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

[select] Select the checked service and go to the search page.

JVO JAPANESE VIRTUAL OBSERVATORY

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 momory = 117026kB (75%)

JVOQL

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


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 momory = 124641kB (80%)

Obs. Name	Process ID	Server	Flag	Elapsed Time (sec)	Status
OBS_20050111233330	proc_0001	irsa.ipac.caltech.edu		3.341	OK VOTable URLs

JVO JAPANESE VIRTUAL OBSERVATORY

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 momory = 117013kB (75%)

Obs. Name : OBS_20050111233330

Show Selected Images | Reset

check	download		POS_EQ_RA_MAIN	POS_EQ_DEC_MAIN		
☐	Download	1600	84.7414377	20.0273093	05h 38m 57.95s	+20d 01m 38.3s
☐	Download	1601	84.7414377	20.0273093	05h 38m 57.95s	+20d 01m 38.3s
☐	Download	1602	84.7414377	20.0273093	05h 38m 57.95s	+20d 01m 38.3s
☐	Download	1603	84.7414377	20.0273093	05h 38m 57.95s	+20d 01m 38.3s

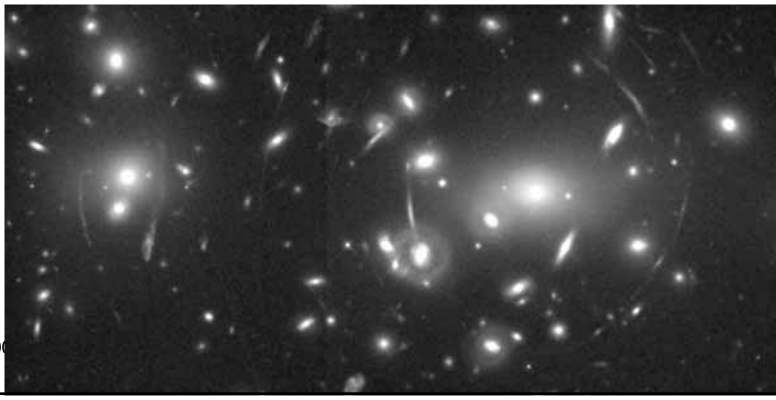
Show Selected Images | Reset

Cosmic Strings?

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

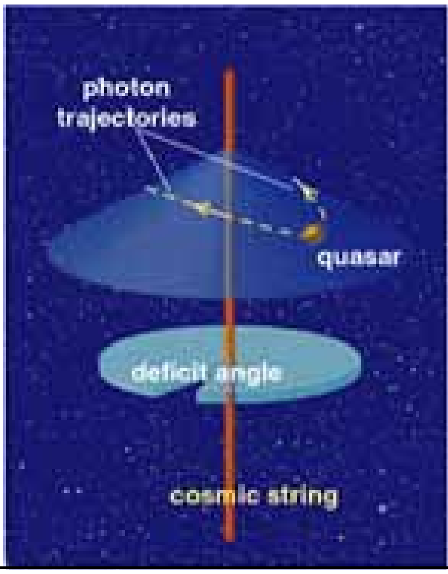
Gravitational Lens

- 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.

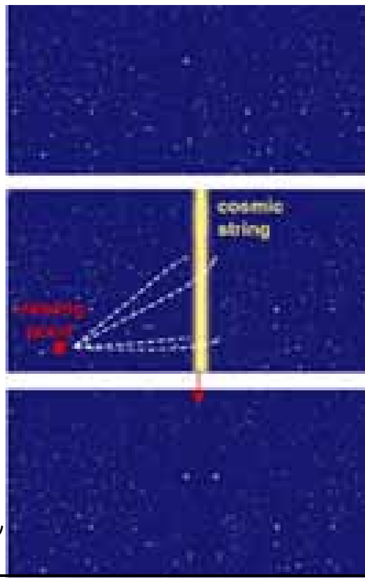


Mar. 22, 20

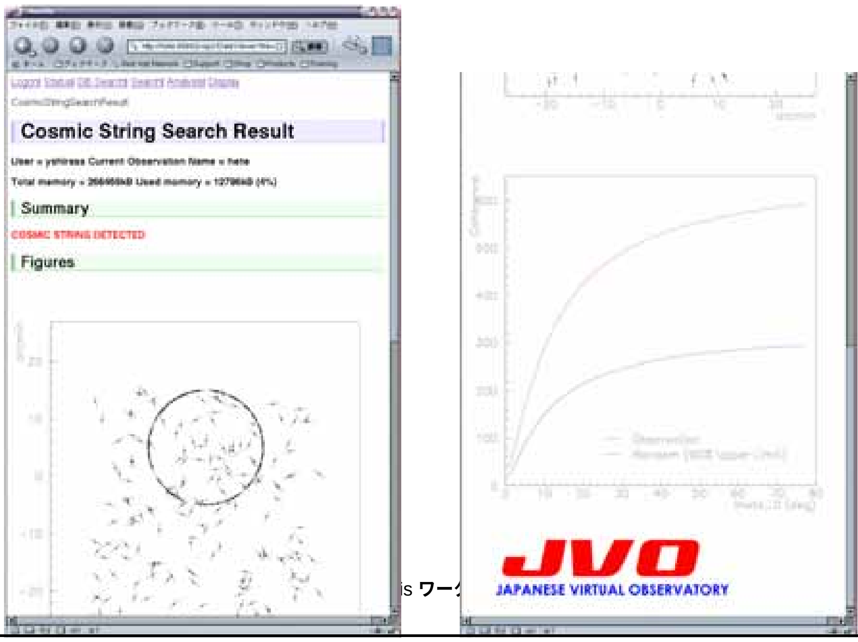
Gravitational Lens by Cosmic Strings



ワークショッ



Cosmic String Search Result

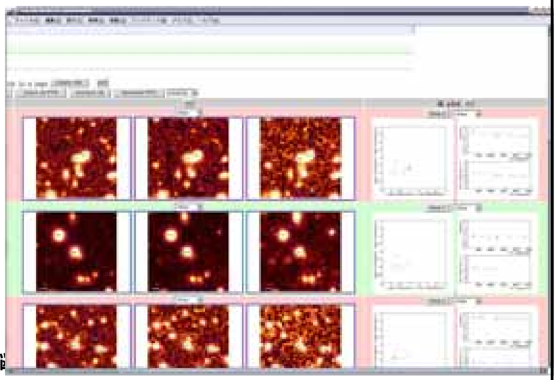


Work flow for Gravitational lens search

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.

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.



QSO銀河探査のフロー

- 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

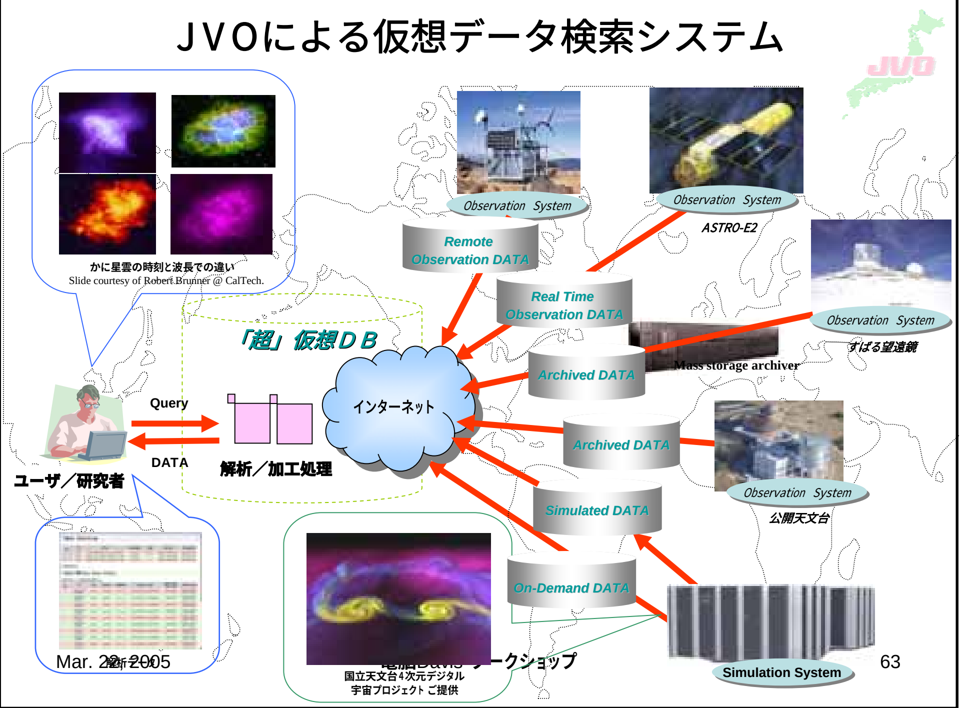
Science results (EU)

- 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)

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

JVOによる仮想データ検索システム



ALMA in Chile (sub-mm wave)

END