

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*

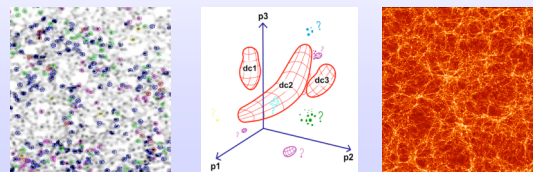
Toward a “new astronomy”

- 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!

The graph illustrates the rapid growth of CCD technology compared to traditional glass-based astronomy. The blue line represents the total area of 3m+ telescopes in m², and the green line represents the total number of CCD pixels in Megapixels. The Y-axis is logarithmic, showing the growth over time. The X-axis shows years from 1970 to 2000. The Z-axis represents the total area of telescopes.

Year	Glass (m²)	CCDs (Megapixels)
1970	~10	~0.1
1975	~20	~0.2
1980	~40	~0.5
1985	~60	~1.0
1990	~80	~2.0
1995	~100	~5.0
2000	~120	~10.0

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

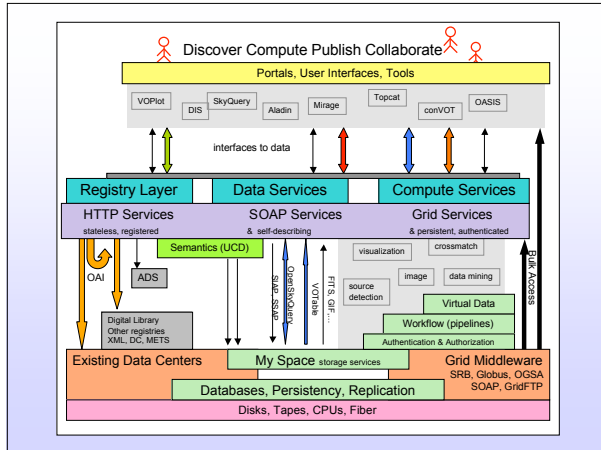
The screenshot shows the VizieR Service interface. The search bar contains 'Johnson V.'. The results table lists 144 different names for Johnson V. The table has two columns: 'Frequency: column name' and 'Frequency: unit'.

Frequency: column name	Frequency: unit
256 Jmag	1163 Jmag
62 J	4 J
21 J0	1 J0 - 17m/nd / sec
13 Jmag	Jmag
6 Jmag1	Jmag1
7 Jmag2	Jmag2
6 Jmag	Jmag
4 JV	JV
6 Jmag2b	Jmag2b
6 J0	J0
6 J1	J1
6 Jmag3	Jmag3
3 J1b1	J1b1
3 Jmag3p	Jmag3p
3 Jmag4	Jmag4
3 Jmag	Jmag
3 Jmag2	Jmag2
3 Jmag	Jmag
2 J42	J42

Below the table, there is a note: 'Observations indicated to 20 lines (out of 144):' followed by a table with two columns: 'Index' and 'Unit'. The 'Index' column contains the number 10, and the 'Unit' column contains the text 'mag'.

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

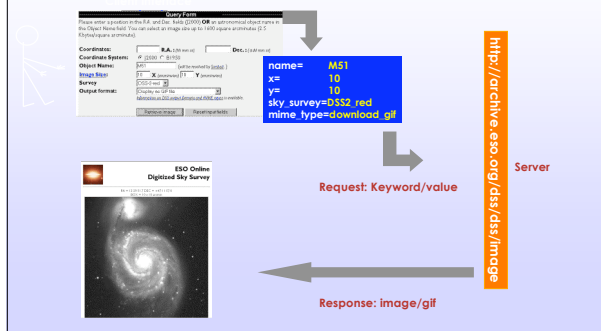


Getting Data: Formats, Services

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 \Rightarrow 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 \rightarrow 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

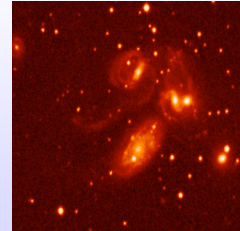
Protocol List

- 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 *UCDs* 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

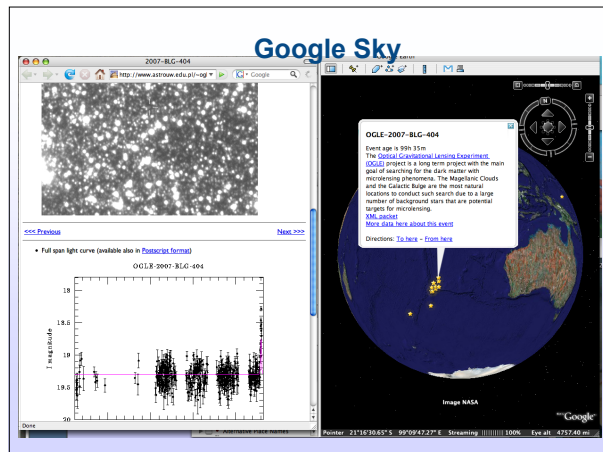
- Registry
 - Collection of resources that users can find included *Cone Search*, *SIA*, *SSA* and *SkyNodes*.
 - User access via *registry access protocol*
 - Registry-registry communication via *OAI*
 - Some publishing registries are meant to be seen only by other registries
- Registry access
 - How 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

- **NESSI**
 - 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, Where, When, Why
 - Citation of previous events
 - Push protocols for immediate delivery



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)


[illegible]

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.

Search
Registry
of services



RCAT
Resource Catalog



STSDJ4U
Registry

HOME
QUEST
PUBLISHER
DEVELOPER
CONTENTS

NVO Home

FAQ

Projects

NVO.WG

Help

Enter keyword (separated by spaces (e.g cool star))

Text Search (History: [previous](#))

☐ AND ☒ OR ☐ Keywords

ResourceType(scent) - click to go to this subject - mouseover for more info.

[CODE](#)
[NVO](#)
[INAPACHE](#)
[INVERING](#)
[JABLA](#)
[JABLAKEYWORDS](#)

997 resources. Showing 1 to 20 [table](#)

	Actions	Title / Description	Subject
View	XSL	Candidata Network for Observational Cosmology(CNOC1)	
Edit		The CNOC1 Cluster Survey targeted 108 X-ray selected galaxy clusters at 0.17 < z < 0.55. Observations were made from Jan 1991 to Feb 1999 using the Chandra	Galaxies, Clusters of Galaxies, Abundance
Copy		Multi-Polymorphic Spectroscopy (MPGS) - Gung et al	
Text		X-ray Emission Line Abund. Clusters of Galaxies (1991-1993): A Survey of the Characteristics of X-ray Emission Lines of the Clusters (X-ARXIV:199103010)	
XSL		The results of RCAT 40-keV Survey (RA40), see url (we have used to investigate the X-ray properties of a complete sample of Abell clusters with X-ray isophote depth regular in high galactic latitude: the mean redshift of the sample is 0.17)	Clusters of galaxies
Copy		Low Surface Brightness galaxies in Centaurus (A1934-3593): Properties of galaxies in the distant Centaurus cluster (X-ARXIV:199304030)	
XSL		Using the entire frame of the cluster, we have studied the low surface brightness galaxies in two regions in the Centaurus cluster: one in the cluster core and the other one observed to the west. The observations were carried out with a rapid response time (keV/bins) and with a 3-arcsec seeing. Our survey is complete down to 0.25 and 0.12 Jy/dB/arcsec. The central field covers an extent of about 60x60 arcmin (21 versus 104 arc of the reference field) significant at one sigma level only. However, the characteristic depth is nearly doubled by the presence of faint (0.05-0.20 Jy/dB/arcsec), neutral colour (X-0.60 to X-0.65) of galaxies at low surface brightness (<0.20 Jy/dB/arcsec). These characteristics are quite consistent with those of the sample in the Local Group which have typically N<= 100 - 1500 galaxies per field and a surface number density of N<= 0.001 galaxies per arcmin². This characteristic is quite similar to the findings of the	Clusters of galaxies, Centaurus
Copy		Low Surface Brightness galaxies in Centaurus (A1934-3593): Properties of galaxies in the distant Centaurus cluster (X-ARXIV:199304030)	
View		Using the entire frame of the cluster, we have studied the low surface brightness galaxies in two regions in the Centaurus cluster: one in the cluster core and the other one observed to the west. The observations were carried out with a rapid response time (keV/bins) and with a 3-arcsec seeing. Our survey is complete down to 0.25 and 0.12 Jy/dB/arcsec. The central field covers an extent of about 60x60 arcmin (21 versus 104 arc of the reference field) significant at one sigma level only. However, the characteristic depth is nearly doubled by the presence of faint (0.05-0.20 Jy/dB/arcsec), neutral colour (X-0.60 to X-0.65) of galaxies at low surface brightness (<0.20 Jy/dB/arcsec). These characteristics are quite similar to the findings of the	Clusters of galaxies, Centaurus
XSL		Low Surface Brightness galaxies in Centaurus (A1934-3593): Properties of galaxies in the distant Centaurus cluster (X-ARXIV:199304030)	
Edit		Using the entire frame of the cluster, we have studied the low surface brightness galaxies in two regions in the Centaurus cluster: one in the cluster core and the other one observed to the west. The observations were carried out with a rapid response time (keV/bins) and with a 3-arcsec seeing. Our survey is complete down to 0.25 and 0.12 Jy/dB/arcsec. The central field covers an extent of about 60x60 arcmin (21 versus 104 arc of the reference field) significant at one sigma level only. However, the characteristic depth is nearly doubled by the presence of faint (0.05-0.20 Jy/dB/arcsec), neutral colour (X-0.60 to X-0.65) of galaxies at low surface brightness (<0.20 Jy/dB/arcsec). These characteristics are quite similar to the findings of the	

NVO Update Registry

Read only view of data.

Title?	Abell Catalog of Rich Clusters of Galaxies		
Horwested From			
Shortname?	Abell CatG	Identifier?	http://vizier.u-strasbg.fr/VizieR/
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 m0 to m0+2 (m0 is the magnitude of the third brightest cluster member) and each with a nominal redshift less than 0.1.		
Instrument?	Schmidt telescope		
ContactName?	Frances Ochoaibein	ContactEmail?	frances@vizier.u-strasbg.fr
Date?	11/12/2004 9:20:39 AM	Version?	NOT PROVIDED
ReferenceURL?	http://vizier.u-strasbg.fr/viz-bin/Cat?	ServiceURL?	http://vizier.u-strasbg.fr/viz-bin/votable/-dhd/-out_addr_RAJ2000_DEJ2000&-source=VU/110A
Type?	Catalog		
RegionOfInterest?		CoverageSpatial?	Circle (PKS, 0.0, 0.0, 180.0)
CoverageSpectral?	Optical	CoverageTemporal?	NOT PROVIDED
ContentLevel?	Research	Facility?	Palomar Observatory
ModificationDate?	12/16/2004 11:23:59 AM	ServiceType?	CONE
Footprint?	Circle J2000 0.0 0.0 180.0		
Comments?			
VOTableAddress?			
HasSB?	NO	HasRecords?	NO

Datascope
another mashup

US National Virtual Observatory

NVO - Facilitating Scientific Discovery

NVO's objective is to enable new science by greatly enhancing access to data and computing resources. NVO makes it easy to locate, retrieve, and analyze data from archives and catalogs worldwide.

Start Using NVO

Browse NVO-Ready Data Collections to locate source catalogs, image archives, and other astronomical resources registered with the NVO.

Query Databases and Cross-Match Object Lists from some of the largest on-line catalogs in astronomy (Open SkyQuery).

Perform Source Extraction and Object Identification by detecting objects in your own images and matching them with objects in the major survey catalogs (WISE).

Explore the Multiresolution Sky in the Vicinity of Transient Events that have recently been observed (VOTableView).

Register Image Coordinates to images with inaccurate or misaligned coordinate systems. (MASS, DPOSS, or SDSS images (Montage)).

Make mosaics from 2MASS, DPOSS, or SDSS images (Montage).


Analyze or simulate your VOTable (TOPCAT).

NVO VO DataScope: Digitized Sky Survey 2 - Red

New Query - Summary - Parameters - Raw VOTable - VOPlus - VOSat - Feedback - DataScope Help

National Virtual Observatory Hosted at NASA/HARPC

VOTable for resource D553R near Abell 1656



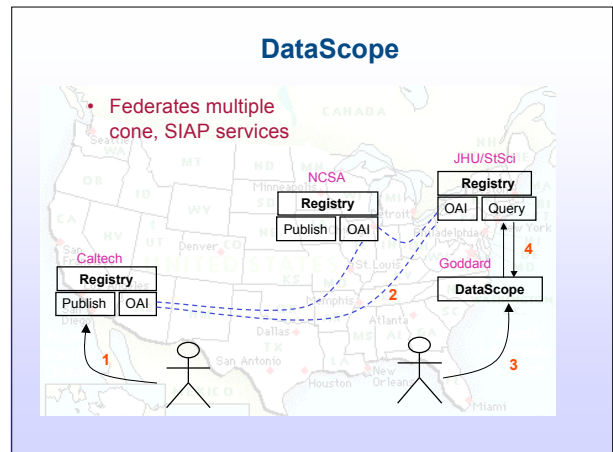
Quicklook Image

New Feature Note: Try the "I" links to see where a given image lies in the region requested.

VOTable: Digitized Sky Survey 2 - Red

Parameter	Value	Units	Info. Key	Info Value
INPUTPOS	194.9305427 180694	RA	QUERY STATUS	OK
INPUTSIZE	0.25	NA		
INPUTrequestID	http://vizier.u-strasbg.fr/viz-bin/votable/-dhd/-out_addr_RAJ2000_DEJ2000&-source=VU/110A	NA		

title	plateid	width	height	size	RA
(link) VOXImage_Title	VOXImage_Pix_Width	VOXImage_Pix_Height	VOXImage_Size	POS_EQ_RA	
10757701	Digitized Sky	A02K	891	893	795463 12 39 48.73

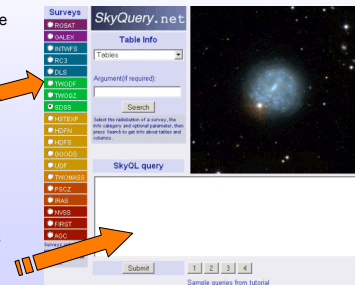


SkyQuery database queries and crossmatch

- 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**.

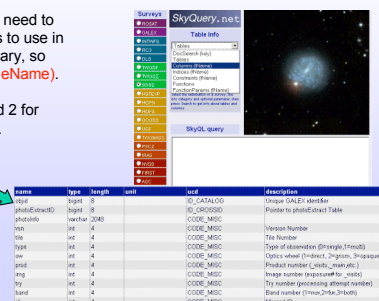


SkyQuery: consult a SkyNode's parameters

- 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).

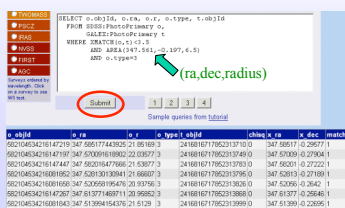
Columns from PhotoPrimary

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"](#)

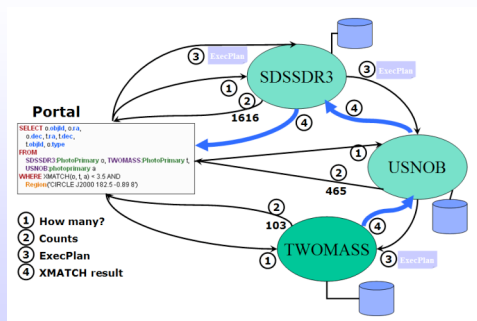


- To submit query, click **Submit**. Cross correlation results appear!

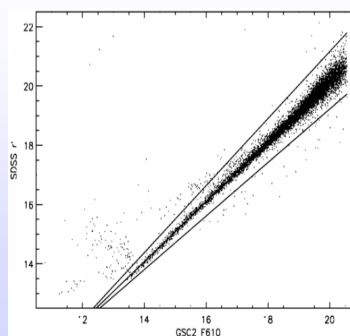
Open SkyQuery.net



Execution



Crossmatch result



Compute Services

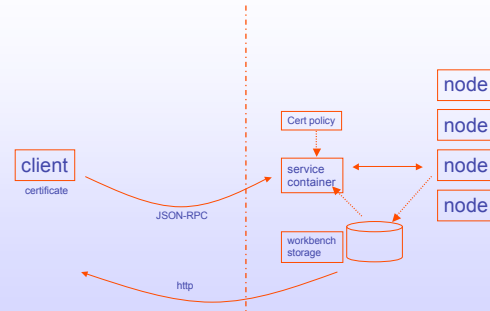
Compute Services

- **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

Step 1:

Connect Messages will appear here.

Connect to the remote Nessal server by clicking this button.

[Connect to Nessal](#)

Step 2:

Run Messages will appear here.

Enter the parameters for the run in the form below, then click this button.

Defaults are for the 751 galaxy.

Enter the size and location (degrees in Equatorial (2000))

Right Ascension:

Declination:

Right Ascension Width:

Declination Width:

Filter:

Background Correction:

[Run Service](#)

Step 3:

Monitor Messages will appear here.

Click here to monitor. When the popup says it is finished, you can download the result from the link below.

[Job Monitor](#)

```

nessiServer=nessi.client('https://envoy.c
acr.caltech.edu:8443/clarens/',debug=0)

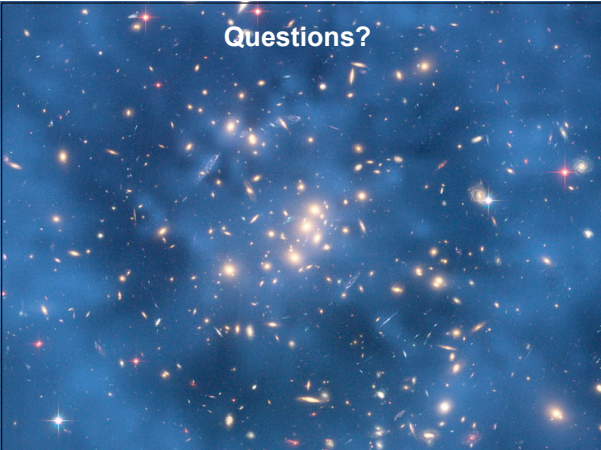
mosaic_loc = "-ra 49.1 -dec 60.1 -rawidth
0.5 -dcwidth 0.5 -filt f -bgcorr 0"

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

msg = dsbvr.dpossMosaic.monitor(session)
print msg

```

Repeat the monitoring



Questions?