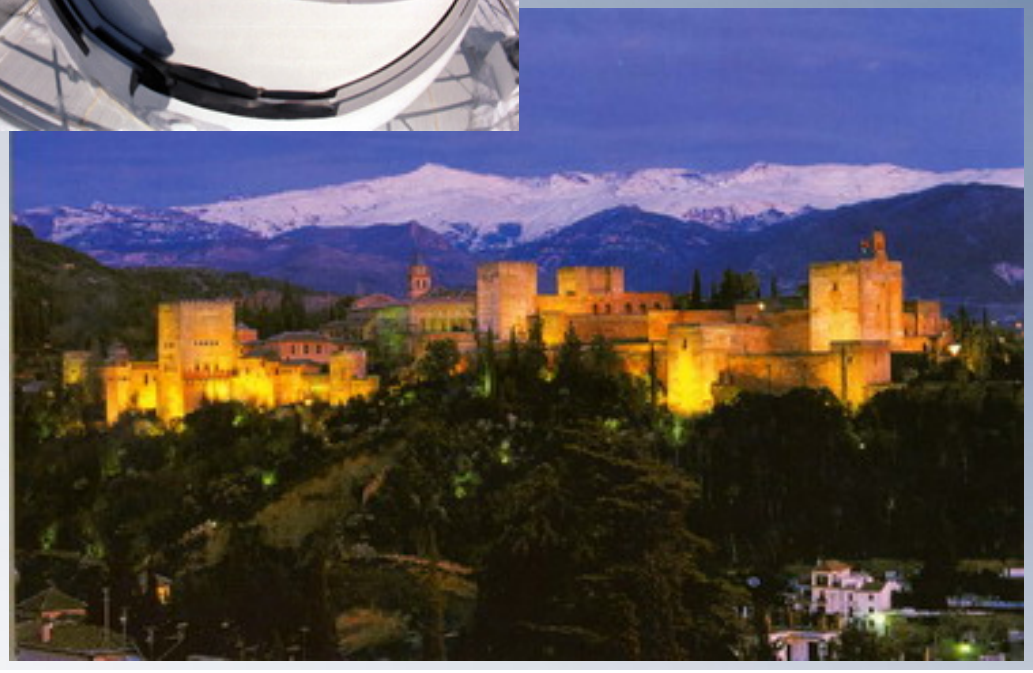
VO Web-services-based astronomy workflows

Jose Enrique Ruiz
IAA - CSIC

Manchester 13th July 2011

IAA - CSIC



AMIGA

Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

Curating and preserving collaborative digital experiments

1. Intelligent Software Components (ISOCO, Spain)
2. University of Manchester (UNIMAN, UK)
3. Universidad Politécnica de Madrid (UPM, Spain)
4. Poznan Supercomputing and Networking Centre (PSNC, Poland)
5. Universisty of Oxford (OXF, UK)
6. Instituto Astrofísica Andalucía (IAA-CSIC, Spain)
7. Leiden University Medical Centre (LUMC, NL)

MANCHESTER 1824

The University of Manchester

SUPERCOMPUTING POZNAN NETWORKING

INSTITUTO de ASTROFISICA de ANDALUCIA CSIC I·A·A

iSOCO enabling the networked economy

SEVENTH FRAMEWORK PROGRAMME

European Commission Information Society and Media

Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

The AMIGA Group

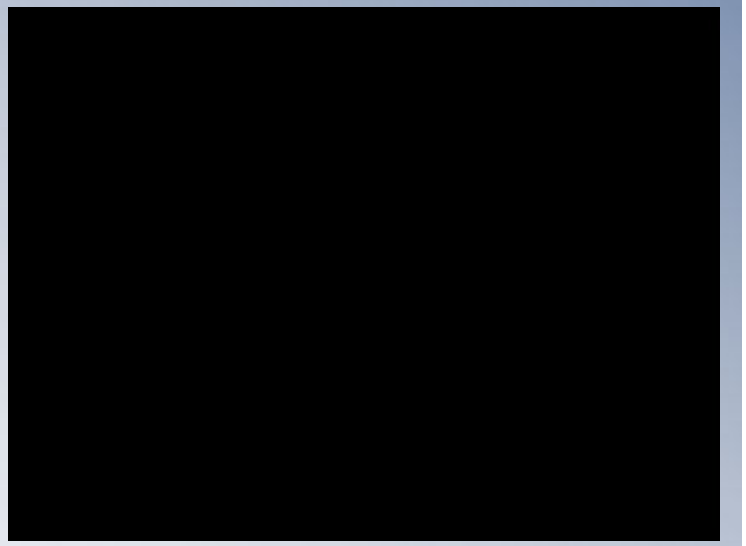
Analysis of the interstellar Medium of Isolated Galaxies

Statistical baseline of isolated galaxies to compare with the behaviour of galaxies in denser environments

Multi λ study of ~1000 galaxies

Instituto Astrofisica de Andalucia - CSIC
Univ . Granada, Obs. Marseille, Obs. Paris,
NAOJ, FCRAO, UNAM, Univ. Edinburgh,
IRAM, ESO, Kapteyn Astronomical Institute.

P.I. Lourdes Verdes-Montenegro
http://amiga.iaa.es



VO Virtual Observatory

- International Virtual Observatory Alliance (IVOA)
- Interoperability and Discovery
- Publishing and Accessing Data
- Service Oriented Architecture (SoA)
- Integration of Software and Data
- Distributed Resources
- Panchromatic Astronomy

- Data Models
- Web Services
- Semantics

Who are you ?

VO Virtual Observatory

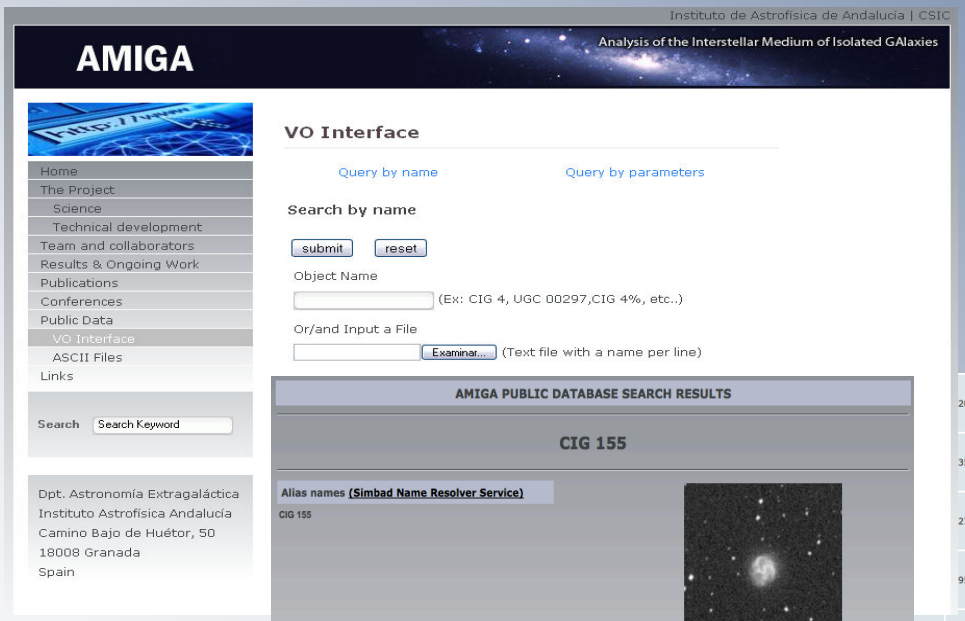
Analysis of the Interstellar Medium of Isolated GAlaxies

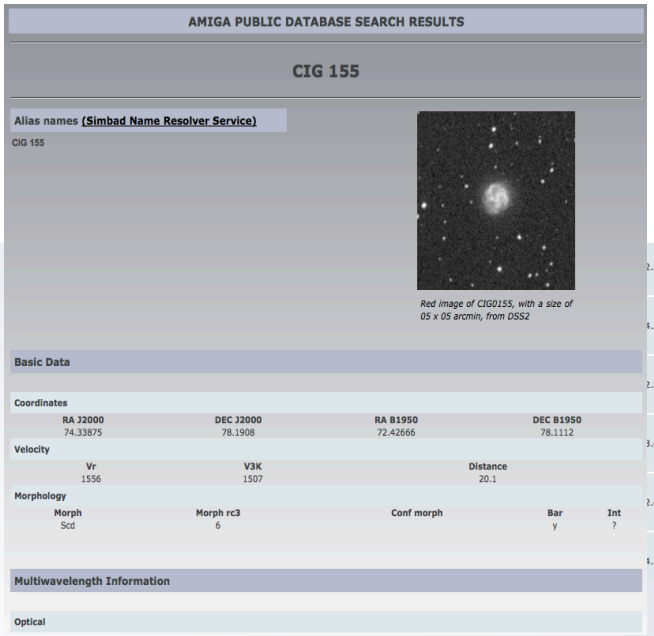
AMIGA

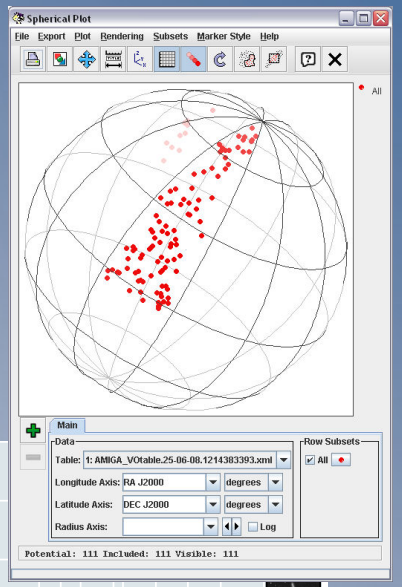
Who are you ?

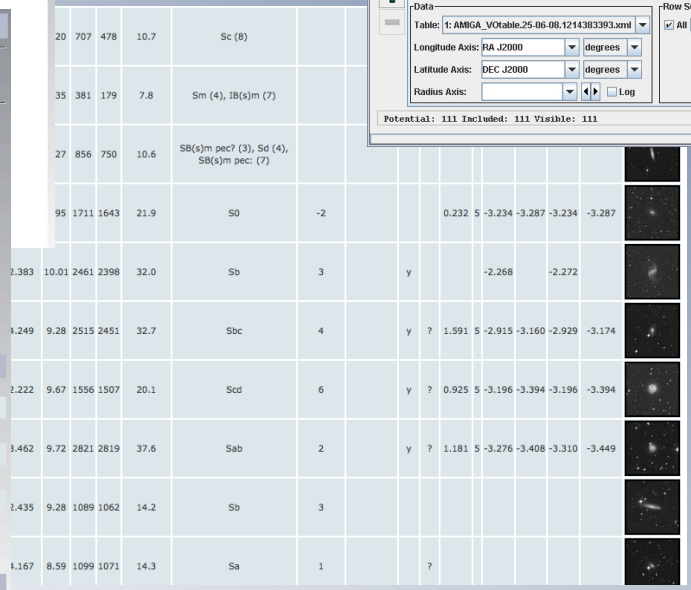
The AMIGA VO Catalog

The Data Provider





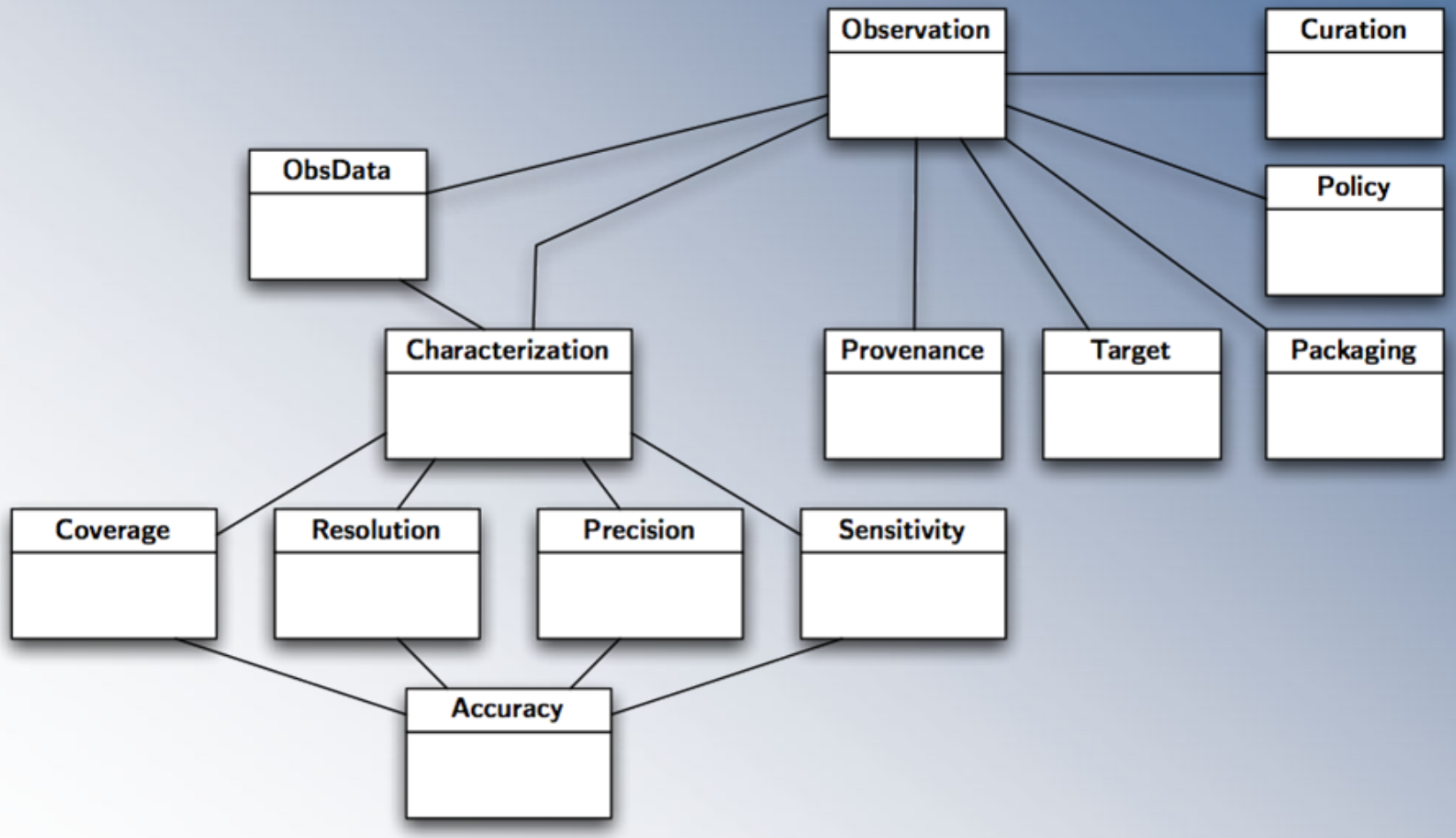




Who are you ?

RADAMS

Radio Astronomy Data Model for Single-Dish telescopes

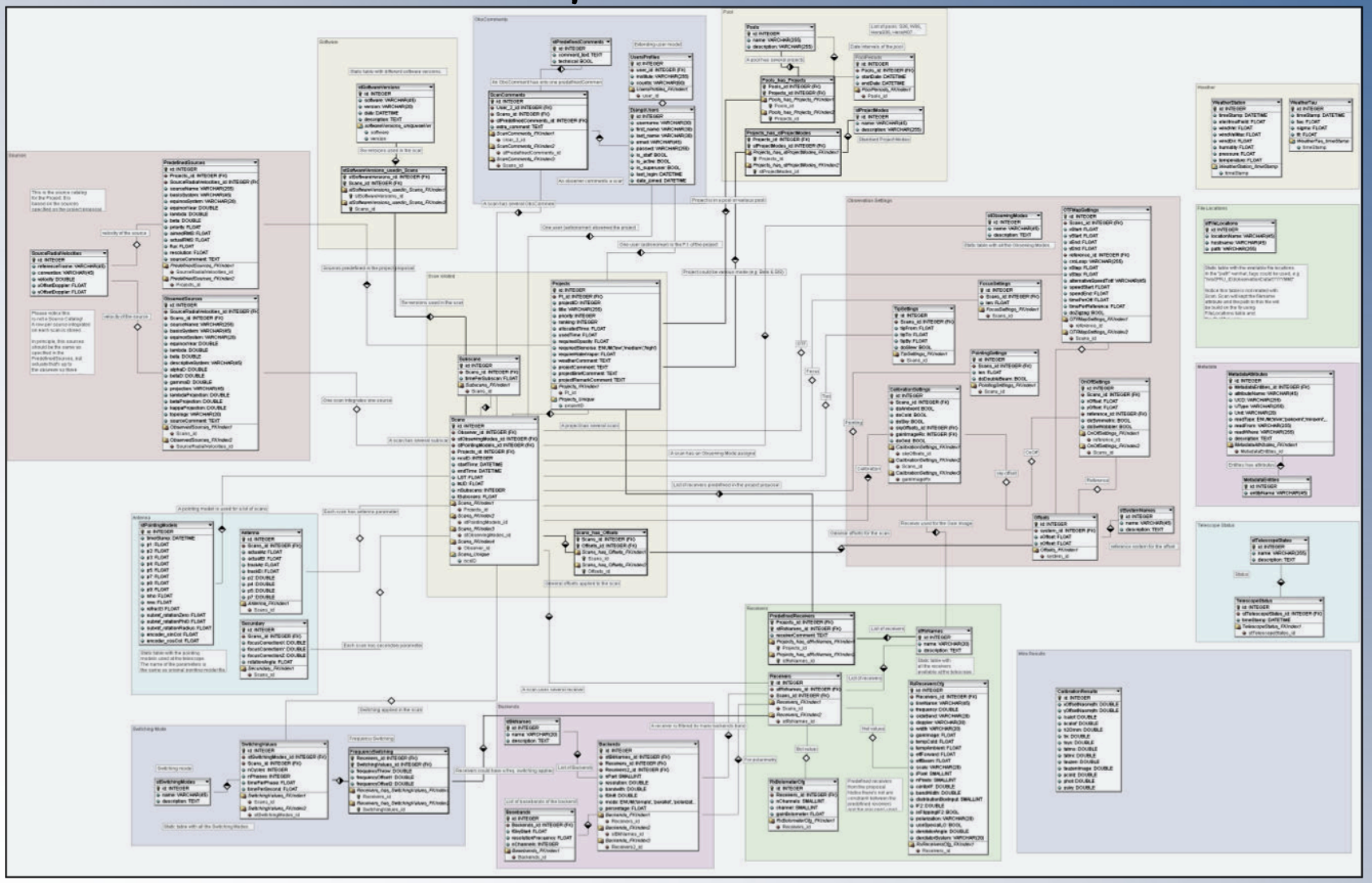


Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

Who are you ?

RADAMS Implementation



VO Archives Developments

Robledo DSS-63

- Madrid Deep Space Communication Complex (MDSCC)
- 70m single dish in Robledo de Chavela (Madrid)
- 5% operational time for observations
- K band Spectra (18 - 26 GHz)
- H2O Masers, methanol, NH3,..

iram
Institut de Radioastronomie Millimétrique

TAPAS – IRAM 30m

- Telescope Archive for Public Access System
- Bolometric observations, maps, spectra
- Rotational molecular transitions
- ~200 scientific projects / year, 1TB

Radio Astronomy DAta Model for Single-dish telescopes

Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

The AMIGA Group

Analysis of the interstellar Medium of Isolated Galaxies

Statistical baseline of isolated galaxies to compare with the behaviour of galaxies in denser environments

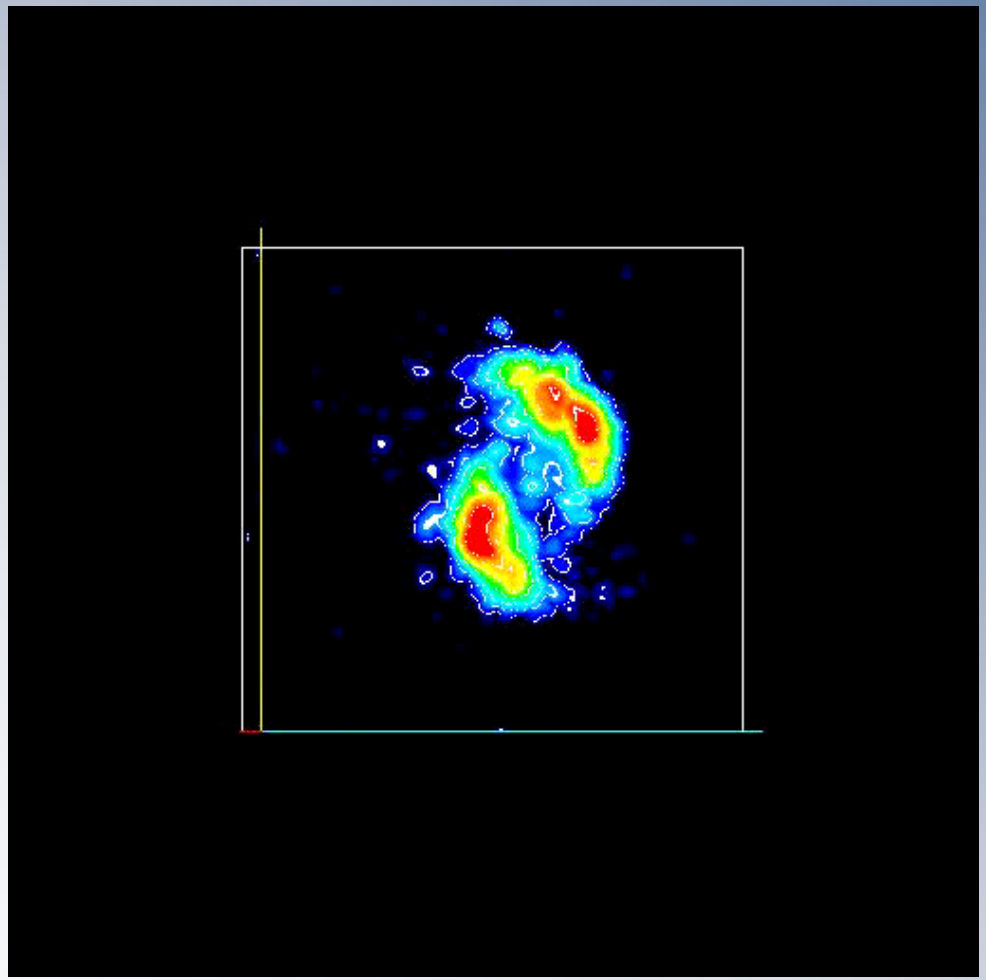
Multi λ study of ~1000 galaxies

+

Need of intensive and complex analysis of 3D data

2D spatial + 1 Velocity

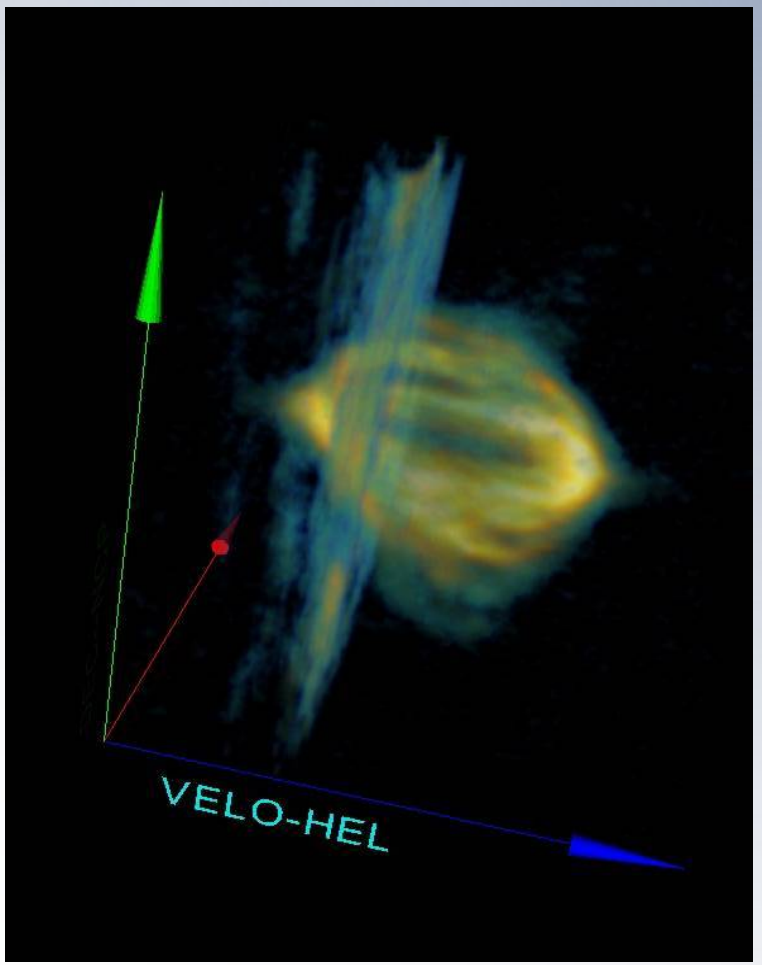
Velocity Datacubes



Who are you ?

GIPSY

Groningen Image Processing SYstem



Connectivity

- VO Archives
- VO Software

Accessibility

- Usability GUI
- VO Web Services

Kapteyn Astronomical Institute
IAA - CSIC

Who are you ?

BODEGA Below 0 DEgrees GAlaxies

P.I. : D. Espada
Legacy project of Submillimiter Array interferometer (SMA)
http://bodega.iaa.es

IAA-CSIC
CfA (Harvard-Smithsonian Center for Astrophysics)
ASIAA (Institute of Academia Sinica Astronomy and Astrophysics)

Molecular gas properties of a survey of nearby galaxies.

30 processed and reduced datacubes of galaxies

Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

Who are you ?

The BODEGA 3D VO Catalog

The Data and Service provider

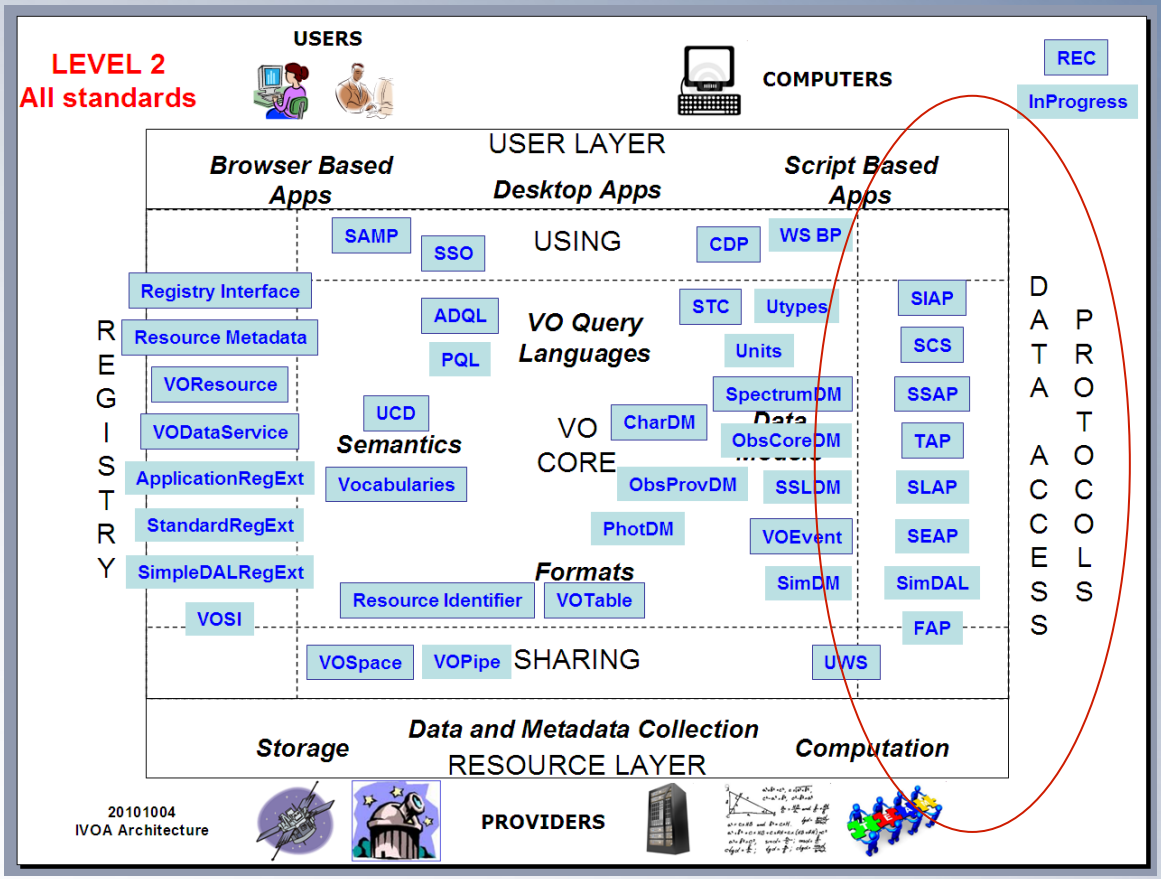
Virtual Data

The Virtual Observatory

Infrastructure of *interoperable* data and services. *Standards* for:

- Providers to *share* data and services
- Developers to *discover* the services, *find* and *access* the data

Goal: astronomers to use this *infrastructure* in a seamless way



Standards for Web Services

- Most of the Web Services in Astronomy
- They are registered and curated
  - VO Registry
- WS for Humans
  - Data discovery and data access
  - Accessed with local software (Europe)
  - Integrated in web portals (USA)
- WS for Machines
  - Storage, transport, authentication, etc.

The VO Registry

- If you are not registered, you are not in the VO
- Web forms to register services
- Three VO Registries
  - Euro-VO
  - National Virtual Observatory (USA)
  - AstroGrid (UK)
- Harvesting among registries
- A VO Registry register resources
  - Organizations
  - Authorities
  - Data collections
  - Services

WS for Humans

- Most WS provide "just" Data Discovery and Access
- Associated to a very specific Archive
- Designed to discover
  - VO Services
  - Catalogs
  - Images
  - Spectra
- Parameters-based -> Standards
- Responses are always VOTables
  - Characterization of data
  - Actual data values
    - List of services
    - Spreadsheets for catalogues
    - Links to binaries for images and spectra

WS for Humans

- Sesame name resolver is one of the most used
  - Resolves objects names into coordinates
  - Provided by Centre de Données de Strasboug (CDS)
- Data Discovery and Access (RESTful)
  - ConeSearch
  - Simple Image Access
  - Simple Spectra Access
    - Parameters: RA, DEC, SIZE
  - Table Access Protocol (TAP), OpenSkyQuery, SkyNodes
    - Astronomical Data Query Langage (ADQL) requests
- Sparse complex services (SOAP)
  - Mosaicing of images, footprint of regions, spectral building and fitting, principal components analysis in spectra..
  - Common Execution Architecture (AstroGrid)- not took off

WS for Machines

- Implementation in progress
  - More standards than implemented services
- Universal Worker Service (Grid oriented)
  - asynchronous
  - stateful
  - job oriented services
- VOSpace
  - distributed storage
  - will be provided for Big Data archives
- Single Sign-On and Credential Delegation
- Registry Interfaces: services acting on the Registry

VOSI

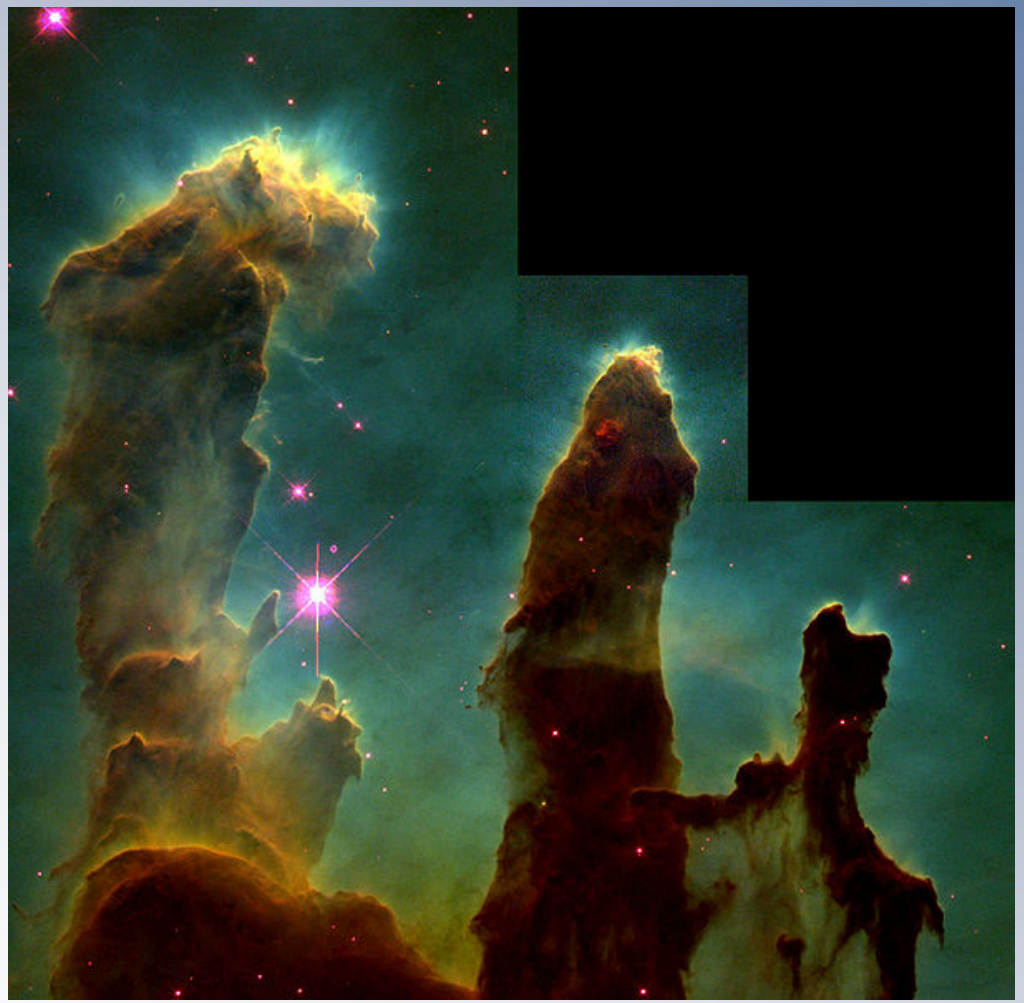
- VO Services Support Interface (REST binding)
- In progress of implementation
- Provides interoperability among services
- Common Contract for all VO services
- Self-descriptive services
  - operations and data
    /capabilities /tables
  - state of the service
    /availability /upSince /downAt /backAt /note
- XML/VOTable VOSI files
- VOSI files stored in service provider server
- Files are scanned by VO Regrisries
- Provide also state of the service

VOTables

XML Format

- Characterization of Data
  - Semantics
    - UCDs (Universal Content Descriptors)
  - Data Models
    - UTypes
- Actual Data
  - Tabular data
  - Links to binary data

Ontologies, SKOS Vocabularies

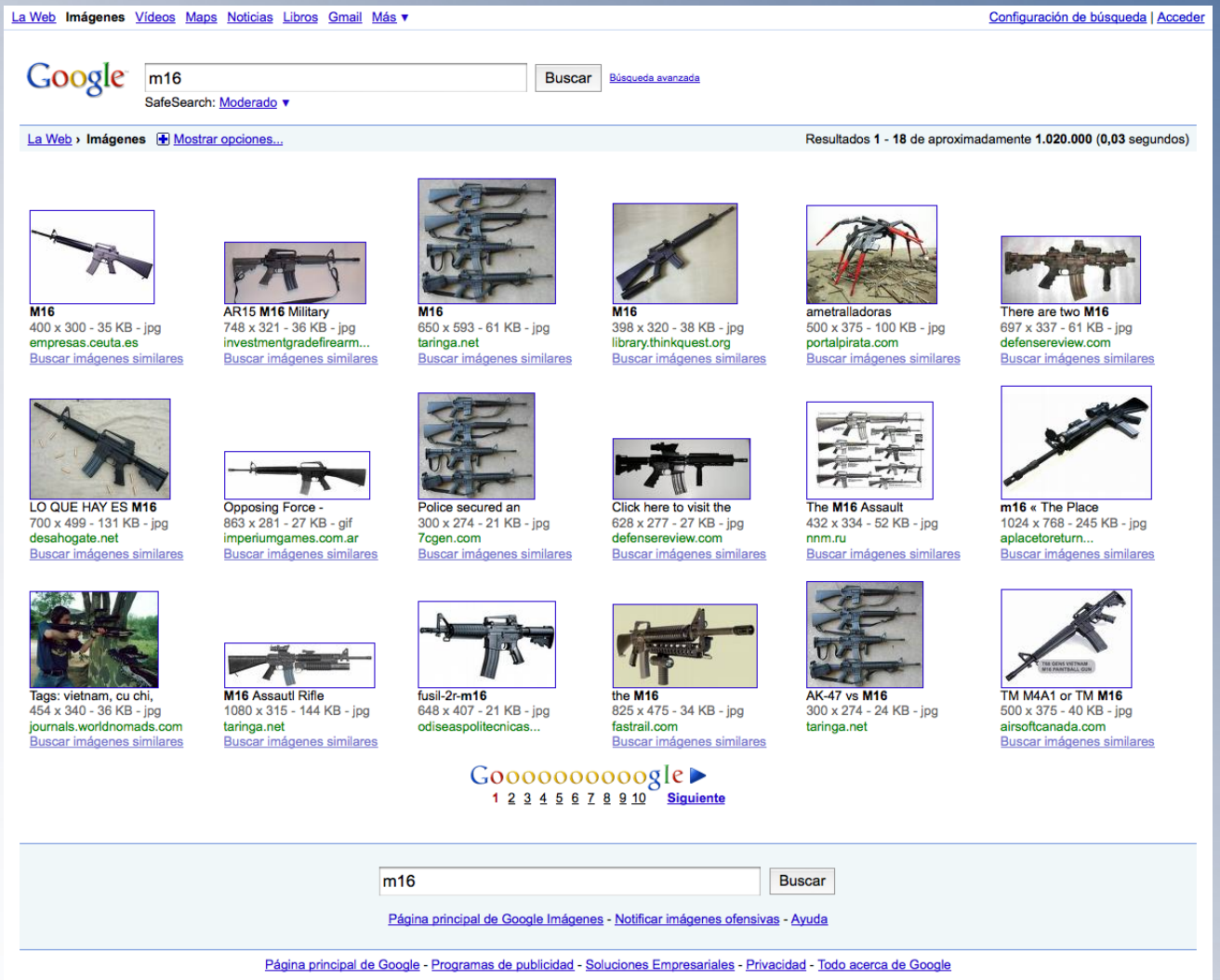


M16

Analysis of the Interstellar Medium of Isolated GAlaxies

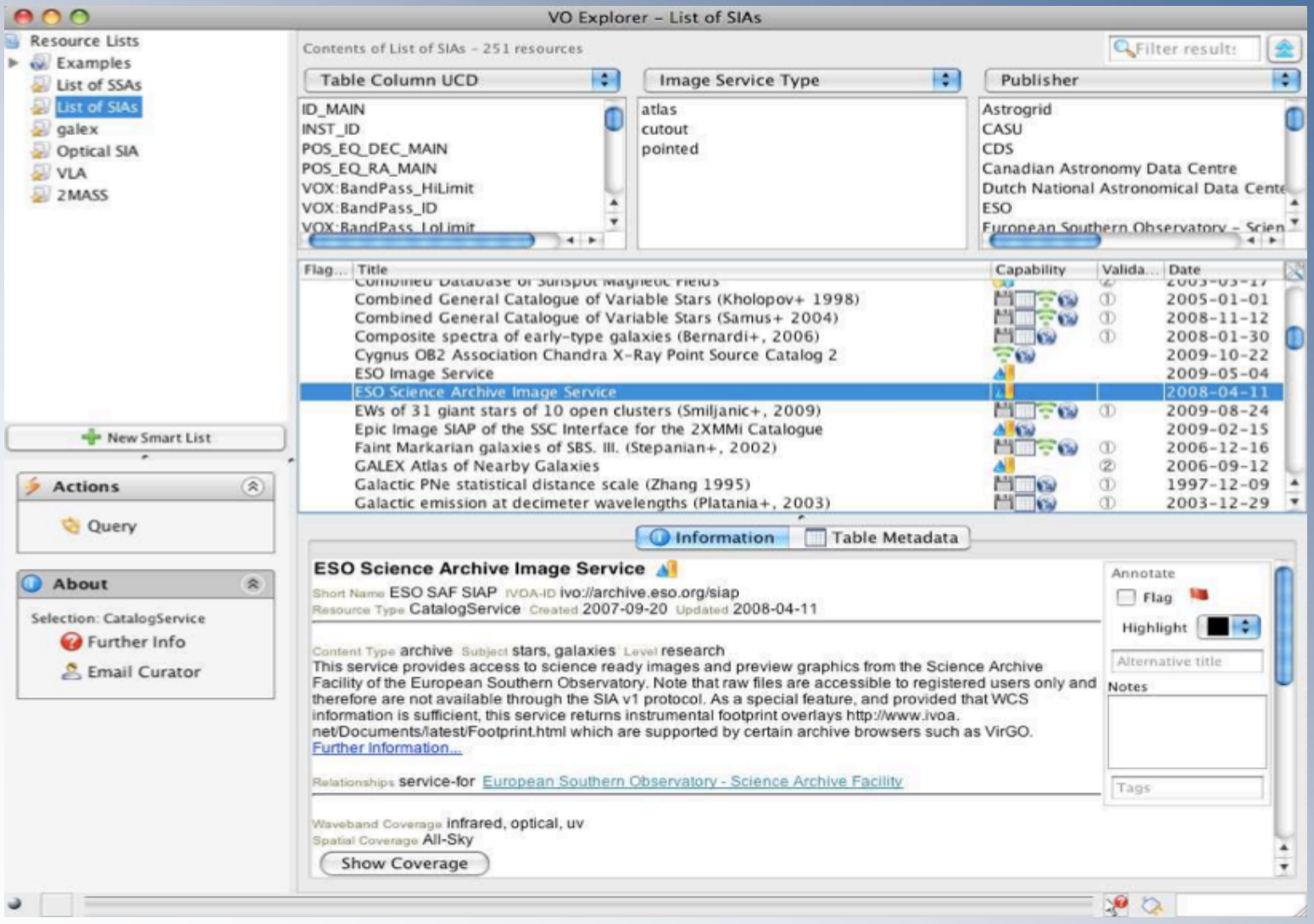
AMIGA

Ontologies, SKOS Vocabularies

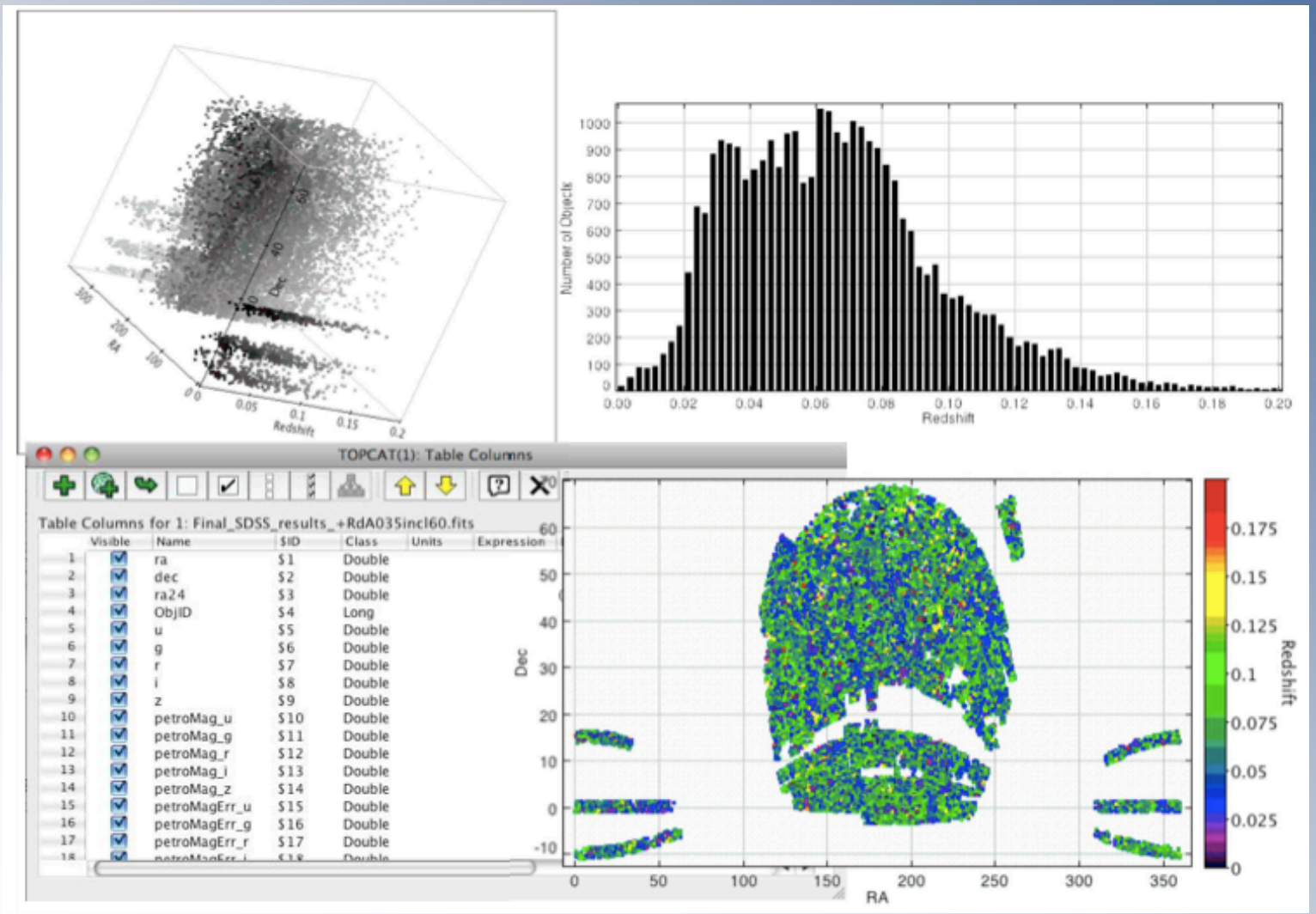


VO Software

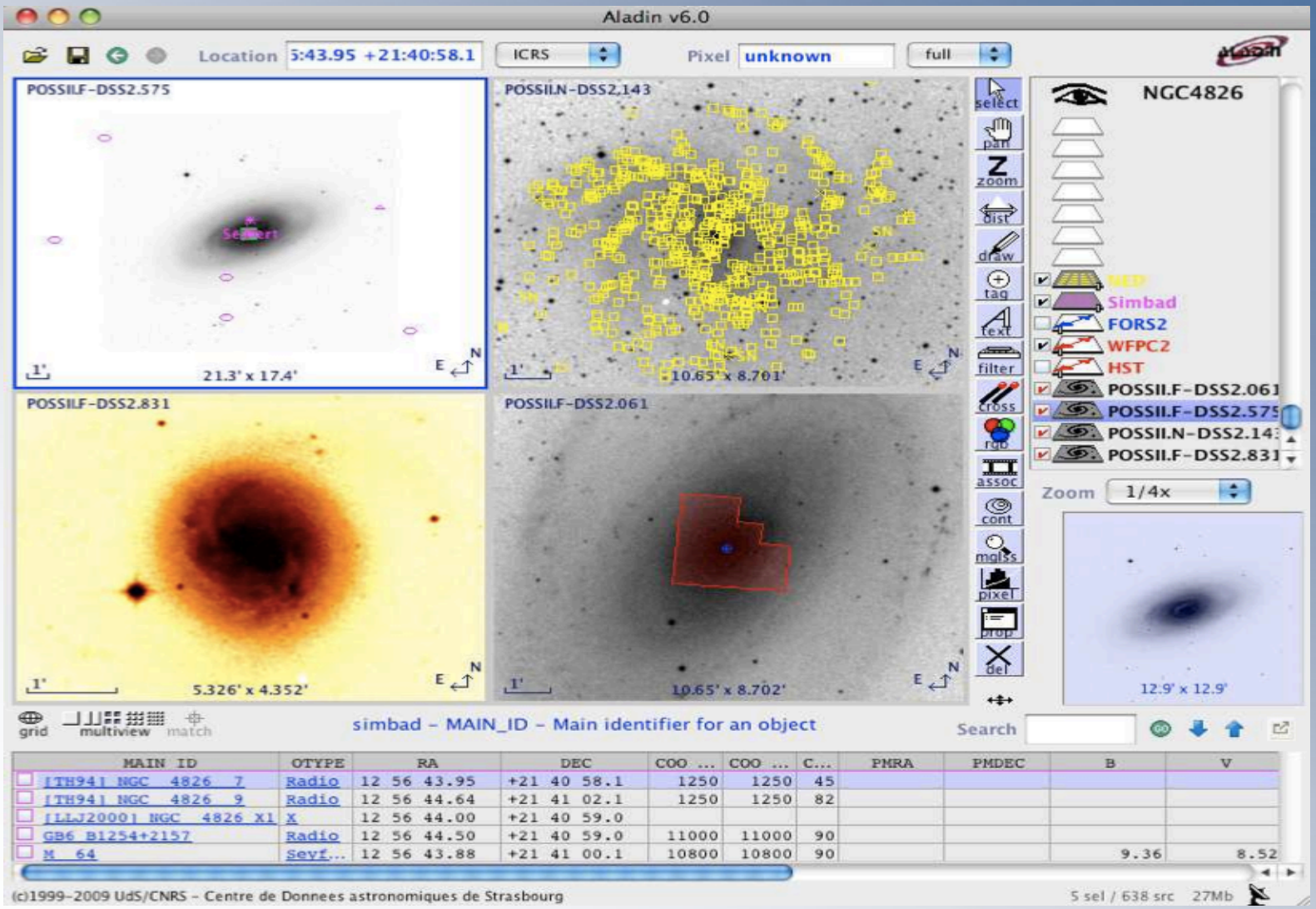
VODesktop



TopCat

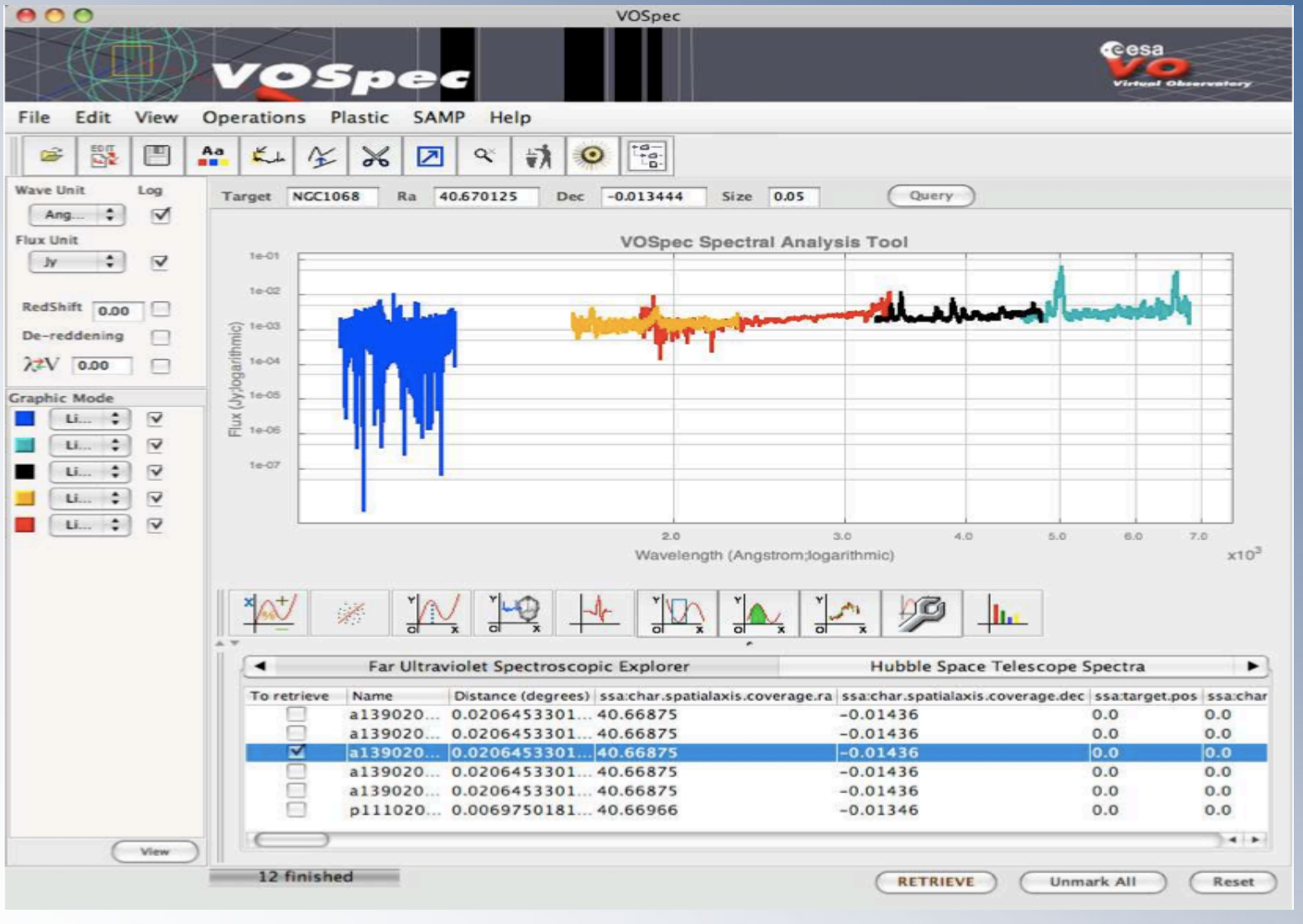

Redshift
RA
Dec
Number of Objects
Redshift
TOPCAT(1): Table Columns
Table Columns for 1: Final_SDSS_results_+RdA035incl60.fits
Visible | Name | $ID | Class | Units | Expression
1 ra $1 Double
2 dec $2 Double
3 ra24 $3 Double
4 ObjID $4 Long
5 u $5 Double
6 g $6 Double
7 r $7 Double
8 i $8 Double
9 z $9 Double
10 petroMag_u $10 Double
11 petroMag_g $11 Double
12 petroMag_r $12 Double
13 petroMag_i $13 Double
14 petroMag_z $14 Double
15 petroMagErr_u $15 Double
16 petroMagErr_g $16 Double
17 petroMagErr_r $17 Double
Dec
RA
Redshift

Aladin Sky Atlas



VO Software

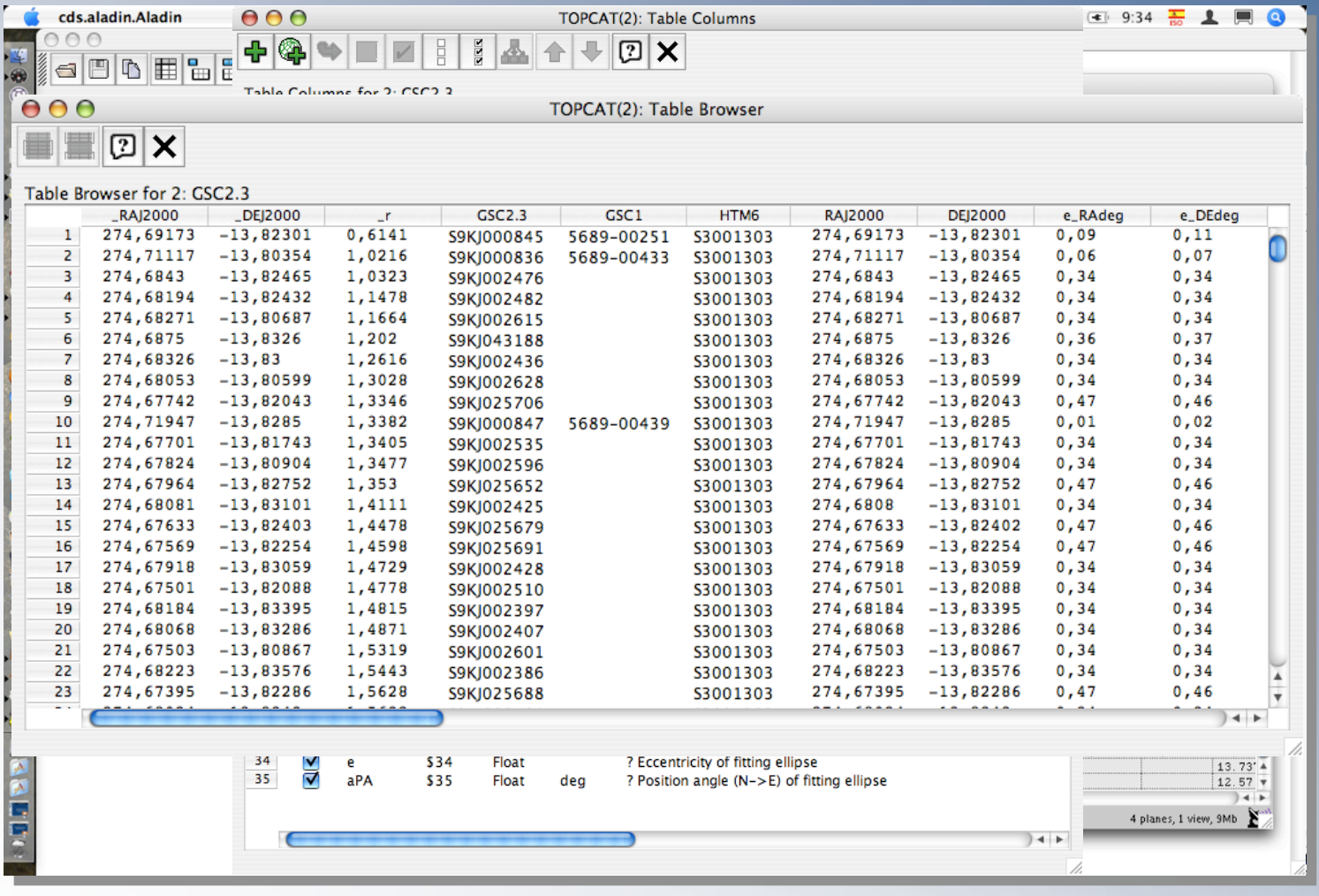
VOSpec



AMIGA

VO Software

SAMP/WebSAMP



	_RAJ2000	_DEJ2000	_r	GSC2.3	GSC1	HTM6	RAJ2000	DEJ2000	e_RAdeg	e_DEdeg
1	274,69173	-13,82301	0,6141	S9KJ000845	5689-00251	S3001303	274,69173	-13,82301	0,09	0,11
2	274,71117	-13,80354	1,0216	S9KJ000836	5689-00433	S3001303	274,71117	-13,80354	0,06	0,07
3	274,6843	-13,82465	1,0323	S9KJ002476		S3001303	274,6843	-13,82465	0,34	0,34
4	274,68194	-13,82432	1,1478	S9KJ002482		S3001303	274,68194	-13,82432	0,34	0,34
5	274,68271	-13,80687	1,1664	S9KJ002615		S3001303	274,68271	-13,80687	0,34	0,34
6	274,6875	-13,8326	1,202	S9KJ043188		S3001303	274,6875	-13,8326	0,36	0,37
7	274,68326	-13,83	1,2616	S9KJ002436		S3001303	274,68326	-13,83	0,34	0,34
8	274,68053	-13,80599	1,3028	S9KJ002628		S3001303	274,68053	-13,80599	0,34	0,34
9	274,67742	-13,82043	1,3346	S9KJ025706		S3001303	274,67742	-13,82043	0,47	0,46
10	274,71947	-13,8285	1,3382	S9KJ000847	5689-00439	S3001303	274,71947	-13,8285	0,01	0,02
11	274,67701	-13,81743	1,3405	S9KJ002535		S3001303	274,67701	-13,81743	0,34	0,34
12	274,67824	-13,80904	1,3477	S9KJ002596		S3001303	274,67824	-13,80904	0,34	0,34
13	274,67964	-13,82752	1,353	S9KJ025652		S3001303	274,67964	-13,82752	0,47	0,46
14	274,68081	-13,83101	1,4111	S9KJ002425		S3001303	274,6808	-13,83101	0,34	0,34
15	274,67633	-13,82403	1,4478	S9KJ025679		S3001303	274,67633	-13,82402	0,47	0,46
16	274,67569	-13,82254	1,4598	S9KJ025691		S3001303	274,67569	-13,82254	0,47	0,46
17	274,67918	-13,83059	1,4729	S9KJ002428		S3001303	274,67918	-13,83059	0,34	0,34
18	274,67501	-13,82088	1,4778	S9KJ002510		S3001303	274,67501	-13,82088	0,34	0,34
19	274,68184	-13,83395	1,4815	S9KJ002397		S3001303	274,68184	-13,83395	0,34	0,34
20	274,68068	-13,83286	1,4871	S9KJ002407		S3001303	274,68068	-13,83286	0,34	0,34
21	274,67503	-13,80867	1,5319	S9KJ002601		S3001303	274,67503	-13,80867	0,34	0,34
22	274,68223	-13,83576	1,5443	S9KJ002386		S3001303	274,68223	-13,83576	0,34	0,34
23	274,67395	-13,82286	1,5628	S9KJ025688		S3001303	274,67395	-13,82286	0,47	0,46

34	e	$34	Float		? Eccentricity of fitting ellipse
35	aPA	$35	Float	deg	? Position angle (N->E) of fitting ellipse

A Cloud of Services

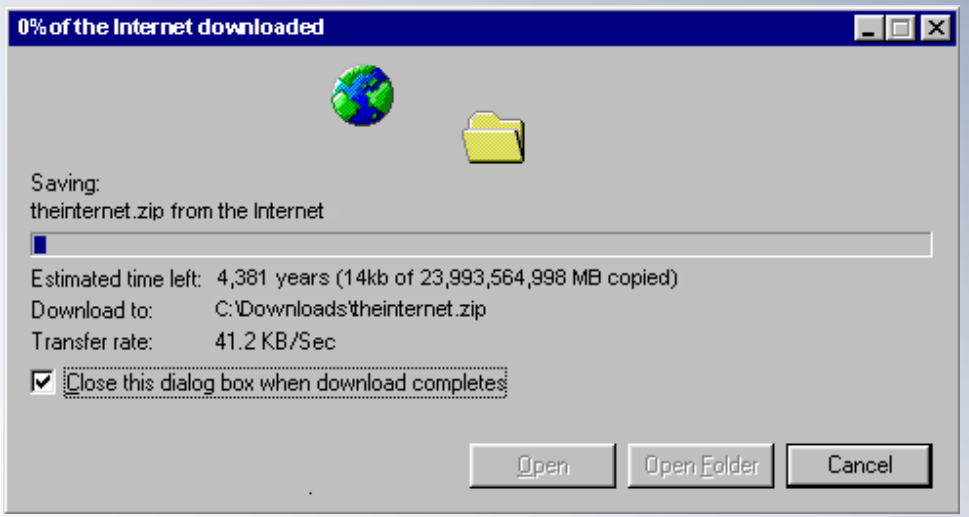
The next generation of archives

Much wider FoV and spectral coverage

- Large volumes for an observed datacube
- Subproducts are Virtual Data generated on-the-fly

Automated surveys

- Huge amounts of tabular data
- Services for Knowledge Discovery in Databases



Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

Cube sizes

	Low Res		High Res		Extreme Res	
Number	4 Bytes	4B	4 Bytes	4B	4 Bytes	4B
Resolution	2,048 x 2,048	16MB	8,192 x 8,192	268MB	12,288 x 12,288	603MB
Channels	16,384	0.27TB	16,384	4.39TB	16,384	9.8TB
Stokes & Weighting	1	0.27TB	1	4.39TB	4 + 1	49.5TB

ASKAP Cubes

Prof. Kevin Vinsen

The overall picture

Distributed, scalable and flexible infrastructure

- Grid + Cloud may solve storage and processing
- Bandwidth is the issue

Big Data Science performance is highly dependent upon I/O data rates (local and transfer)

The data is the infrastructure

- Interconnected and interoperable archives
- Distributed, multi-wavelength and multi-facilities

Archives speaking Web Services

ALMA, LSST, ASKAP, MeerKAT, LOFAR, Apertif,...

The overall picture

We are moving into a world where

- computing and storage are cheap
- data movement is death

Archives should evolve from data providers into virtual data and services providers, where web services may help to solve bandwidth issues.

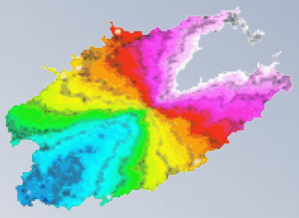
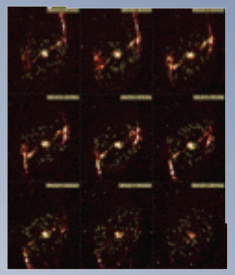
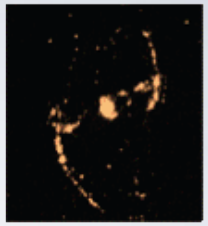
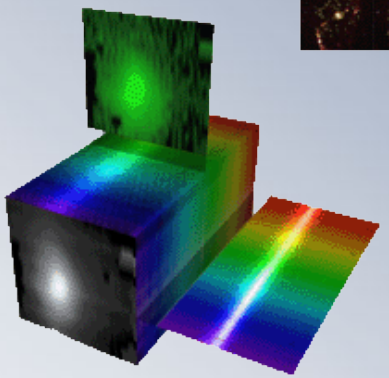
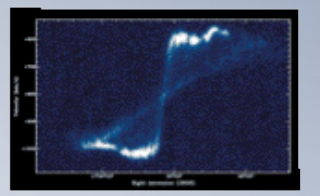
Web Services

- Smaller virtual data subproducts
- Distributed, multi-archive, multi-wavelength astronomy
- Workflows as a disruptive working methodology

A Cloud of Services

3D Data Services

- Cutout
- Resample
- Spectrum extraction
- 2D slice extraction
- Dimensional reduction
- Filtering/Flagging
- 2D Moments
- Complex transformations

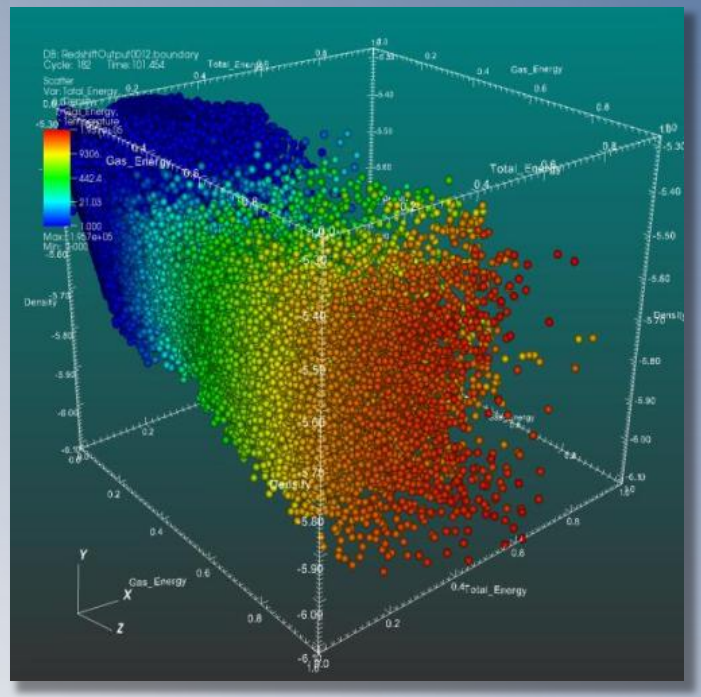






Exploration services

KDD - Knowledge Discovery in Databases

Understand what information is contained within the data in order to know how we can efficiently extract it

- Anomaly detection
- Cross-matching data
- Dimensionality reduction



VisIt software

Extraction of scientifically relevant information from a multidimensional parameter space.

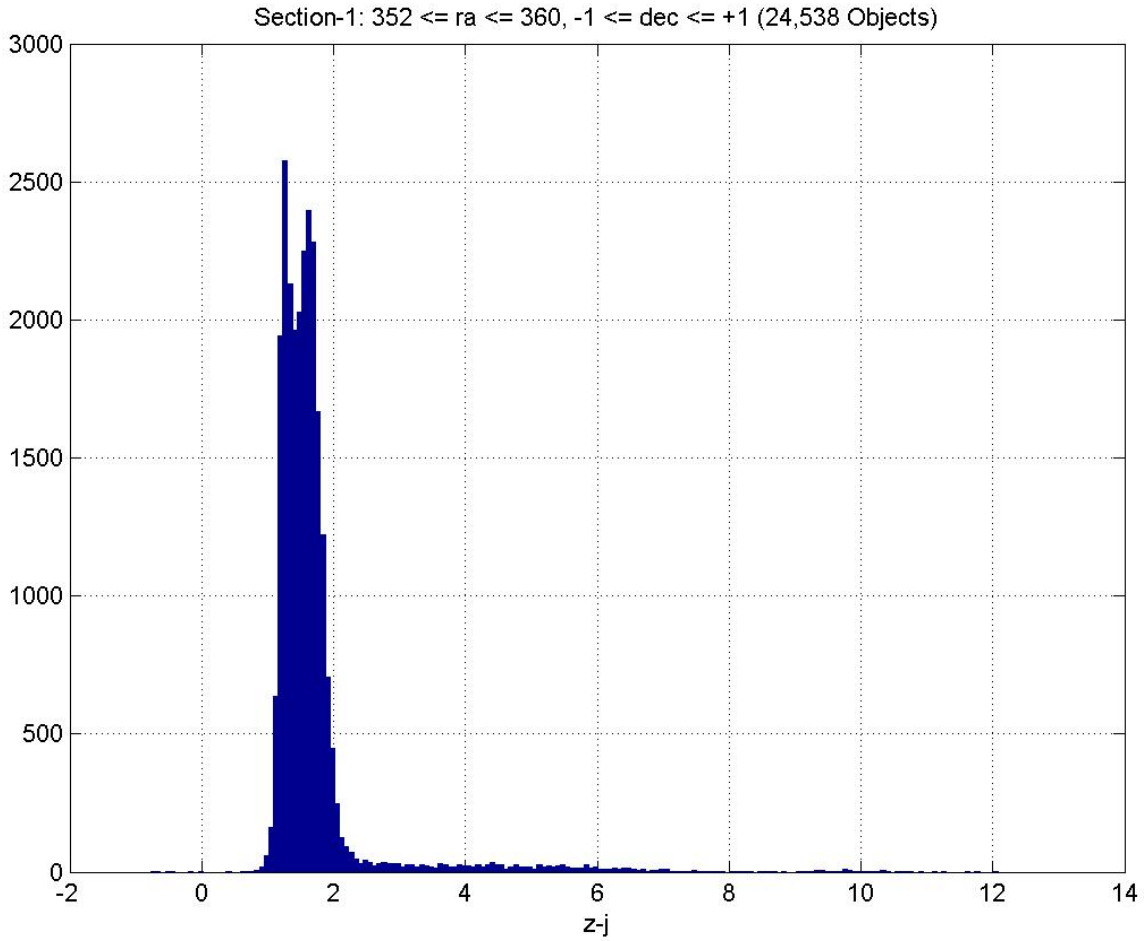
Data Mining

Some key astronomy problems that can be addressed with data mining techniques:

- Cross-Match objects from different catalogues
- The distance problem (e.g., Photometric Redshift estimators)
- Star-Galaxy Separation
- Cosmic-Ray Detection in images
- Supernova Detection and Classification
- Morphological Classification (galaxies, AGN, gravitat. lenses, ...)
- Class and Subclass Discovery (brown dwarfs stars, ...)
- Dimension Reduction = Correlation Discovery
- Learning Rules for improved classifiers
- Classification of massive data streams
- Real-time Classification of Astronomical Events
- Clustering of massive data collections
- Novelty, Anomaly, Outlier Detection in massive databases

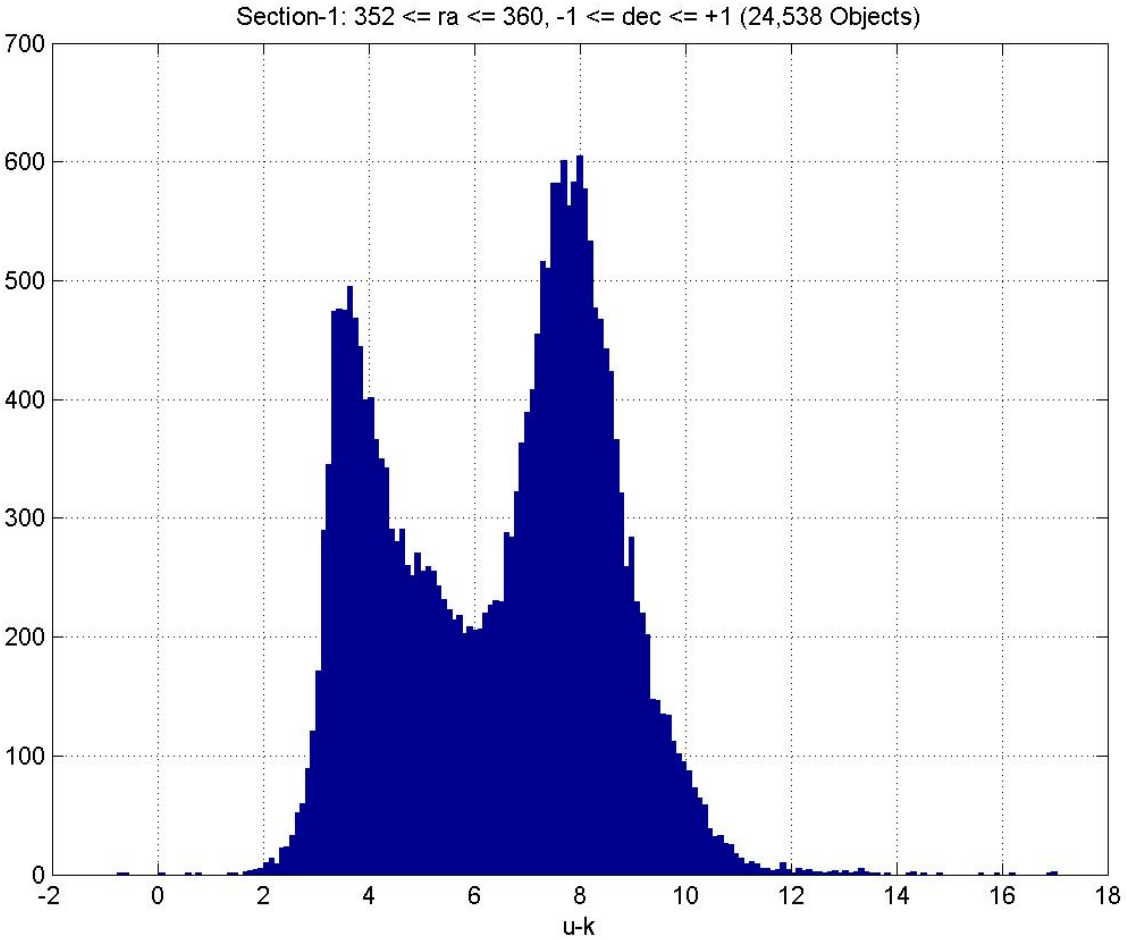
Scientific use cases

Clustering



Scientific use cases

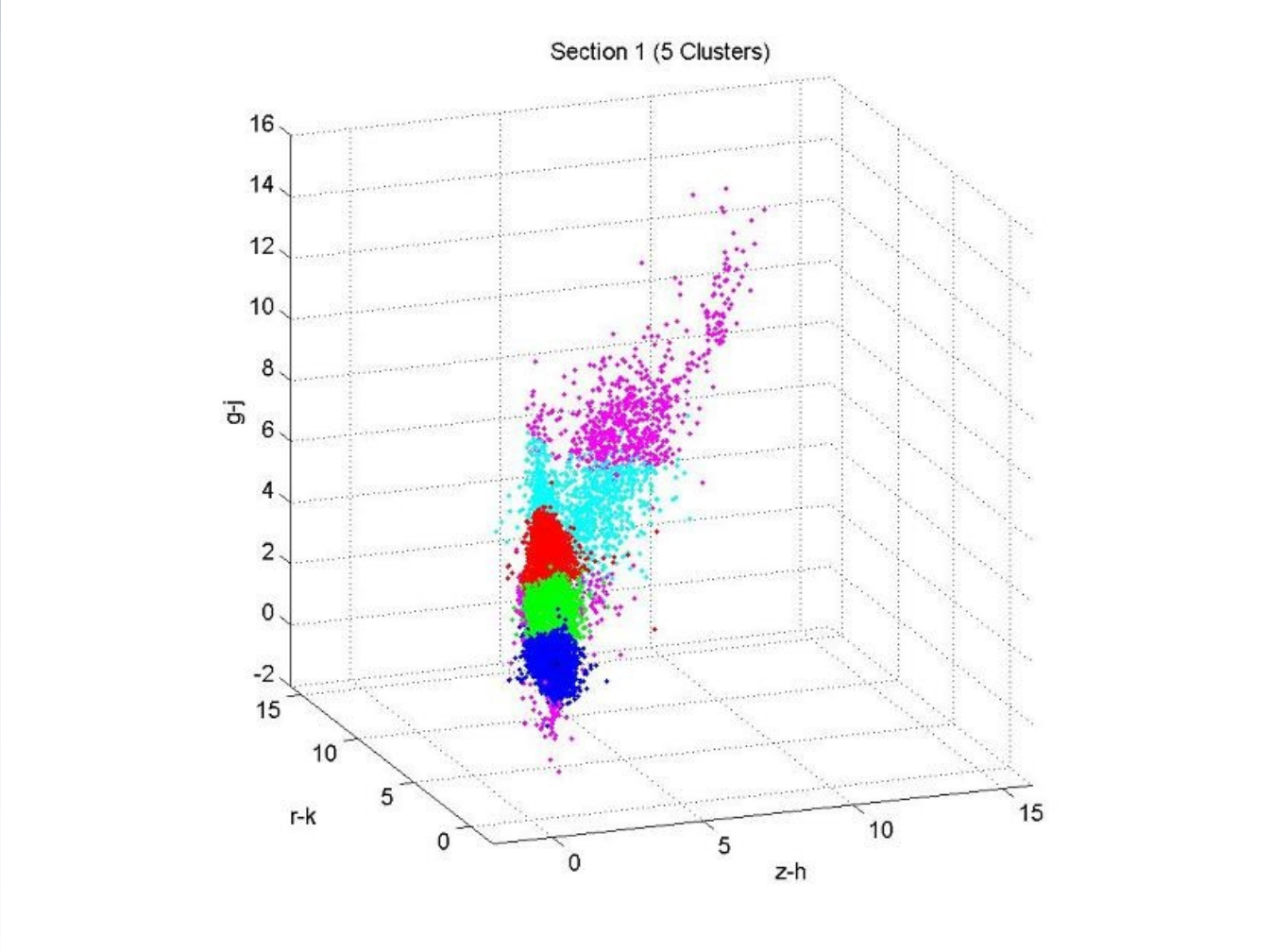
Clustering



Analysis of the Interstellar Medium of Isolated GAlaxies

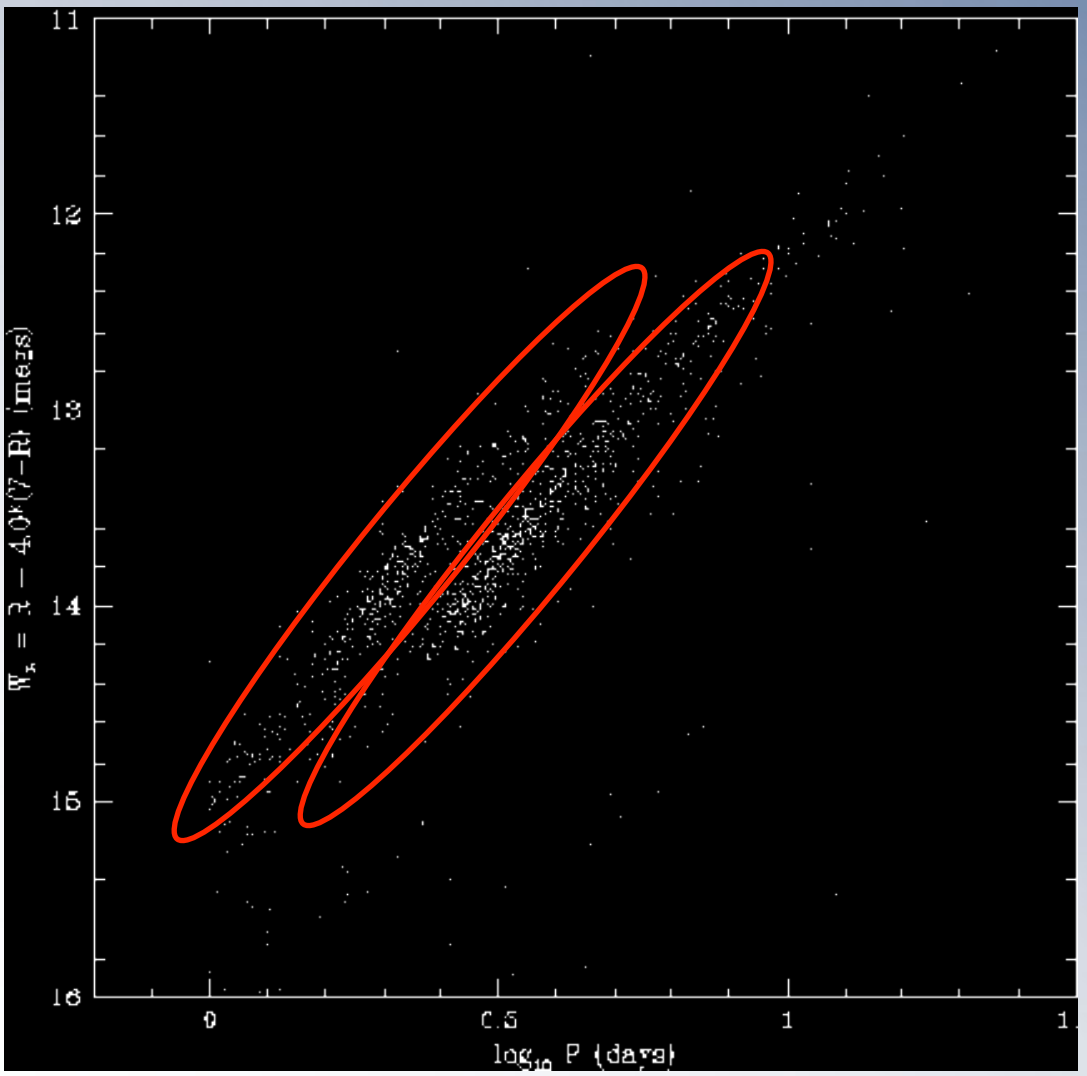
AMIGA

Multidimensional Clustering



Scientific use cases

Clustering



Cepheid variables
Cosmic yardsticks

-- One Correlation
-- Two Classes!

Scientific use cases

Outlier detection

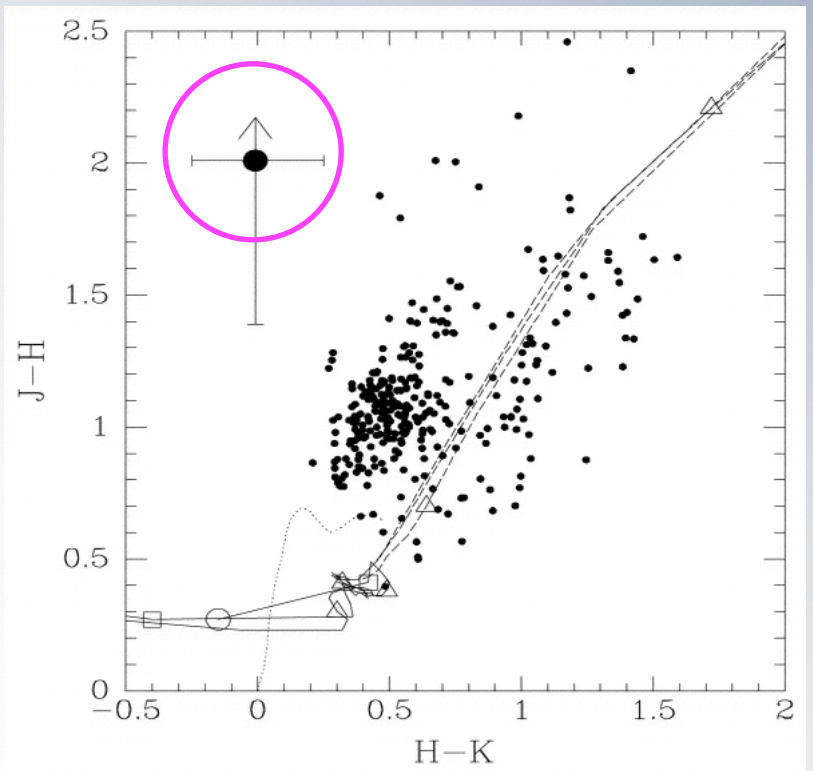
THE NEW YORK TIMES **NATIONAL** TUESDAY, JUNE 1, 1999

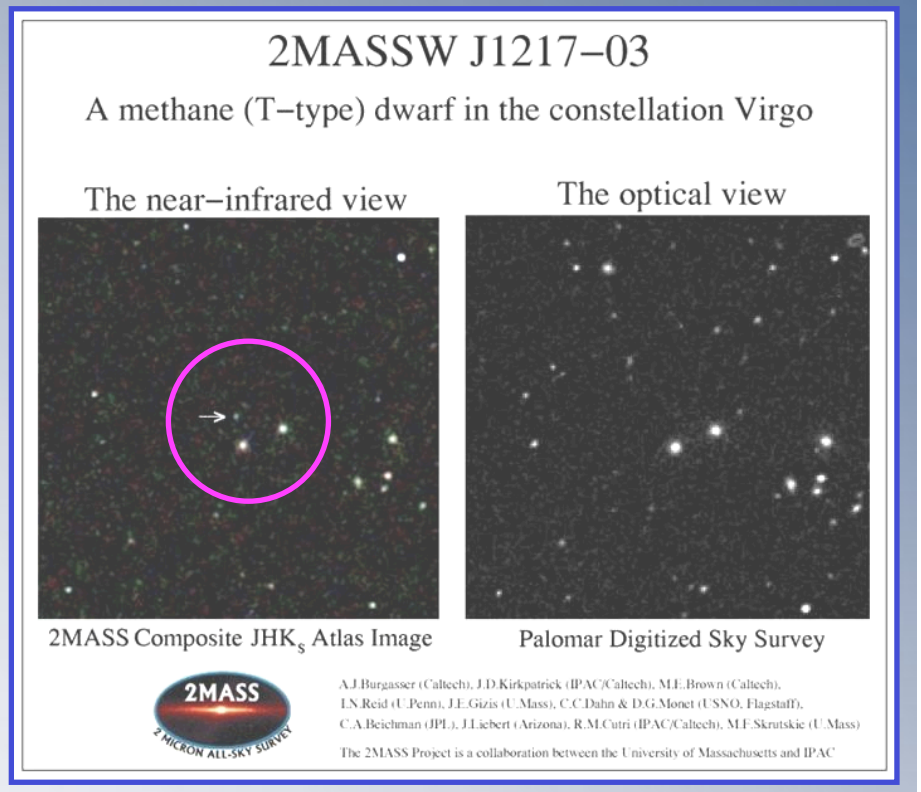
Astronomers Detect New Category of Elusive 'Brown Dwarfs'

By JOHN NOBLE WILFORD

CHICAGO, May 31 — Ambitious Apache Point, N.M. Dr. Michael Strauss and a graduate student, was a brown dwarf, but was not associated with a star companion. have been possible in hotter, younger objects. An estimate of their mass dim methane dwarfs.

The detection of four oth objects was reported separa scientists at the California I of Technology and the Jet Pro





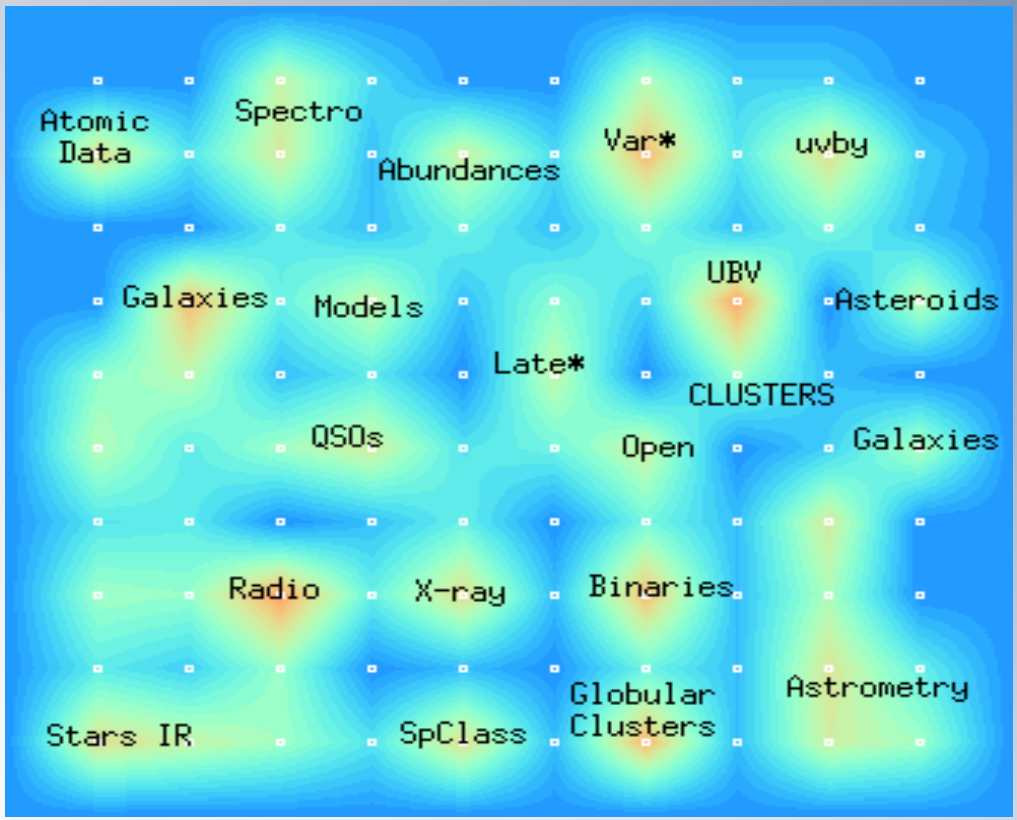
Scientific use cases

Self Organizing Map

Organizing information in complex data collections

Find hidden relationships and patterns

Based on links among keywords and metadata

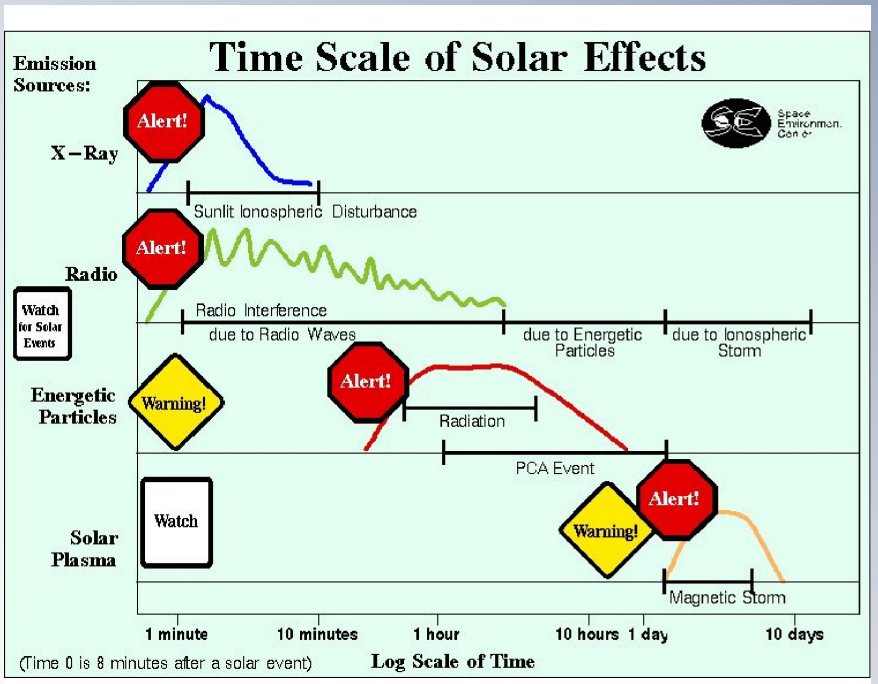


Scientific use cases

The time domain

- VO Sky Event reporting metadata
- What, Where, Who, How ?
- Stars flares ,GRBs, solar, atmospheric particle bursts,..

The Helio-VO Project



The VO-Experiment

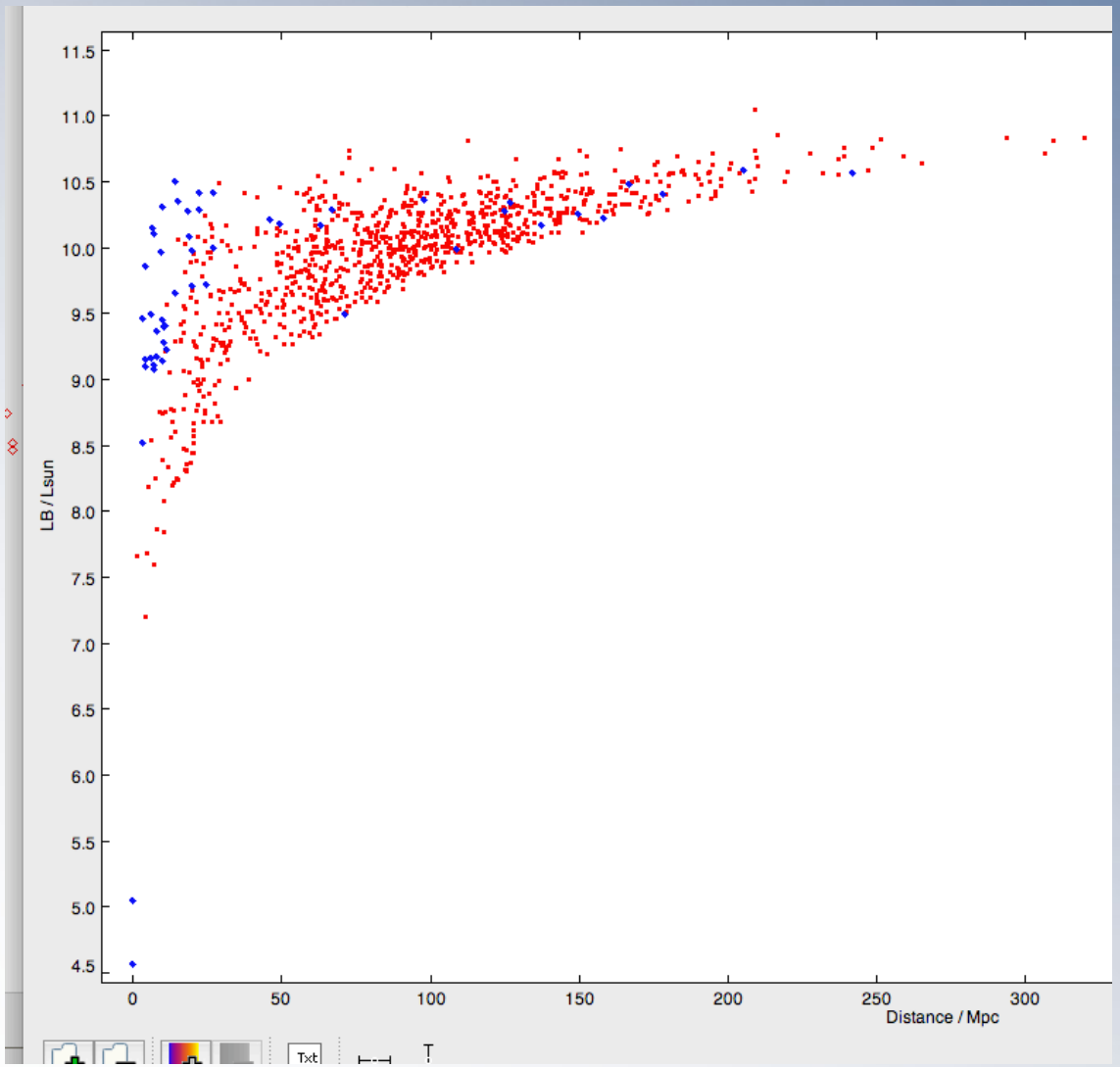
- Data Mining Oriented
- VO Services
  - Discovery
  - Access
  - Waiting for analysis services
- Local software (also some Web portals)
  - Crossmatching
  - Inspection
  - Visualization
- Web services associated to archives of big facilities
  - Hinders cross-boundary science

XMM Observations of the AMIGA Sample

TopCat Hands-On
Let's do some science !

Scientific use cases

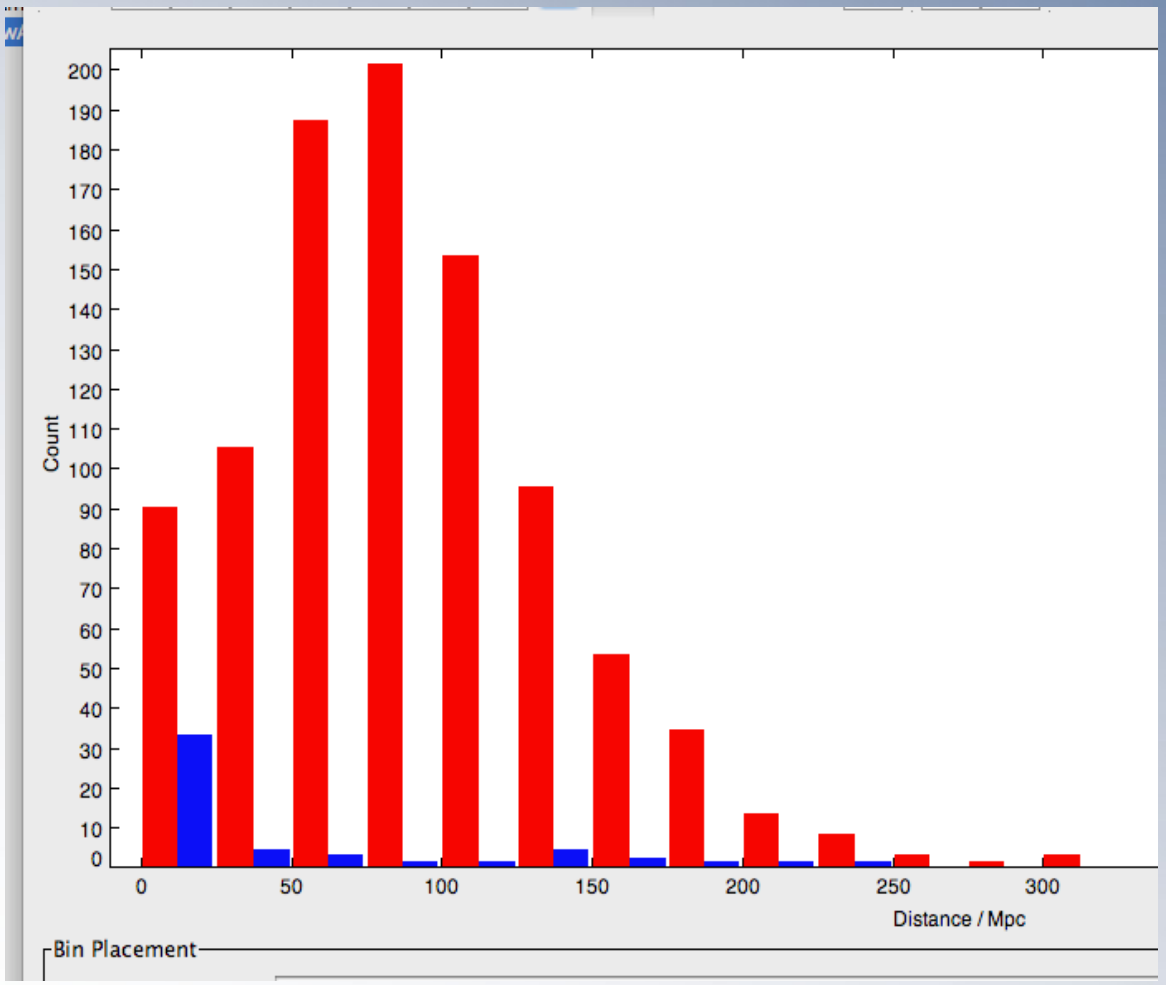
XMM Observations of the AMIGA Sample



Slightly brighter

Scientific use cases

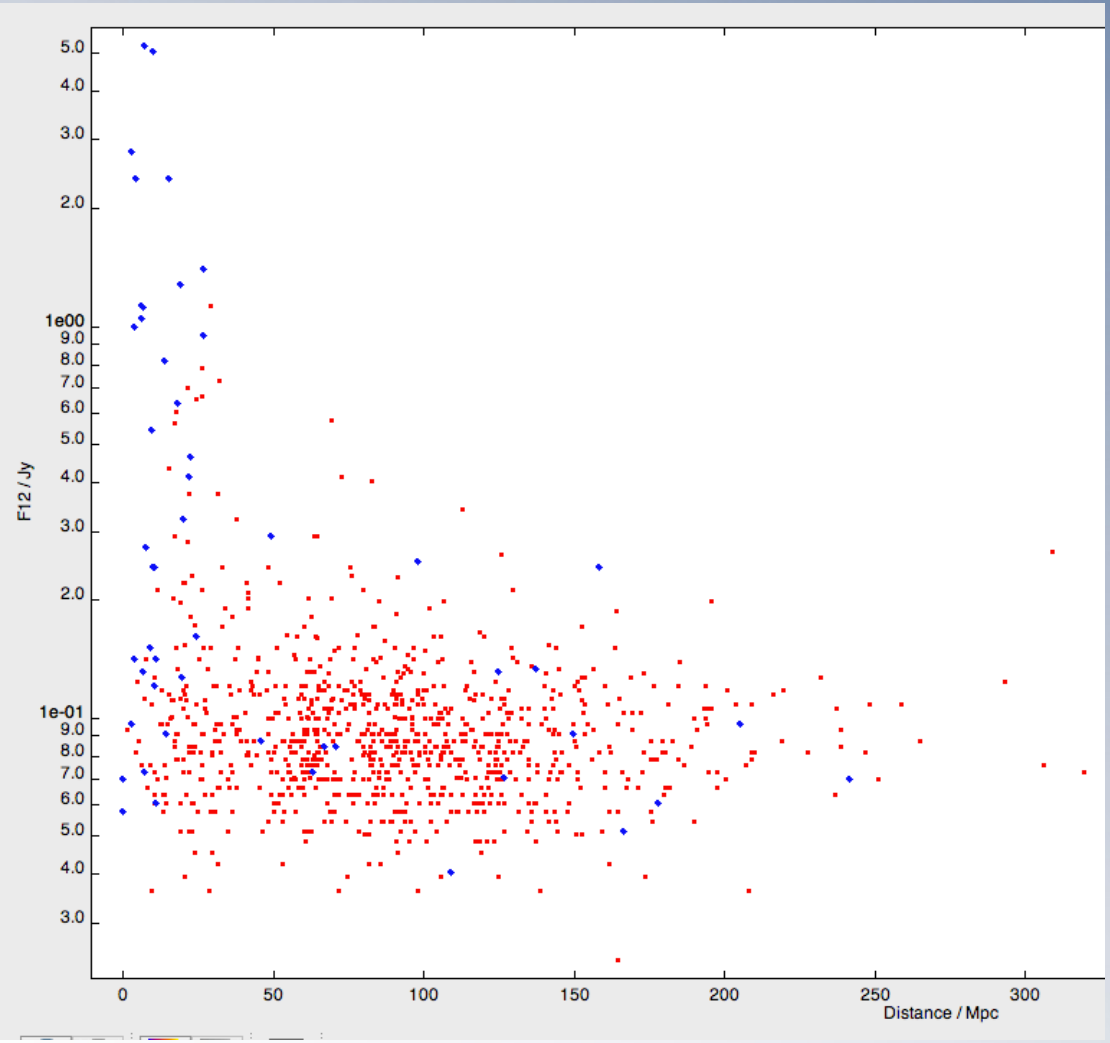
XMM Observations of the AMIGA Sample



Slightly brighter
Closer

Scientific use cases

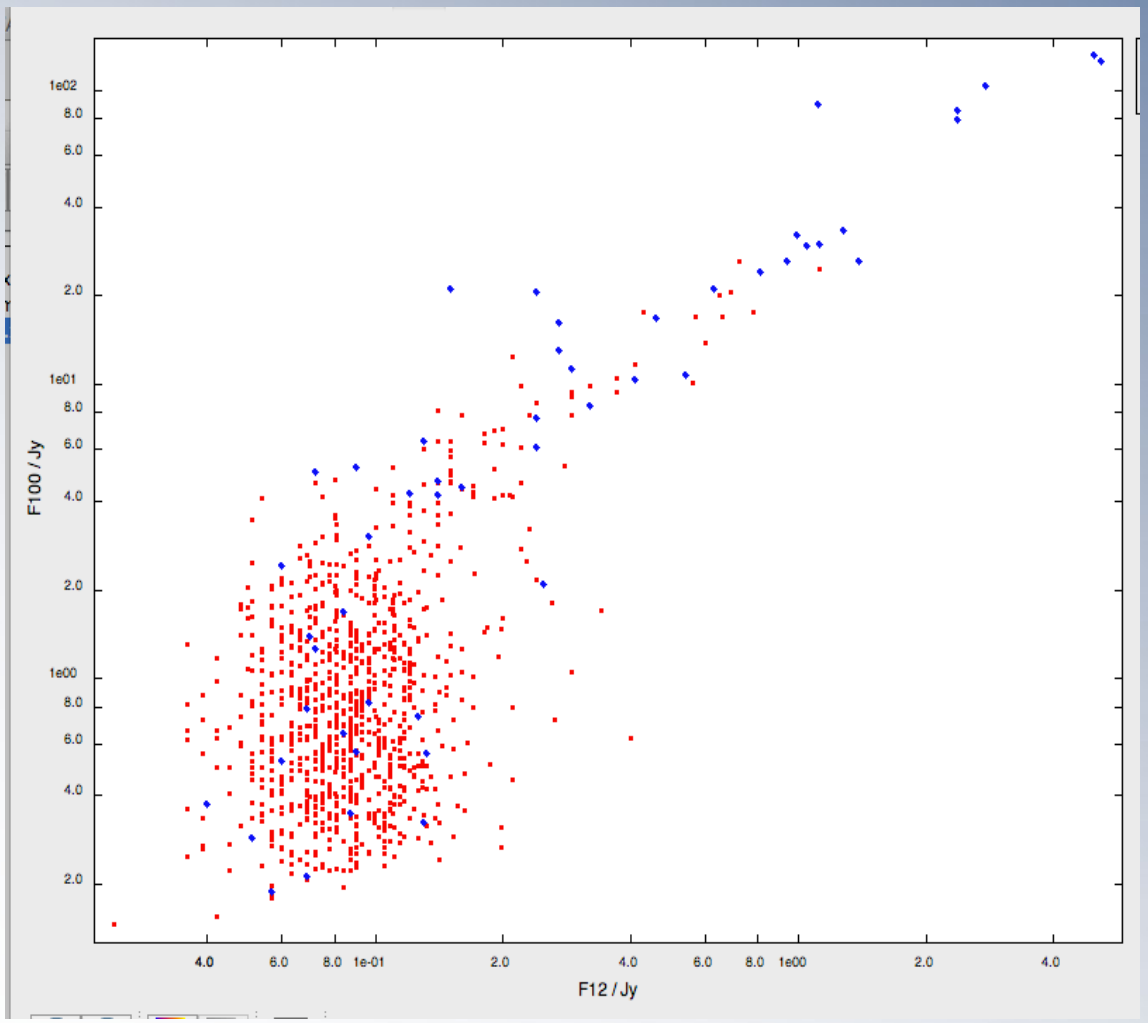
XMM Observations of the AMIGA Sample



Slightly brighter
Closer
Brighter in FIR

Scientific use cases

XMM Observations of the AMIGA Sample



Slightly brighter
Closer
Brighter in FIR
Excess in longer λ

Why Workflows ?
Web-services-based vs. Pipelines

- Expose the scientific methodology
- Keep the provenance
- Pack the experiment
- Enable
  - repeatable results
  - reproducibility
  - reuse, repurpose
  - cross-boundary science
  - preservation

formative

collaborative

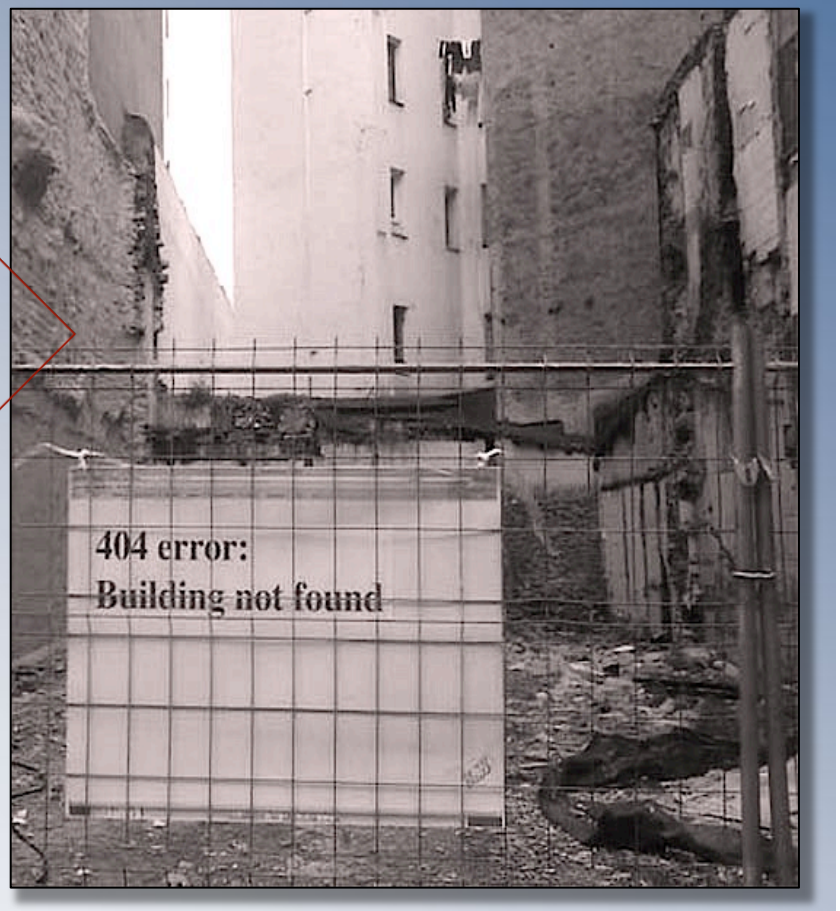
Workflows Preservation

All components related to the research lifecycle should be available.

Preserved and easily retrievables

- Proposals
- Data
- Processes
- Workflows
- Publications

RESEARCH OBJECT



Open questions for Web Services In the Virtual Observatory

- Curation and preservation (identifiers)
- Discovery (semantics) of web services
- Characterization: input, outputs, functionality, etc.
- Copies (authenticity) or similar used as alternates
- Permissions (authentication), licenses, platform, costs,..
- Metrics for quality: popularity, use stats, logs uptime, etc.
- Versioning and authoring (referenced and acknowledged)

In a cloud of services and data, Web Services should benefit of the same privileges acquired by Data.

IVOA Note on Workflows

Announcement about Workflow future plans IVOA | X Wf4ever | X

Andre Schaaff para interop mostrar detalles 1 jul (hace 12 días) Responder a todos

Dear all,

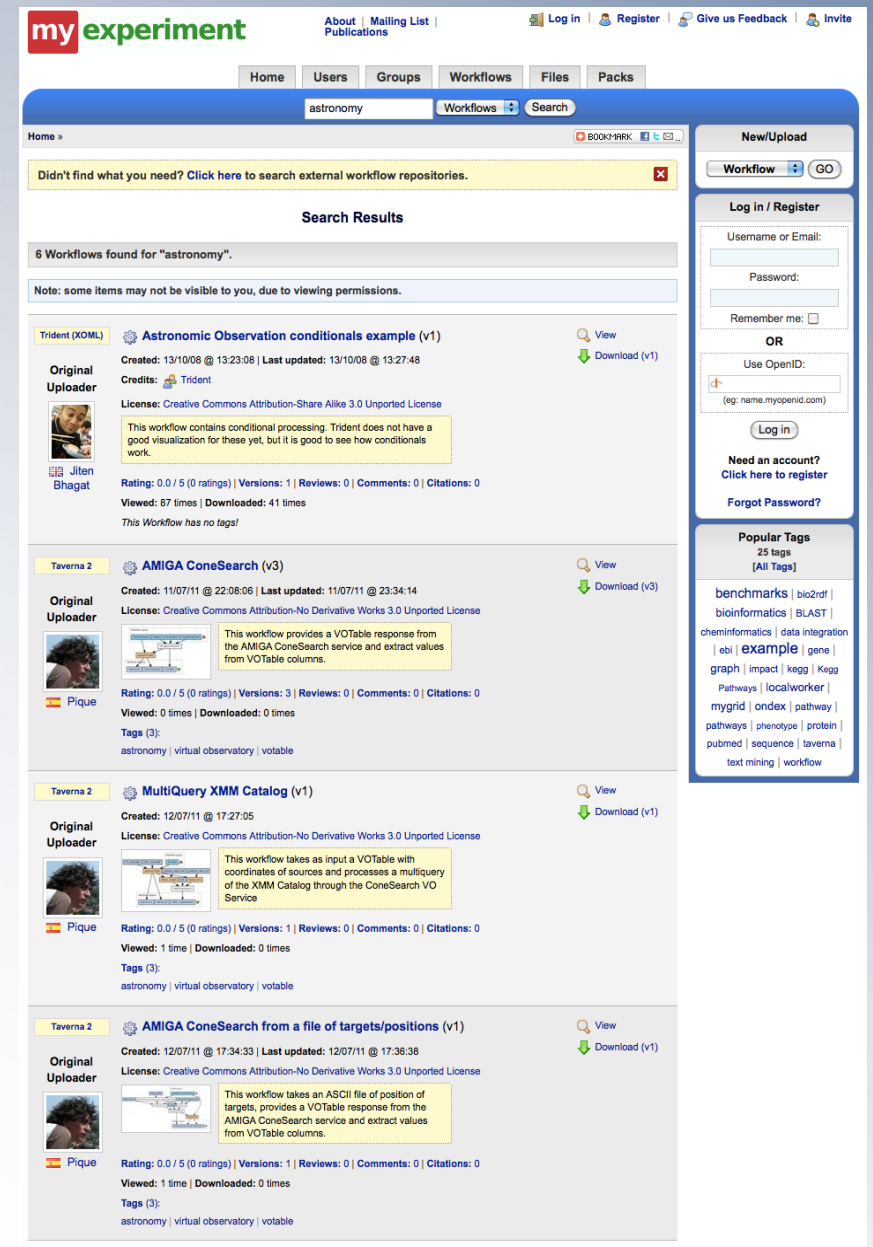
As you know we will soon be facing a new generation of facilities and archives dealing with huge amounts of data (ALMA, LSST, panSTARRS, LOFAR, SKA pathfinders,..) where scientific Workflows will play an important role in the working methodology of astronomers. A detailed analysis about the state of the art of workflows in the frame of the VO involves languages, design tools, execution engines, use cases, etc. A major topic is also the preservation of the workflows and the capability to replay a workflow several years after its design and implementation. Several talks concerning these issues have been presented during the past IVOA Interop meetings (see references below).

In order to undertake this task within our community we think that as a first step a Note should be written. Participation is welcome, in particular, we would like to collect experiences (including use cases, tools, etc.), references, remarks, etc. We plan the Note to be published around September and discussed at the Pune Interop meeting. We should then decide on starting a working draft describing those aspects and possibly in a new Interest Group on Workflows.

Best regards,

André Schaaff and Jose Enrique Ruiz

MyExperiment



Astronomy

- No VO services-based Wfs
- Helio Project Wfs
- VOTables parsing
- Internal services

Amiga

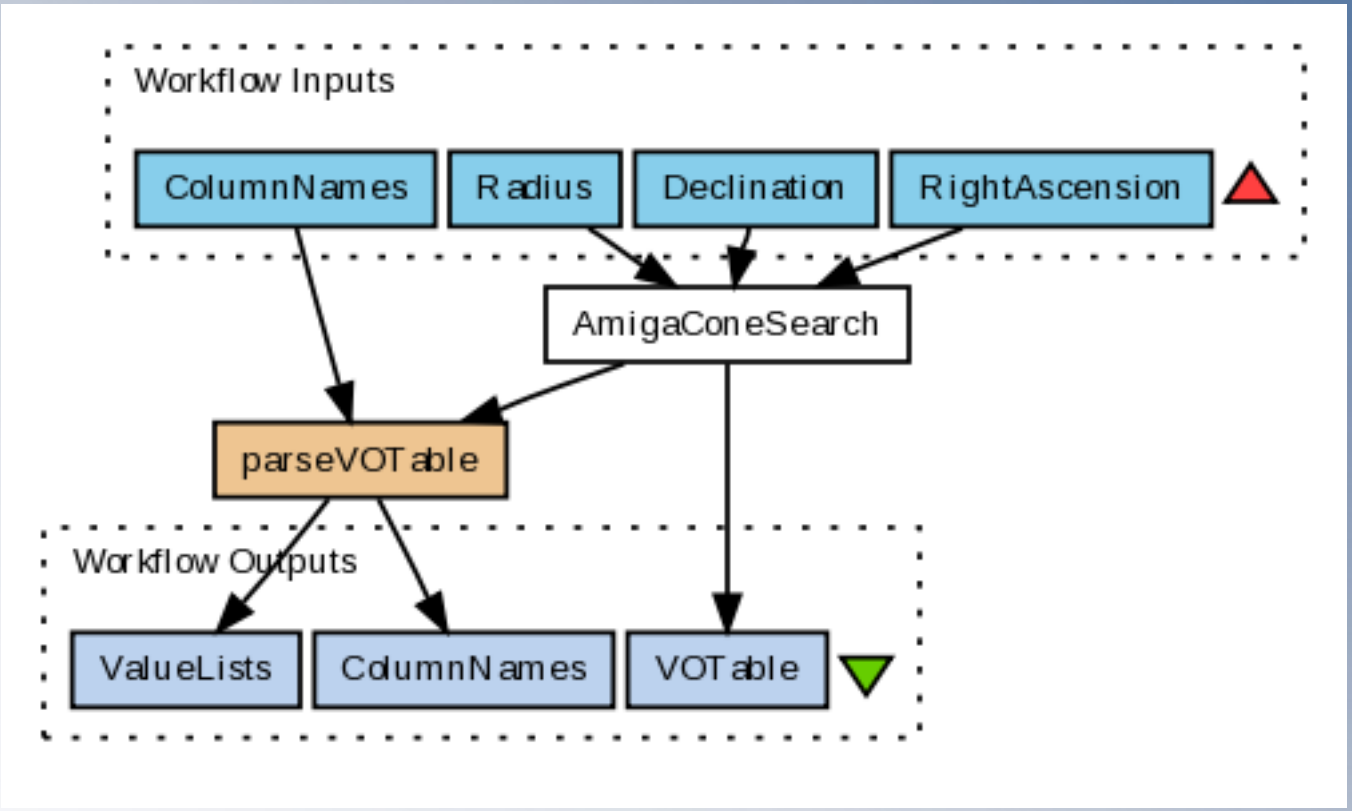
- Querying Catalogue

Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

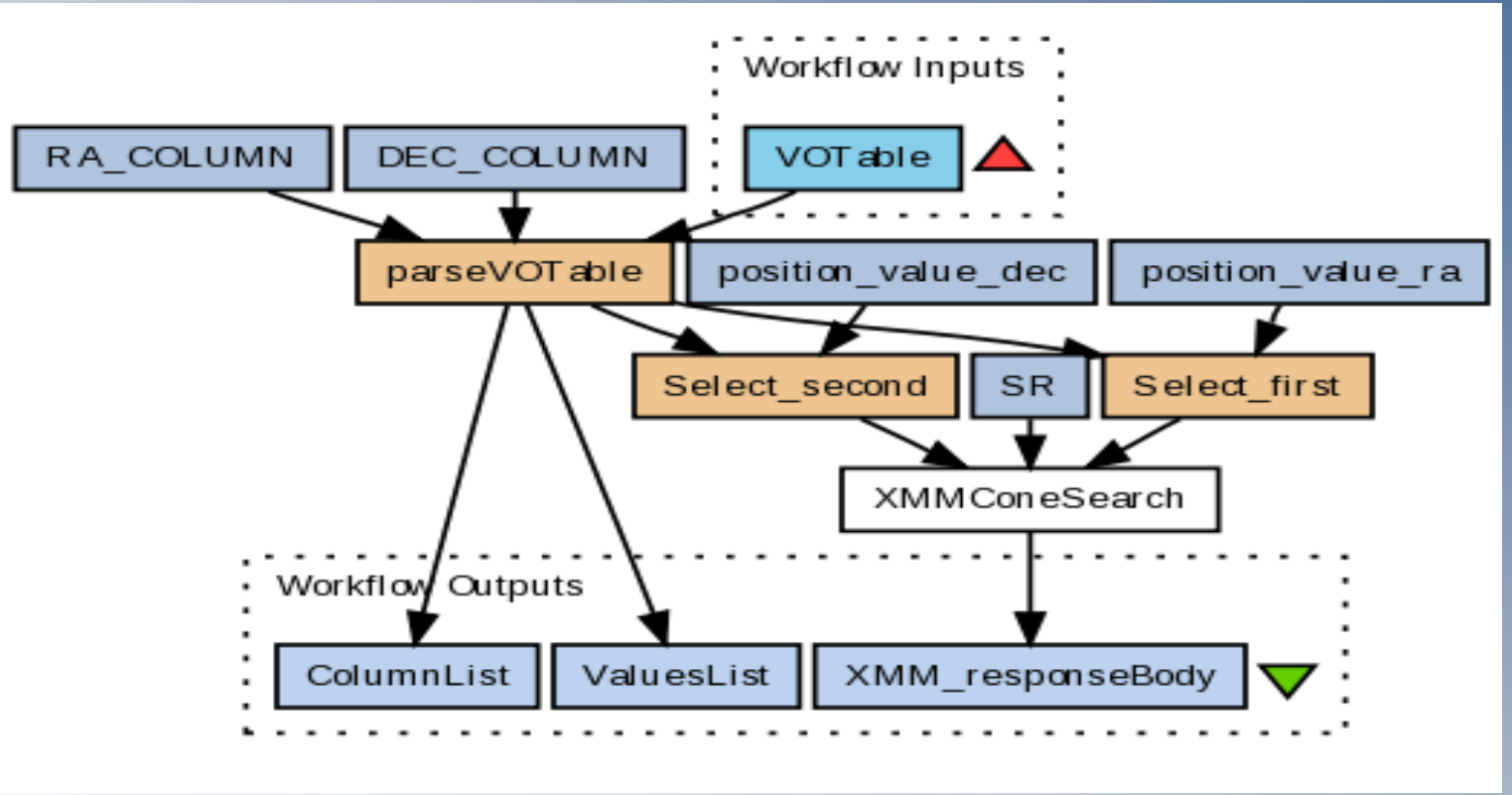
Working with the v2.3

Simple AMIGA ConeSearch



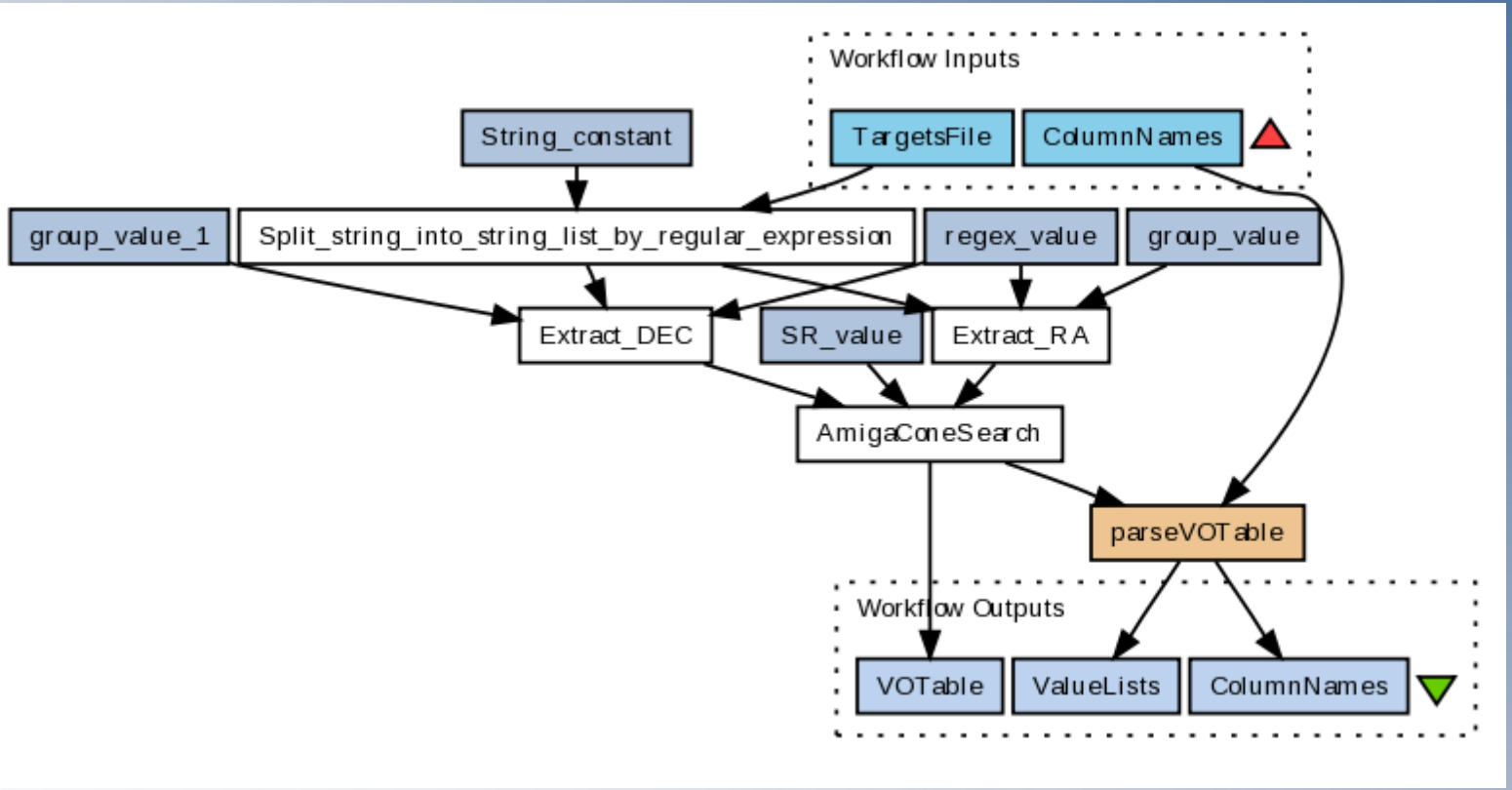
- Xpath plugin not a useful for extracting info from VOTable
- Helio-VO beanshell used instead (Thanks !)
- Visualization of results.. (VOTables)

XMM Multi-ConeSearch



- Lot of previous VOTable parsing ..
- The response is 1051 VOTables !
- VOTable merging tool needed

AMIGA Multi-ConeSearch



- Lot of beanshells for VOTabl and CSV parsing ..
- Beanshells development needed for splitting lists into values
- STILTS Library needed for VOTable crossmatching

The VO-experiment

- Discover Services
- Multi-query
- Crossmatching
- Inspection
- Visualization and Comparison

Proposed shortcuts for Taverna

- VORegistry Access Perspective
- STILTS VOTable Library
- SAMP (Connectivity with VO Software)
- Python based beanshells
- Simple standard astronomy functions

Thanks !

Wf4Ever @ Manchester

- Carole Goble
- Sean Bechhofer
- Jiten Baghat
- Stian Soiland-Reyes
- Kalid Belhajjame

Helio-VO

- John Brooke
- Donal Felows
- Anja Leblanc

And all the MyGrid Team !

Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

Thanks !

Thanks !

Thanks !

Thanks !