

Workshop
The Growing Demands on Connectivity and Information Processing in Radio Astronomy from
VLBI to the SKA

Instituto de Telecomunicações, Aveiro, Portugal
24th - 25th May 2011

web services based workflows to deal with 3D data

The next generation of archives

Jose Enrique Ruiz del Mazo
IAA - CSIC



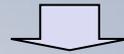
Index

- AMIGA e-Science developments
- Archives Infrastructure
- Web services to the rescue
- Scientific workflows
- Wf4Ever: Preserving digital experiments

The AMIGA Group

Analysis of the interstellar **M**edium of **I**solated **G**alaxies

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



Multi λ study of ~ 1000 galaxies

IAA-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>



The Virtual Observatory

- Interoperability and Discovery
 - Publishing and Accessing Data
 - Service Oriented Architecture (SOA)
 - Integration of Software and Data
 - Distributed resources
 - Panchromatic science
- Data Models
 - Web Services
 - Semantics



The AMIGA VO Catalog

The Data Provider

AMIGA
Instituto de Astrofísica de Andalucía | CSIC

Analysis of the Interstellar Medium of Isolated GALAXIES

- Home
- The Project
- Science
- Technical development
- Team and collaborators
- Results & Ongoing Work
- Publications
- Conferences
- Public Data
- VO Interface
- ASCII Files
- Links

Search

Dpt. Astronomía Extragaláctica
 Instituto Astrofísica Andalucía
 Camino Bajo de Huétor, 50
 18008 Granada
 Spain

VO Interface

[Query by name](#) [Query by parameters](#)

Search by name

Object Name (Ex: CIG 4, UGC 00297, CIG 4%, etc..)

Or/and Input a File (Text file with a name per line)

AMIGA PUBLIC DATABASE SEARCH RESULTS

CIG 155

Alias names (Simbad Name Resolver Service)
 CIG 155

Red image of CIG155, with a size of 05 x 05 arcmin, from DSS2

Basic Data

Coordinates			
RA J2000	DEC J2000	RA B1950	DEC B1950
74.33875	78.1908	72.42666	78.1112
Velocity			
Vr	V3K	Distance	
1556	1507	20.1	
Morphology			
Morph	Morph rc3	Conf morph	Bar Int ?
Scd	6		

Multiwavelength Information

Optical

20	707	478	10.7	Sc (8)															
35	381	179	7.8	Sm (4), IB(s)m (7)															
27	856	750	10.6	SB(s)m pec? (3), Sd (4), SB(s)m pec: (7)															
95	1711	1643	21.9	S0	-2			0.232	5	-3.234	-3.287	-3.234	-3.287						
2.383	10.01	2461	2398	32.0	Sb	3													
4.249	9.28	2515	2451	32.7	Sbc	4		y	?	1.591	5	-2.915	-3.160	-2.929	-3.174				
2.222	9.67	1556	1507	20.1	Scd	6				y	?	0.925	5	-3.196	-3.394	-3.196	-3.394		
3.462	9.72	2821	2819	37.6	Sab	2				y	?	1.181	5	-3.276	-3.408	-3.310	-3.449		
2.435	9.28	1089	1062	14.2	Sb	3													
4.167	8.59	1099	1071	14.3	Sa	1													

Spherical Plot

File
Export
Plot
Rendering
Subsets
Marker Style
Help

Main

Data: Table: 1: AMIGA_VOTable.25-06-08.1214383393.xml

Longitude Axis: RA J2000 degrees

Latitude Axis: DEC J2000 degrees

Radius Axis: Log

Row Subsets

All

Potential: 111 Included: 111 Visible: 111

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,..



TAPAS - IRAM 30M

- Telescope **A**rchive for **P**ublic **A**ccess **S**ystem
- Bolometric observations, maps, spectra
- Rotational molecular transitions
- ~200 scientific projects / year, 1TB

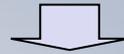
Radio **A**stronomy **D**ATA **M**odel for **S**ingle-dish telescopes

Who are you?

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

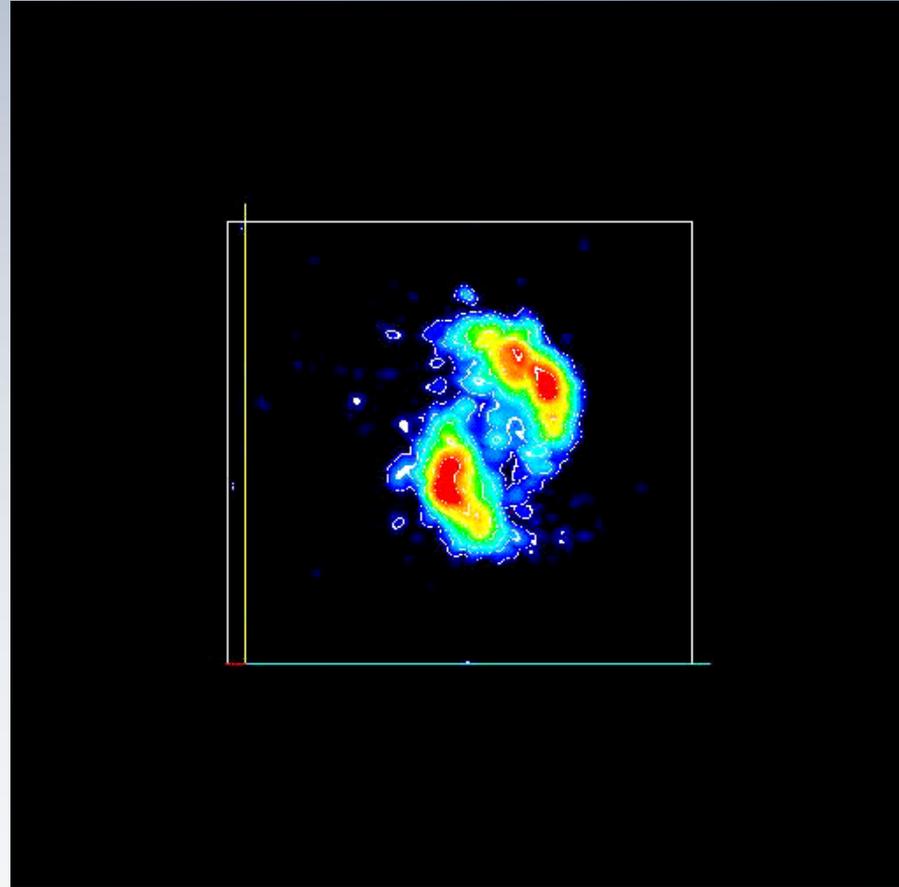


Multi λ study of ~ 1000 galaxies
+

Need of intensive and complex analysis of 3D data
2D spatial + 1 velocity

Who are you?

Velocity Datacubes

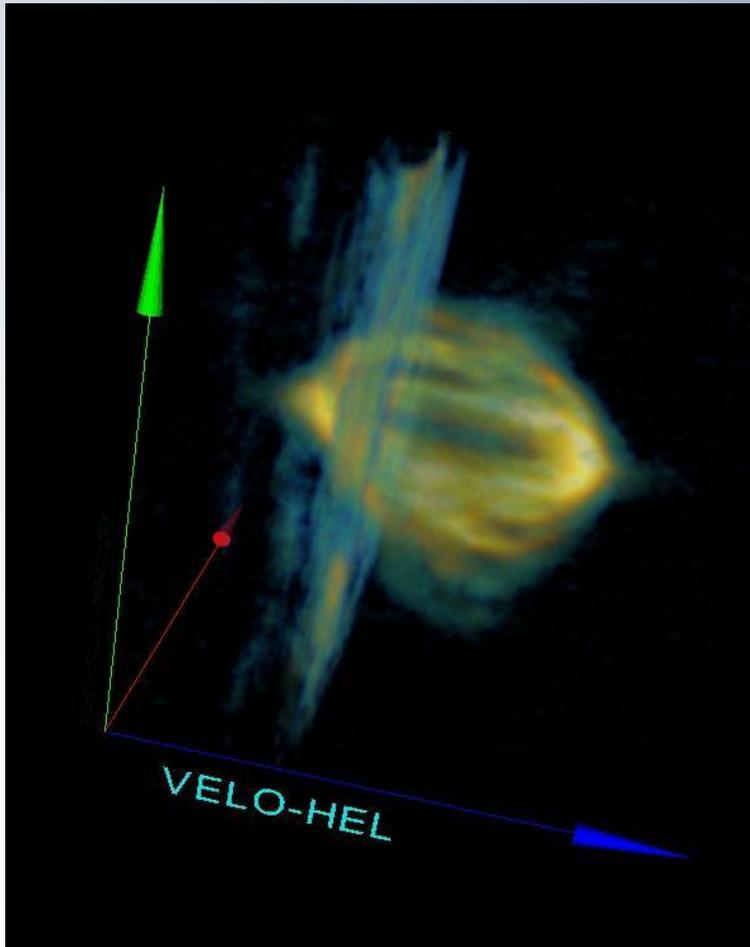


M. Kríps - ESO 3D2008 Workshop - Garching

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 Submillimeter 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

The BODEGA 3D VO Catalog

The Data and *service* provider

BODEGA

Below zero degrees galaxies

About
Team
Publications

Basic data

Target

- Name: **NGC5247**
- Class: **Galaxy**

Coordinates

- RA J2000: **13:38:3.00** hh:mm:ss.ss
- DEC J2000: **-17.88** deg

Velocity

- V: **1319.98** Km/s
- Redshift: **0.00440299**

Extended data

Provenance

- Telescope: **SMA**
- Bandpass: **Millimeter bandwidth**
- Beam Major Axis: **0.000982176** deg
- Beam Minor Axis: **0.000892319** deg
- Beam Position Angle: **-66.64** deg

Spatial

- Aperture angular size (width x height): **0.025 x 0.025** deg
- Spatial bin size (width x height): **8.3e-05 x 8.3e-05** deg

Spectral

- Spectral coord value: **1319.98** Km/s
- Width of spectrum: **999986.33** Km/s
- Start in spectral coordinate: **579.99** Km/s
- Stop in spectral coordinate: **1579.98** Km/s

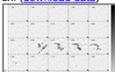
Flux

- Flux min : **0.0175** Jy/Beam
- Flux Support Extent (max): **0.7208** Jy/Beam
- Flux Support Extent (min): **0.0175** Jy/Beam

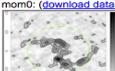
Images

12CO21

ch: [\(download data\)](#)



mom0: [\(download data\)](#)



mom1: [\(download data\)](#)



sp: [\(download data\)](#)



diatrad: [\(download data\)](#)



Download Fits file



Right click and "Save Link As" to download

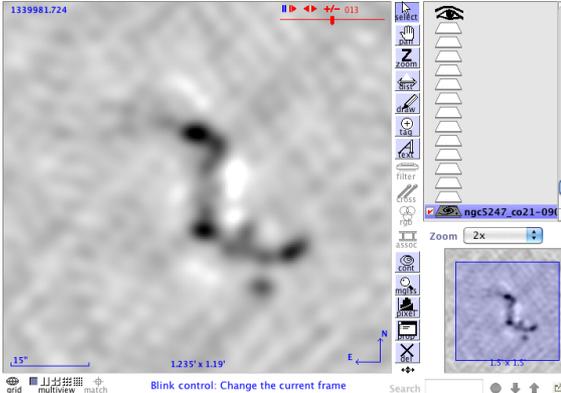
Number of points: 2250000
 Size: 9011520 Kbs
[Open this with Aladdin Applet](#)





File Edit Image Catalog Overlay Tool View Interop Help
Install

Location:



Aladin VO Software

©1999-2009 Uds/CNRS - Centre de Données astronomiques de Strasbourg 0 sel / 0 src 17Mb

Virtual Data

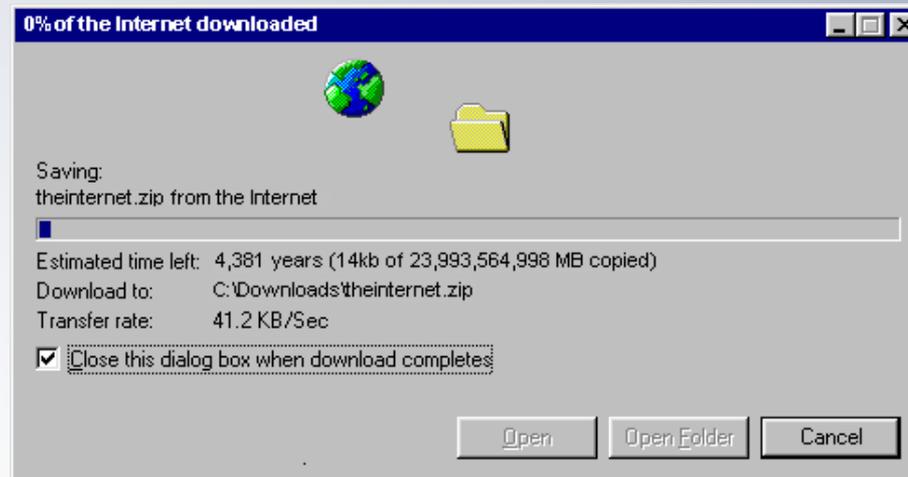
The next generation of archives

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**



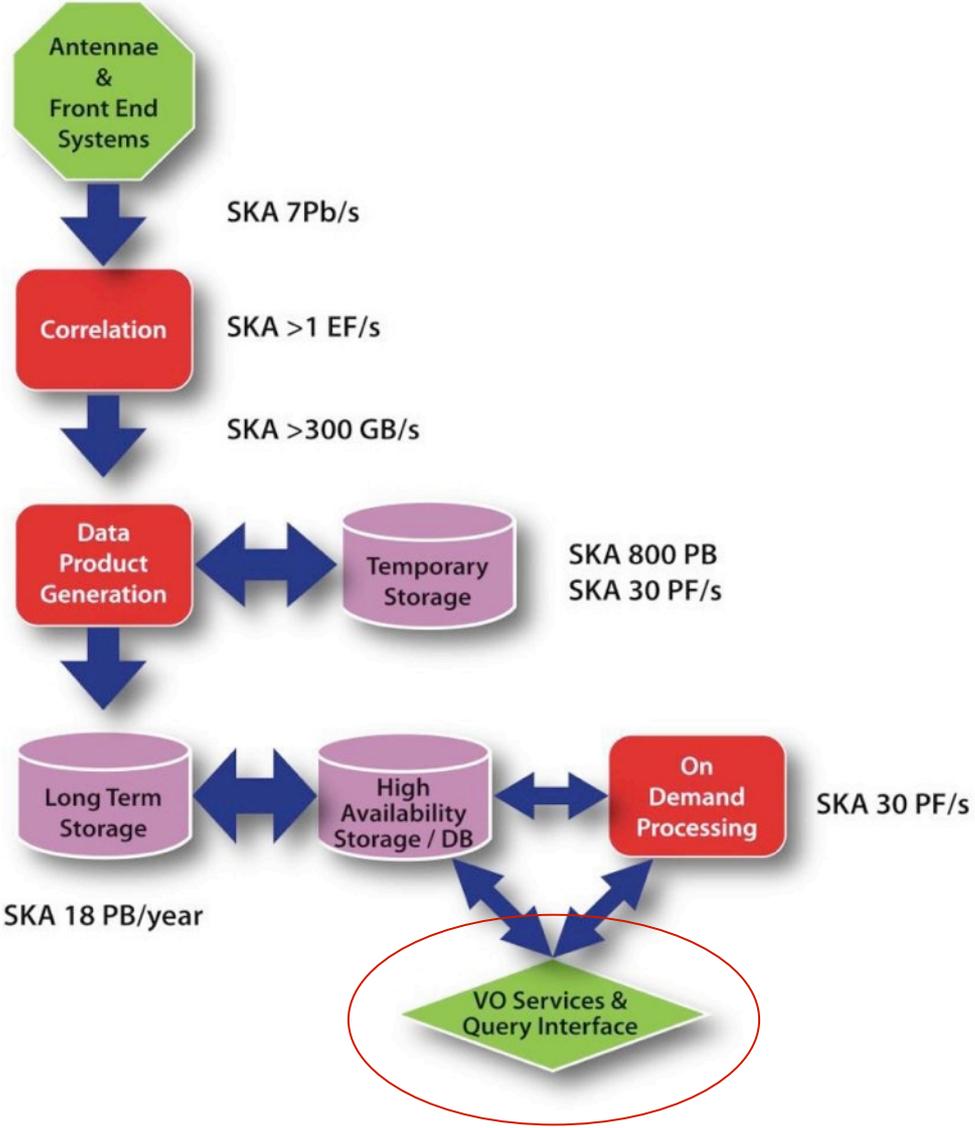
Cube sizes

	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

Infrastructure



Processing



Storage



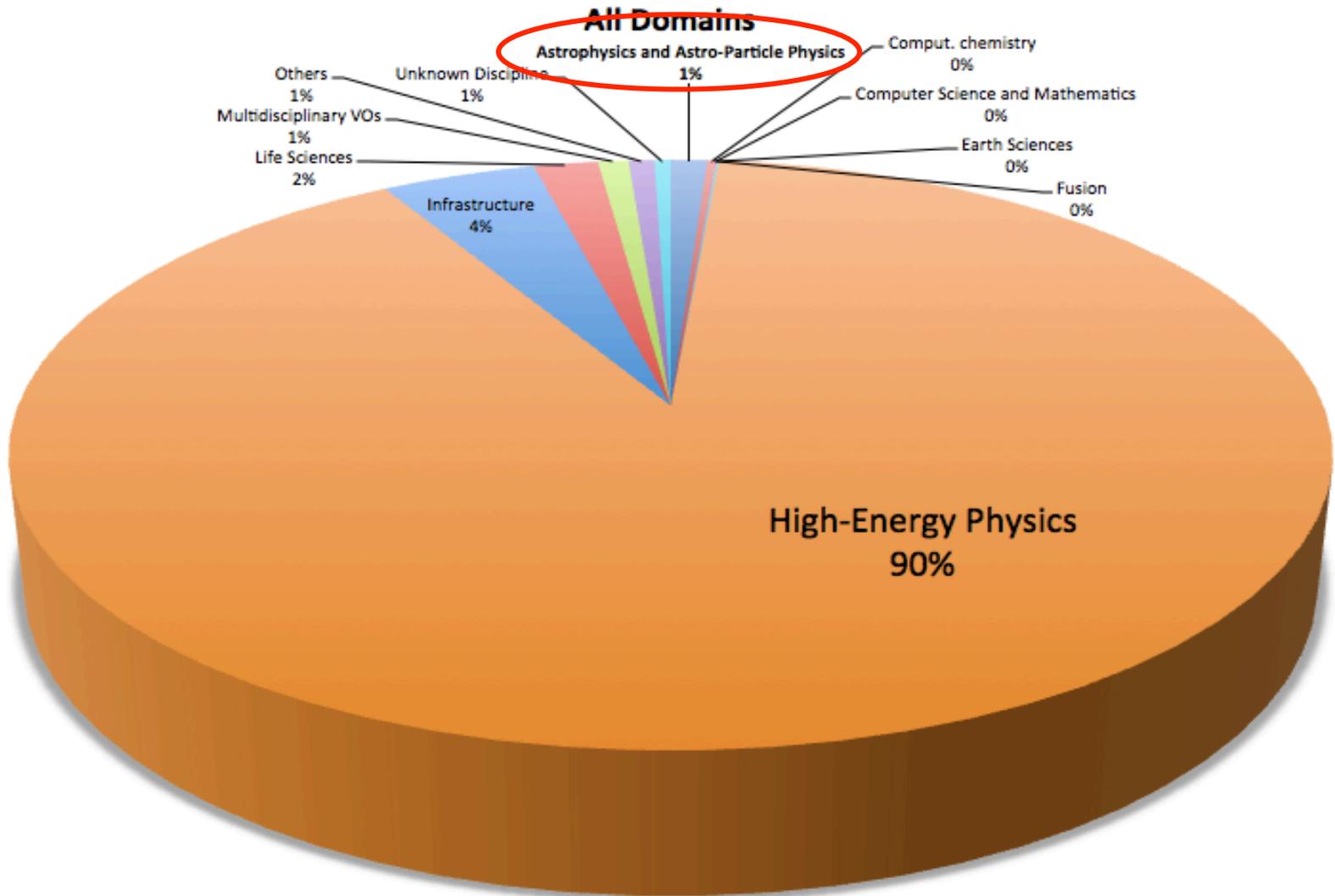
Bandwidth



Power

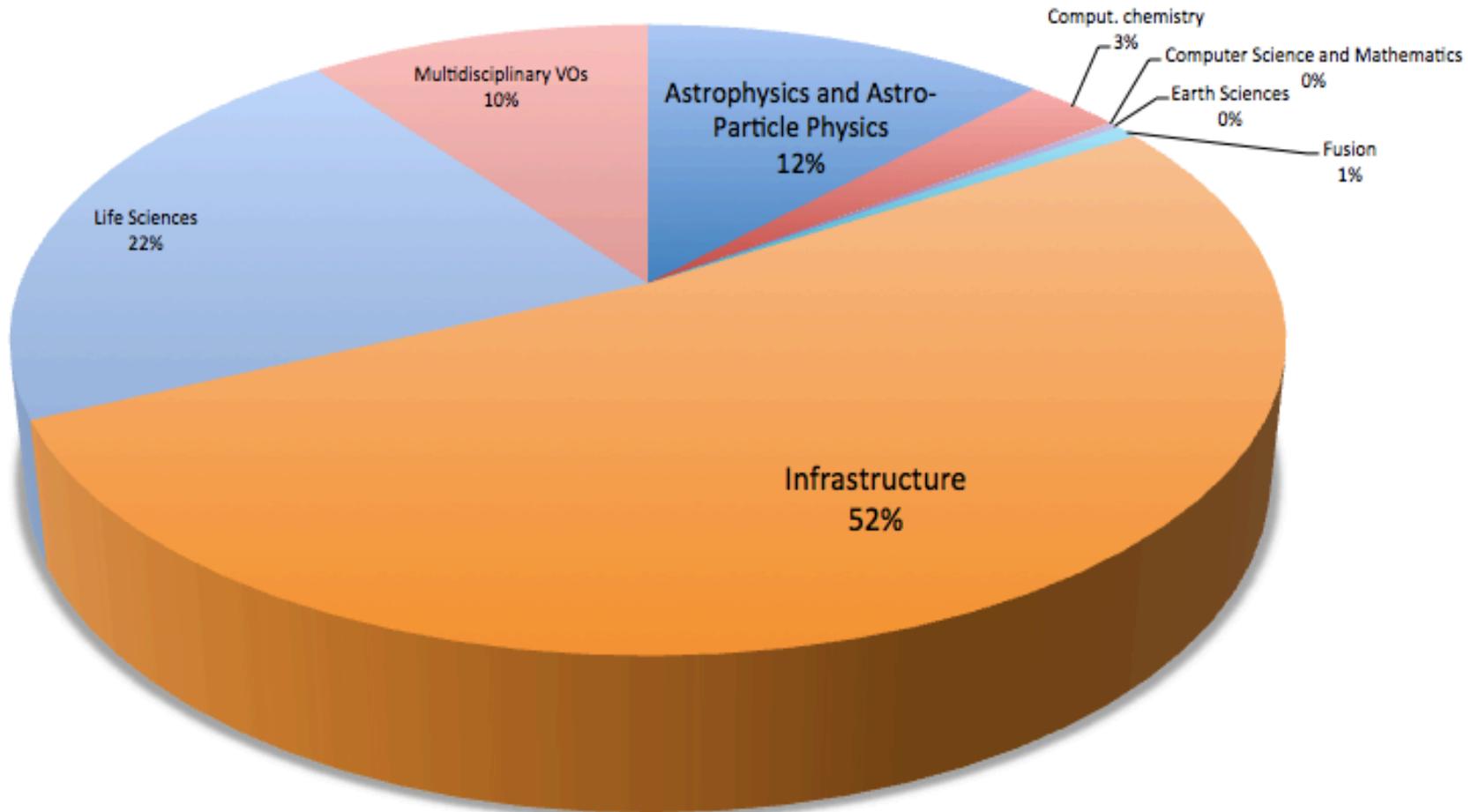


Use of the European Grid Infrastructure May 2010 - May 2011



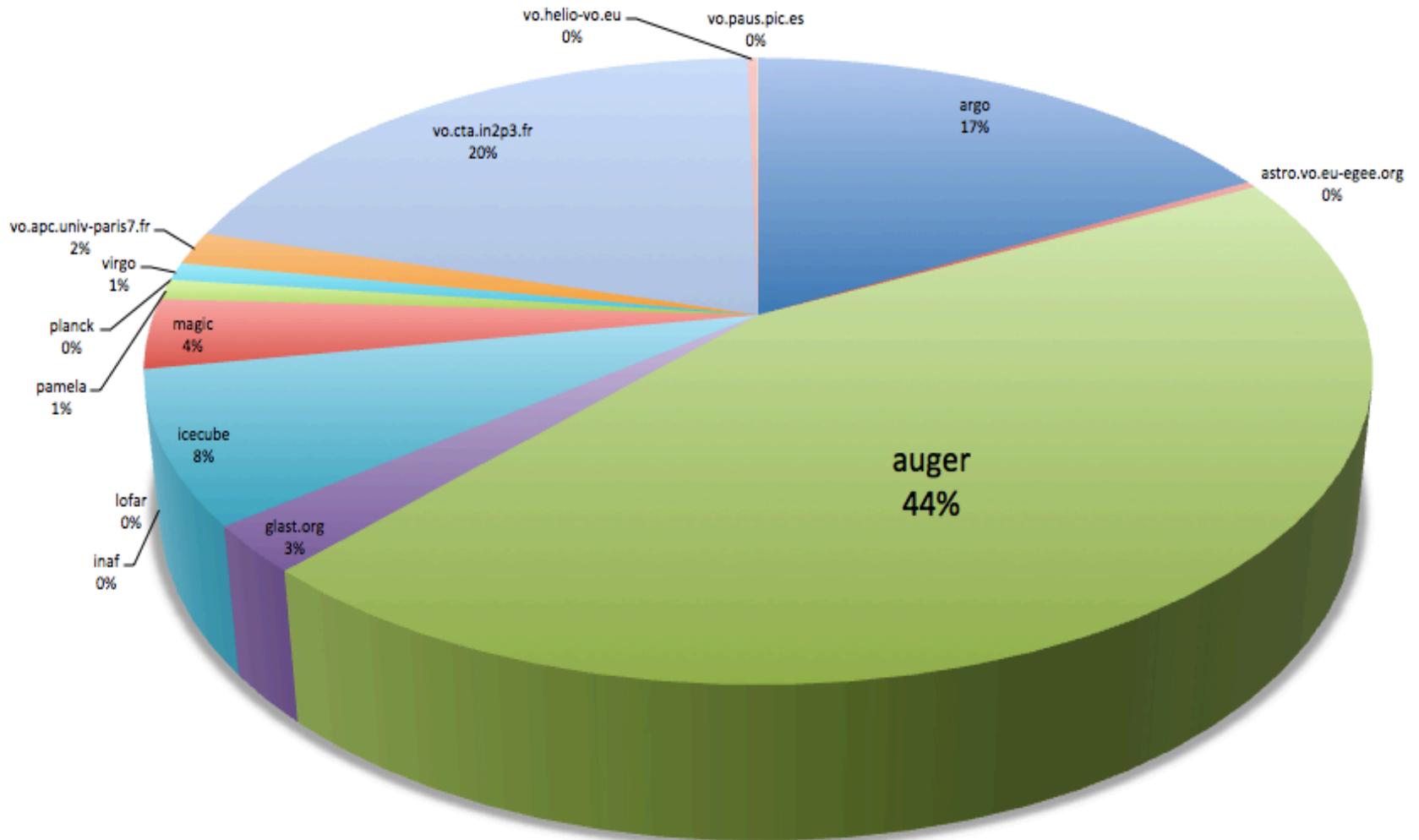
Use of the European Grid Infrastructure May 2010 – May 2011

High Energy Physics Excluded



Use of the European Grid Infrastructure May 2010 - May 2011

Astronomy & Astrophysics



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

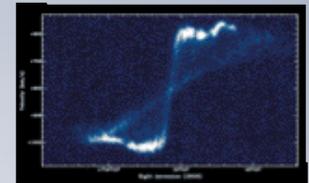
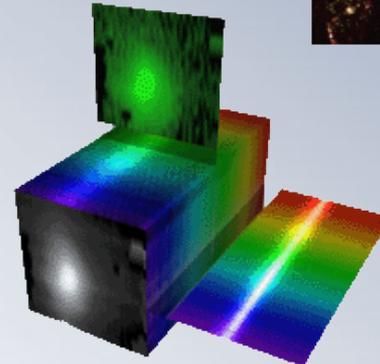
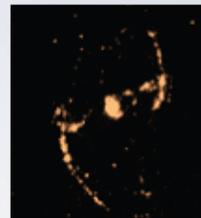
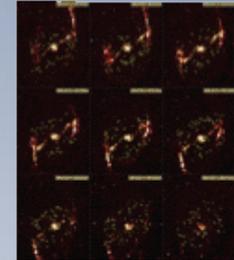
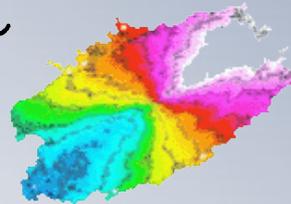
- Interconnected and interoperable archives
- Distributed, multi-wavelength and multi-facilities

Archives speaking **Web Services**

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

3D Data Services

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



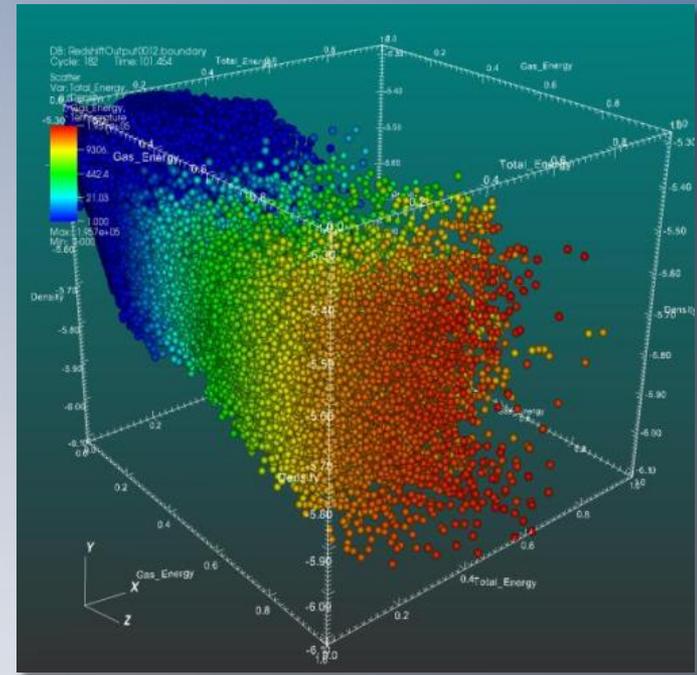
Exploration services

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

Extraction of scientifically relevant information from a multidimensional parameter space.



visit software

Scientific Workflows

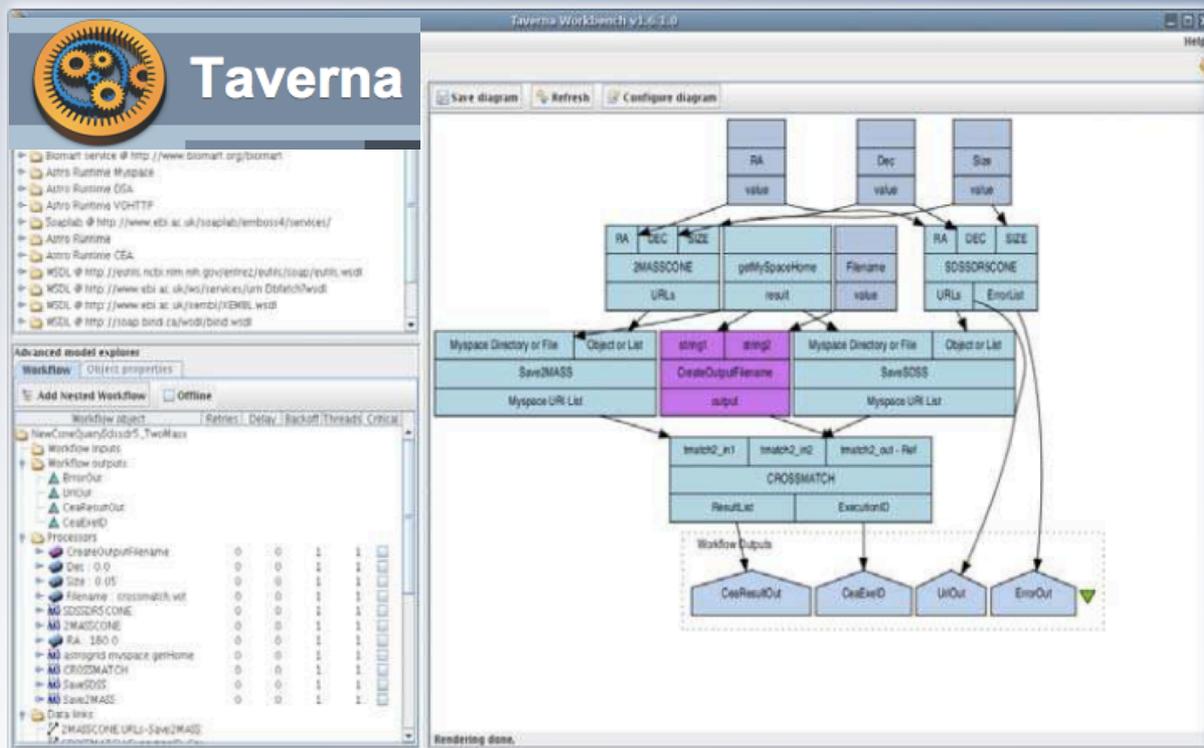
Web-services-based users oriented workflows **ARE NOT PIPELINES**

- Archived science-ready data
- Automation
- Repeatable results
- Reproducible
- Encourage best practices
- Modular nature allows
 - Re-use
 - Re-purpose
- Expose the scientific method
- Formative



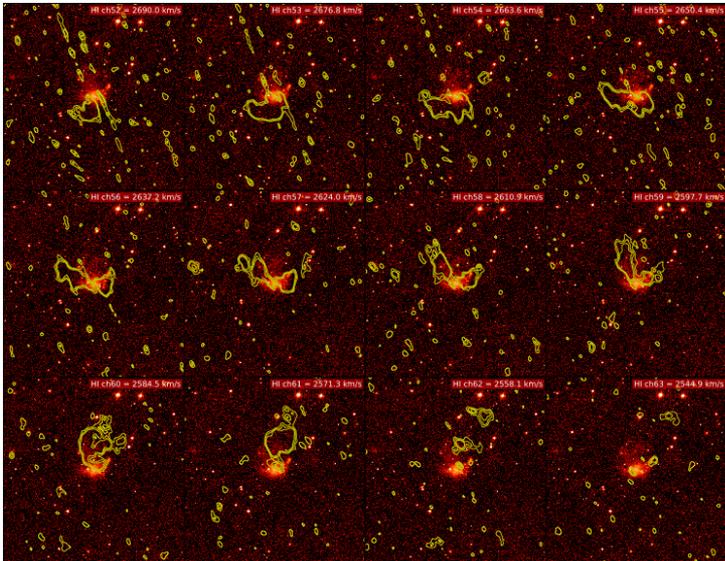
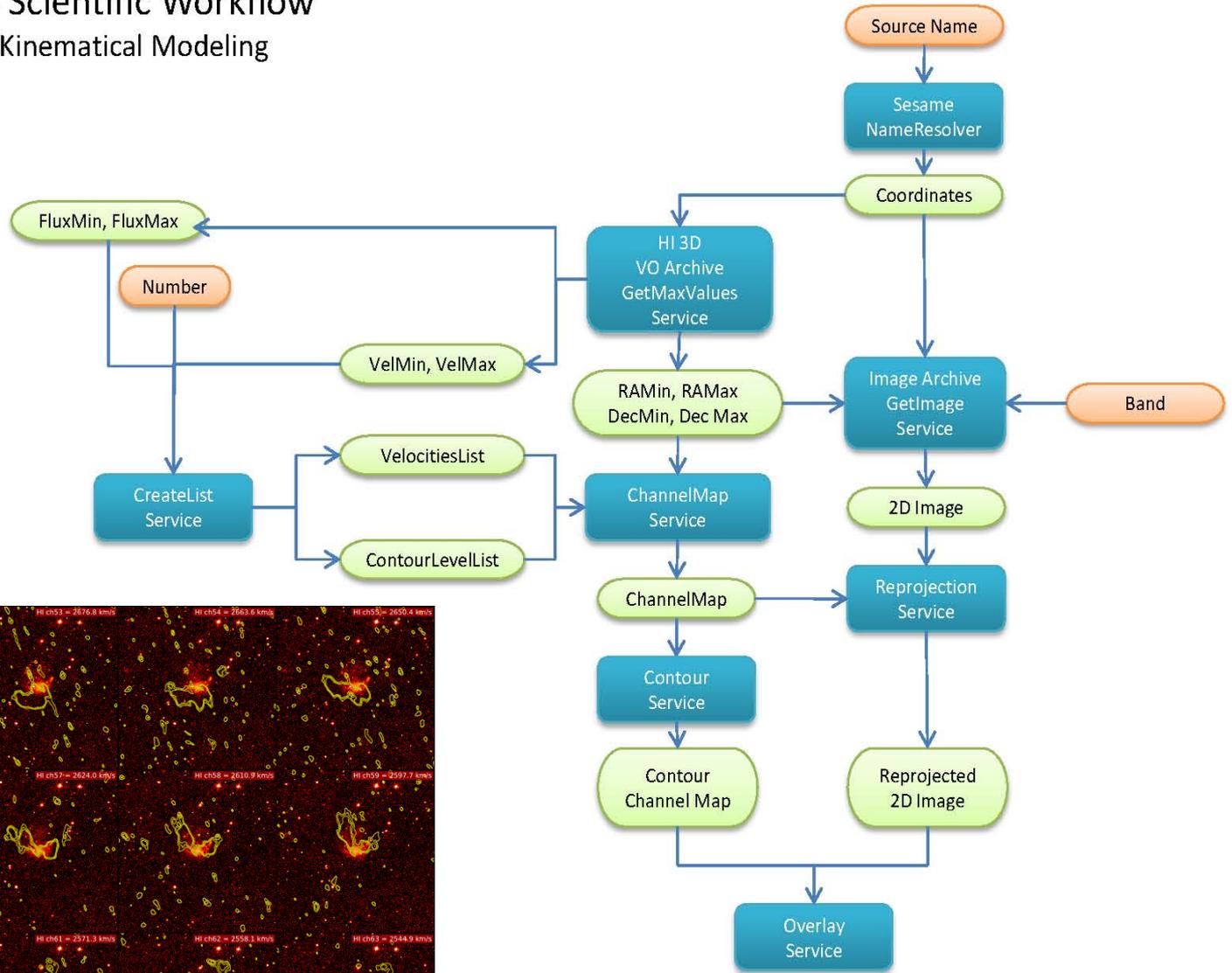
Taverna

A workflow enactment and management system



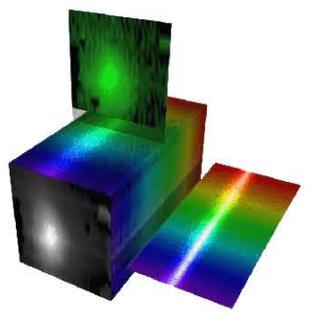
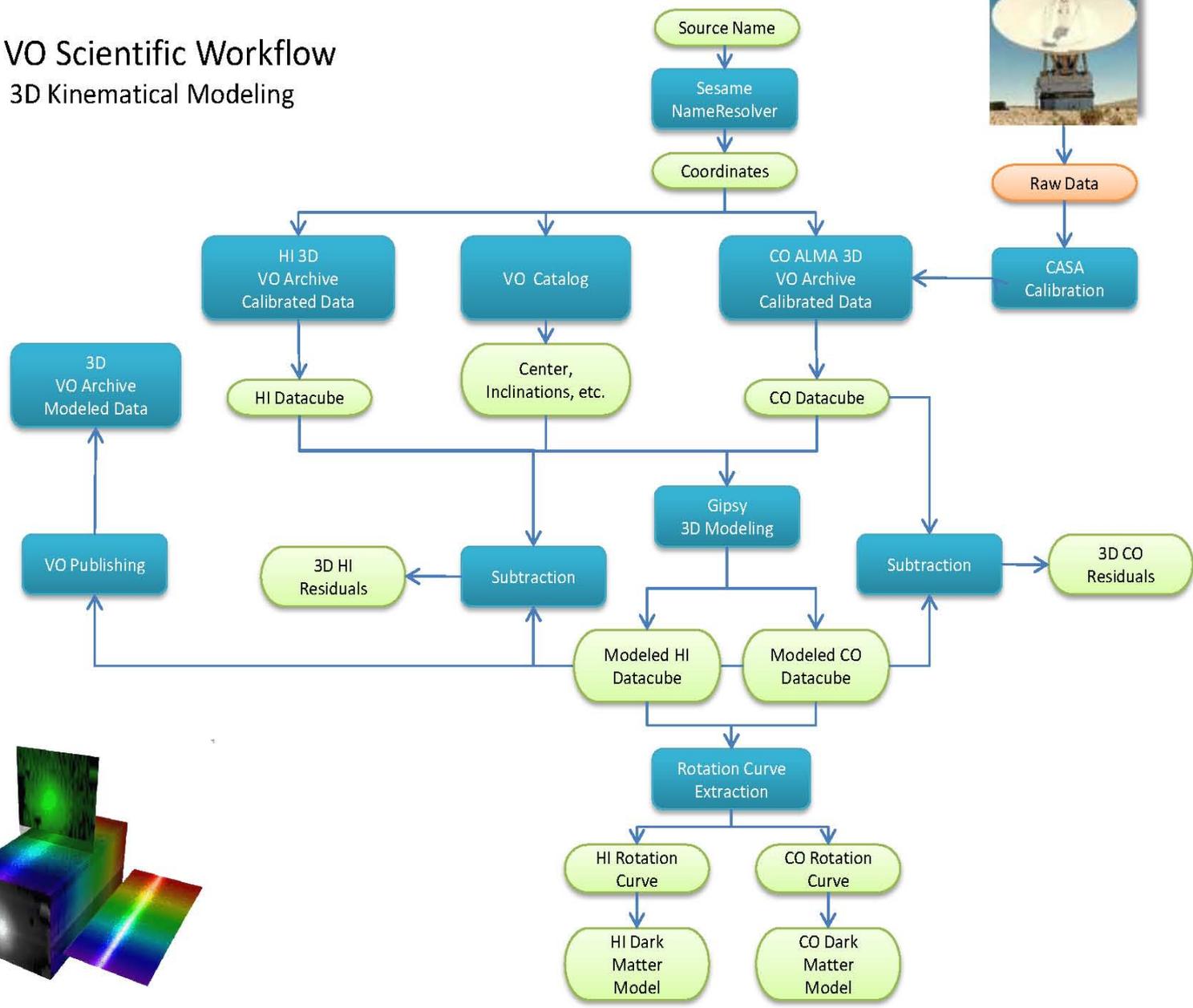
University of Manchester

VO Scientific Workflow 3D Kinematical Modeling



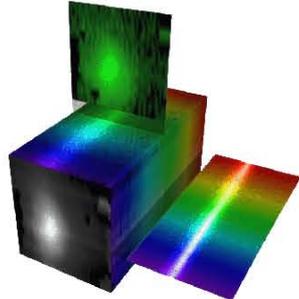
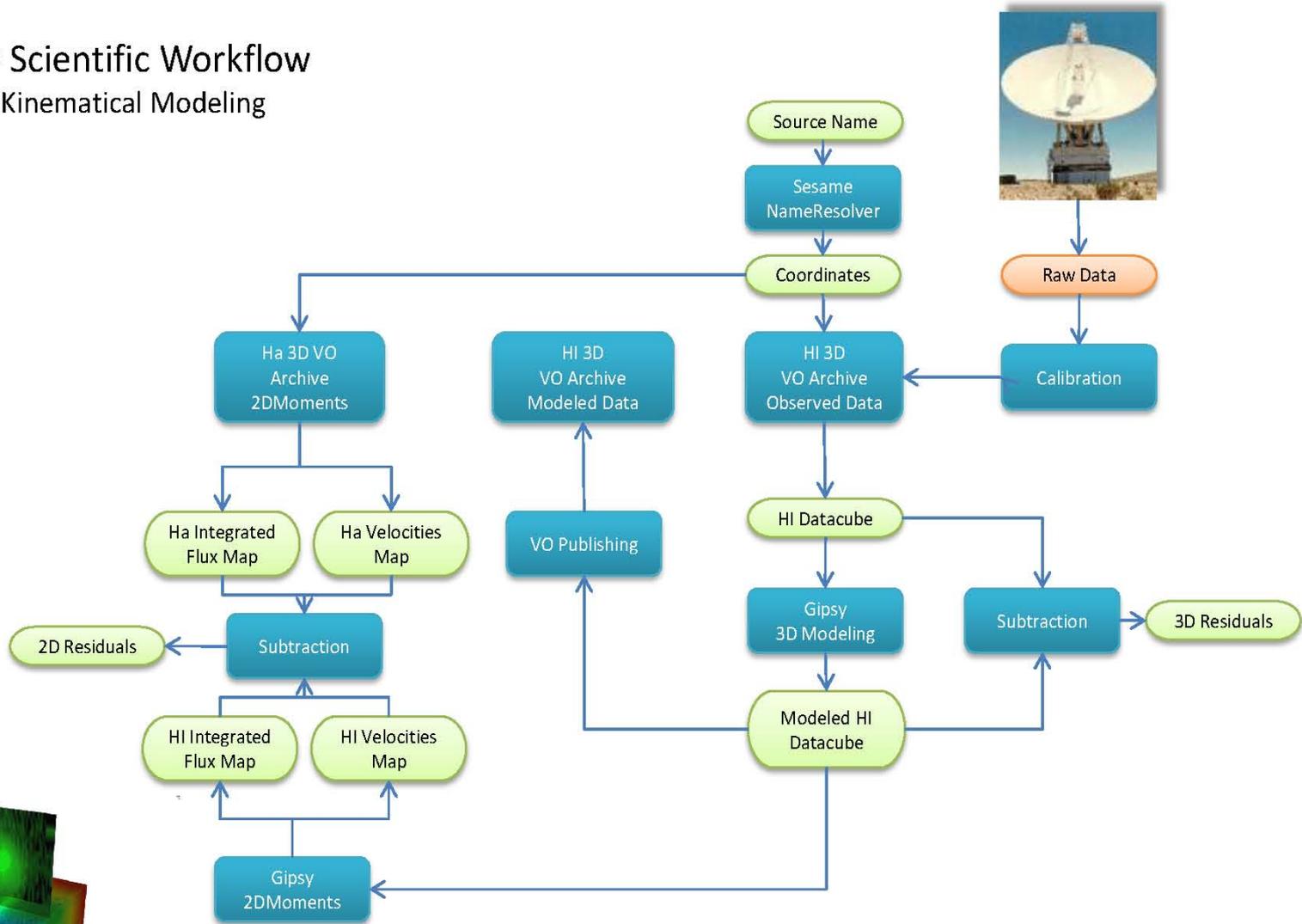
VO Scientific Workflow

3D Kinematical Modeling



VO Scientific Workflow

3D Kinematical Modeling



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. University of Oxford (OXF, UK)
6. Instituto Astrofísica Andalucía (IAA-CSIC, Spain)
7. Leiden University Medical Centre (LUMC, NL)

All components related to the research lifecycle
available, preserved and easily retrievable.

- Proposals
- Data
- *Processes*
- *Workflows*
- Publications

RESEARCH OBJECT



<http://wf4ever-project.org/>

The recipes store

- Find workflows
- Share workflows and files
- Find people
- Build communities
- Publish packages
- Tag workflows
- Score and rate workflows
- Comment on workflows
- Write reviews
- Access rights

The screenshot shows the 'myexperiment' website interface. The main content area displays a list of workflows under the heading 'Workflows'. The first workflow is 'Mapping OligoNucleotides to an assembly (v7)' by Wassinki, created on 13/02/09. It includes a description, a license (Creative Commons Attribution-Share Alike 3.0), and a rating of 0.0/5. The second workflow is 'Add Mesh String to Biological Process (v2)' by Paul Fisher, created on 03/10/07. It also includes a description, a license (Creative Commons Attribution-No Derivative Works 3.0), and a rating of 0.0/5. The interface includes search filters, a search bar, and a sidebar with 'Filter by type' and 'Filter by user' options.

Oxford university

Astronomy research is entirely digital Time has come to go "Beyond the PDF"

- Preserved experiments
- Methodology "in action"
- All data exposed
- Reproducible
- Repeatable
- Re-usable
- Re-purposeable
- Collaborative
- Formative



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.*

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

Thanks !

