Technology for the Virtual Observatory

BRAVO Lecture Series, INPE, Brazil July 23-26, 2007

1. Virtual Observatory Summary
2. Service Architecture and XML
3. Building and Using Services
4. Advanced Services

The Virtual Observatory

Roy Williams

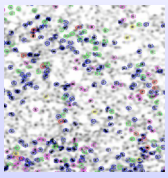
California Institute of Technology
Co-Director, US Virtual Observatory
Technical Lead, Int'l Virtual Observatory

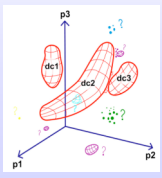
Toward a "new astronomy"

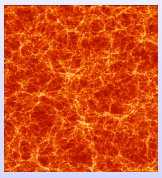
- Past:
  Observations of small, carefully selected samples

Toward a "new astronomy"

- Future: Multi-wavelength data for millions of objects, allowing us to:
  - Discover significant patterns from the analysis of statistically rich and unbiased image/catalog databases
  - Understand complex astrophysical systems via confrontation between data and sophisticated numerical simulation

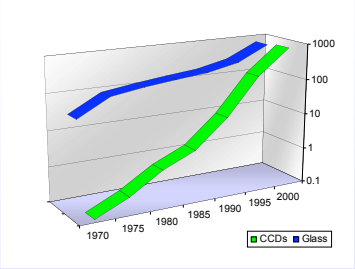






Why the VO?

Astronomy is becoming Data-Centric!



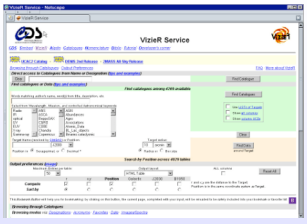
Total area of 3m+ telescopes in the world in m², total number of CCD pixels in Megapixels, as a function of time. Growth over 25 years is a factor of 30 in glass, 3000 in pixels.

International collaboration

http://ivoa.net

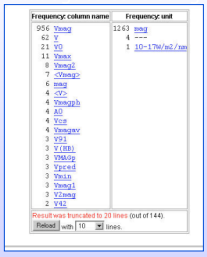
- IVOA now has 16 member projects
  - Fall 2007 Interop: Cambridge UK, Sept 2008
  - Spring 2008: Trieste IT, May 2008
  - Fall 2008: Baltimore US, Sept 2008
- Adopted a standards process based on W3C
  - Forum for discussion and sharing of experience

Interoperability



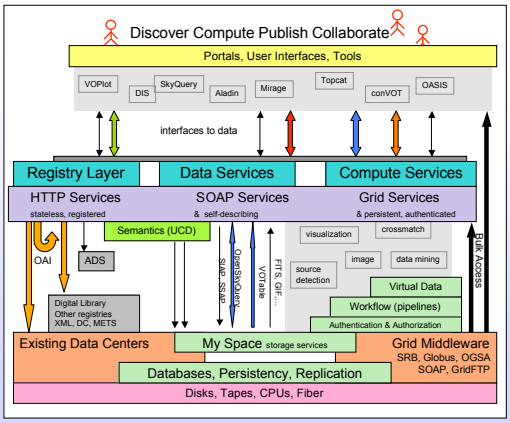
"Give me all tables containing the V magnitude in the Johnson system."

144 different names for Johnson V.



Interoperability challenges

- Data formats and delivery
- Data service request/response
- Metadata standards
- Publishing and discovery (Registry)
- Content descriptors and Units
- Database query language
- Web services
- Authentication and authorization

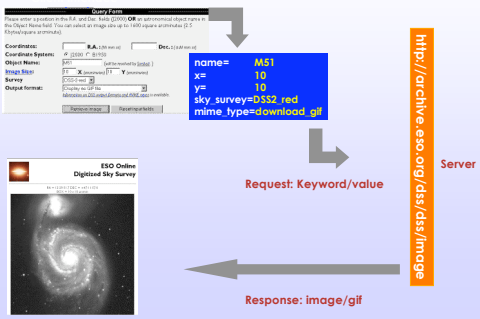


Getting Data:
Formats, Services

Data Standards

- VOTable
  - Rich metadata
  - XML
- Space-Time Coordinates
  - Location and regions
  - Celestial frames, surface of Mars, etc etc
  - Sophisticated and comprehensive
- Unified Content Descriptors
  - Standard names for astronomical data types
  - Mined from thousands of published catalogs
  - Johnson V magnitude:
    `phot.mag;em.opt.V`

Example Service



Web Services

- Principle: **Click or Code**
- Protocol: simple REST/**GET** or sophisticated **SOAP**
- Standards
  - Basic service profile
    - formal description (WSDL), input parameters, output formats, capabilities
    - VOResource (who and what is it, for the Registry)
  - Security (single sign-on, authentication and authorization)
  - Asynchronous (batch) services
  - Distributed data storage (VOSpace)
- Coordinated with IVOA, Open Grid Forum

Data Services

- Cone Search
  - First standard NVO service:
    - radius+position ⇒ list of objects
    - encoded as VOTable
- Simple Image Access Protocol
  - "cone search for images"
  - images are referenced by URL
- Simple Spectrum Access Protocol
  - spectra have subtleties → protocol more complicated

Data Services

- Astronomical Data Query Language
  - For database queries
  - Core SQL functions plus astronomy-specific extensions
    - Sky region, Xmatch
- SkyNode
  - Exposes relational databases
  - Accepts ADQL query
  - "Full" SkyNodes support positional cross-match function
  - OpenSkyQuery portal
    - show database structure
    - query tools
- Inventory services
- Footprint services

Protocol List

- VOTable
- FITS
- Cone Search Protocol
- Simple Image Access (SIA) Protocol
- Simple Spectral Access (SSA) Protocol
- SkyNode
- Astronomical Data Query Language (ADQL)
- Uniform Column Descriptors (UCDs)
- Space-Time Coordinates (STC)
- IVOA Identifiers
- VO Registry
- Resource Metadata
- Registry Access Protocols
- OAI
- VOEvent
- NESSSI
- VOSpace
- PLASTIC
- Common Execution Architecture

Formats

- VOTable
  - XML format for tabular information
    - Catalog query results
    - List of images or spectra
    - Lists of resources
    - Includes *UCD*s for columns
- FITS
  - Standard astronomical binary data format
  - Images, spectra, …

Semantic content

- IVOA identifiers
  - Scheme for providing unique identifiers for any thing that needs to be distinguished. Services, requests, …
  - ivo://authority/authority_controlled_string#sub-resource
  - Used in *registries* and other protocols
- Registry resource metadata
  - Standardized format for the description of a service, data collection, facility.
  - XML format based on the Dublin-core metadata descriptions for published resources.
  - Used inside *registries*.
- Uniform Column/Content Descriptors (UCDs)
  - Standardized names for table columns with relatively high degree of specificity.
  - pos.eq.ra or phot.flux;em.ir
  - Used in *VOTables* and by *Cone search* and *SIA*
- Space/Time Coordinates
  - Way of specifying points or regions of space-time.
  - Can be simple or very complex
  - Used in *ADQL* and *VOEvent*

Discovering Resources

- Registry
  - Collection of resources that users can find included *Cone Search*, *SIA*, *SSA* and *SkyNodes*.
  - User access via *registry access protocol*
  - Registry-registry communication via *OAI*
  - Some publishing registries are meant to be seen only by other registries
- Registry access
  - How users query the *registry*
  - In standards process
- Open Archive Interface (OAI)
  - Library community standard
  - *Registry-registry* communication to ensure that all *registries* are up to date.
- VOEvent
  - VO protocol for publishing and receiving notice of transient/urgent events
  - Describes events using *STC*

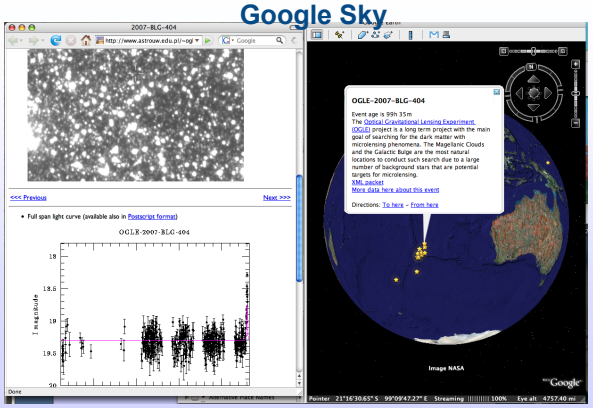
Accessing and Controlling Computer Resources

- NESSSI
  - NVO Extensible, Scalable, Secure Service Infrastructure
  - Provide access to protected resources, especially Grid computational resources.
  - Make it easy to do simple things.
  - First implementations just coming on-line
- VOSpace
  - Access to storage resources in a standardized way.
- PLASTIC
  - Platform for Astronomical Task InterCommunication
- Common Execution Environment (CEA)
  - Script/pipeline environment with security and such incorporated (Astrogrid)

VOEvent

- Protocol for notification of transient events
  - Gamma-ray burst, Supernova, Asteroid, M-dwarf flare, etc
- Enables immediate, automated robotic followup
- XML format:
  - Who, What, How, Wherewhen, Why
  - Citation of prevous events
  - Push protocols for immediate delivery

Google Sky



Registry

- publish -- find -- bind
- Registry Metadata
  - Descriptions of
    - data collections
    - data delivery services
    - organizations, etc.
  - Based on Dublin Core with astronomy-specific extensions
  - Represented as XML schema; extensible
  - Contents stored in Resource Registries
    - exchange metadata records through the Open Archives Initiative Protocol (OAI-PMH)

Distributed Registry

Ongoing harvesting March 07
(CFA, ESO, NOAO soon)

NVO Registry Portal

Find source catalogs, image archives, and other astronomical resources registered with the NVO

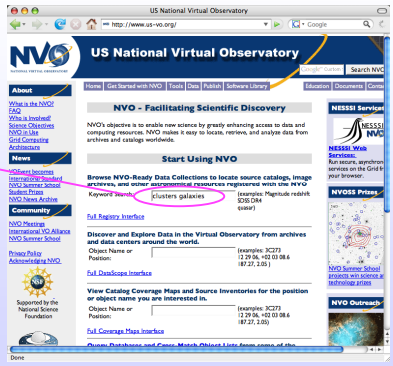
A Registry is a distributed database of Virtual Observatory resources: access services for catalog, image, and spectral data; descriptions of organizations and data collections.

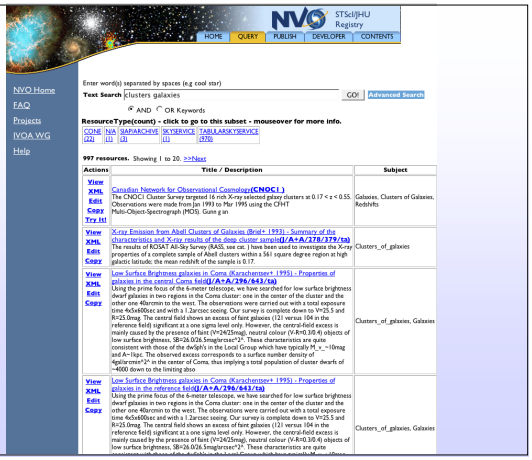
Coordinated registry implementations that share information by harvesting each other's resources.

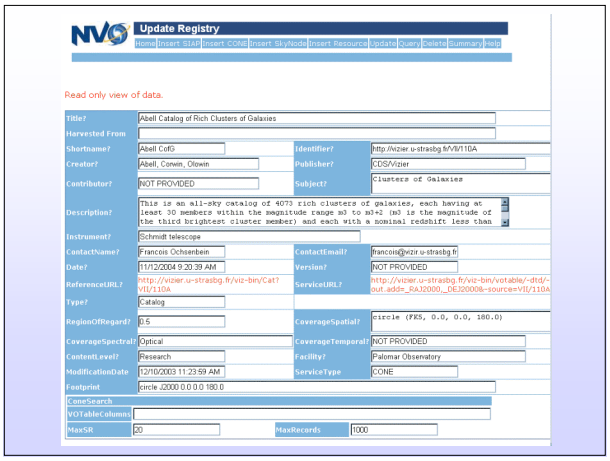
Searches for resources can be done by keyword, or advanced queries can be expressed in the SQL language. The registry is open for humans through web forms, or machines through SOAP web services.

Registry

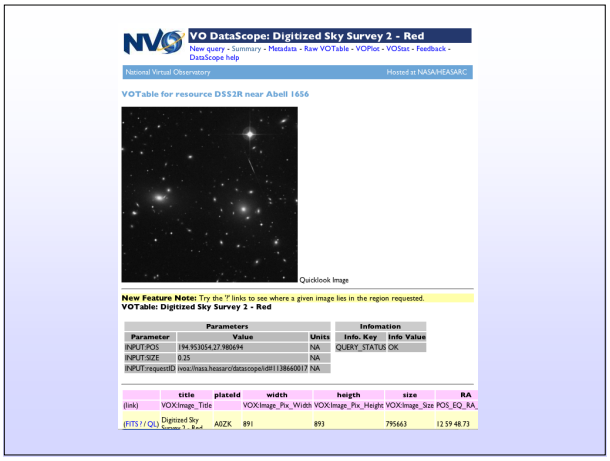
Search Registry of services

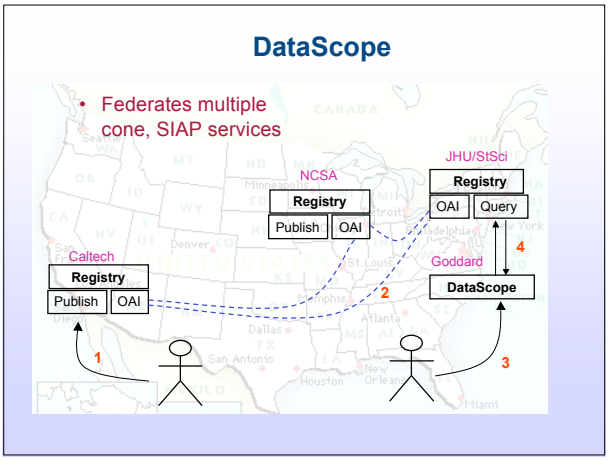





Update Registry
Home Insert SIAP Insert CONE Insert SkyNode Insert Resource Update Query Delete Summary Help
Read only view of data.
Title? Abell Catalog of Rich Clusters of Galaxies
Harvested From
Shortname? Abell CofG
Identifier? http://vizier.u-strasbg.fr/VII/110A
Creator? Abell, Corwin, Olowin
Publisher? CDS/Vizier
Contributor? NOT PROVIDED
Subject? Clusters of Galaxies
Description? This is an all-sky catalog of 4073 rich clusters of galaxies, each having at least 30 members within the magnitude range m3 to m3+2 (m3 is the magnitude of the third brightest cluster member) and each with a nominal redshift less than
Instrument? Schmidt telescope
ContactName? Francois Ochsenbein
ContactEmail? francois@vizir.u-strasbg.fr
Date? 11/12/2004 9:20:39 AM
Version? NOT PROVIDED
ReferenceURL? http://vizier.u-strasbg.fr/viz-bin/Cat?VII/110A
ServiceURL? http://vizier.u-strasbg.fr/viz-bin/votable/-dtd/-out.add=_RAJ2000,_DEJ2000&-source=VII/110A
Type? Catalog
RegionOfRegard? 0.5
CoverageSpatial? circle (FKS, 0.0, 0.0, 180.0)
CoverageSpectral? Optical
CoverageTemporal? NOT PROVIDED
ContentLevel? Research
Facility? Palomar Observatory
ModificationDate 12/10/2003 11:23:59 AM
ServiceType CONE
Footprint circle J2000 0.0 0.0 180.0
ConeSearch
VOTableColumns
MaxSR 20
MaxRecords 1000


Datascope
another mashup
DataScope
US National Virtual Observatory
NVO - Facilitating Scientific Discovery
Start Using NVO


VO DataScope: Digitized Sky Survey 2 - Red
New query - Summary - Metadata - Raw VOTable - VOPlot - VOStat - Feedback - DataScope help
National Virtual Observatory
Hosted at NASA/HEASARC
VOTable for resource DSS2R near Abell 1656
Quicklook Image
New Feature Note: Try the 'i' links to see where a given image lies in the region requested.
VOTable: Digitized Sky Survey 2 - Red
Parameters
Parameter Value Units
INPUT:POS 194.953054,27.980694 NA
INPUT:SIZE 0.25 NA
INPUT:requestID ivoa://mass.hesarc/dsascopeId!dfl!138660017 NA
Infomation
Info. Key Info Value
QUERY_STATUS OK
title plateId width heigth size RA
(link) VOX:Image_Title VOX:Image_Pix_Width VOX:Image_Pix_Height VOX:Image_Size POS_EQ_RA_
(FITS / QL) Digitized Sky A0ZK 891 893 795663 12 59 48.73

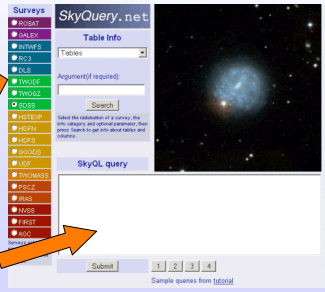

DataScope
• Federates multiple cone, SIAP services
NCSA
Registry
Publish
OAI
JHU/StSci
Registry
OAI
Query
Caltech
Registry
Publish
OAI
Goddard
DataScope
1
2
3
4

SkyQuery
database queries and crossmatch

- At the heart of SkyQuery is spatial matching
  - find the same objects listed in different catalogs.
- Astronomical archives double every year. Increasingly, data are stored in on-line archives. Instrumental surveys covering a large fraction of the sky are a major part of this revolution.
  - SkyQuery relies on surveys stored at different geographic locations.

SkyQuery: Interface

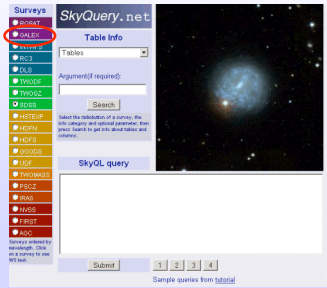
- The SkyQuery interface window is divided in 2 parts:
- The list of SkyNodes can be queried by marking checkboxes
- A SQL-like text box is used to submit queries



SkyQuery: consult a SkyNode's parameters. Step 1

The contents of each SkyNode can be browsed.

- Select a survey and its list of tables or functions. Click Search.
- From the entries returned, enter any column/function of interest in Argument. Click Search.

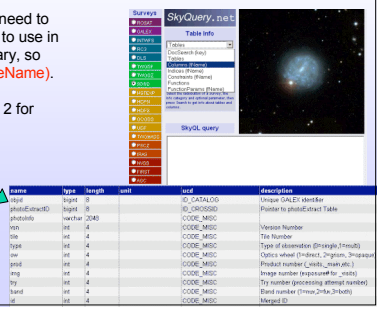


SkyQuery:
consult a SkyNode's parameters

- To build a query, we need to know what keywords to use in table e.g. PhotoPrimary, so select Columns(TableName).
- Now run Steps 1 and 2 for another SkyNode(s).

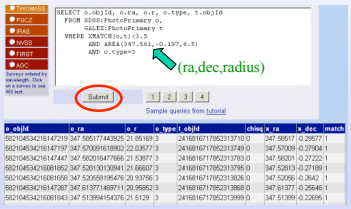
Columns from PhotoPrimary

Now we are ready to run the query…



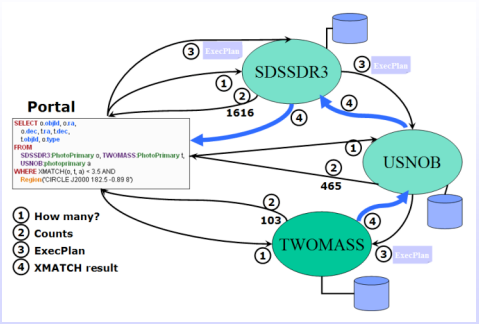
Run an ADQL (SQL-like) query

- Running a query is simple. Write it in the SQL-like query box or …
- Click and customize any of 4 sample queries, #1, #2 …
- See also "tutorial"
- To submit query, click Submit. Cross correlation results appear!

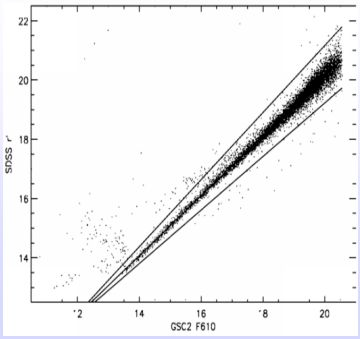


Open SkyQuery.net

Execution



Crossmatch result



Compute Services

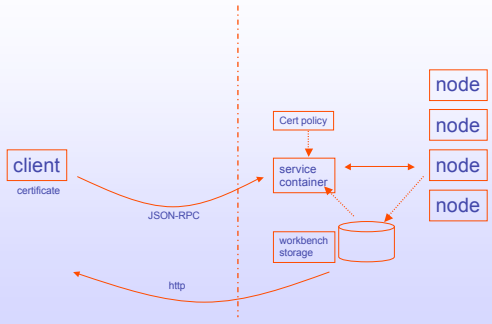
Compute Services

- Asynchronous
  - May not get immediate answer
    - just get a place to check back
- Security
  - Expensive resources, big requests, sequestered data
  - Strong or Weak or None
- Scalable
  - Graduated path to powerful computation and big data

Compute Services

- Web-based Portals
  - User interacts with web portal
  - Portal runs code on Grid
  - Service requests forwarded to grid resources
- Scripted service call
  - User writes code to submit and monitor jobs
  - User manages certificate

Compute Service Architecture

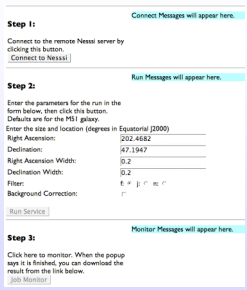


Graduated security

- **No** certificate gets 15 CPU-minutes from community account
  - Just switch on Javascript or connect to portal
- **Weak** certificate gets 1 CPU-hour from community account
  - In exchange for registering name/email
  - Can do scripted access
- **Strong** (gridmapped) certificate gets infinity from own account
  - Get this one from TeraGrid HQ

```
"nesssi_strong_cert_max_time" : 216000,
"nesssi_weak_cert_max_time" : 3600,
"nesssi_anon_max_time" : 900,
"nesssi_anon_user"     : "nvo",
"nesssi_weak_user"     : "nvo",
```

Mosaic Service




```
nesssiServer=nesssi.client('https://envoy.c
acr.caltech.edu:8443/clarens/',debug=0)

mosaic_loc = "-ra 49.1 -dec 60.1 -rawidth
0.5 -decwidth 0.5 -filt f -bgcorr 0"

session =
nesssiServer.dpossMosaic.mosaic(mosaic_loc)
print "Your session ID is %s." % session

msg = dbsvr.dpossMosaic.monitor(session)
print msg
```

Repeat the monitoring