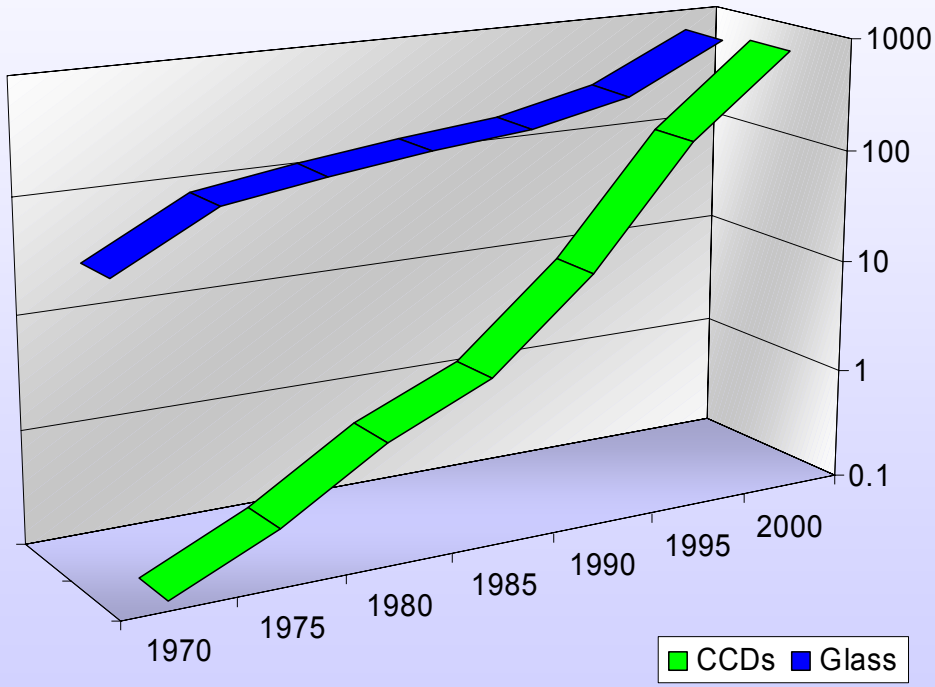
Virtual Observatory Architecture

Data Services
Registry Services
Compute Services

Roy Williams
Caltech
US-VO co-director

Trends

- Future dominated by detector improvements



- Moore's Law growth in CCD capabilities
- Gigapixel arrays on the horizon
- New Detector Technologies (*e.g.*, STJ)
- Improvements in computing and storage will track growth in data volume
- Investment in software is critical, and growing

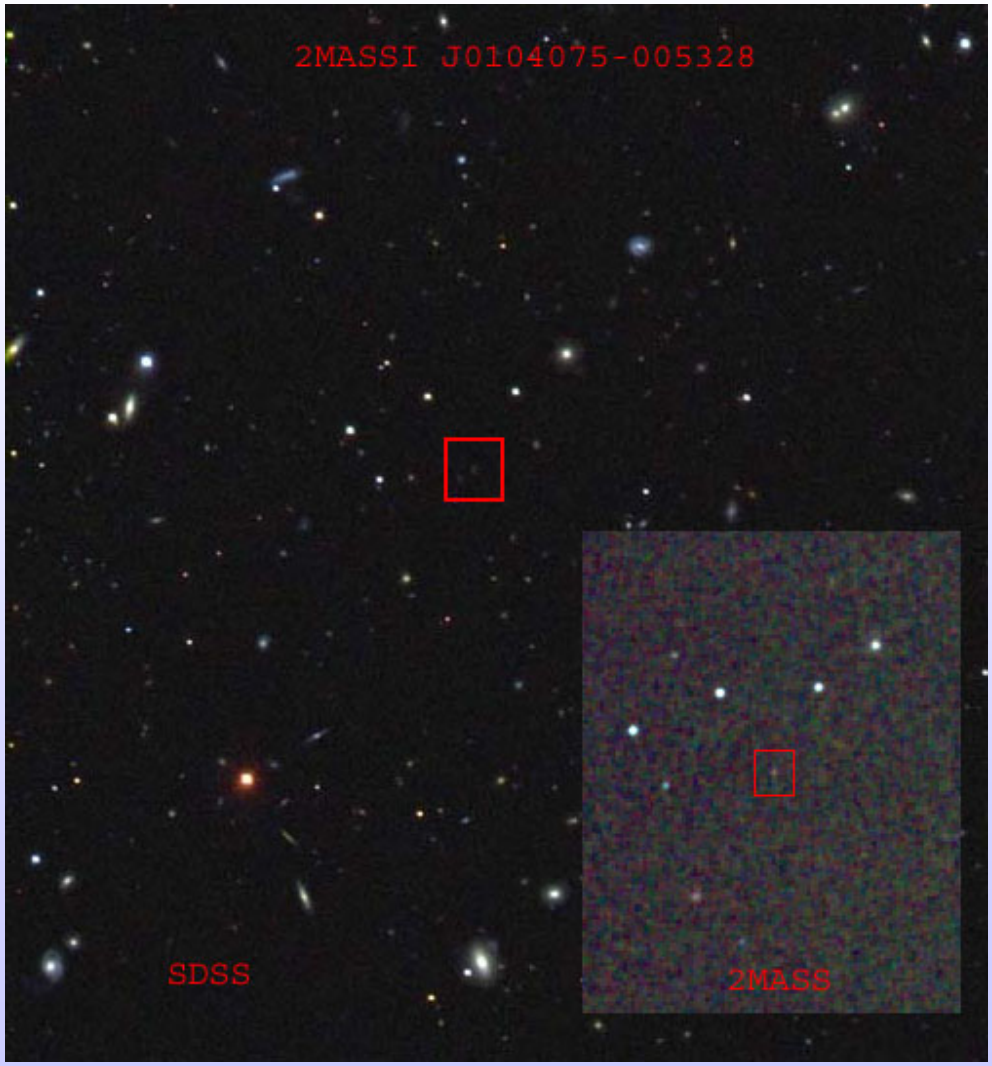
Total area of 3m+ telescopes in the world in m^2, total number of CCD pixels in Megapixels, as a function of time. Growth over 25 years is a factor of 30 in glass, 3000 in pixels.

Astronomical Data

- Image
  - Standard file format: FITS
    - Standardized c.1980
    - Keyword-value dictionary + binary block
- Catalog
  - Derived from image
    - Connected set of bright pixels
  - “Table of stars”
  - Standard format: VOTable
    - Standardized 2002
    - XML with remote binary
- Spectrum, Time series, ...

new instruments
new astronomy
new requirements
-> more **DIVERSITY**

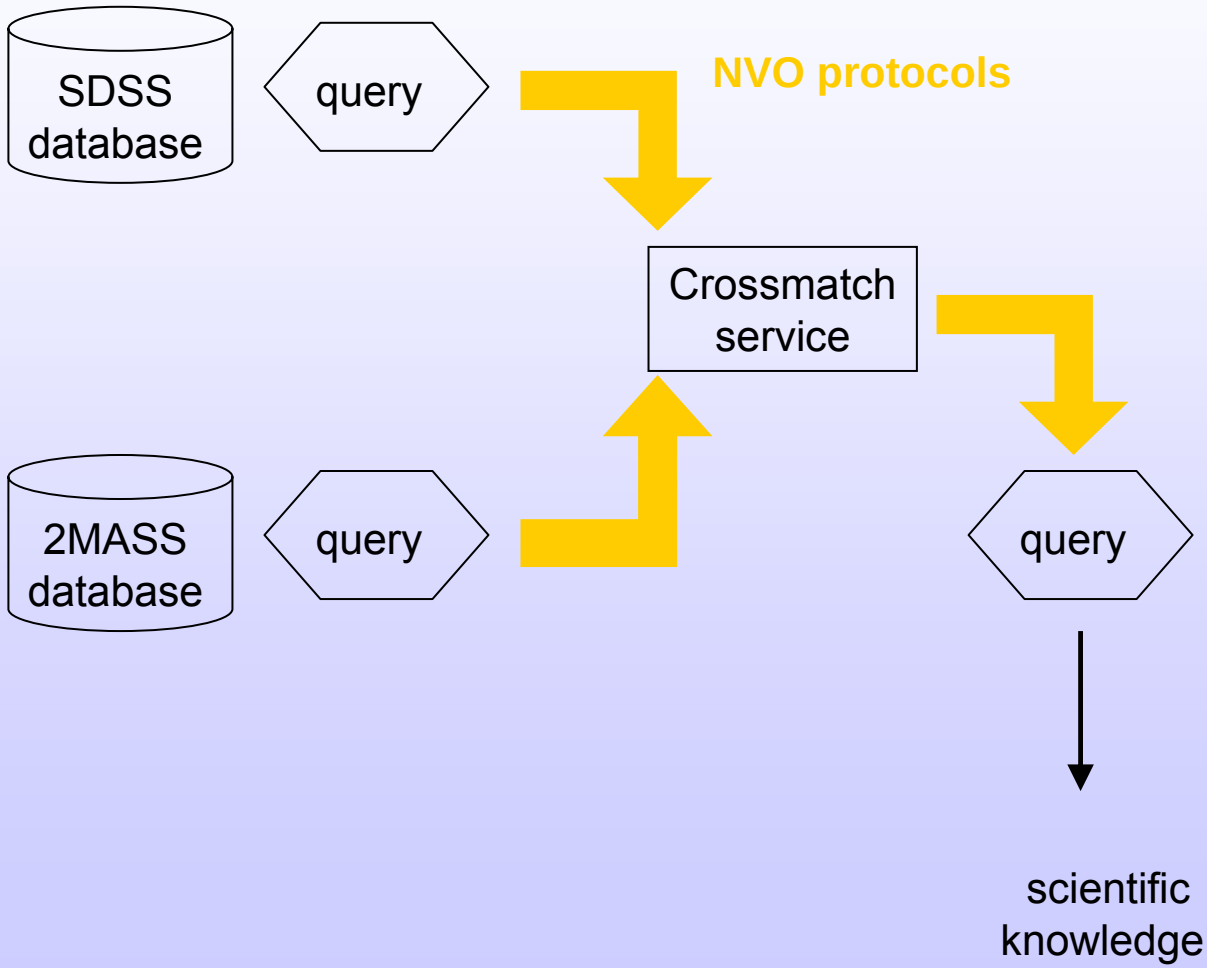
First NVO Discovery



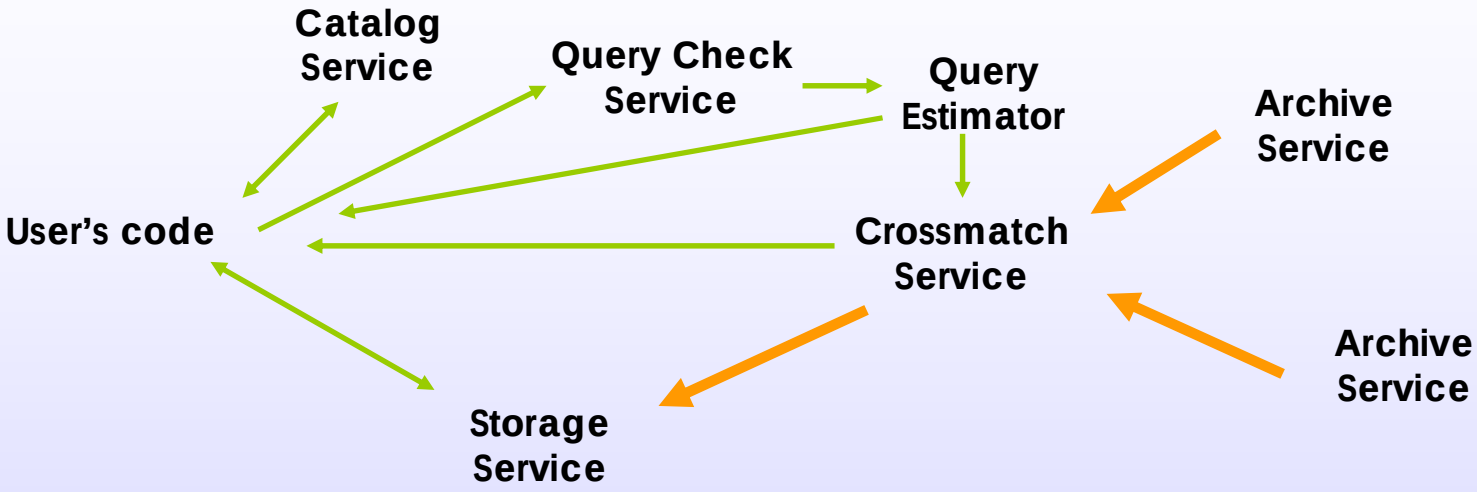
Database crossmatch of two massive databases creates new science

"The sum is greater than the parts"

Crossmatch Services



Networks of Services



What is the *meaning* of the service?

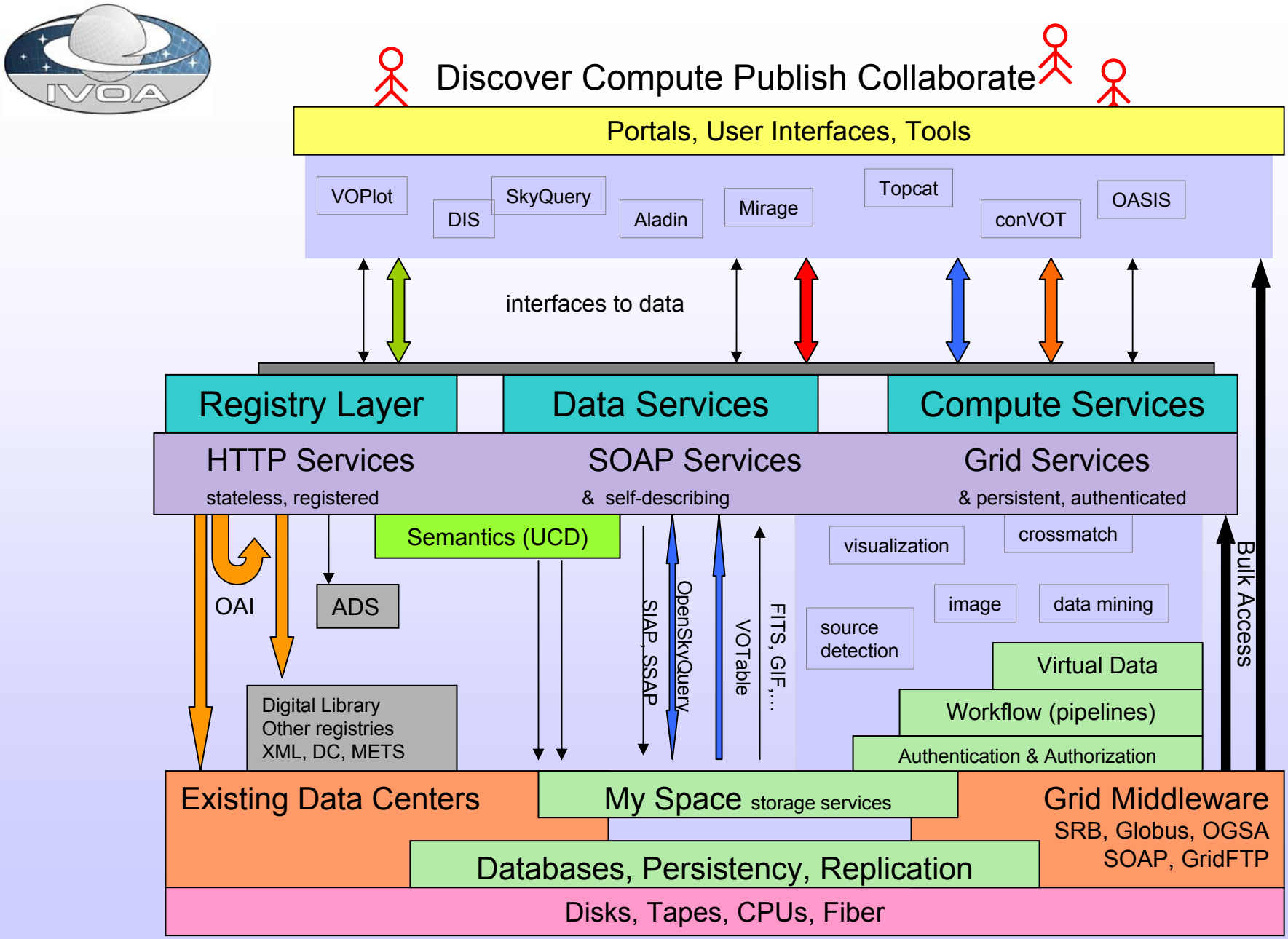
Who is responsible?

How can I use the service from Perl/Java/C++/IDL/IRAF?

Is there a simple web client for the service?

What is the request and response syntax?

What authentication do I need?


Discover Compute Publish Collaborate
Portals, User Interfaces, Tools
VOPlot
DIS
SkyQuery
Aladin
Mirage
Topcat
conVOT
OASIS
interfaces to data
Registry Layer
Data Services
Compute Services
HTTP Services
stateless, registered
SOAP Services
& self-describing
Grid Services
& persistent, authenticated
Semantics (UCD)
visualization
crossmatch
OAI
ADS
SIAP, SSAP
OpenSkyQuery
VOTable
FITS, GIF,...
image
data mining
source detection
Bulk Access
Virtual Data
Workflow (pipelines)
Authentication & Authorization
Digital Library
Other registries
XML, DC, METS
Existing Data Centers
My Space storage services
Grid Middleware
SRB, Globus, OGSA
SOAP, GridFTP
Databases, Persistency, Replication
Disks, Tapes, CPUs, Fiber

VOTable

- Full metadata representation
- Hierarchy of RESOURCEs
  - containing PARAMs and TABLEs
- UCD (unified content descriptor)
  - a has **unit** meter
  - a has **UCD** *ORBIT_SIZE_SMAJ* (Semi-major axis of the orbit)
- Can reference remote and/or binary streams
  - Table can be
    - Pure XML
    - "Simple Binary"
    - FITS Binary Table

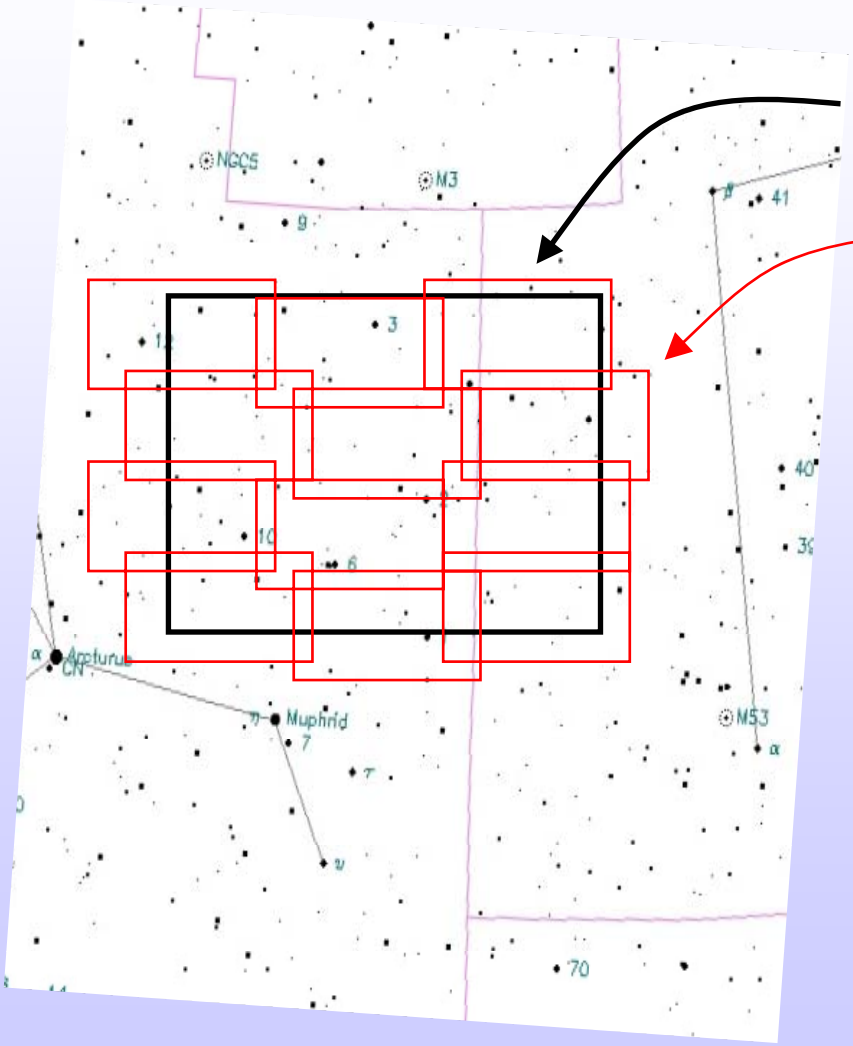
Cone/SIAP/SSAP

Data Services

- Simple, pragmatic solutions
  - quickly Specified, Created, Registered, Utilized!
- Cone
  - request is cone, response is VOTable with RA, Dec
  - many of these since 2/02
- SIAP
  - request is cone, response is VOTable of image links
- SSAP
  - under development

Simple Image Protocol

Data Services



- Specify box by position and size
- SIAP server returns relevant images
  - *Footprint*
  - *Logical Name*
  - *URL*

Can choose:

standard URL:
http://.....

SRB URL
srb://nvo.npaci.edu/…..

Unified Content Descriptors

Data Services

- UCD is a “semantic type”

PHOT.INT-MAG.B	Integrated total blue magnitude
ORBIT.ECCENTRICITY	Orbital eccentricity
STAT.MEDIAN	Statistics Median Value
INST.QE	Detector's Quantum Efficiency

- Can be resolved by web service
  - to description, examples, etc
- Base + Specifiers
  - eg error in default right ascension
  - **POS.EQ.RA, MAIN, ERROR**

OpenSkyNode

Data Services

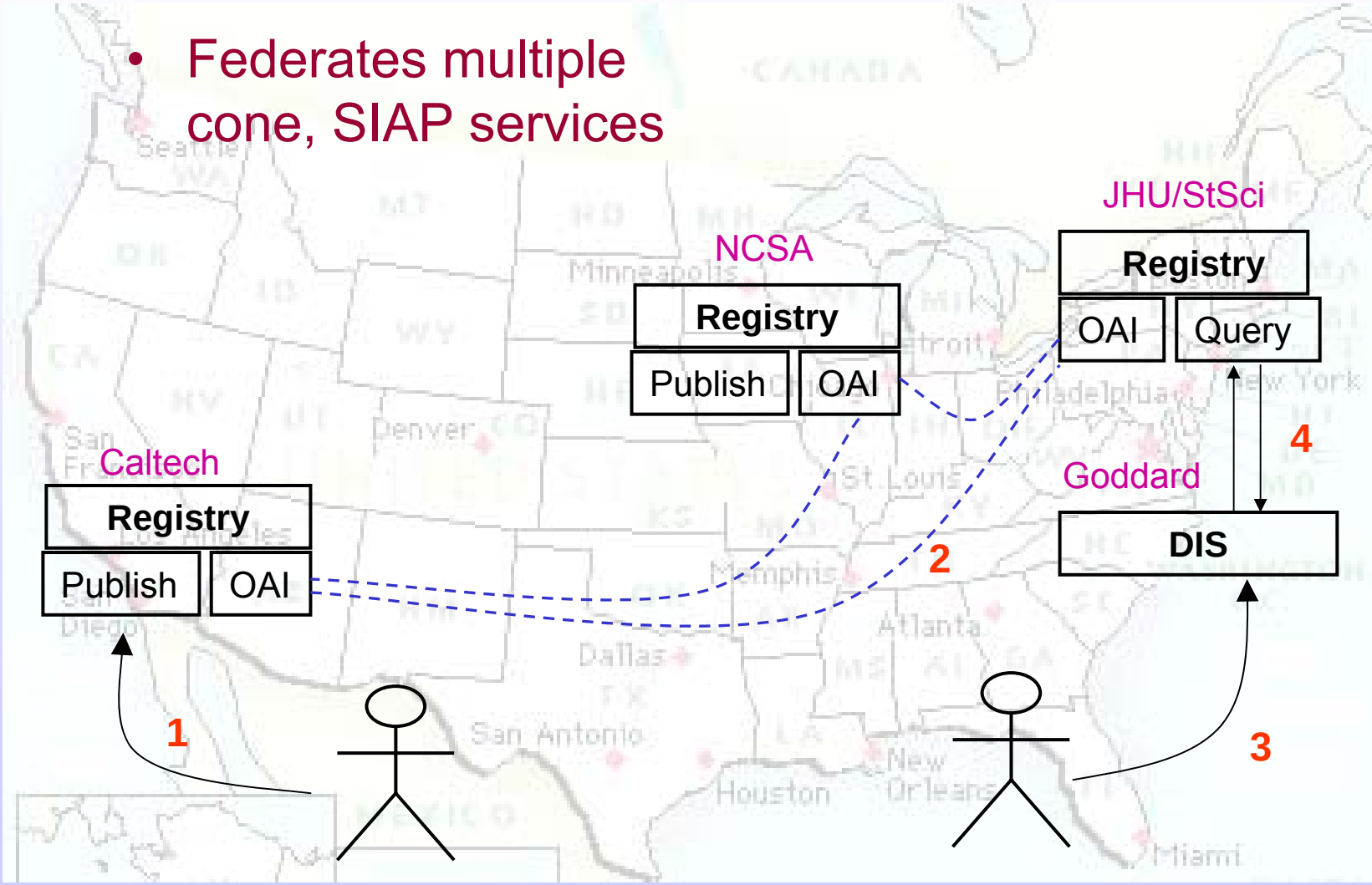
- Exposes a relational DB
  - select* from tables
  - select * from columns of table
  - select a,b,c where d>3 and e<4
  - select ra, dec where REGION(ra, dec, ...)
  - select from Xmatch (SDSS, 2MASS) .....

Registry Services

- Publish
  - Caltech, NCSA registries
- Query
  - ADQL (borrow from OpenSkyNode)
  - XQuery/XPath
- Harvest
  - OAI from NCSA, Caltech, JHU, Vizier (France)
- What entities are described by registry?
  - Service
    - VO standard or arbitrary
  - Project, Data Collection
  - (person, community, VD object, etc etc ...?)

Data Inventory Service

- Federates multiple cone, SIAP services



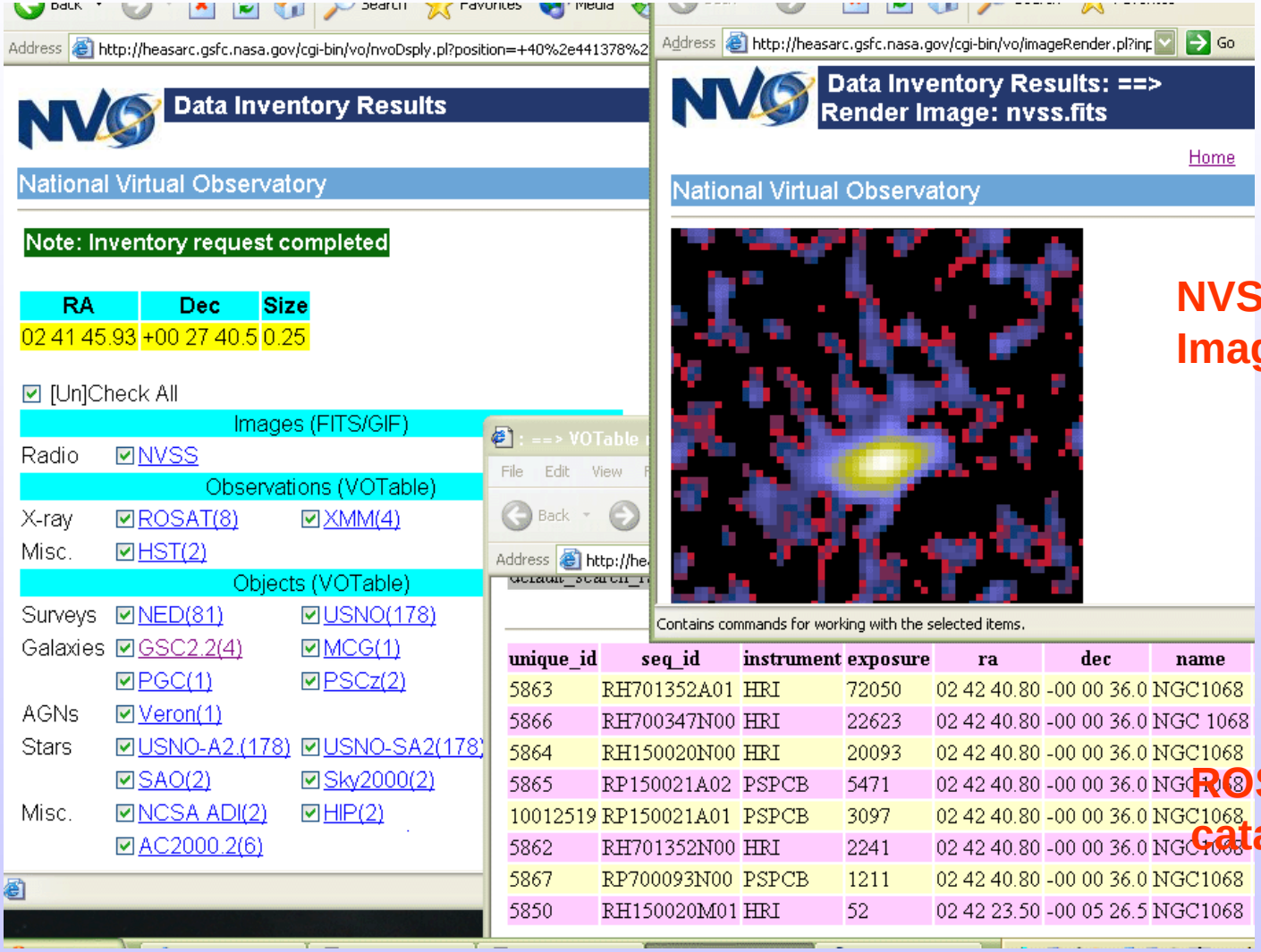
Data Inventory Service

Data Services
Registry Services

Relevant Images and Catalogs

NVSS Image

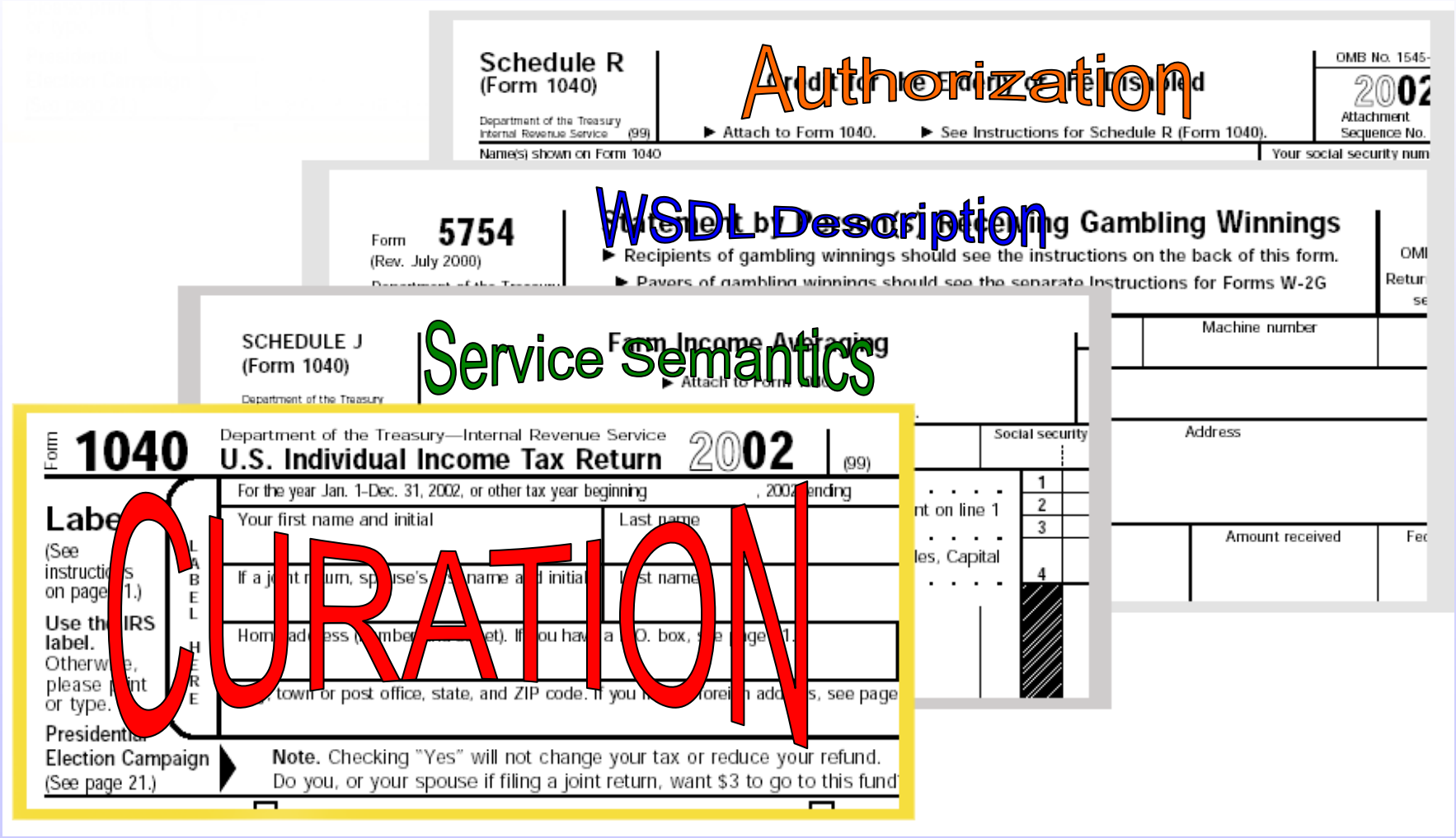
ROSAT catalog



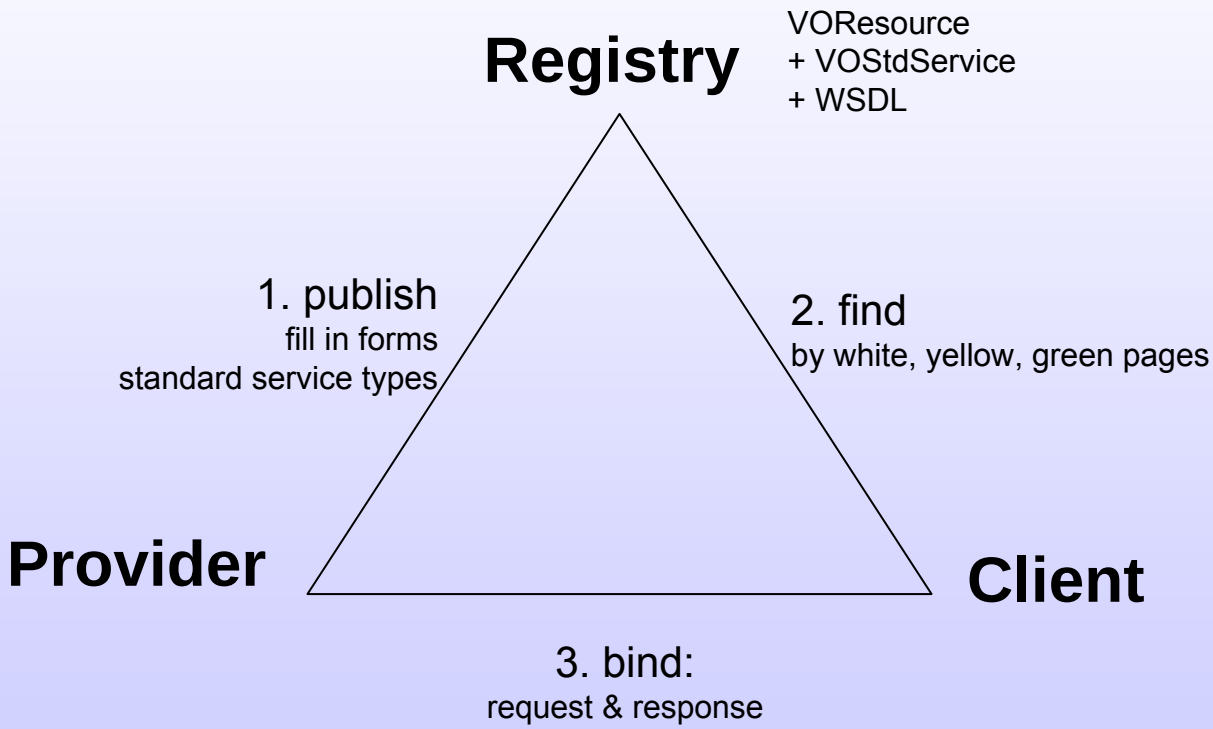
VOResource

Registry Services

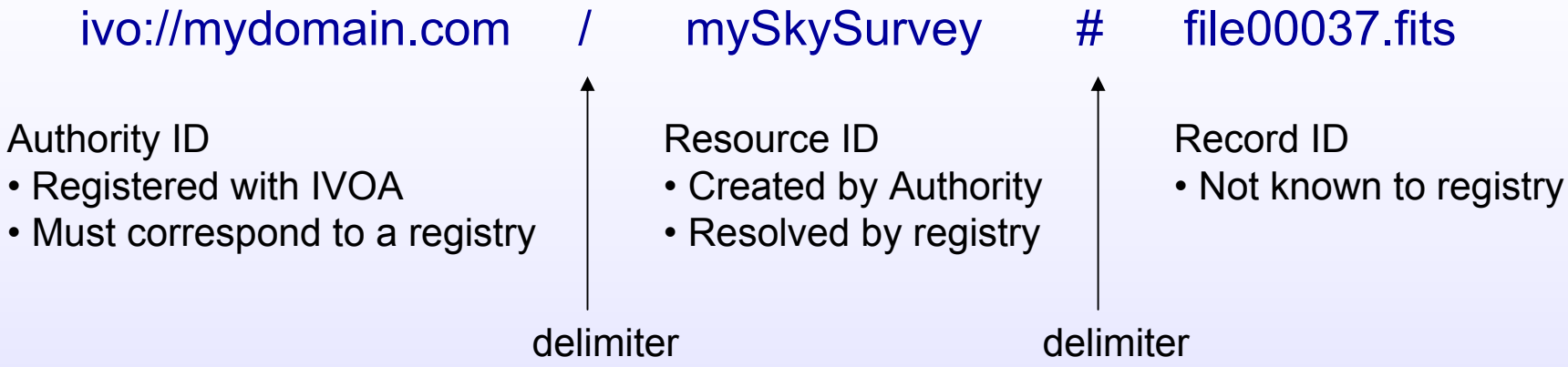
A mandatory form plus other supporting forms



Service Paradigm



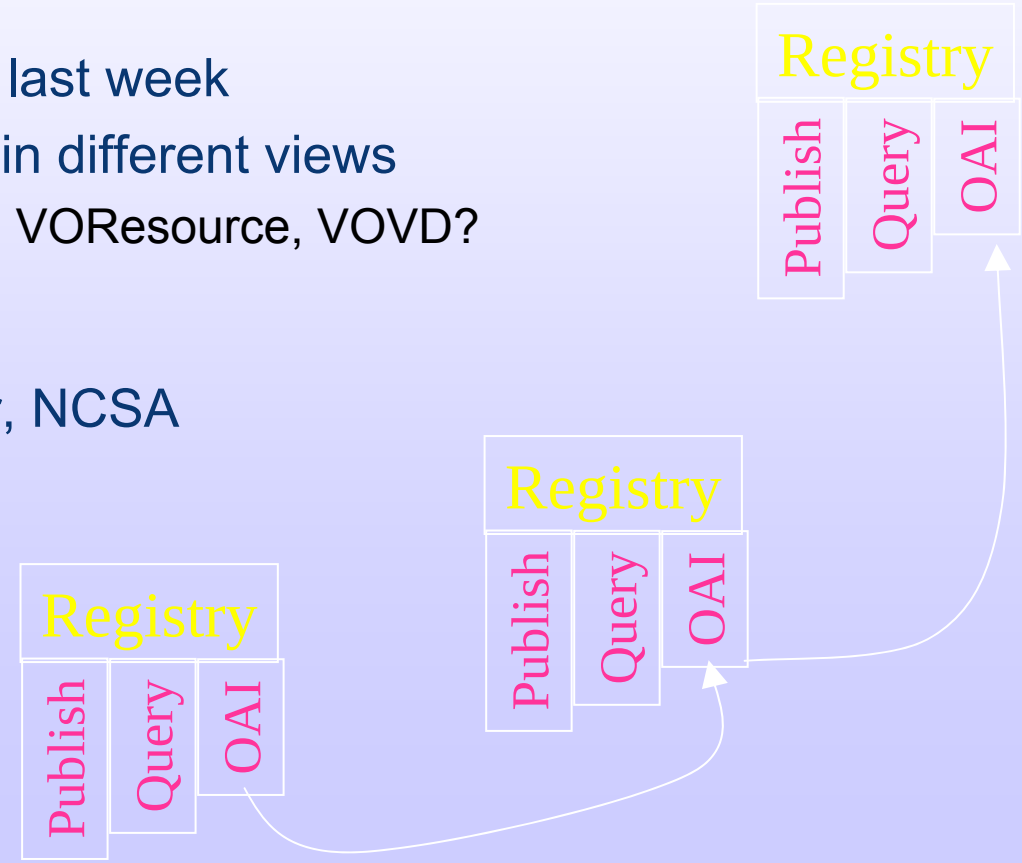
VO Identifiers



- URI form
- Resolved by registries

Open Archives Initiative (OAI)

- For harvesting registries
  - allows distributed control/fault-tolerance
- Queries
  - Changes since last week
  - Give metadata in different views
    - Dublin Core, VOResource, VOVD?
- Examples
  - Heasarc, Vizier, NCSA



Services

- Why services
  - distributed & relocatable
  - workflow components
  - described by request/response protocols
  - self-describing (WSDL etc)
  - architecture independent
- Questions
  - What's wrong with remote objects?
  - Security framework
  - On-ramp for the normal human?
  - Toolkits for services
  - Bulk data and SOAP

The Sky is a Database: Catalog Space

Compute Services

\char DATABASE='SQL: FROM tmassr.pts_samp_cat (2MASS Sampler Point Source Cata
\char EQUINOX ='J2000'
\char SKYAREA ='No constraint'
\char WHERE ='SQL: WHERE k_m BETWEEN 12.0 AND 12.1'
\char SELECT ='SQL: SELECT (Selected column names follow in next row.)'

ra	dec	j_m	j_msig	h_m	h_msig	k_m	k_msig	rd_flg	cc_flg	extd_fl
346.995577	27.962376	12.282	0.031	11.990	0.035	12.000	0.034	222	000	
40.202781	28.413649	12.679	0.030	12.152	0.033	12.000	0.030	222	000	
40.956840	29.456654	12.645	0.029	12.165	0.032	12.000	0.031	222	000	
106.165728	20.975992	12.[illegible]								
106.556947	20.133484	12.								
134.320346	14.391330	12.								
135.038423	17.798285	12.								
344.606603	29.220026	12.								
344.850216	25.119240	12.								
346.759144	28.661133	12.								
43.087673	29.482285	12.3								
43.960498	28.390268	12.3								
106.135655	19.268063	12.								
106.334142	17.934162	12								

☑	ra	right ascension (J2000 decimal deg)
☑	dec	declination (J2000 decimal deg)
☐	err_maj	major axis of position error ellipse
☐	err_min	minor axis of position error ellipse
☐	err_ang	position angle of error ellipse major axis (E of N)
☑	j_m	J selected "default" magnitude or 95% confidence upper limit
☑	j_msig	J "default" mag uncertainty
☑	h_m	H selected "default" magnitude or 95% confidence upper limit
☑	h_msig	H "default" mag uncertainty
☑	k_m	K selected "default" magnitude or 95% confidence upper limit
☑	k_msig	K "default" mag uncertainty
☑	rd_flg	source of JHK "default" mags (AKA "read flag")
☐	bl_flg	indicates # JHK components fit to source (each digit=0|1|2)
☑	cc_flg	indicates artifact contamination and/or confusion
☑	extd_flg	indicates src associated with or contaminated by an ext. src

Catalog Space: The Final Frontier

Compute Services

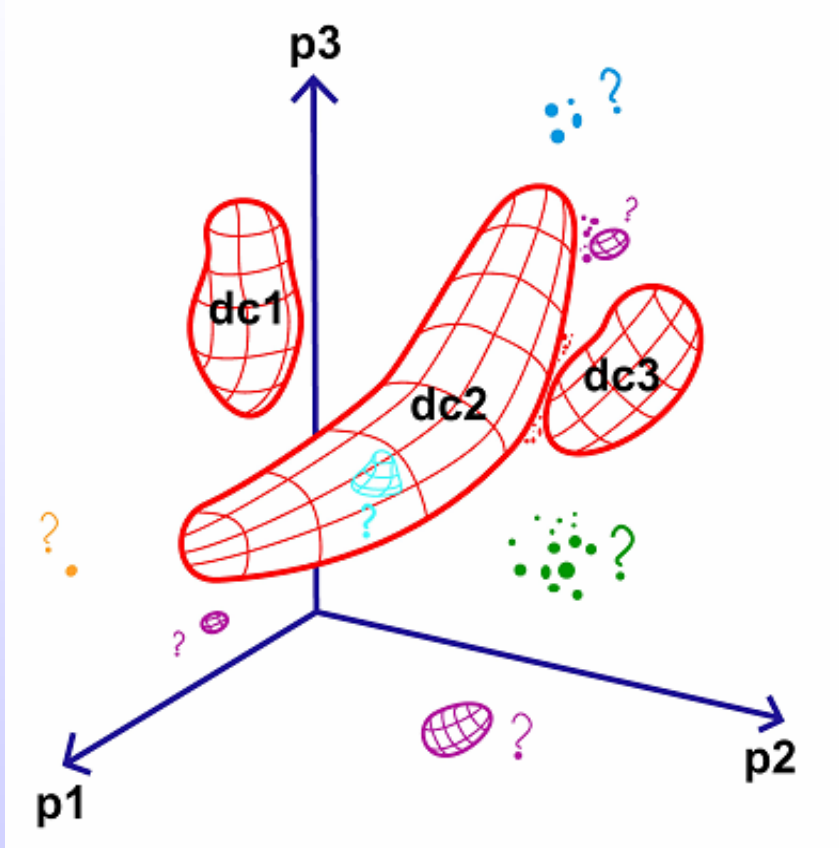
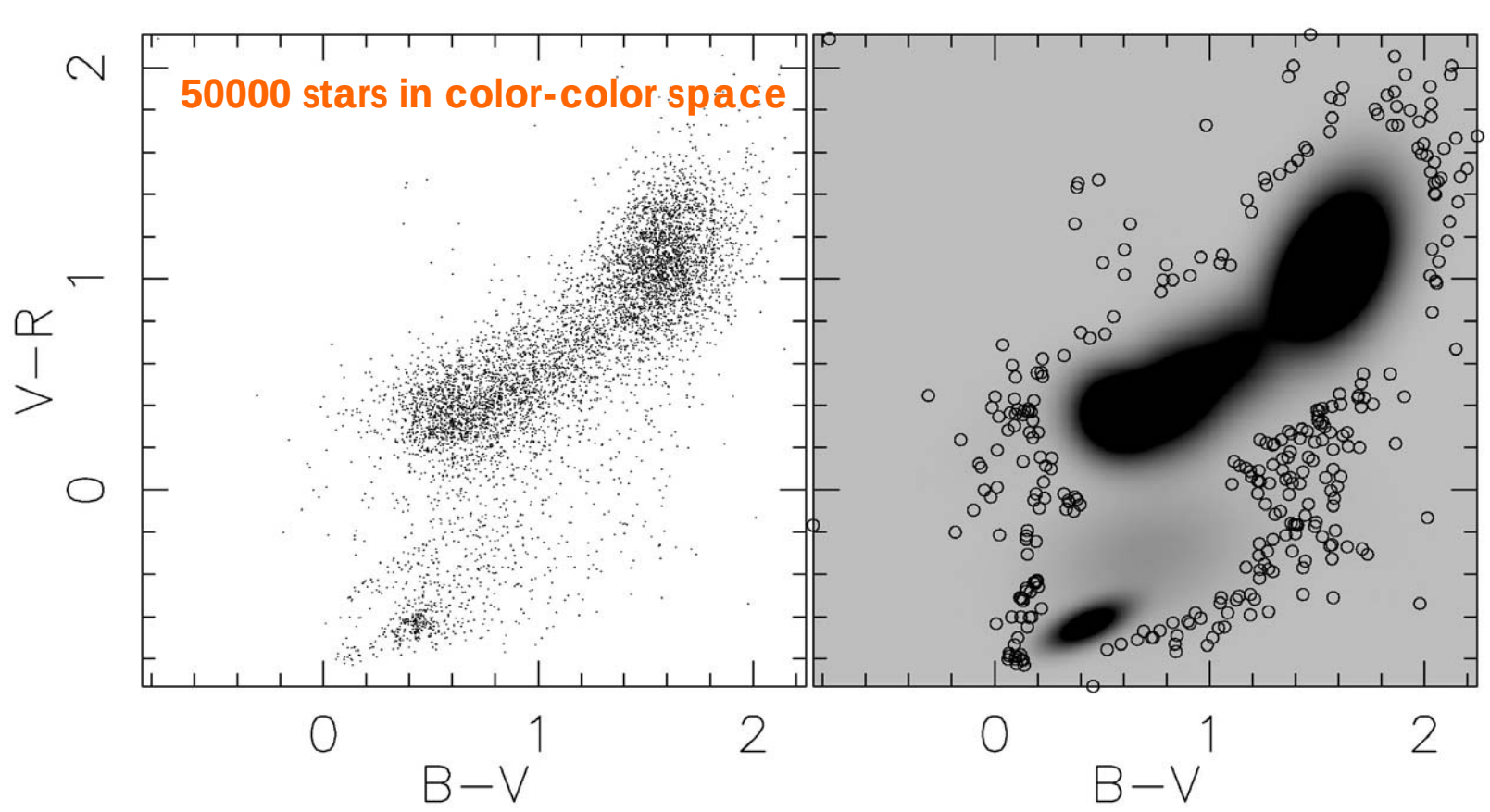


Figure 1. A schematic illustration of the problem of clustering analysis in some parameter space. The outliers, which do not fit well into any of the existing clusters, may be the new, previously unknown, and interesting objects.

Statistical Services

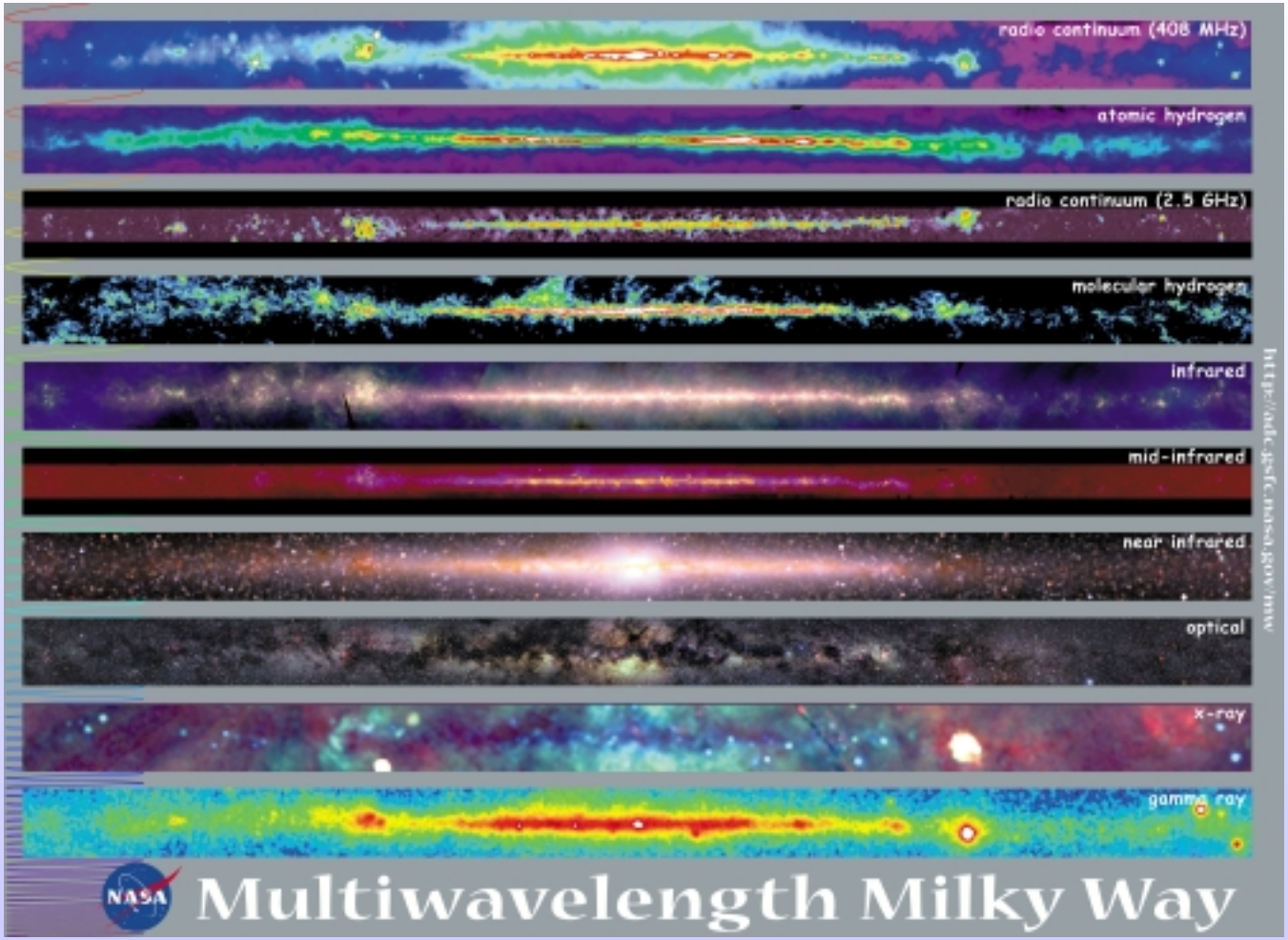
Compute Services

Convert pointset to density plus outliers



Compute Services

Image Federation



Multispectral Imagery

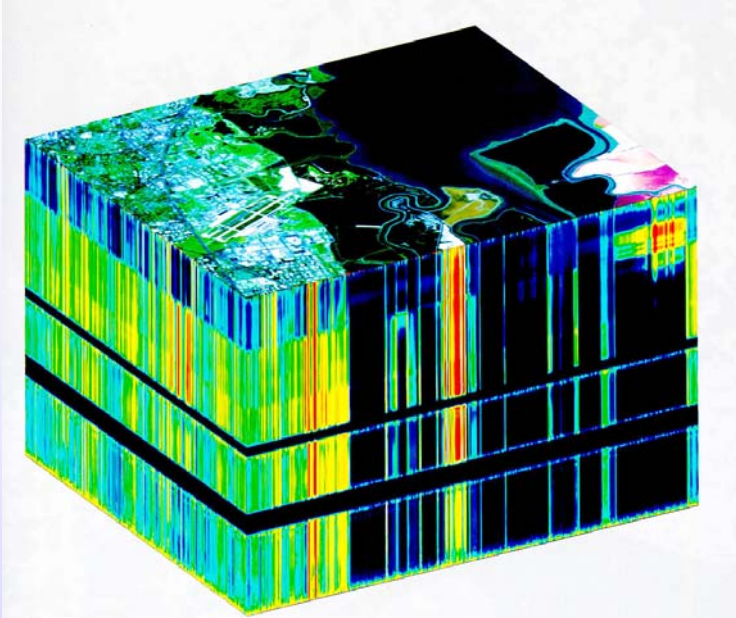
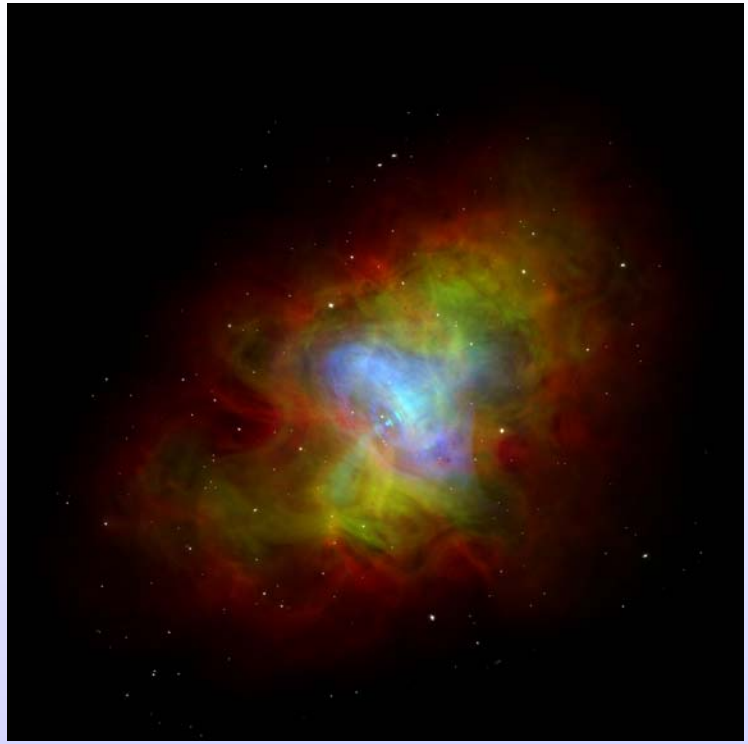


Figure 2-42 An AVIRIS dataset acquired over Moffet Field, California. AVIRIS acquires images in 224 bands each 10 nm wide, in the 400 to 2,500 nm region (Green, 1994). Three of the 224 spectral bands of data were used to produce the color composite on 'top' of the 'hyperspectral datacube'. The black areas in datacube represent atmospheric absorption bands. (Courtesy R. O. Green, JPL)

Moffet Field California.
224 channels from 400 nm to 2500 nm

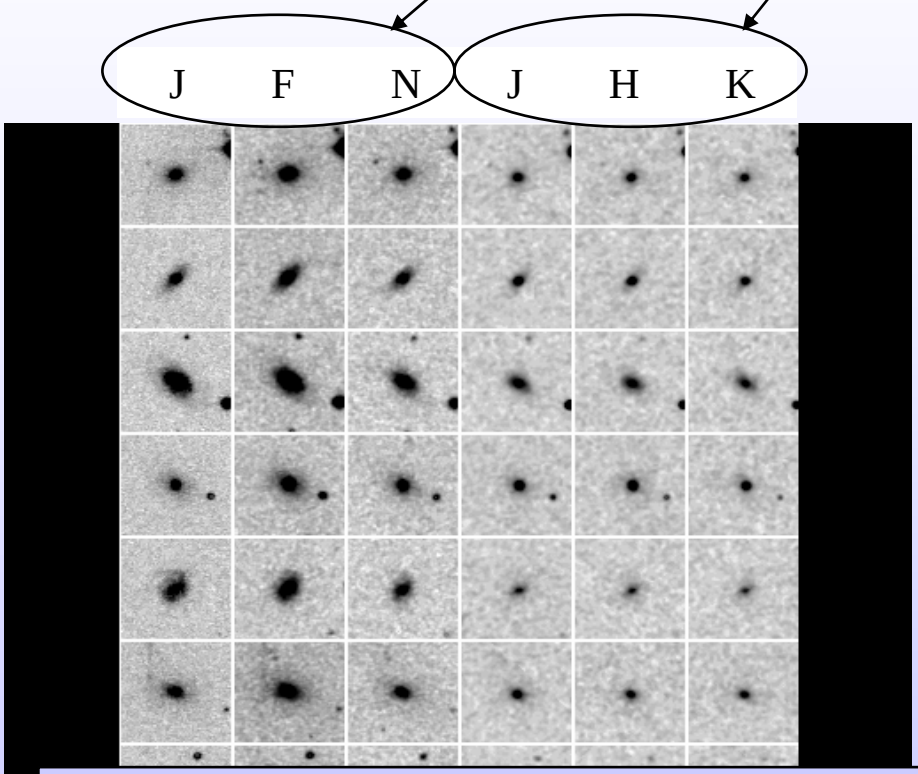


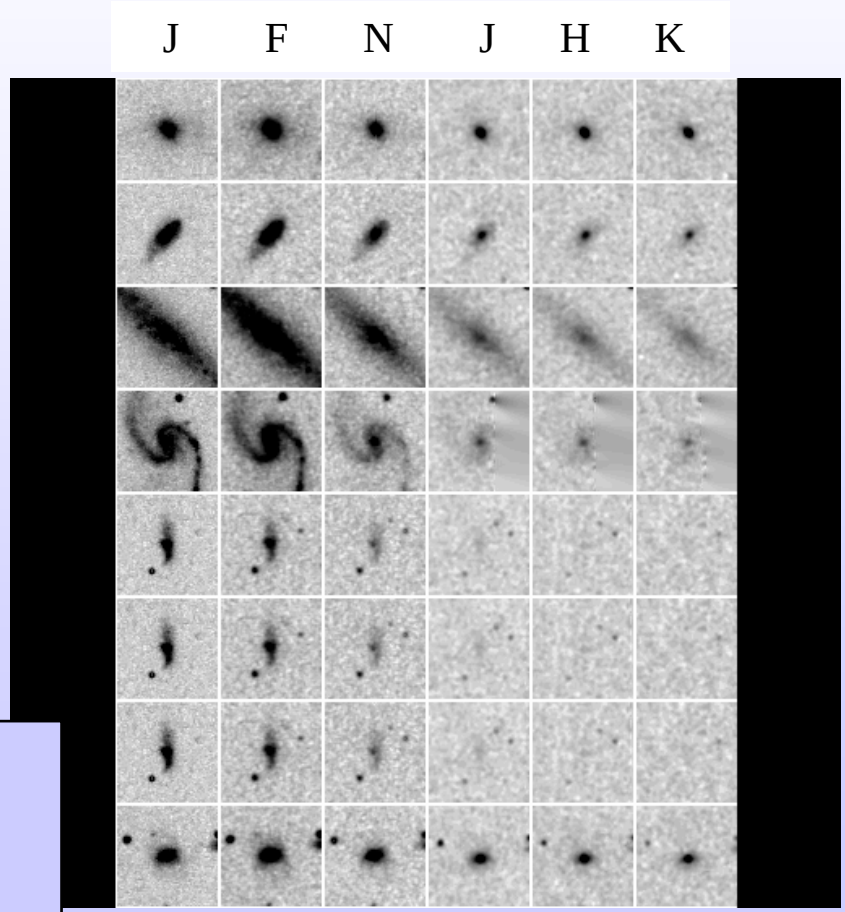
Crab Nebula.
3 channels: X-ray in blue, optical in green, and radio in red.

Multi-Wavelength Image Morphology

Compute Services

DPOSS-2MASS Image Mosaics

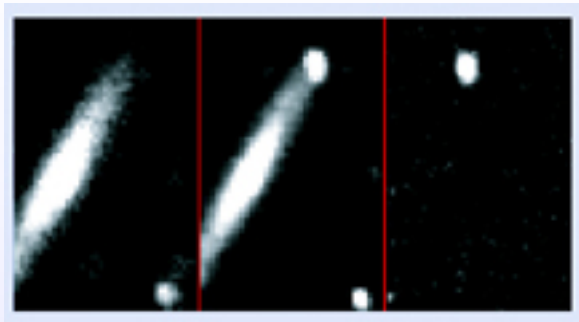




Galaxy identifcation, galaxy clusters
Pattern matching with shape AND color

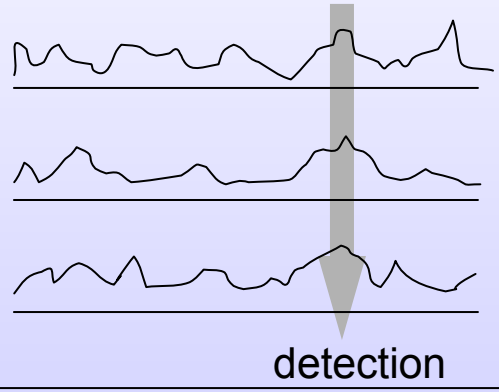
Image Federation

It's A New Window!



Images of the same galaxy taken several days apart are automatically subtracted from one another, and remaining bright spots may be supernova candidates. (NEAT project)

Stacking allows detection of faint sources. A 1-sigma detection in each of many bands becomes a 3-sigma detection.



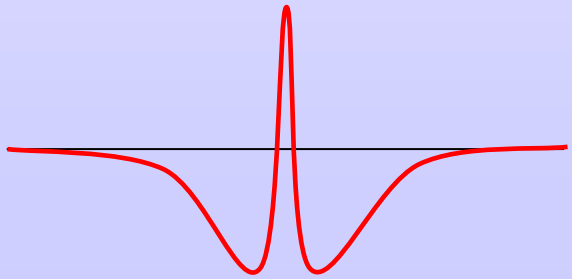
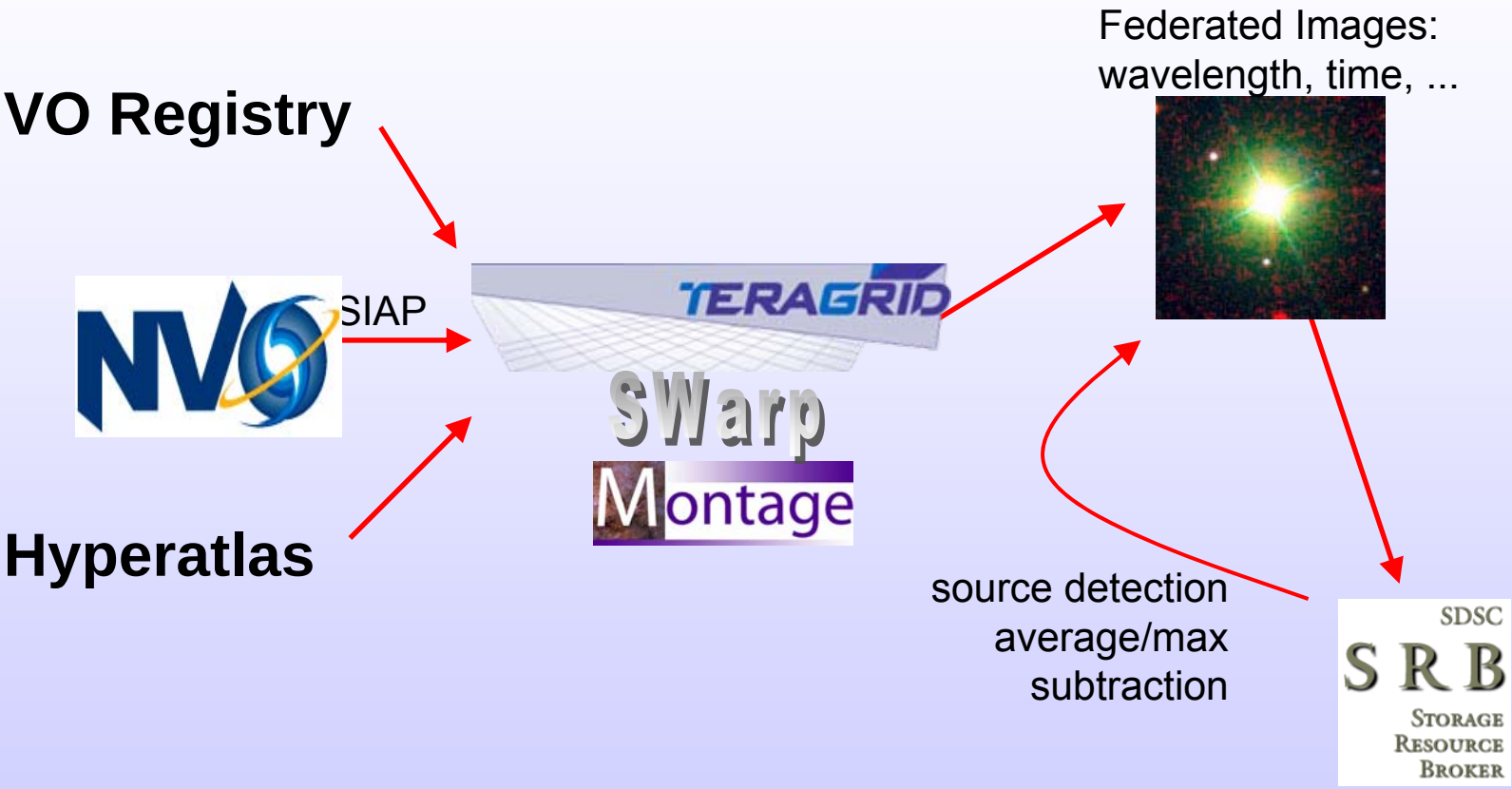


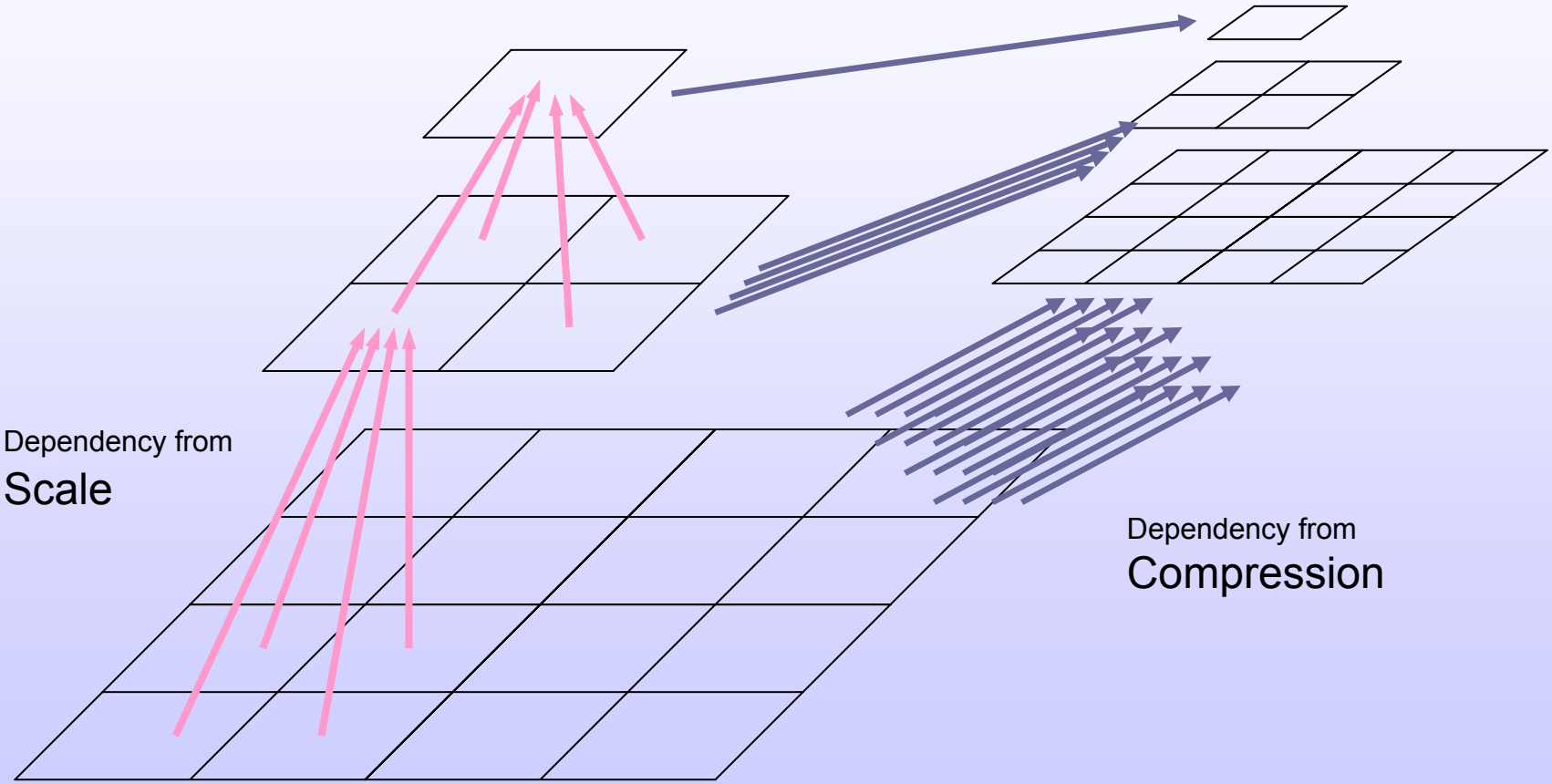
Image subtraction allows detection of narrow-line features that are not also wide-band (eg Hα but not R-band)

Compute Services

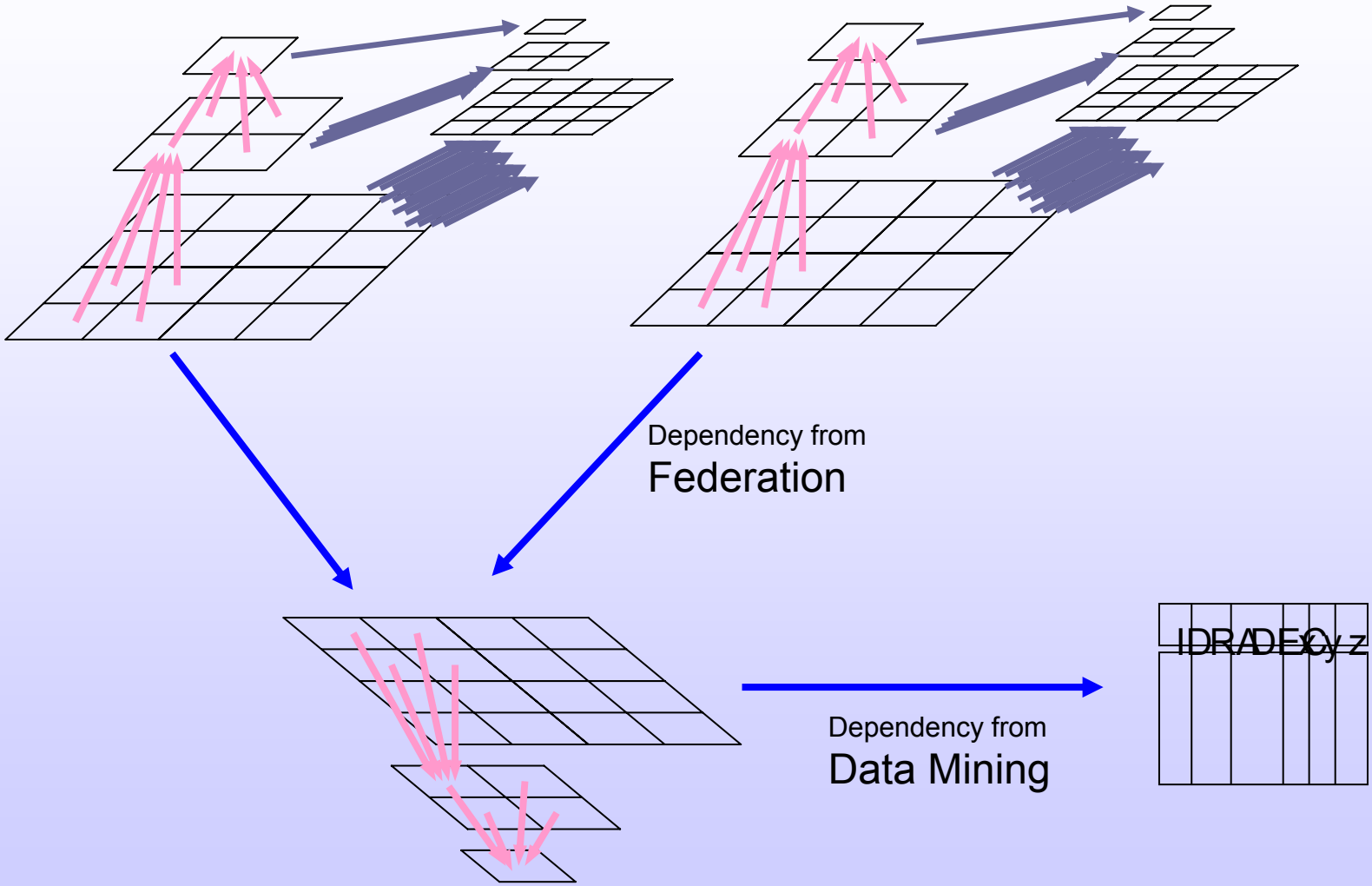
Atlasmaker



Atlasmaker DAG

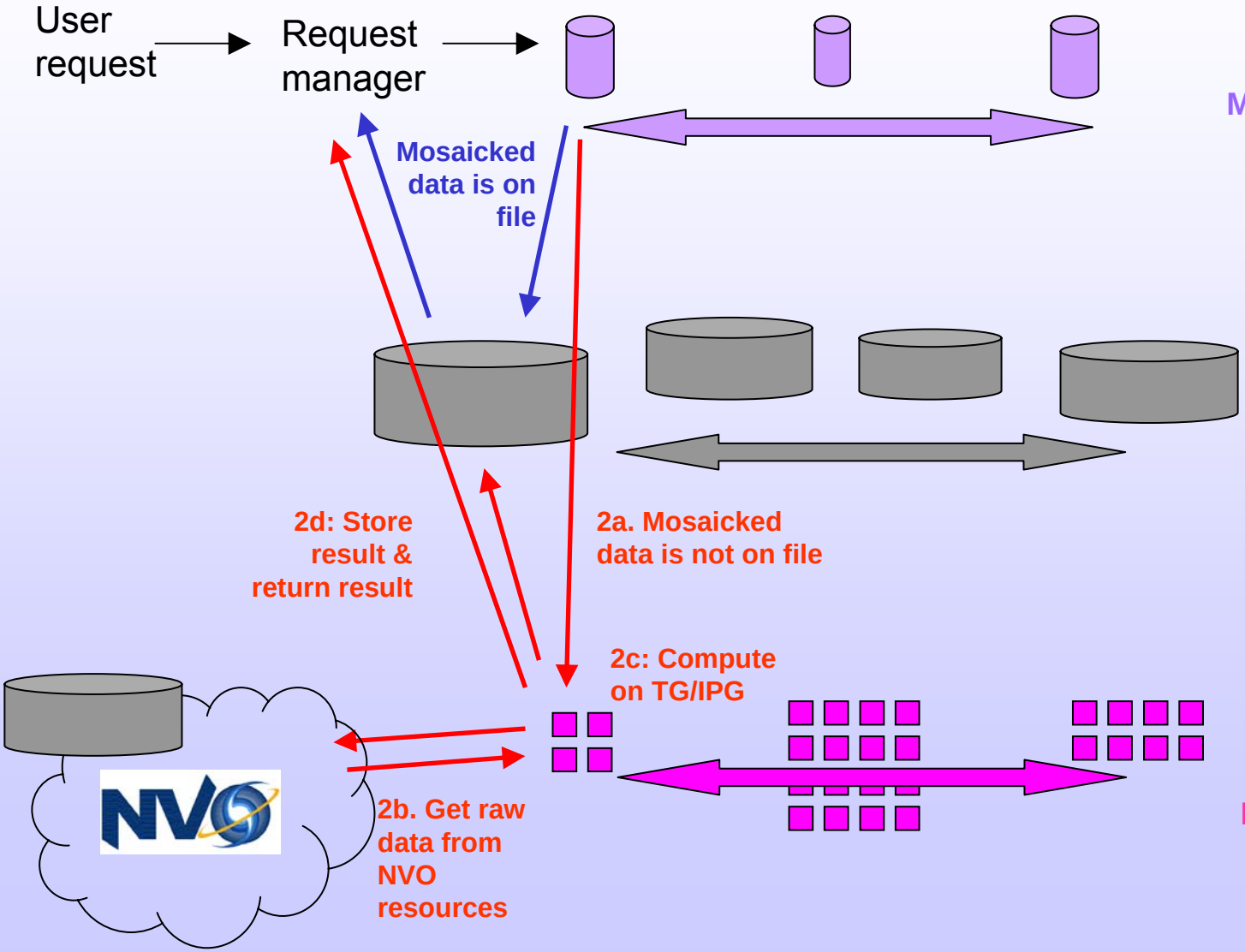


Atlasmaker DAG



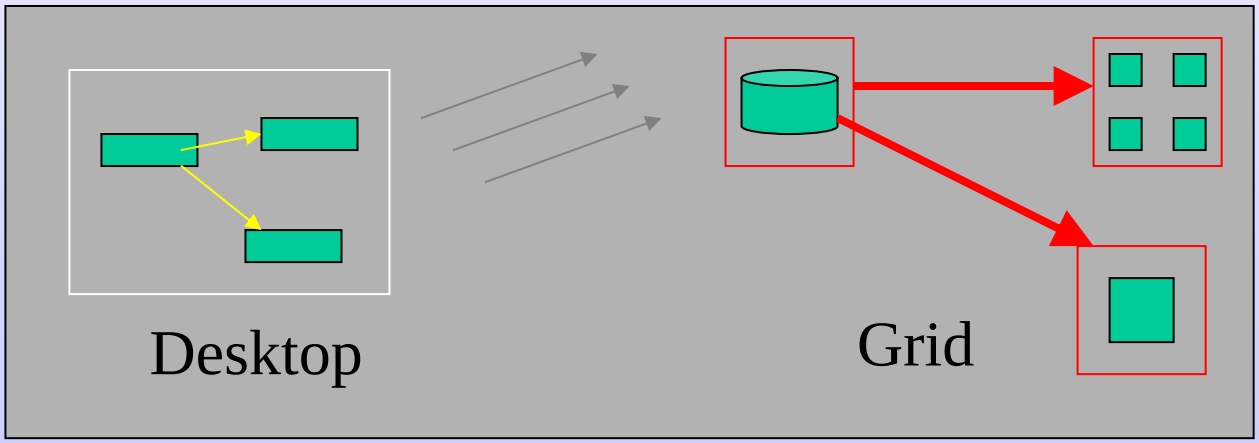
Atlasmaker
Virtual Data System

Compute Services



Grist

- Grid Data Mining for Astronomy
  - Williams, Djorgovski, Graham, Jacob, Katz, Mahabal, Miller et al
- Architecture
  - Persistent Grid Services
  - VO Registry
  - Virtual Data
  - Distributed file system



Triana, Viper, ...?

Teragrid, UK Grid, ...?

Compute Services

GRIST Objectives

- Workflow
  - » Portal, Batch, Grazing
  - » Virtual Data
- VO Data services
  - » OpenSkyNode for crossmatch
  - » Palomar-Quest exposure
  - » SIAP exposure
- Mining Palomar-Quest
  - » Hi-z Quasar candidates
  - » Cluster/outlier/correlation
- Image Processing
  - » Hyperatlas library
  - » Faint source detection
  - » Education
- Grid computing and massive data
  - » Teragrid