# SOFTWARE, ARCHIVES AND NVO

David De Young
NOAO

# The Data Avalanche

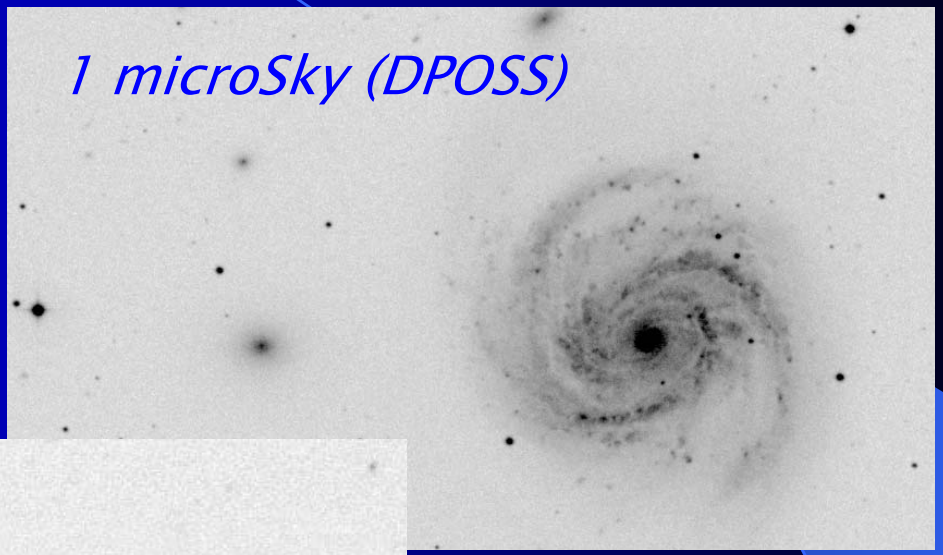
- Current Holdings:
  - ~ 20 Collections in Excess of 1 TB Each
    - 2MASS, Chandra, DENIS, DPOSS, DSS I&II, GSC II, HST, LONEOS, LOTIS, MACHO, Mosaic N., NEAT, NOAO Surveys, NVSS, OGLE II, SDSS (early), SUMSS, VLA, VLBA
  - Plus Many Other Smaller Collections
  - Ground Based Surveys: ~ 50 TB

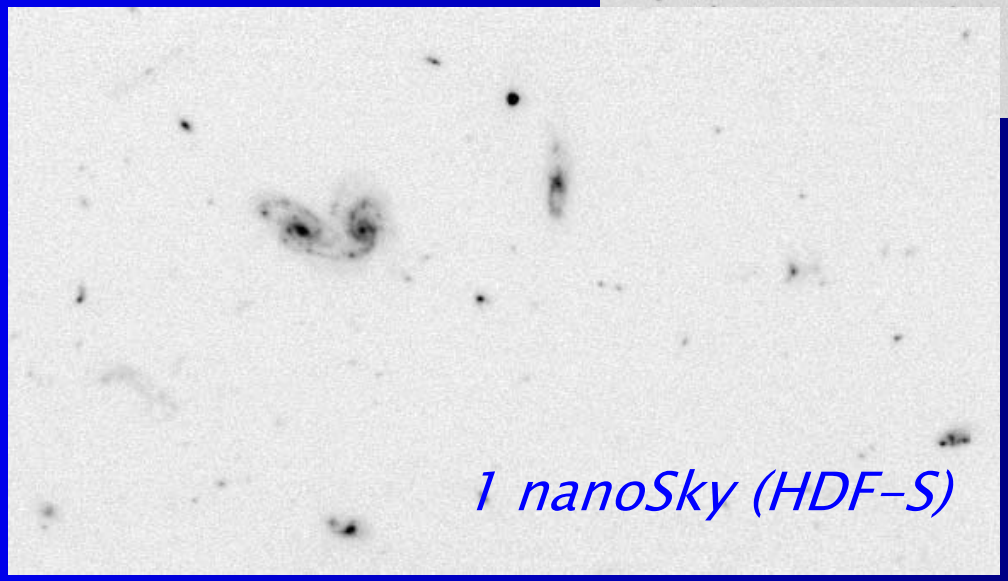
# The Data Avalanche

- Some Comparisons:
  - Ground Based Surveys: ~ 50 TB; Current Total > 100 TB and Growing
  - Human Genome: < 1 GB
  - Library of Congress: ~ 20 TB

# The Data Avalanche: The Future

Multi-Terabyte (soon: multi-Petabyte) sky surveys and archives over a broad range of wavelengths




Billions of detected sources, hundreds of measured attributes per source

# The Data Avalanche

- The Future:
  - SDSS Final (~ 20 TB)
  - Future Missions and Facilities (GALEX, Spitzer, GSMT, …)
  - LSST
    - ~ 6 TeraPixels/ Night (~ 2MASS)
    - Total Data Archive in Petabyte Range
    - Largest Projected Single Contributor of Data

# The Data Avalanche

- The Future:
  - Digital Libraries:
    - ADS, astro-ph, NED, CDS, NSSDC
      - (ADS Already at ~ 400 GB)
  - ~ 1 TB/Sky/Band/Epoch

    (1 Arcsec Pixels; 1 Byte per Pixel)

# Archived Data

- Can Provide New Science
  - New Ideas
  - Statistical Revelations
  - Multi-Wavelength Synthesis
  - Time Domain Analysis
- Can Be a Unique Educational Resource
- Can Enhance Public Education/Outreach

# Archived Data

- Is Essentially Useless Unless:
  - It Is Easily Accessible
    - Meta Data, Standards
    - Protocols, Bandwidth
    - Simplicity of Procedures
  - Many Useful Tools Are –
    - Readily Available to All
    - Simple to Use

# The Solution…

- The Virtual Observatory
  - "Single Point" User Access to Multiple Datasets
  - Standardized Formats for Metadata and Access Protocols
  - Provision of Tools for Retrieval and Analysis
  - Provision of High Bandwidth and Grid Computing
  - National (NVO) and International Initiative

# The US NVO: History

- 1990s: NASA establishes science archive centers
- Apr 1999: Decadal Survey Panel on Theory, Computation, and Data Discovery (LANL)
  - Szalay, Prince, and Alcock coin the name "National Virtual Observatory"
- Nov 1999: NVO organizational workshop at JHU
- Feb 2002: 2nd NVO workshop - NOAO-Tucson
- Jun 2000: 1st NVO Conference - Caltech

# The US NVO: History

- National Academy of Sciences Decadal Survey:

  *" Several small initiatives recommended by the committee span both ground and space. The first among them—the National Virtual Observatory (NVO)—is the committee's top priority among the small initiatives. "*

  *—Astronomy and Astrophysics in the New Millennium, p. 14*

# The US NVO: History

- April 2001: Proposal Submitted to NSF ITR Program- 17 Collaborating Organizations
- September 2001: NSF Announces Proposal Selection - $10 M Award
- January 2003: First NVO Science Prototypes Shown at Seattle AAS

# What the Virtual Observatory Is…

- A suite of international standards for discovery, exchange, and analysis of data
- A data access and analysis environment that utilizes developments in ITR
- A framework for data processing that enables development and propagation of useful algorithms

# What the Virtual Observatory Is…

- A tool for science planning: New missions, instruments, and experiments
- A catalyst for world-wide access to astronomical archives
- A routinely used tool of the research astronomer
- A vehicle for education and public outreach

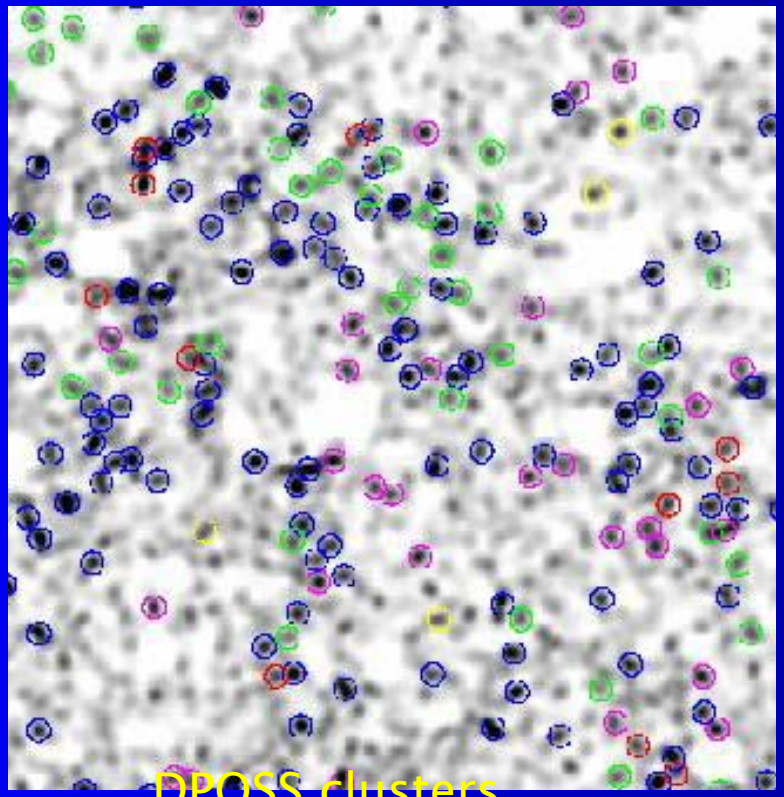
# What the Virtual Observatory is *Not*…

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

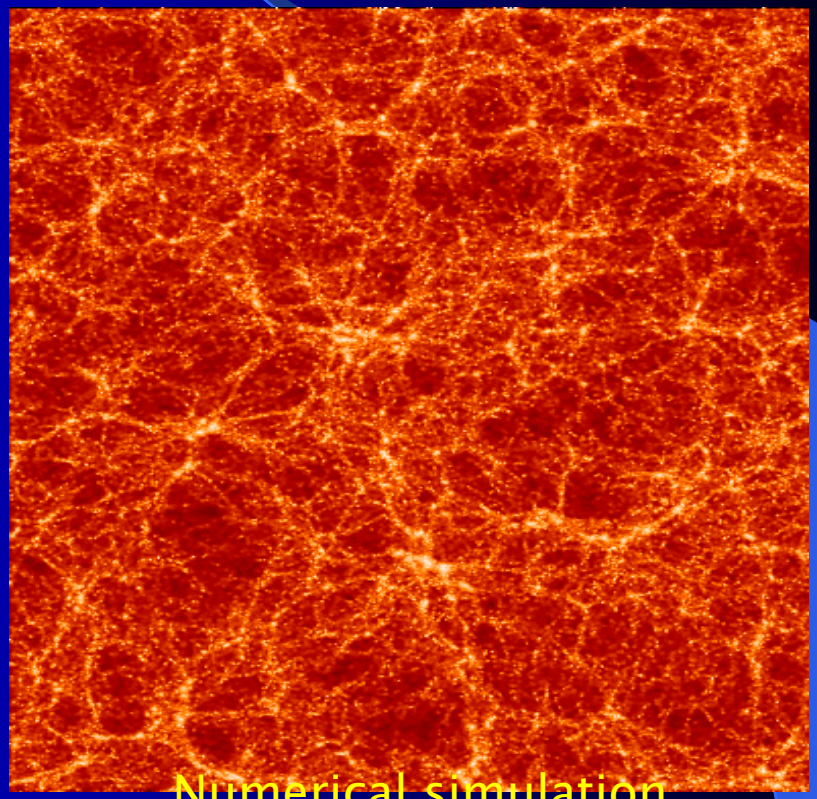
# Science of a Qualitatively Different Nature

- Statistical astronomy done right
  - Cosmology, Galactic structure, stellar astrophysics
  - Discovery of significant patterns and multivariate correlation
- Systematic exploration of the observable parameter spaces
  - Searches for rare or unknown types of objects and phenomena
  - Low surface brightness universe, the time domain
  - Confronting massive numerical simulations with massive data sets

# Cosmology: A Marriage of Theory and Observations

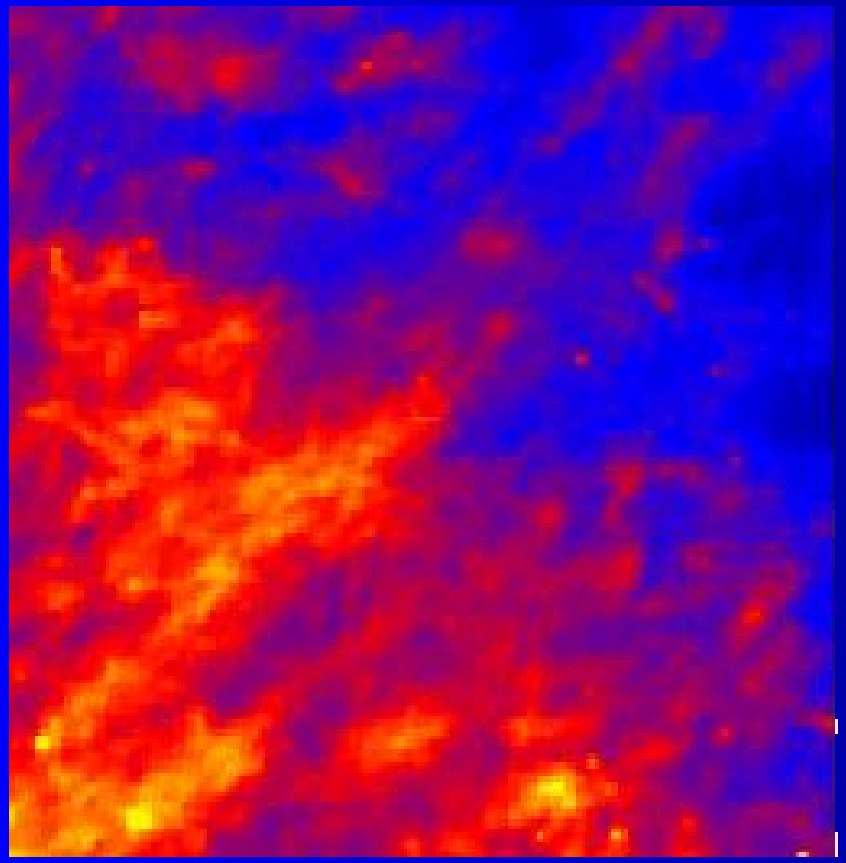


DPOSS clusters

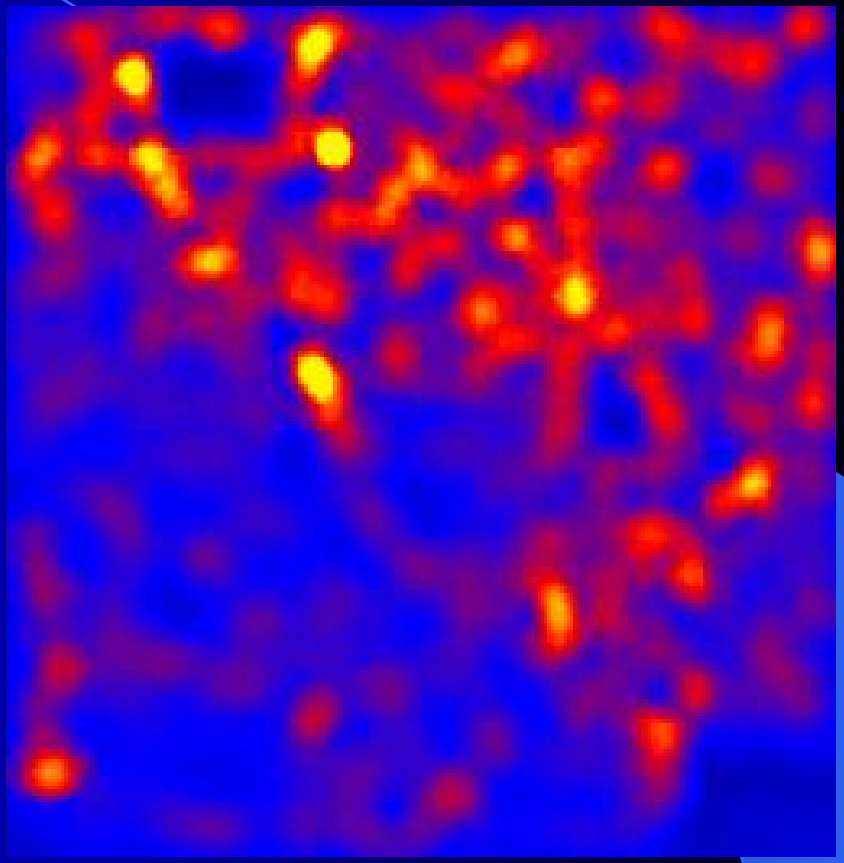


Numerical simulation

# Multi-Wavelength Data paint a more complete (and more complex!) picture of the universe



Infrared emission from interstellar dust



Smoothed galaxy density map

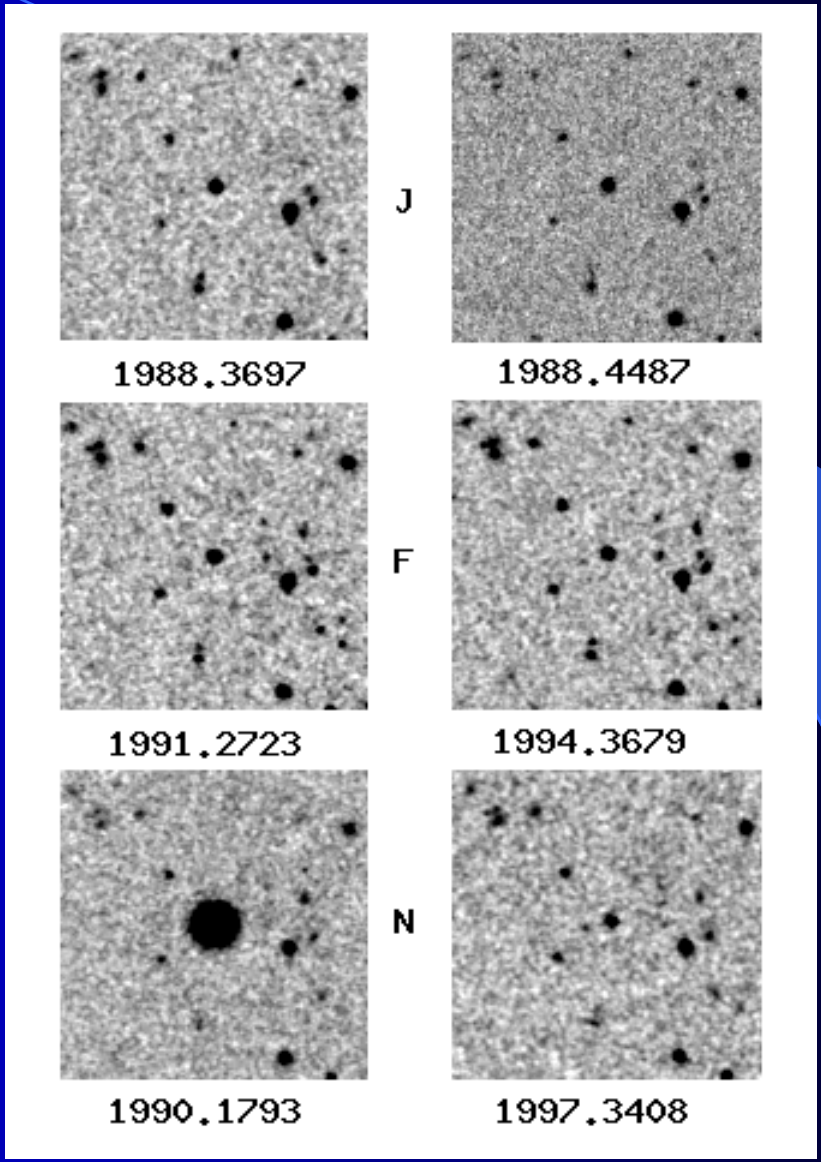
# A Panchromatic Approach to the Universe…

…reveals
**a more complete physical picture**

The resulting complexity of data translates into increased demands for data analysis, visualization, and understanding

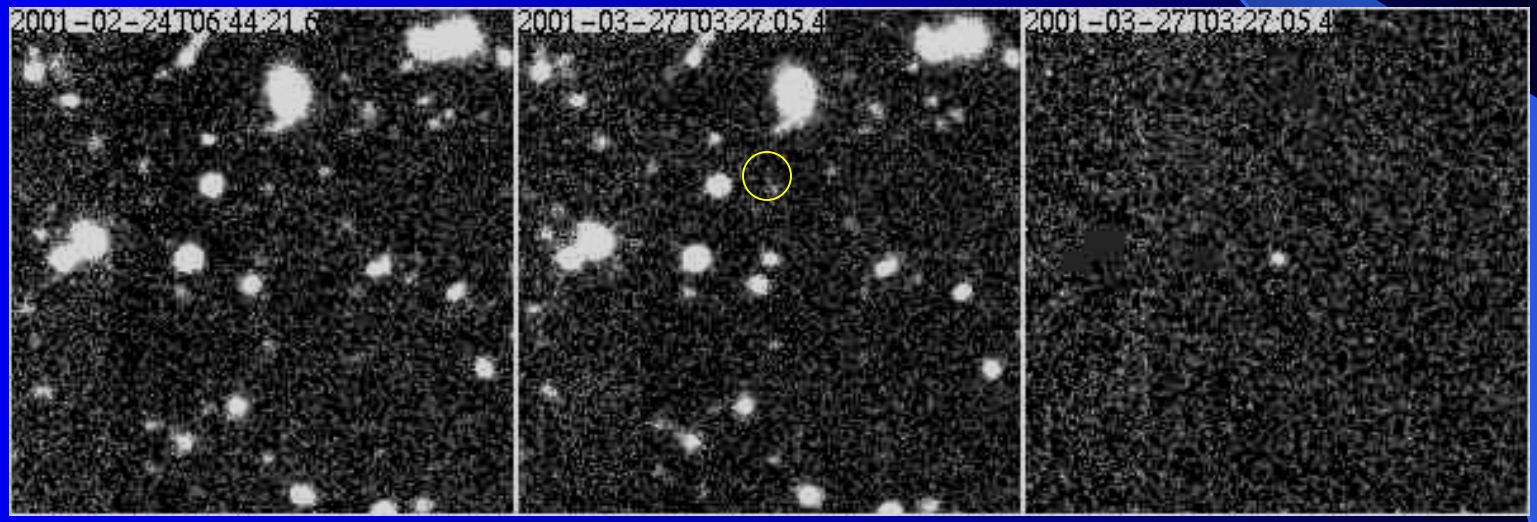
# The Time Domain

Megaflares on normal main sequence stars (DPOSS)

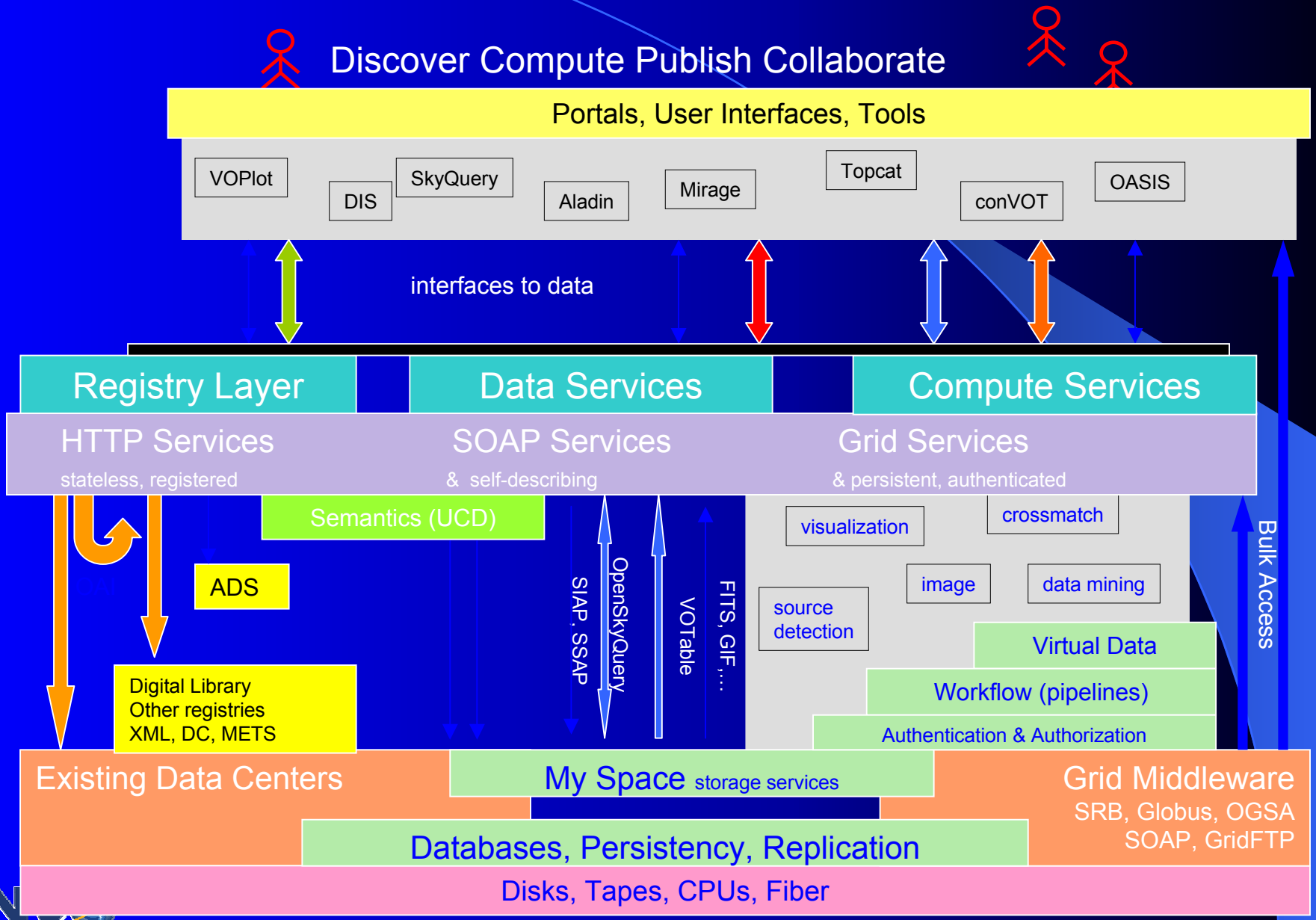


# The Time Domain

Faint, Fast Transients



# NVO Architecture



# VOTable

- Reached international agreement on VOTable V1.0 specification in April 2002
- XML-based standard with in-line data or links to external data
- Utilized for basic catalog and image access protocols
- Merges “AstroRes” heritage with XML flexibility
- Complements FITS
- Multiple I/O libraries available (Java, Perl, C++, C#)

# Data Models

- Data modeling effort aimed at defining basic data types and relationships among them
- High-level entities: image, spectrum, time series, catalog
- Low-level entries: quantity, resolution, time of observation
- Interfaces and protocols for other VO services derived from DM relationships

# Data Access Layer

- Data Access Layer is mediator between NVO data requests and data delivery
- Defined "Cone Search" protocol and have ~100 implementations
- Defined Simple Image Access Protocol (SIAP) and have 20+ implementations
- Specification for Simple Spectral Access Protocol in development

# Metadata and Registry

- Metadata describes NVO data collections, services; Metadata is collected into a Registry
- Resource Identifiers are component of resource metadata; have agreed on syntax
- Using Open Archive Initiative protocols for metadata collection
- Now focusing on query mechanisms and general updating/synchronizing options
- Prototype registry utilized in science demonstration, Data Inventory Service

# Unified Content Descriptors

- UCDs provide common data dictionary for describing contents of catalogs
- CDS initiative; now broadened to international VO discussion
- Current discussion focusing on structure and extensibility

# Unified Content Descriptors

I/220 The HST Guide Star Catalog, Version 1.1 (Lasker+ 1992) Query this catalogue in VizieR

| UCD | Associated to colum | With unit | Column description |
|---|---|---|---|
| CLASS_OBJECT | Class | | [0,3] Class of object (0=star; 3=non-stellar) |
| CODE_MULT_FLAG | Mult | | [TF] True if multiple object / False otherwise |
| DATA_LINK | Versions | | Get parameters of that star from all GSC versions |
| ERROR | PosErr | arcsec | Mean error on position |
| ERROR | e_Pmag | mag | Mean error on photographic magnitude |
| ID_MAIN | GSC | | GSC designation |
| ID_PLATE | Plate | | Plate designation |
| NOTE | n_Pmag | | [0,18] Coded passband for magnitude |
| PHOT_PHG_MAG | Pmag | mag | photographic magnitude (see n_Pmag) |
| POS_EQ_DEC_MAIN | DEJ2000 | deg | Declination in J2000, epoch of plate |
| POS_EQ_RA_MAIN | RAJ2000 | deg | Right ascention in J2000, epoch of plate |
| TIME_EPOCH | Epoch | yr | Epoch of plate |

# VO Query Language

- Minimal extensions to SQL to support astronomical queries
- Defining standard query service based on SDSS SkyQuery: OpenSkyNode and OpenSkyQuery
- Investigating higher-level query languages; e.g., “natural” language

# Grid and Web Services

- Increasing number of web services
- Registry services will be implemented as web services
- Prototyped use of Grid in galaxy morphology science demonstration
- Working closely with Grid community to understand progress on Grid services (e.g., OGSA) and to determine best time to adopt

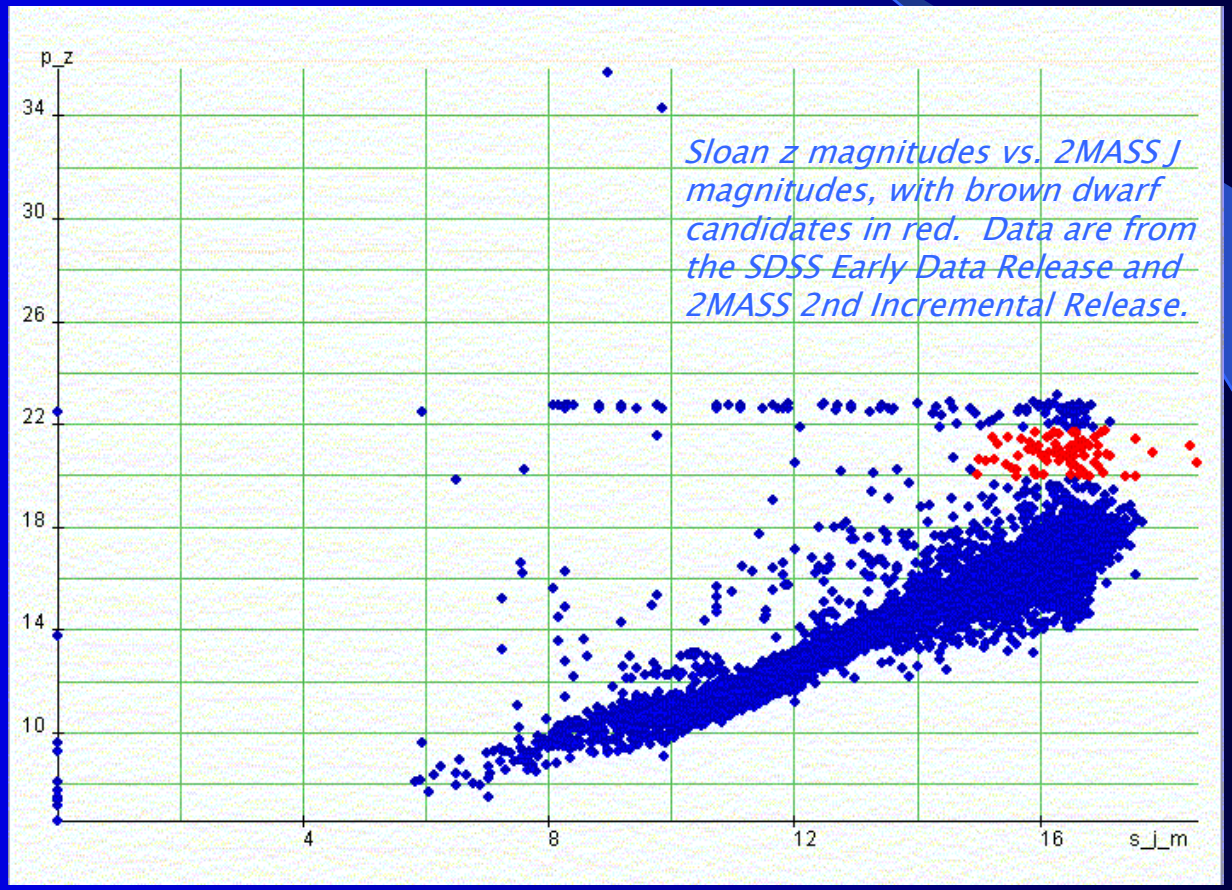
# Science Prototypes

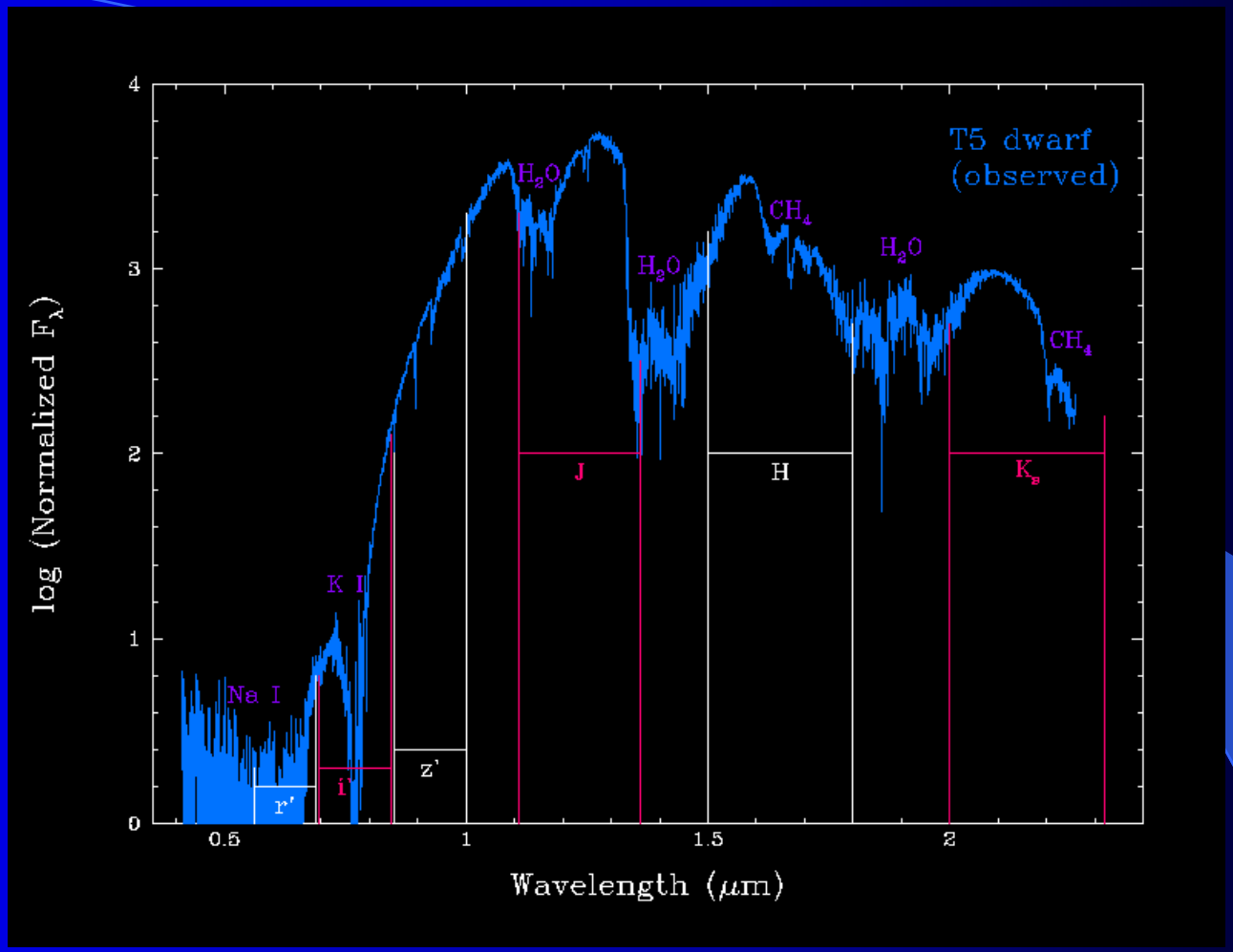
- Goal: Guide and validate technical initiatives
- Year 1
  - Brown dwarf candidate search
  - Gamma-ray burst follow-up
  - Galaxy morphology measurement (utilizing computational grid)
- Year 1.5
  - Data Inventory Service

# Science Prototypes

- Goal: Guide and validate technical initiatives
- Year 2
  - Interface theory simulations and observations
- European VO project
  - Type 2 (obscured) quasars: 40 new candidates found
  - Galactic star formation regions

# Brown Dwarf Candidate Search



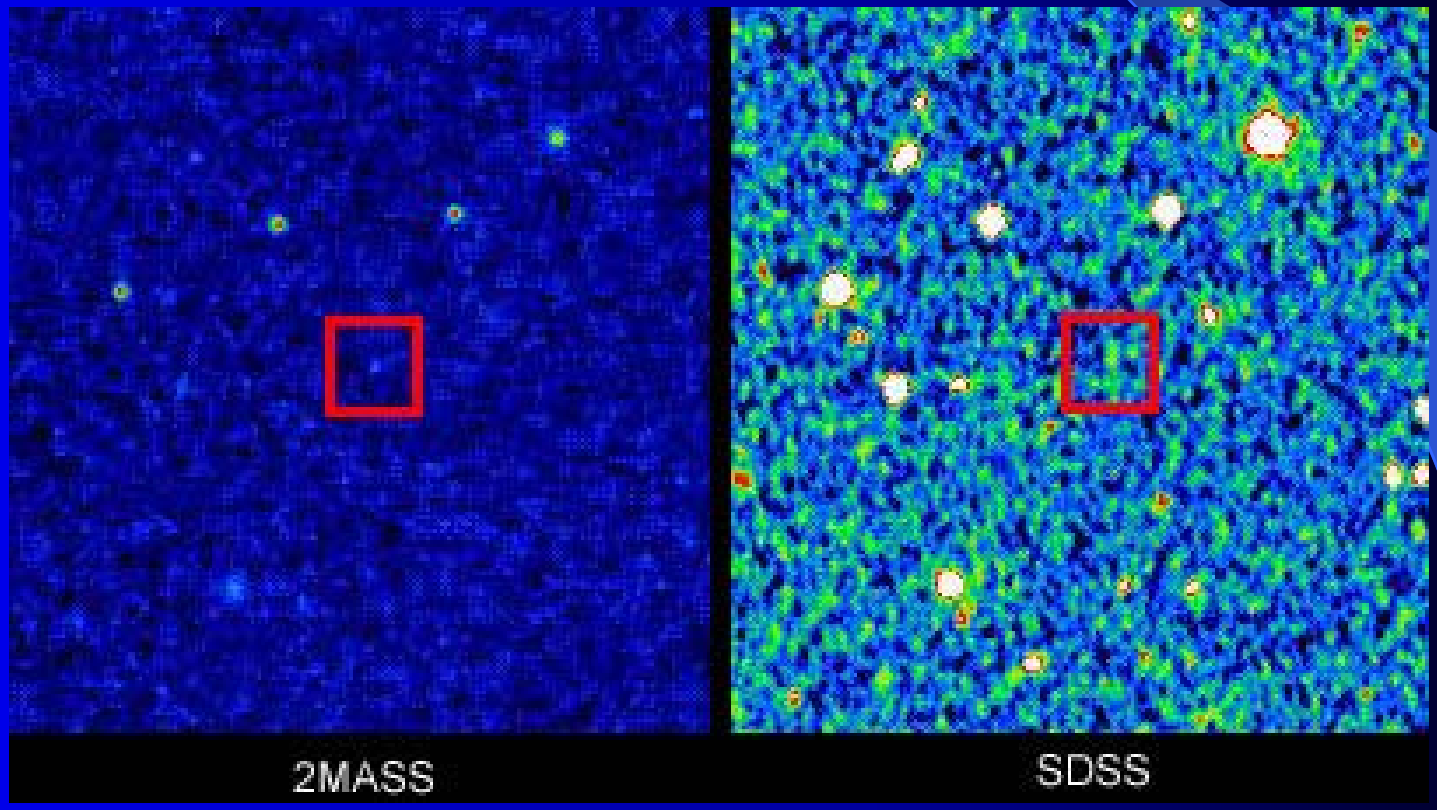



As a T dwarf becomes cooler (i.e., methane and water absorptions increase) or more distant…

- SDSS detects it only at z' band
- 2MASS detects it only at J band

# Demo Leads to Discovery!

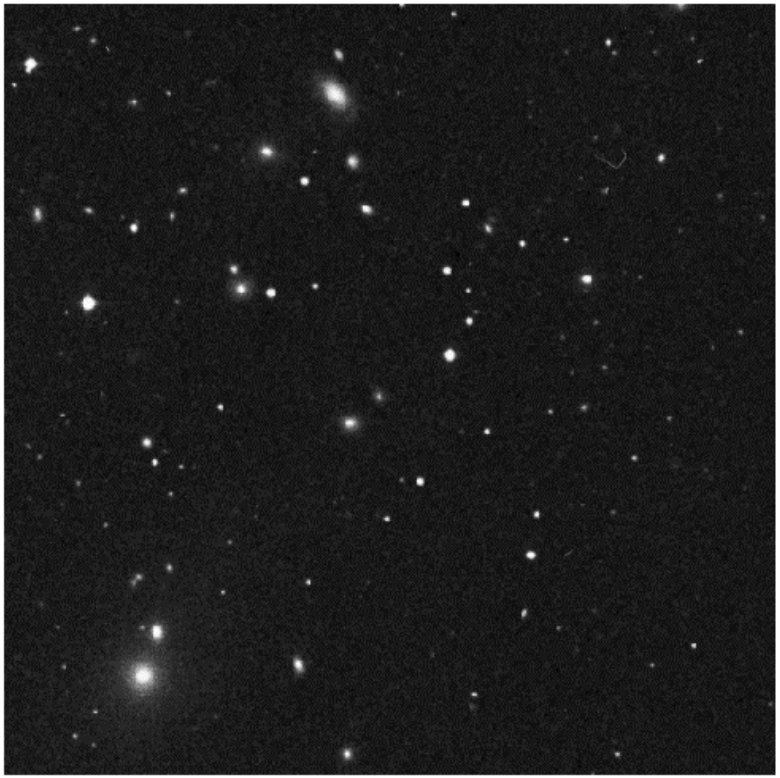
- New brown dwarf candidate confirmed spectroscopically with Keck Observatory



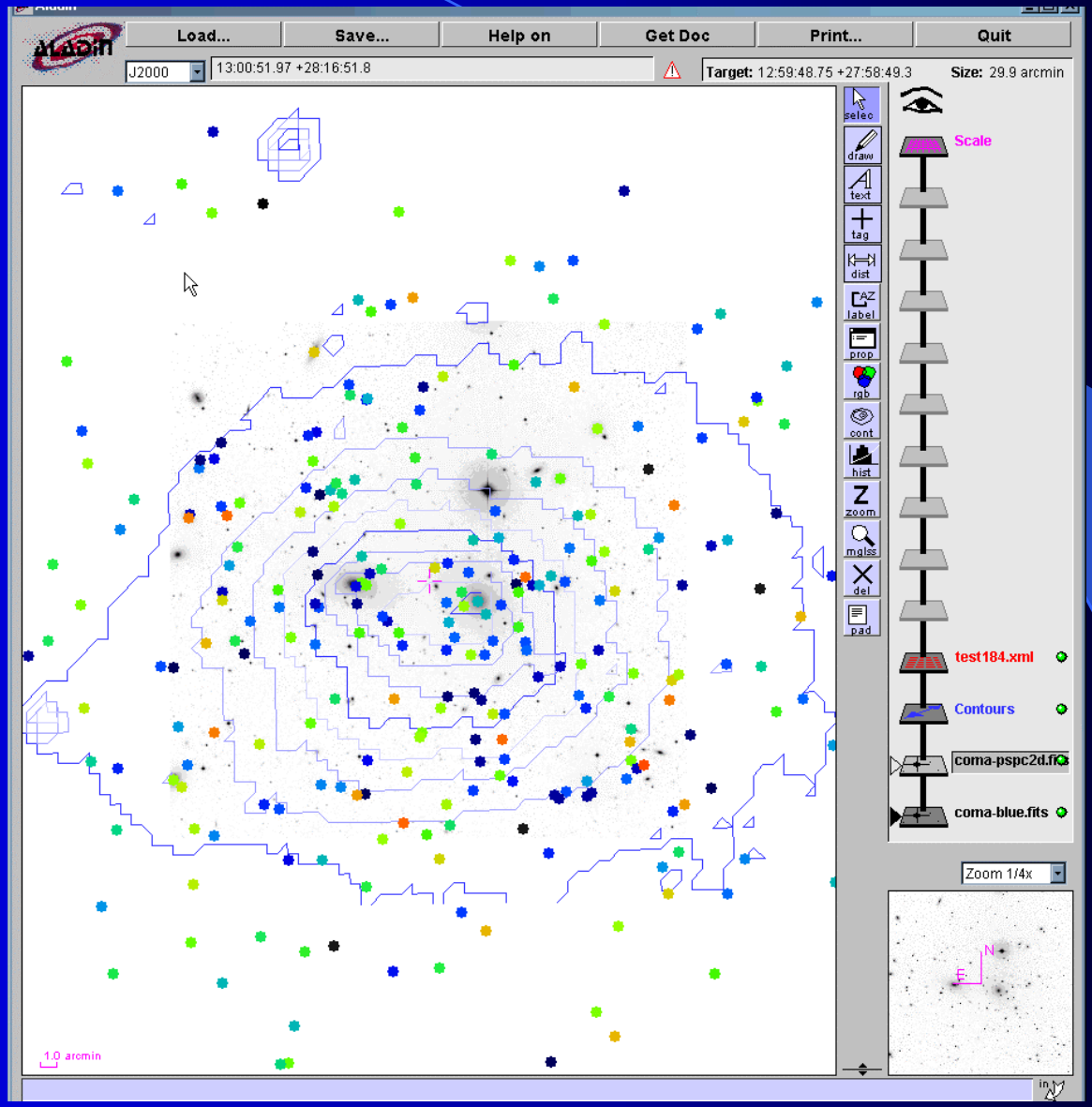
# Galaxy Morphology in Clusters

## Select a Cluster

| | | |
|---|---|---|
| Cluster Name: | Abell 1367 | When you select a cluster, the coordinates will be update and an optical image will appear. |
| Coordinates: | 176.1858333, 19.6997222 | |
| Morph. Service: | Analyze Cluster on Grid | Click the "Analyze" button to launch the Grid calculations. |
| Results: | ☐ CACHED Results | Check box to view previously computed results |
| Cluster Images: | DSS FITS<br>ROSAT FITS<br>CHANDRA FITS | **Check out what's happening under the hood!** |



# Galaxy Morphology in Clusters



# Galaxy Morphology in Clusters

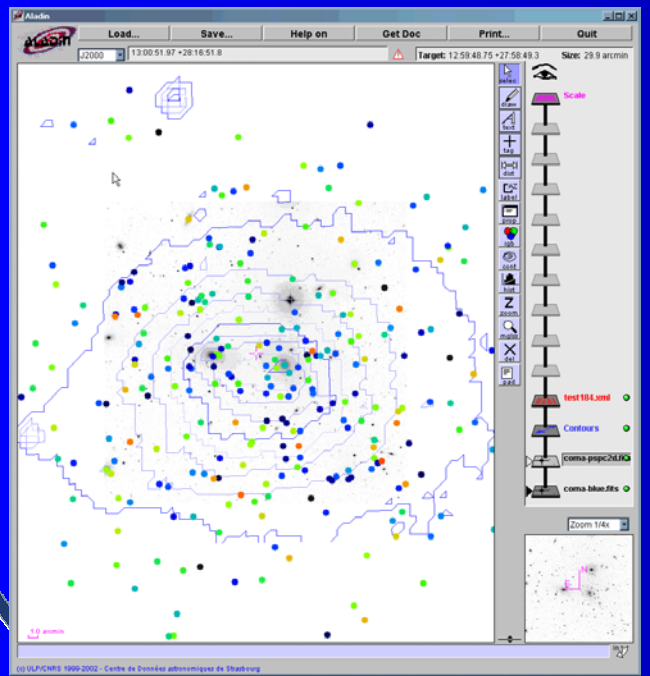
**Scientific Motivation:** Investigate the dynamical state of galaxy clusters and galaxy evolution within the context of large-scale structure. Use galaxy morphology as a probe of dynamical history by calculating, for each galaxy in a cluster:

- Surface brightness
- Concentration index
- Asymmetry index

These parameters are analyzed with other indicators such as magnitude, color, peculiar velocity, position in cluster, and cluster large-scale structure.

**Data Resources:**
Chandra X-ray image (SAO/CXC)
ROSAT image (GSFC/HEASARC)
DSS image (STScI/MAST)
Galaxy cluster catalogs (NED)
CNOC1 cluster images and catalogs (CADC)

**Computing Resources:**
USC/ISI
UW-Madison/NCSA
Fermilab




**What the VO Brings:** Distributed data access and Grid-based computing make possible for the first time effective integration of multiple datasets and real-time computing. Integration of data from diverse sources is enabled by standardized data objects and standardized remote computing services. Flexibility of access means that further NVO-compliant images and catalogs can be added easily. Users can select their visualization portal (Aladin, OASIS, DS9).

**Enabling Technologies:** VOTable, NVO-compliant catalog and image access, standard semantics, Grid computing infrastructure.

**Future Prospects:** Dynamic discovery and selection of image, catalog, and computing resources. User-selection of analysis tools and ability to publish data to the NVO framework.

# Globular Cluster Simulations

**TVO demo AAS203**

**Overview**

*Action:*

Select CMD demo

**1. Choose a cluster:**

- (●) Globular Cluster M30 @ 8.0 kpc:
- ( ) Globular Cluster NGC 104 @ 4.5 kpc:
- ( ) Globular Cluster NGC 6397 @ 2.3 kpc:
- ( ) Old Open Cluster M 67 @ 0.83 kpc

Override default distance (kpc):

**2. Choose a model:**

- (●) AMNH run3: N=100,000, 0% binaries, Z=0.001, small globular cluster with 40k left at 12Gyr
- ( ) AMNH run2: N=27,000 50% binaries, Z=0.02, rich open cluster dead at ~ 6Gyr
- ( ) GRAPE run2
- ( ) GRAPE run3
- ( ) Survey2: choose parameter selections from the appropriate popup menus below
- ( ) Test: for internal use; not for public consumption

Simulation Time (Myr): 12000.0 (Use a range 5000:7000 to select a timerange, or 'all' for all times)

**3. Optional parameters for some of the models:**

**Survey2:** *King W:* 3 *IMF slope:* -2.35 *Binary fraction:* 0.1 *N:* 256 *Realization#:* 0

☐ Animate CMD? (can take 2-3 minutes) ☐ debug

**4. GO:**

Start the simulation/Perform action | Reset all values

# Globular Cluster Simulations

**Survey2:** *King W:* 3 *IMF slope:* -2.35 *Binary fraction:* 0.1 *N:* 256 *Realization#:* 0

☐ Animate CMD? (can take 2-3 minutes) ☐ debug

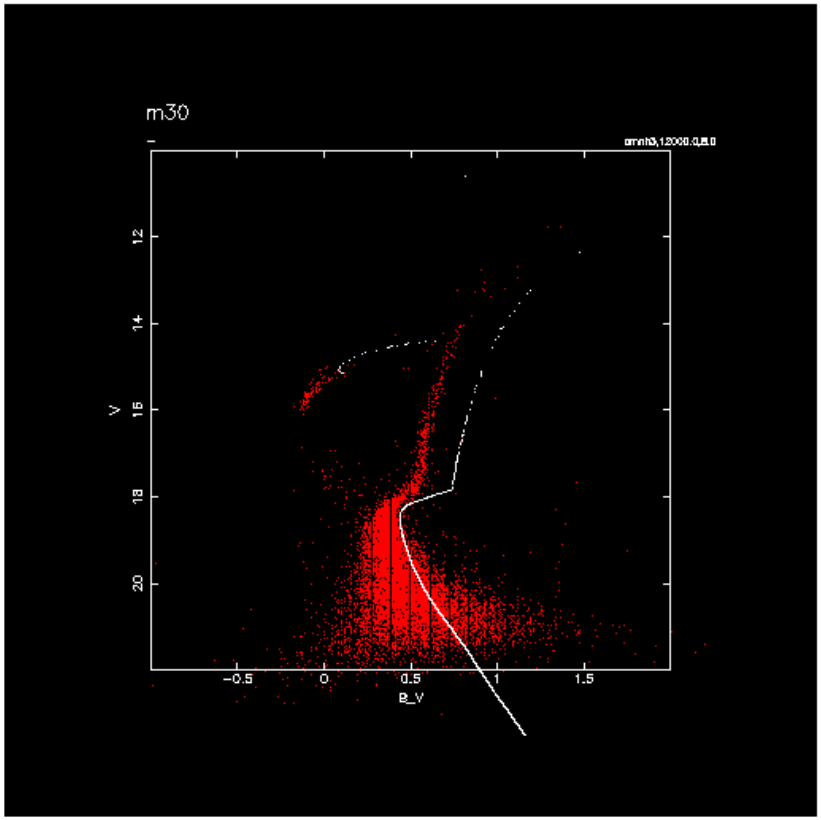
**4. GO:**

Start the simulation/Perform action | Reset all values

Using cluster 'm30' at 8.0 kpc
comparing to AMNH model 'amnh3' at time = 12000.0 Myr

(Working directory = http://bima.astro.umd.edu/nemo/tvo/nvodemo2004//tmp)

CMD:



# NVO Science Applications

The First Steps

# NVO Data Inventory Service

**Scientific Motivation:** Rapid collection of multi-wavelength imaging, catalog and observational data for any location on the sky. Very relevant for transient events.

**Data Resources:** Multi-wavelength data from several sites (currently 13) sampling energies from X-ray to radio; includes images, object lists, and catalogues.

**Role of NVO:** Integration and organization of a variety of data sources into an easily comprehensible information set. Scalability to an arbitrary number of data providers. Integrates data with multiple data visualization services.

# Data Inventory Service

**Data Inventory Service**

National Virtual Observatory: Hosted at the HEASARC

What do we know about regions of sky?
Using new Virtual Observatory protocols we can gather and organize information efficiently on a given region of sky.

Enter a position(or name) and the maximum size of the region of sky you're interested in.

| | | |
|---|---|---|
| Object Position or Name: | cen a | (degrees or sexagesimal) |
| Size: | 0.25 | (in decimal degrees) |

Send Request | Reset Form

☐ Ignore cache! The DIS will reproccess an identical request rather than linking to the existing cache results.

**Example Inputs for the Object Position or Name**

- 13.29, -18.47 [Object Position: Decimal degrees]
- 6 45 10.8, -16 41 58 [Object Position:Sexagesimal format; RA in hours]
- 3c273 [Object name]
- Use a comma to delimit J2000 RA and Dec pair.

**About Data Inventory Service**

1. A user request is broadcast to sites scattered all over the world using two simple common protocols.
2. Catalog data and lists of available images are returned using the new VOTable XML standard.
3. Image, observation and catalog data from these sites are collected and organized for immediate viewing.
4. Data may be analyzed or visualized in Aladin or OASIS

Participating sites currently include: NRAO, NOAO, JHU, ST ScI, HEASARC, NCS, IRSA, CDS, NED, ESO, SDSS, CXC.

# Data Inventory Service

**Data Inventory Results: cen a**

Data missing - Instructions  Home

National Virtual Observatory: Hosted at the HEASARC

**Note: Inventory request completed**

| RA | Dec | Size |
|---|---|---|
| 13 25 27.62 | -43 01 08.8 | 0.25 |

☐ Check All

| | | | | | |
|---|---|---|---|---|---|
| **Images (FITS/GIF)** | | | | | |
| **Optical** | ☐ DSS1 SV | ☑ DSS2 | ☐ DSS2B | ☐ DSS2IR | ☐ DSS2R |
| **Infrared** | ☐ 2MASS-H | ☐ 2MASS-J | ☐ 2MASS-K | | |
| **Radio** | ☑ SUMSS | | | | |
| **X-ray** | ☐ RASS B | ☑ Chandra(6) | | | |
| **Observations (VOTable)** | | | | | |
| **Optical** | ☐ HST(100) | ☐ STIS(100) | ☐ WFPC2(100) | ☐ WFPC1(22) | ☐ HSTG(394) |
| **Infrared** | ☐ NICMOS(100) | | | | |
| **X-ray** | ☐ ASCA(3) | ☐ ROSAT(9) | ☐ ROSPUBLC(10) | ☐ RXTE(23) | ☐ EXOSAT(12) |
| | ☐ CHANMAST(10) | ☐ Einstein(5) | ☐ XMMMAST(3) | ☐ ASCAMAST(3) | ☐ XTEINDEX(5) |
| **Gamma-ray** | ☐ OSSE(29) | | | | |
| **UV** | ☐ FUSE(1) | ☐ FOC(20) | ☐ HUT(2) | ☐ IUE(41) | ☐ UIT(7) |
| | ☐ WUPPE(1) | | | | |
| **Objects (VOTable)** | | | | | |
| **Surveys** | ☐ USNO-A2.0(1197) | ☐ USNO-SA2.0(1197) | ☑ GSC1(289) | ☐ GSC2.2(2259) | ☐ UCAC1(305) |
| | ☐ USNO-A2.0 CDS(999) | | | | |
| **Galaxies** | ☐ SGC(1) | ☐ PGC(1) | ☐ NBG(1) | ☐ RC3(1) | ☐ RNGC(1) |
| | ☐ PSCz(3) | | | | |
| **Stars** | ☐ HIP(1) | ☐ SAO(2) | ☐ WDS(1) | ☐ AC2000.2(30) | ☐ ASCC-2.5(21) |
| | ☐ HD(4) | | | | |
| **Misc.** | ☐ EGRET3(45) | ☐ WGACAT(35) | ☐ Radio Catalogs(69) | ☐ 2MASS-PSC(CDS)(999) | ☐ Veron-Veron(1) |
| | ☐ TYCHO-2(22) | | | | |

# Data Inventory Service

| Event | Time | RA | Dec | Size | Source |
|---|---|---|---|---|---|
| 141 | 2001-02-22 12:00:00 | 14 52 12.00 | +43 01 01.6 | 0.25 | Other |

**Images**

| | | | |
|---|---|---|---|
| Optical | ☑ DSS(Blue) | ☑ DSS(Red) | |
| X-ray | ☑ RASS | | |
| Radio | ☑ NVSS | ☑ WENSS | ☑ FIRST |

**Observations**

| | |
|---|---|
| HST | ☑ HST(141) |
| X-ray | ☑ Chandra(2) |

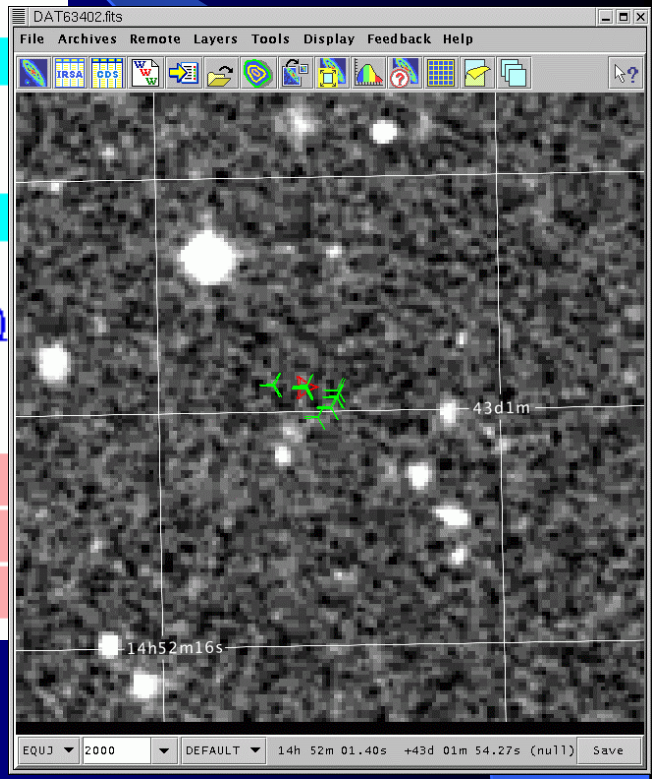
**Objects**

| | | | |
|---|---|---|---|
| Major Catalogs | ☐ NED(25) | ☐ GSC2.2(294) | |
| | ☐ USNO A2(213) | ☐ 2MASS(P)(236) | ☐ 2MASS(X)(2) |
| Clusters | ☑ CEDAG(3) | | |
| Stars | ☐ AC2000.2(1) | | |

Analyze data in Aladin

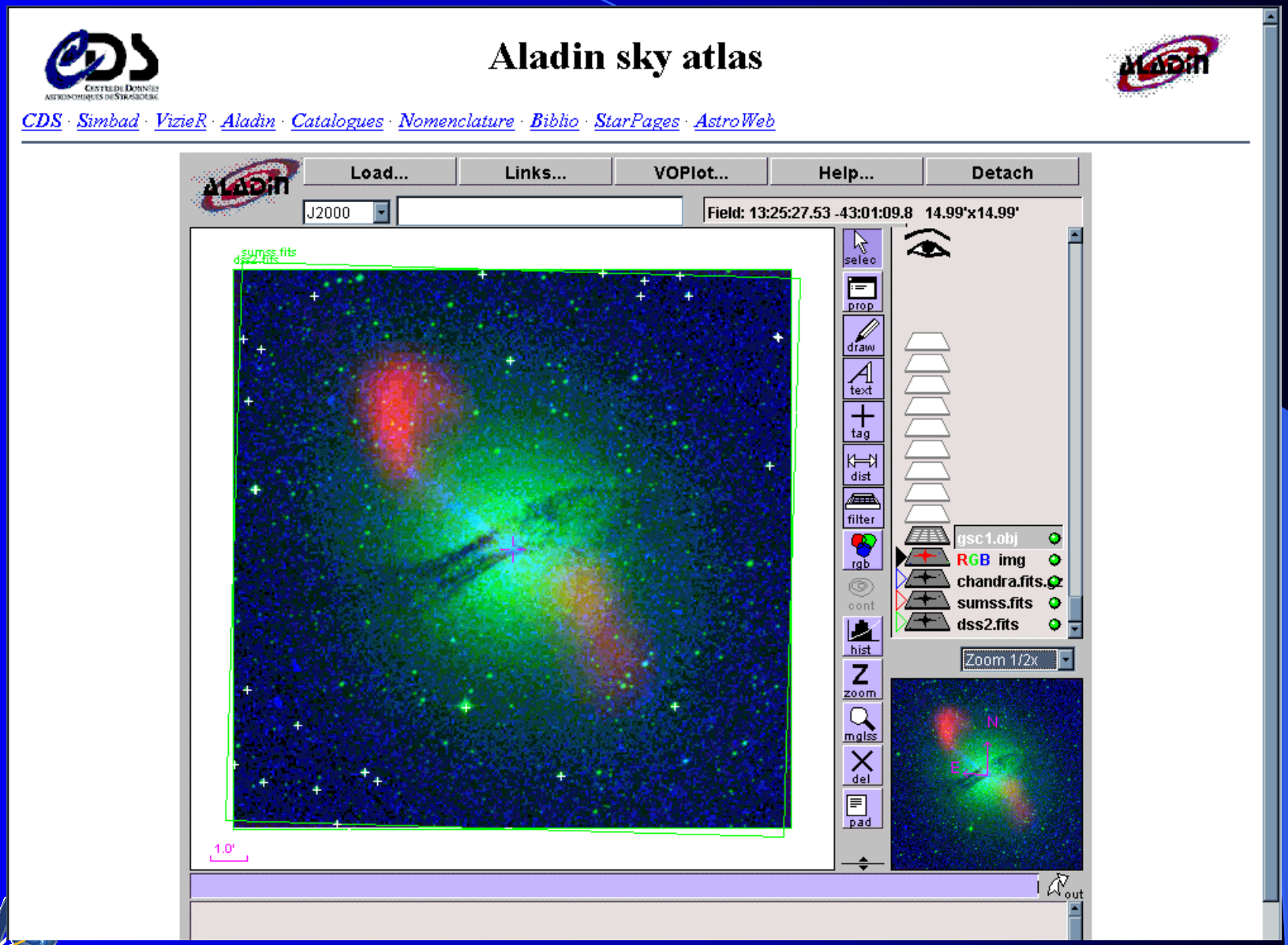
Analyze data in OASIS

Download selected data




*Pointing of HST and Chandra observations for GRB010222*

# Data Inventory Service



# Data Inventory Service

**Enabling Technologies:** Standard protocols to remote services such as Cone Search and Simple Image Access, standardized VOTables for data retrieval transformation, and standardized semantics encoded as Uniform Content Descriptors (UCDs). Resource registry.

**Future Prospects:** Customization and quality control of resources searched; more sophisticated use of metadata.

*Positions of HST and Chandra observations for GRB010222*

# Spectral Database Browser

Spectrum Services for the Virtual Observatory

not logged in
login | register

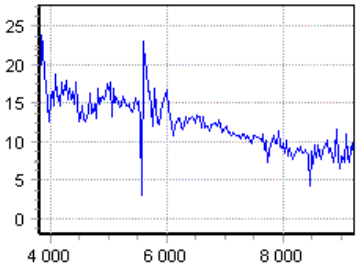
home | filters | spectra | webservices | user

## Spectrum Services for the Virtual Observatory

Welcome to our web site dedicated to spectrum related VO services developed at the Johns Hopkins University. On these pages you will find tools and tutorials on how to access close to 500,000 spectra from the Sloan Digital Sky Survey (SDSS DR1) and the 2 degree Field redshift survey (2dFGRS). In addition to the spectra, these pages also publish the filter profiles of photometric pass bands of many surveys. The services are open to everyone to publish their own spectra in the same framework. Reading the tutorials on XML Web Services, you can learn how to integrate the 45 GB spectrum and passband database with your programs with few lines of code.

### Spectrum catalog

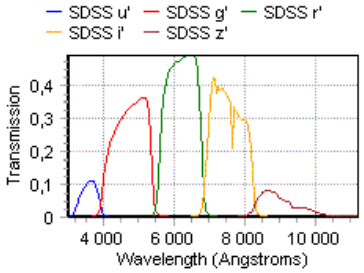
The spectrum catalog has a growing selection of spectra data from several surveys. It currently contains the complete SDSS DR1 and 2dFGRS spectrum catalog. The collection is expandable: you can add your own spectra using the web site or the web services.




The search tools include simple ones e.g. select a redshift range as well as more advanced ones, where the selection of spectra may be refined. Keeping with the VO standard ConeSearch, one can also select objects in a given virtual pointing allowing to find objects by their RA, Dec coordinates.

### Filter Profile catalog

The passband database contains the profiles of almost 100 photometric filters of several instruments including the largest and most recent ones such as the SDSS, HST, etc. The list is easily also expandable, anyone can post additional filter profiles using a web form or automatically using a client program to the web services.




The filter profile catalog can be searched by keyword and by advanced query options including effective wavelength and UCD. The results may be overplotted on each other for easy comparison of photometric systems or between different versions. Give it a try now.

# Spectral Database Browser

Spectrum Services for the Virtual Observatory

not logged in
login | register

home | filters | spectra | webservices | user

## Spectrum Advanced Search

Keyword:

Name:

Description:

UCD:

ra between: (deg) and

dec between: (deg) and

z between: 2.6 and 2.7

$z_{Err}$ between (absolute): and

Object class: QSO

Survey: Any

Measured between: and

Registered between: and

Modified between: and

Min. wavelength between: (A) and

Max. wavelength between: (A) and

Wavelength scale: Any

Flux unit: Any

Display format: List Graph

Search

# Spectral Database Browser

Spectrum Services for the Virtual Observatory

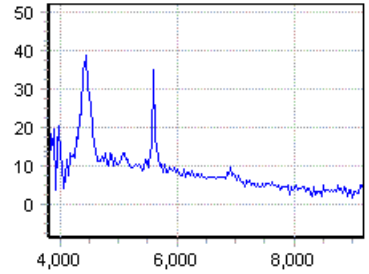
not logged in
login | register

home | filters | spectra | webservices | user

## Spectrum Advanced Search Results

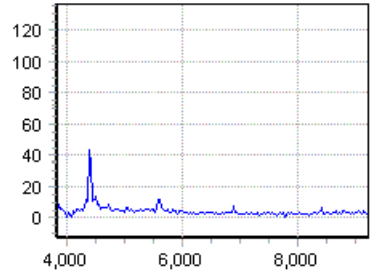
Found 47 objects. Displaying from 1 to 12 switch to list mode 1 [ Next ] [ Last ]

1 . SDSS J173612.70+542009.23
0360-51809-01
class: QSO , Z = 2.6246
ra = 264.052910 , dec = 54.335898



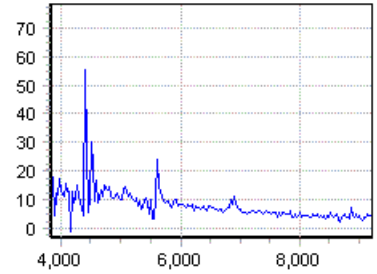

details

2 . SDSS J095334.96+003724.52
0268-51633-01
class: QSO , Z = 2.6050
ra = 148.395680 , dec = 0.623478



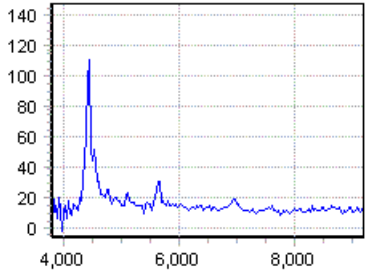

details

3 . SDSS J103152.90+010923.06
0273-51957-01
class: QSO , Z = 2.6190
ra = 157.970400 , dec = 1.156406



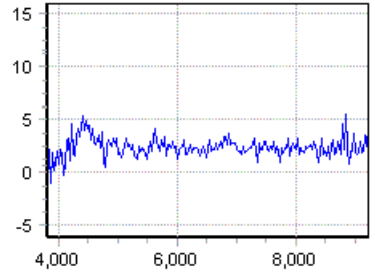

details

4 . SDSS J122848.21-010414.55
0289-51989-04
class: QSO , Z = 2.6554
ra = 187.200890 , dec = -1.070707



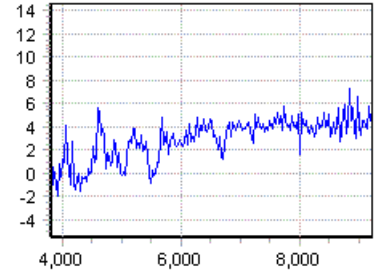

details

5 . SDSS J124157.72+001702.44
0291-51922-02
class: QSO , Z = 2.6028
ra = 190.490510 , dec = 0.284010




details

6 . SDSS J131010.75-003007.20
0294-51986-01
class: QSO , Z = 2.6500
ra = 197.544780 , dec = -0.501999




details

# Spectral Database Browser

Spectrum Services for the Virtual Observatory

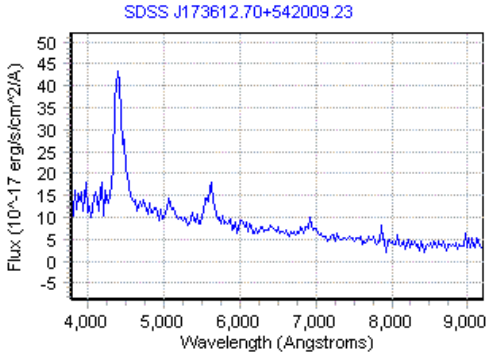
not logged in
login | register

home | filters | spectra | webservices | user

## Spectrum Details

| | |
|---|---|
| Spectrum name: | SDSS J173612.70+542009.23 |
| Description: | 0360-51809-01 |
| UCD: | SPECT_FLUX_VALUE |
| ra = | 264.05291 |
| dec = | 54.335898 |
| z = | 2.62457 |
| $z_{Err}$ = | 0.000679108 |
| Object class: | Qso |
| Wavelength scale: | Logarithmic |

Spectrum graph:

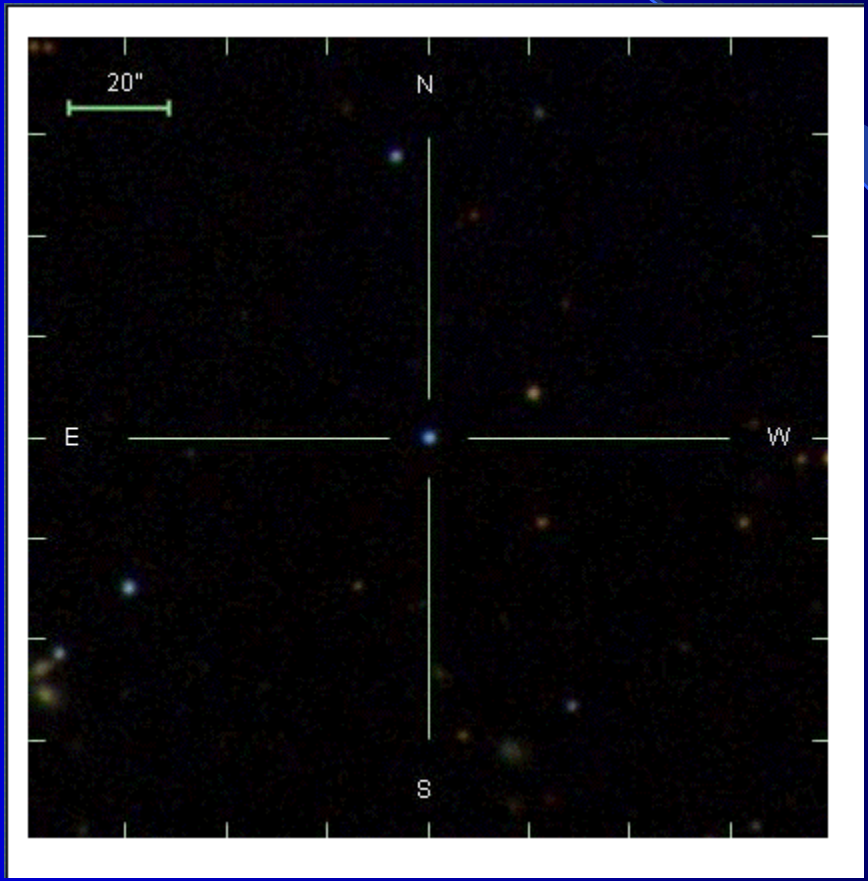



Back | Photo of the object

Get XML Get VOTable Get ASCII Get CSV

Plot a graph

# Spectral Database Browser



# NVO: Broader Context and Vision

# International VO Alliance

# International VO Alliance

- A World-wide Coordination of VO Initiatives
  - Agreements on standards for data access
  - Coordination of development activities
  - Sharing of software, expertise
- 14 participating organizations:
  - Astrogrid, AVO, US-NVO, VO-Australia, VO-Canada, VO-China, VO-France, VO-Germany (GAVO), VO-Hungary, VO-India, VO-Italy (DRACO), VO-Japan, VO-Korea, VO-Russia

# International VO Alliance

ivoa.net / IVOA.WebHome

Edit | Attach | Ref'd By | Printable | More | Advanced Search | Full Text Topic Name Go

THIS WEB

WebHome
WebChanges
WebTopicList
WebStatistics

ALL WEBS

IVOA
Know
TWiki
Test
Tracking
Trash

IVOA.NET

IVOA site
Members
Technical docs

TWiki basics
TWiki tutorial
User registration
Notify me

OFFSITE LINKS

W3C
xml.com
twiki.org

Welcome **TWikiGuest !**

This is the web-based collaboration area of the **International Virtual Observatory Alliance**

- User Registration
- Notification Service
- How TWiki works (TWiki Tutorial)

*Main topics*:

- Member Organizations
- Events
- Interoperability
- Reports & Minutes
- Software Repository
- Forums

*Working Groups*:

- Resource Registry
- Data Modeling
- Content Description (UCD)
- Data Access Layer
- VOTable
- VO Query Language
- Grid & Web Services
- Standards & Processes

*Interest Groups*:

- VO Architecture
- VO Applications
- VO Theory

# International VO Alliance

- A World-wide Coordination of VO Initiatives
- Total Spending Levels:
  - > $7M/yr (US - $2M/yr)
  - ~ $30M over 5 Years (US - $10M)
- VO Initiative has Gained World-wide Momentum, and…
- The Train is Leaving the Station

# VO and Future Large Facilities

- Software and data management are major components of future large facilities—VO infrastructure should help mitigate costs
- Maximum scientific return from future large facilities depends on ease of comparison with other data, availability of standard data products
- Education and outreach enabled by VO provides agencies, taxpayers with visible return on investment

# The VO Vision

- The VO is the "semantic web" for astronomy (Tim Berners-Lee)
- The VO democratizes astronomical research
- The VO brings the universe to your desktop
  - The professional astronomer
  - Graduate students
  - Undergraduates
  - K-12
  - Amateurs
  - The public