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 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

IAA-CSIC
Uuiv . Granada, Obs. Marseille, Obs. Paris, NAOJ, FCRAO, UNAM, Univ. Edinburgh, IRAM, ESO, Kapteyn Astronomical Institute.

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

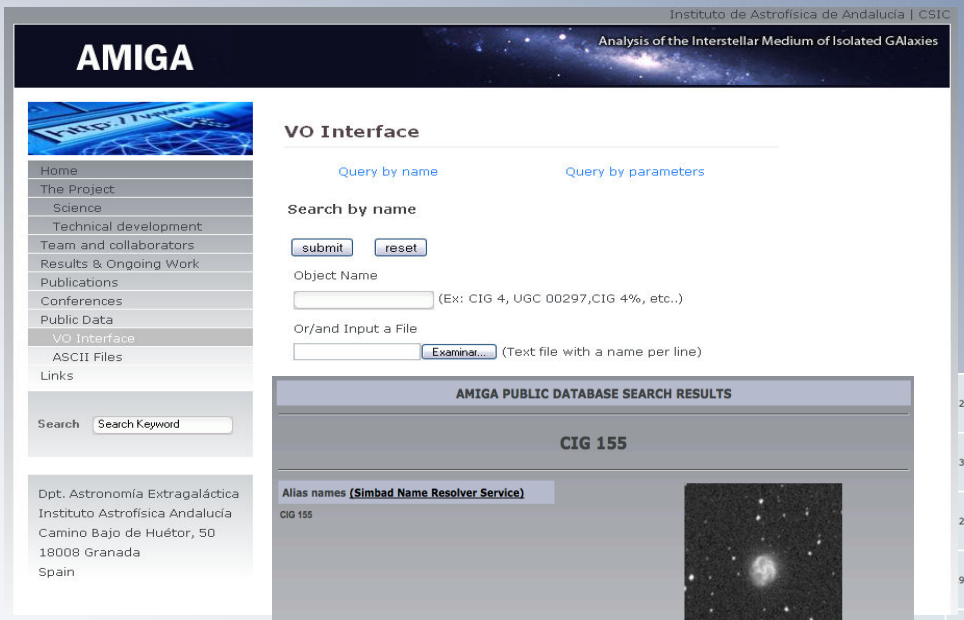
The Virtual Observatory

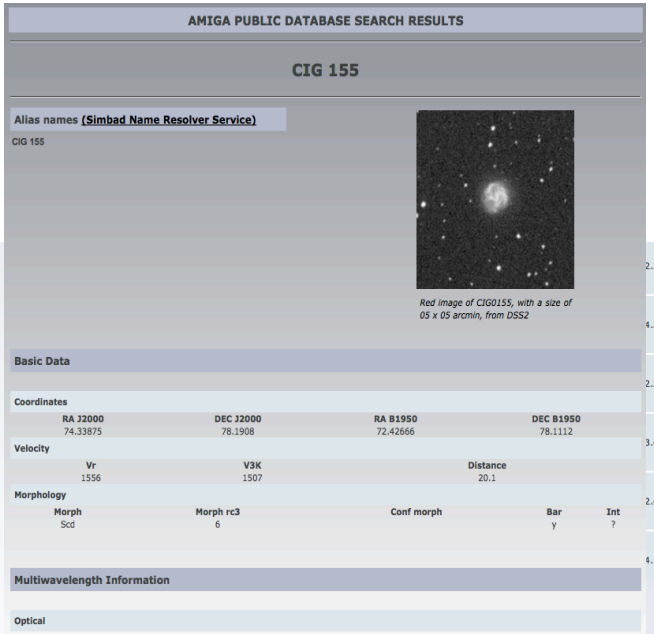
- Interoperability and Discovery
- Publishing and Accessing Data
- Service Oriented Architecture (SoA)
- Integration of Software and Data
- Distributed resources
- Panchromatic science

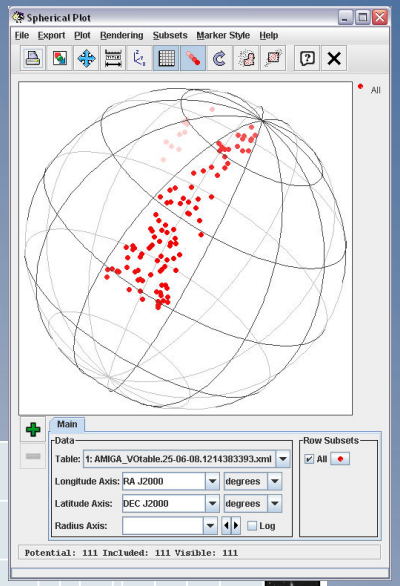
- Data Models
- Web Services
- Semantics

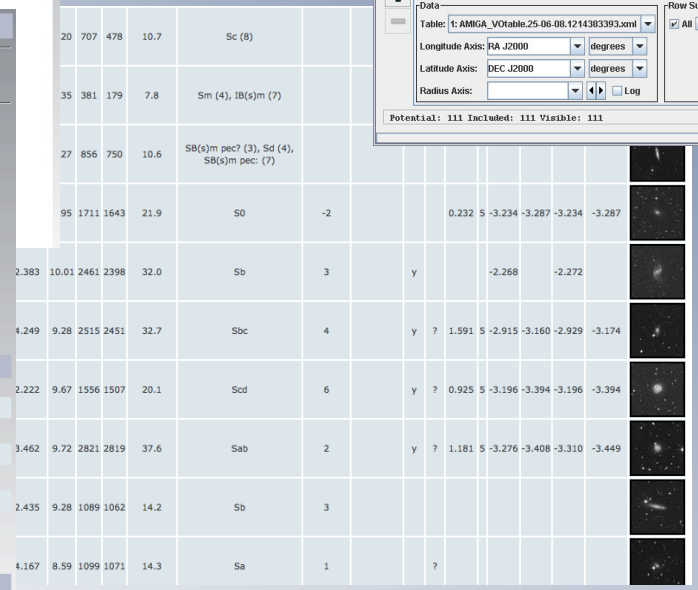
The AMIGA VO Catalog

The Data Provider









Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

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 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

The AMIGA Group

Analysis of the interstellar Medium of Isolated Galaxies

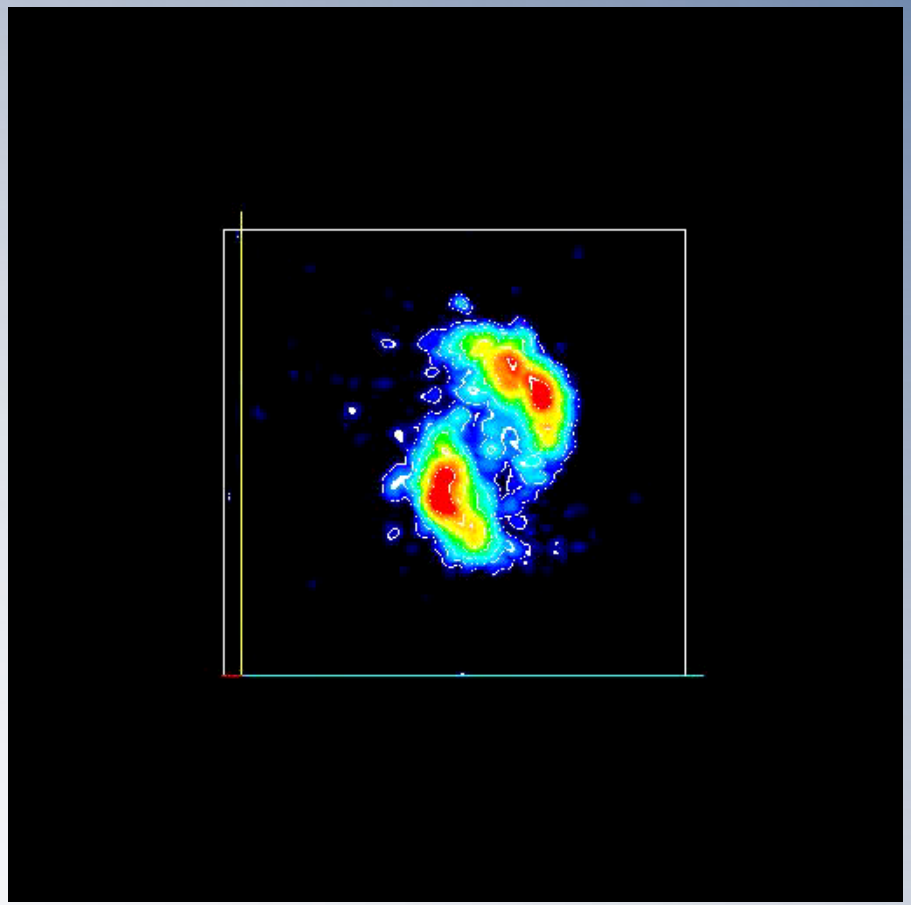
Statistical baseline of isolated galaxies to compare with the behaviour of galaxies in denser environments

Multi λ study of ~1000 galaxies

+

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

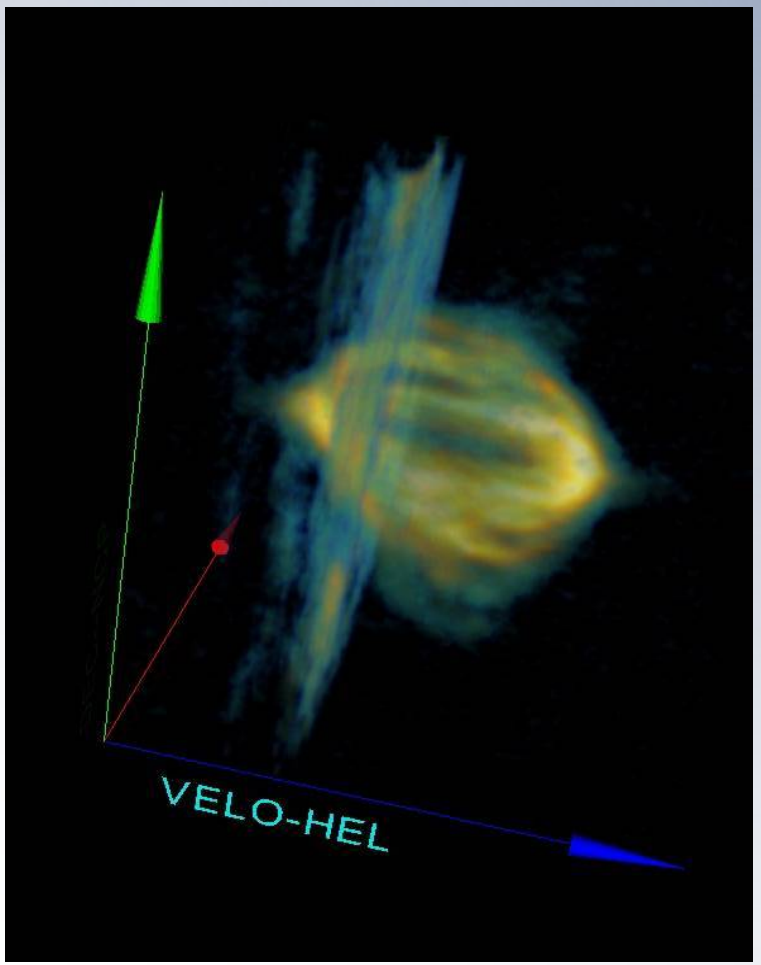
Velocity Datacubes



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

The BODEGA 3D VO Catalog

The Data and Service provider

Virtual Data

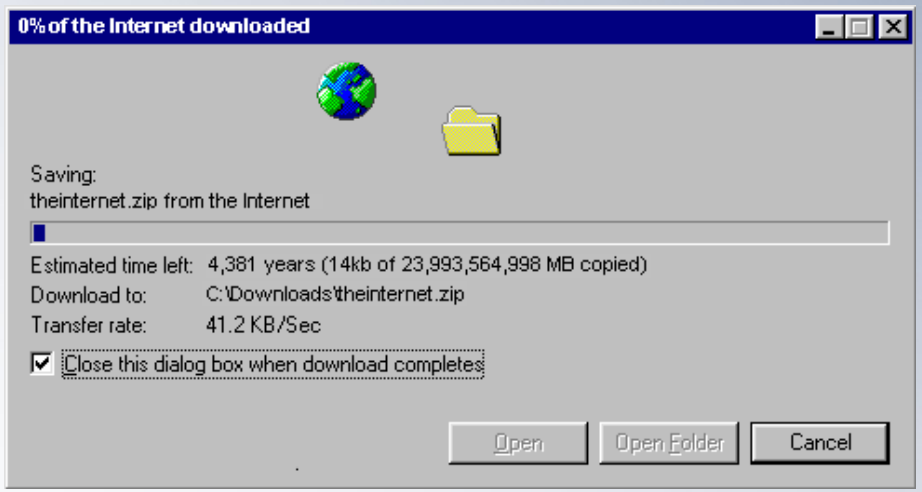
The next generation of archives

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



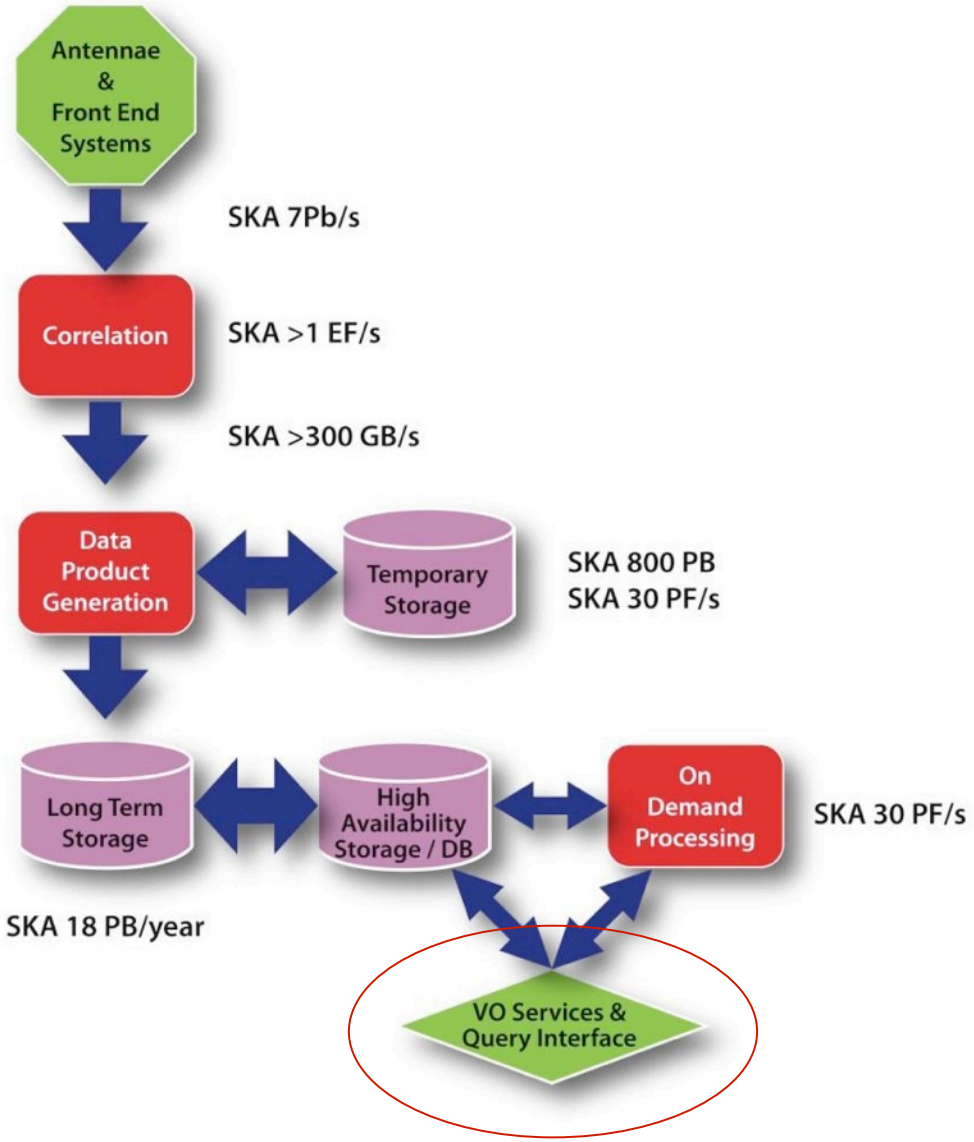
Cube sizes

	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

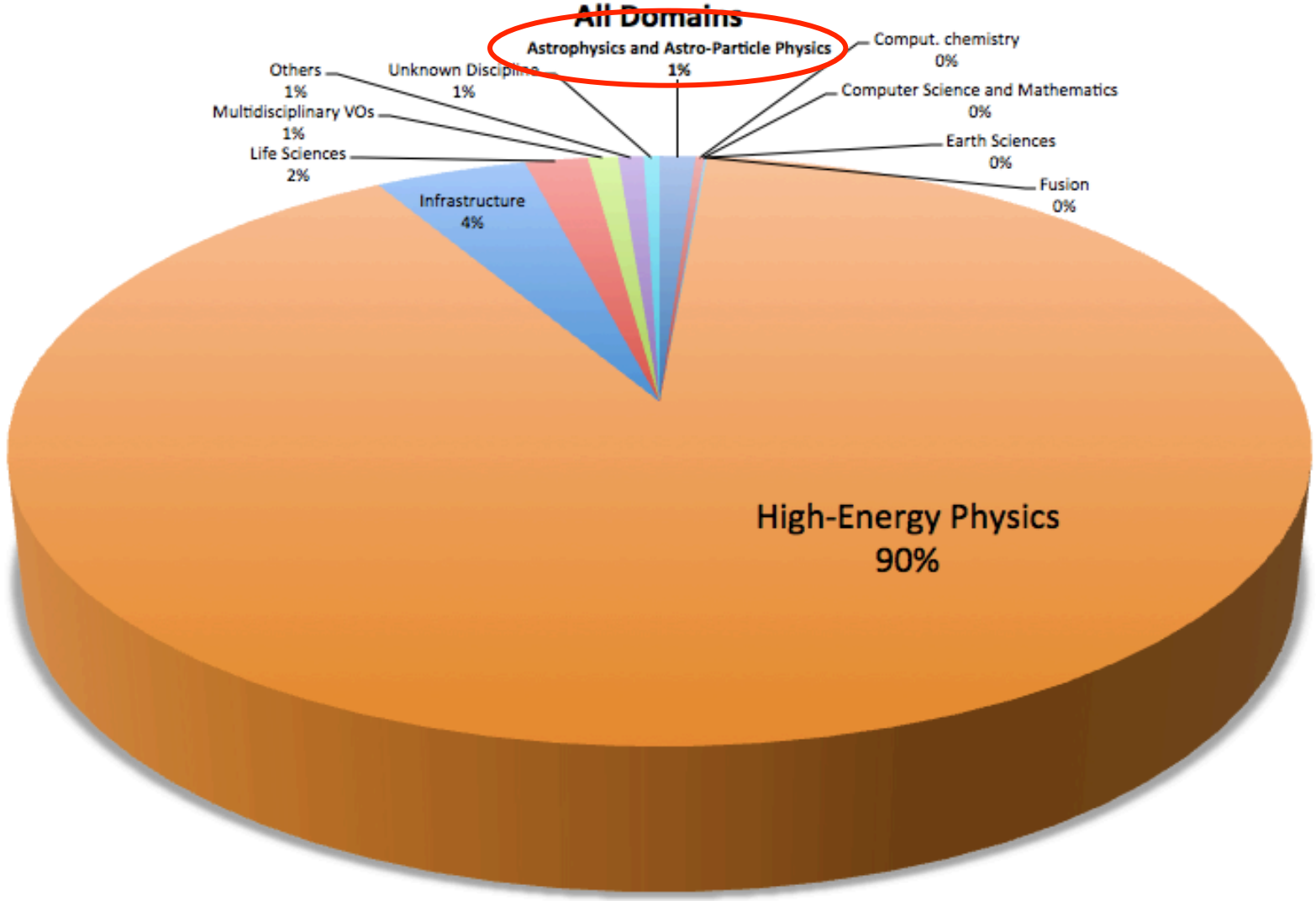
Storage

Bandwidth

Power

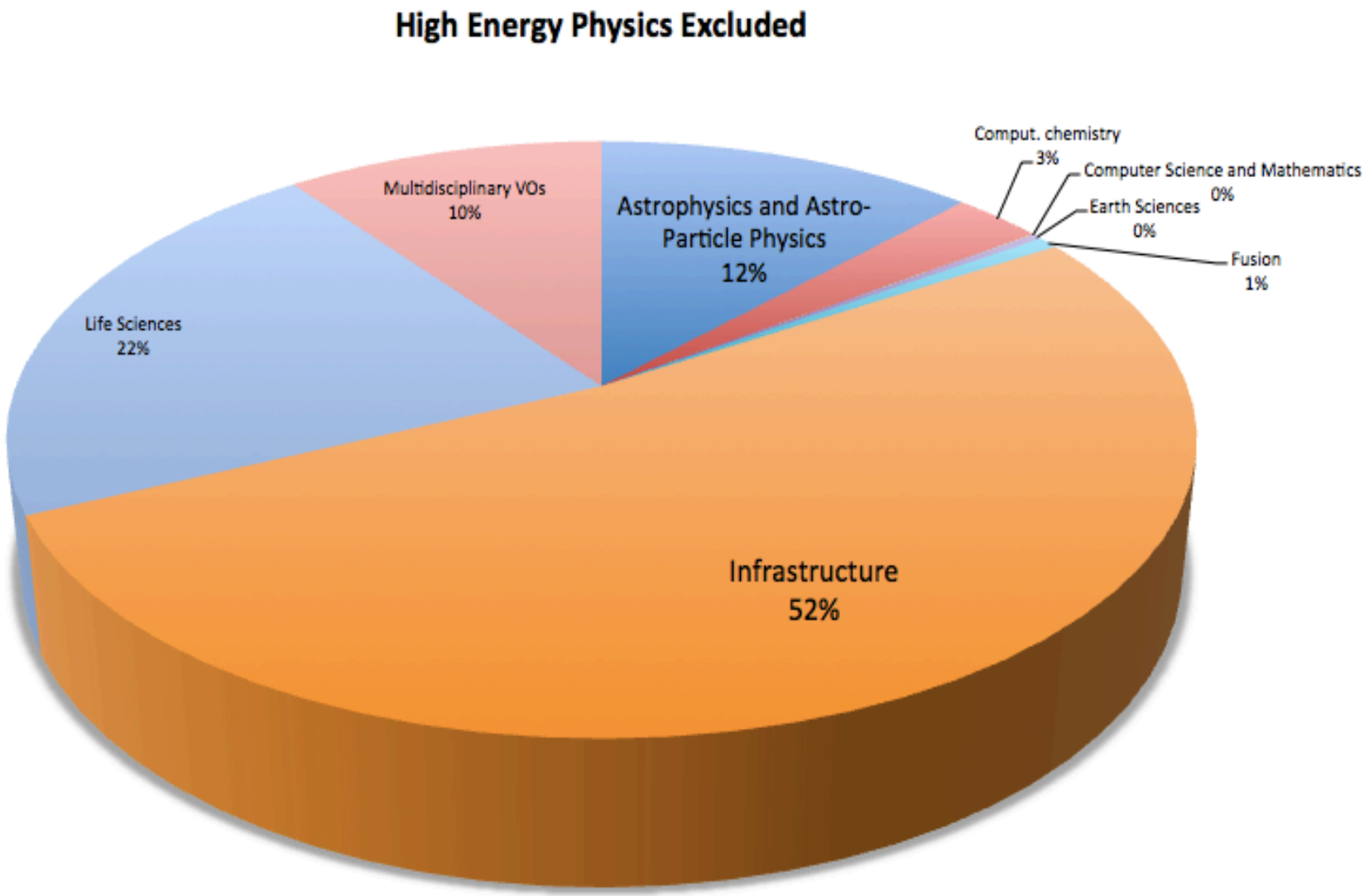
Infrastructure

Use of the European Grid Infrastructure May 2010 – May 2011



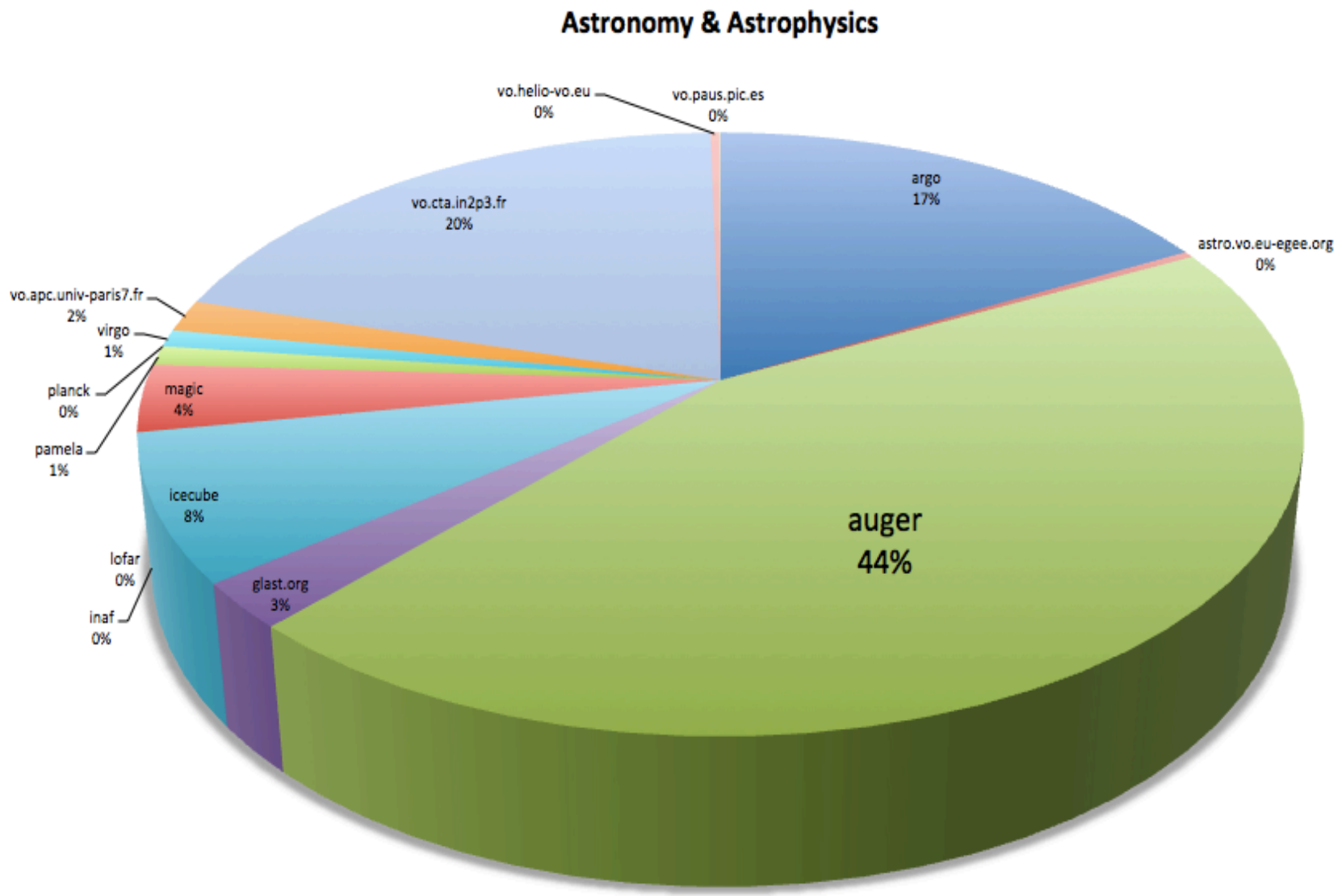
Infrastructure

Use of the European Grid Infrastructure May 2010 – May 2011



Infrastructure

Use of the European Grid Infrastructure May 2010 – May 2011



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

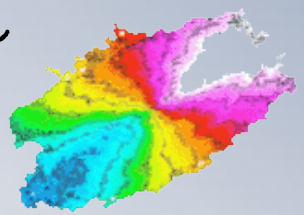
- Interconnected and interoperable archives
- Distributed, multi-wavelength and multi-facilities

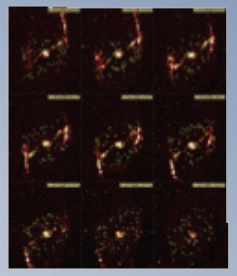
Archives speaking Web Services

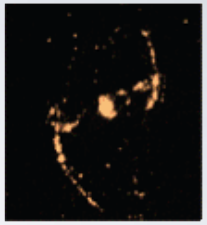
ALMA, LSST, ASKAP, MeerKAT, LOFAR, Apertif,...

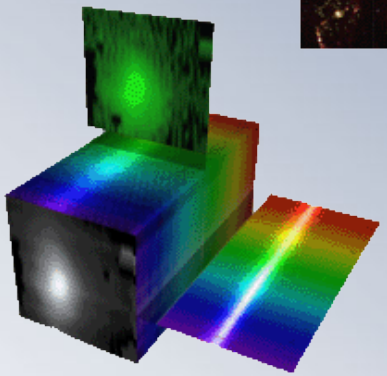
3D Data Services

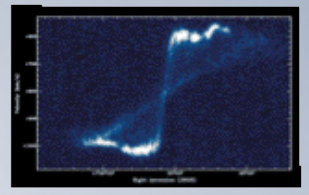
- Cutout
- Resample
- Spectrum extraction
- 2D slice extraction
- Dimensional reduction
- Filtering/Flagging
- 2D Moments
- Complex transformations









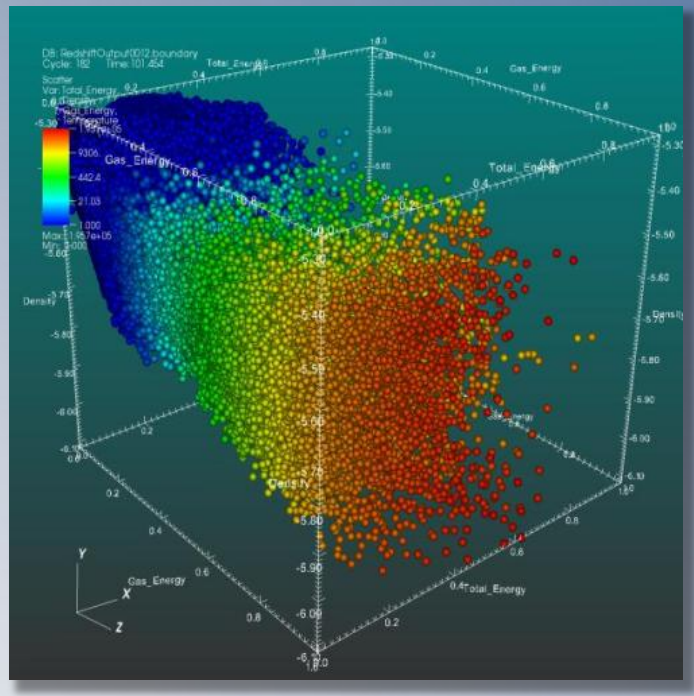


Exploration services

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



VisIt software

Extraction of scientifically relevant information from a multidimensional parameter space.

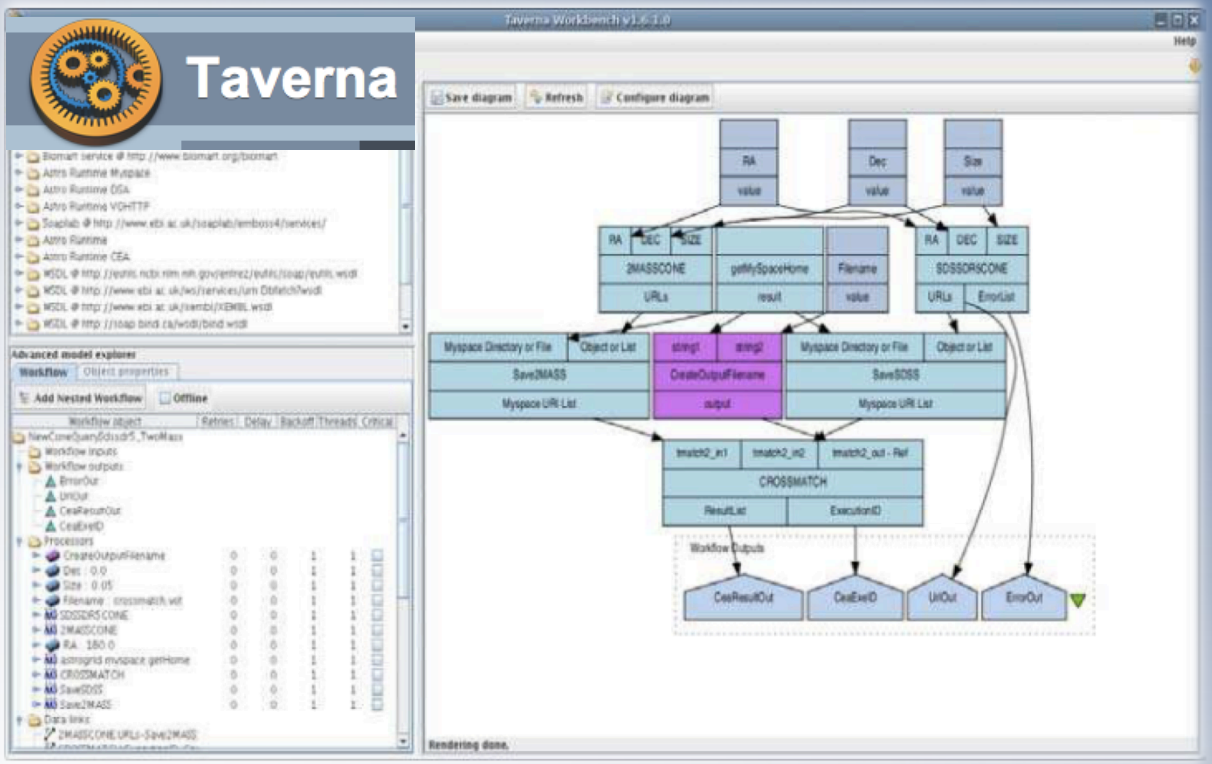
Scientific Workflows

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



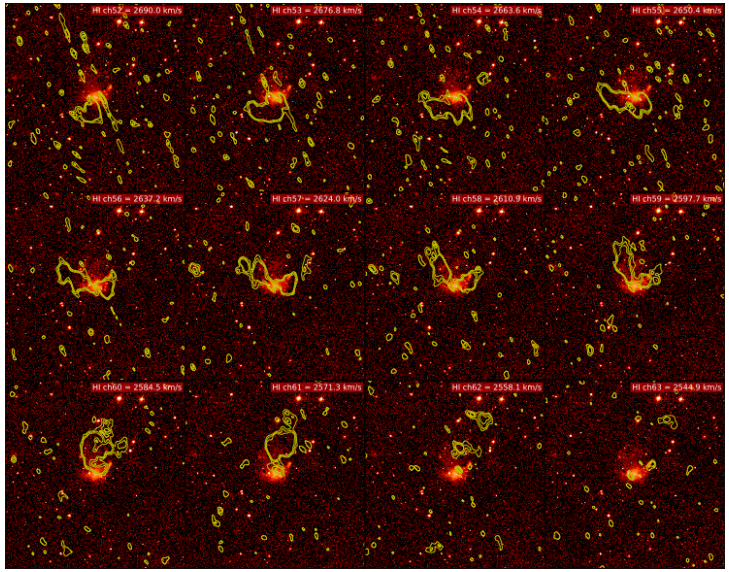
University of Manchester

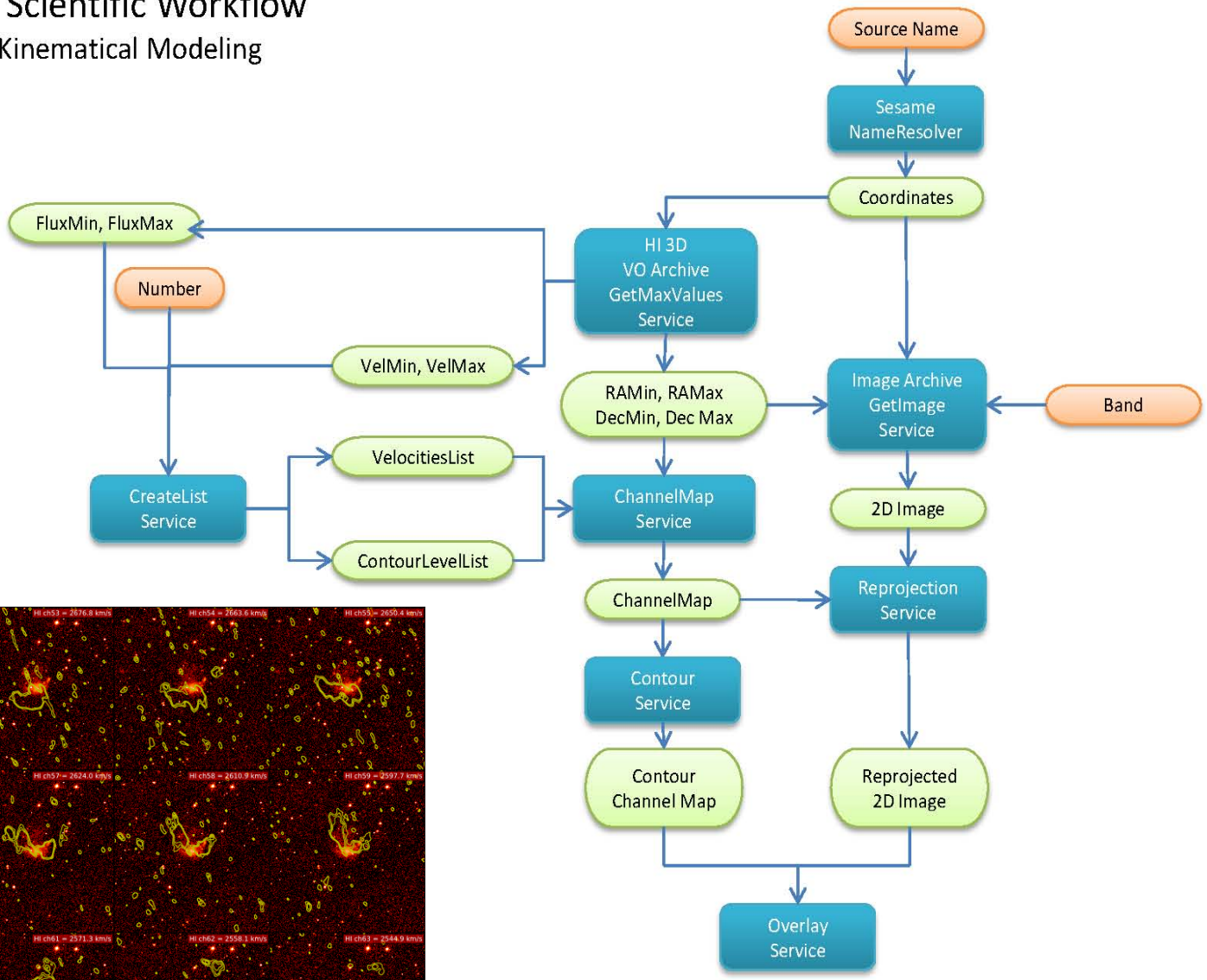
Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

VO Scientific Workflow

3D Kinematical Modeling



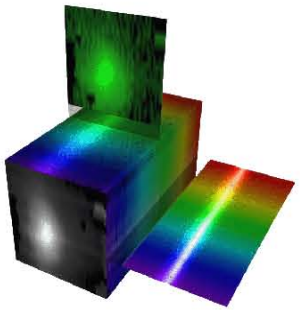


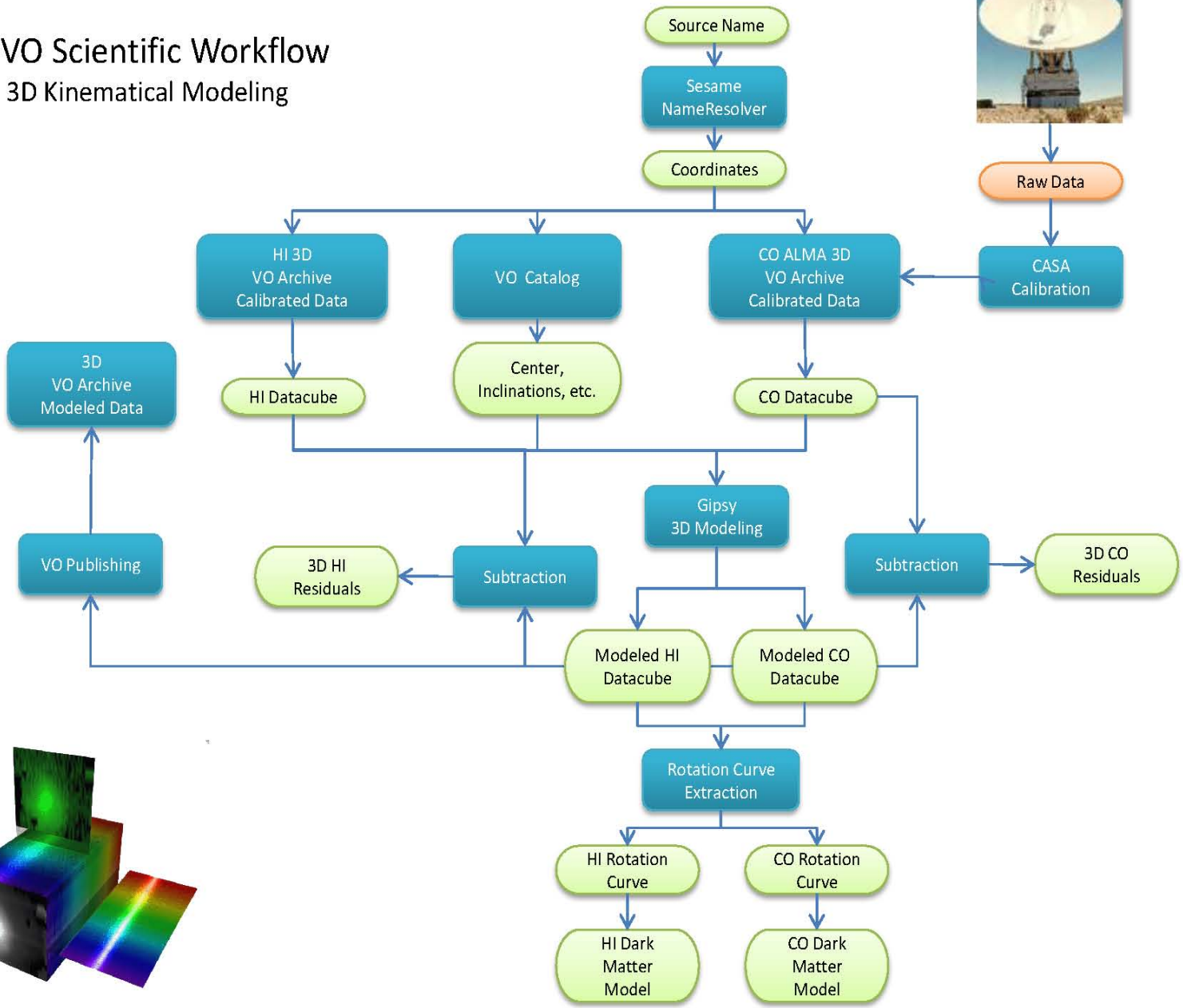
Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

VO Scientific Workflow

3D Kinematical Modeling



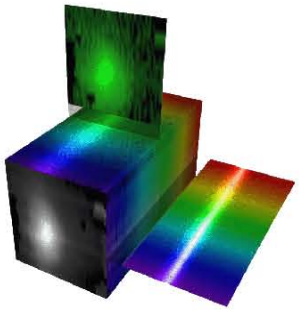


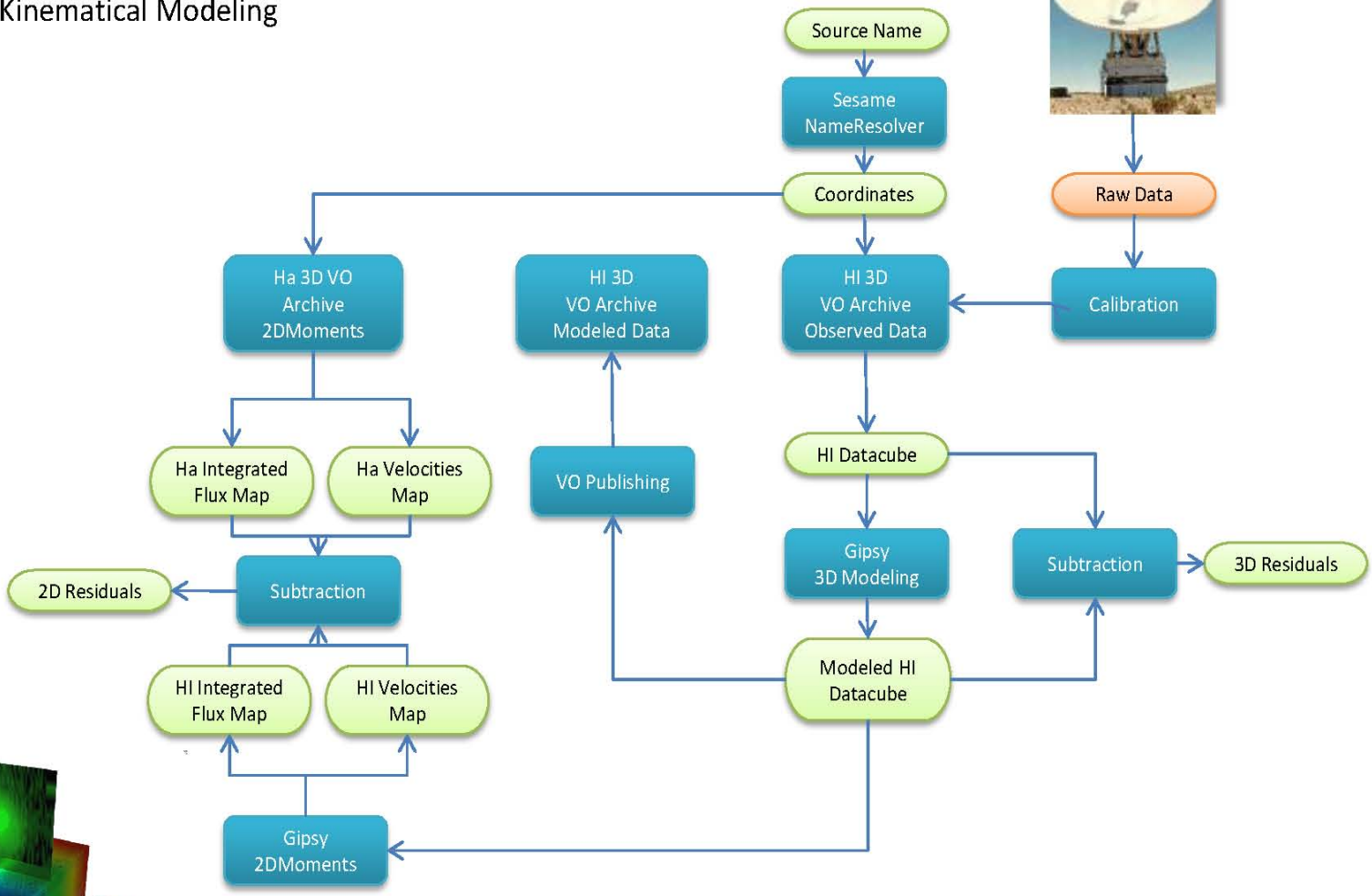
Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

VO Scientific Workflow

3D Kinematical Modeling





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 Astrofisica Andalucia (IAA-CSIC, Spain)
7. Leiden University Medical Centre (LUMC, NL)

MANCHESTER 1824

The University of Manchester

SUPERCOMPUTING POZNAN NETWORKING

UNIVERSITY OF OXFORD

LUMC

INSTITUTO de ASTROFISICA de ANDALUCIA CSIC I·A·A

iSOCO enabling the networked economy

SEVENTH FRAMEWORK PROGRAMME

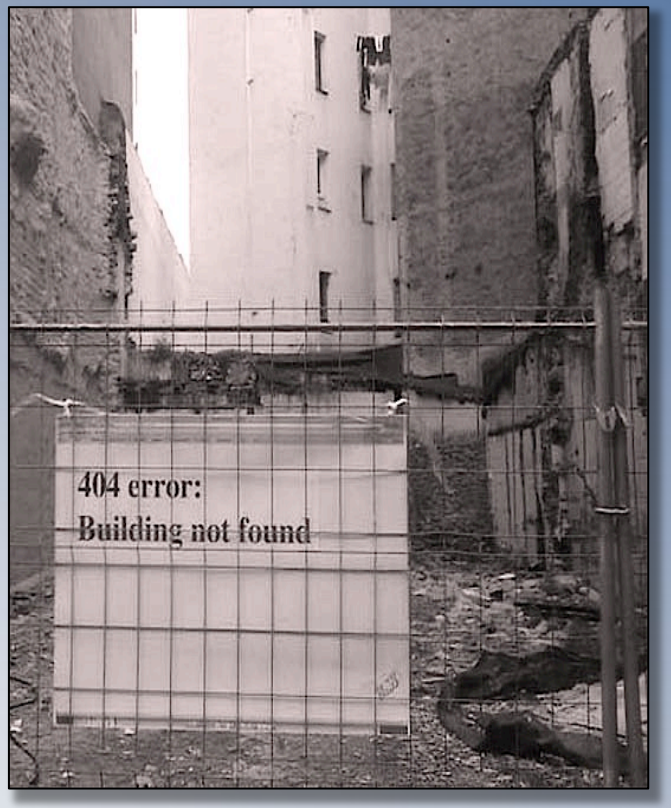
European Commission Information Society and Media

Wf4Ever

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

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.

Conclusions

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 !