

# Workflows and Infrastructures in the Exascale era

S. Sánchez Expósito, J.E. Ruiz, L. Verdes-Montenegro, J.D. Santander-Vela, A. Ruiz-Falcó, R. M. Badia, R. Sirvent, J.M. Martínez

6ª Reunión Plenaria de la Red Española de e-Ciencia, 16 Mayo 2012

# AMIGA

Analysis of the interstellar  
Medium of Isolated GALaxies

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

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

- Statistical baseline of isolated galaxies to compare with the behavior of galaxies in denser environments
- Multi- $\lambda$  study of  $\sim 1000$  galaxies

## AMIGA4GAS: AMIGA for GTC, ALMA y SKA pathfinders

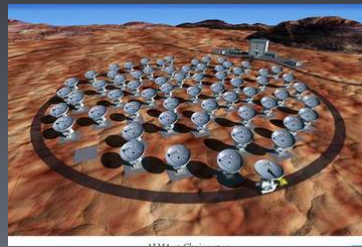
- 4<sup>th</sup> AMIGA project for 2012 – 2014 (we are starting!)
- Coordinated with Fundación Centro de Supercomputación Castilla y León (FCSL)
- A deeper and more detailed study of isolated galaxies using state of the art facilities that generate high data volumes
- Challenge: The efficient exploitation of the very high data volumes

# About the instruments

GTC: Gran Telescopio Canarias 10.4 m



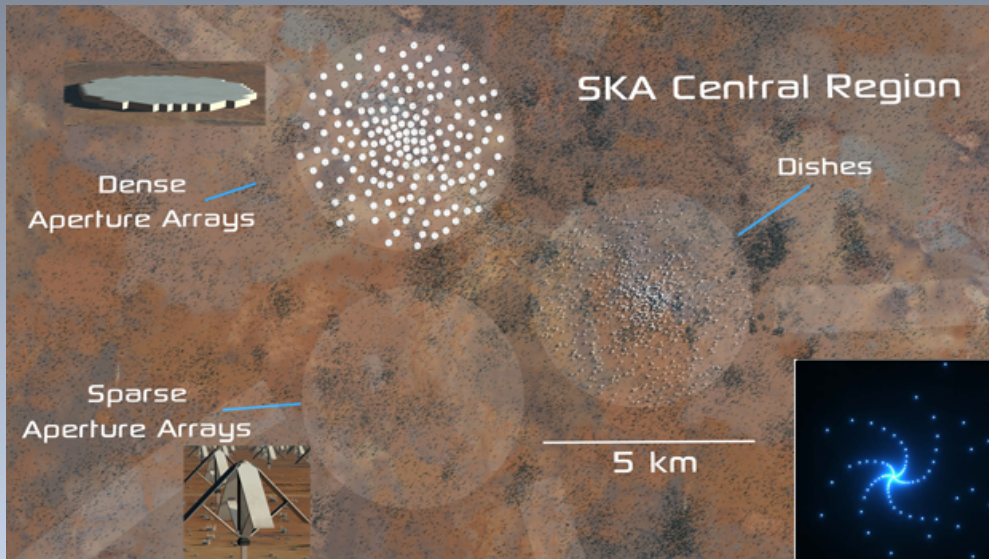
ALMA: Atacama Large (sub)Millimeter Array. (66 Antennas)



EVLA: Expanded Very Large Array (27 Antennas)

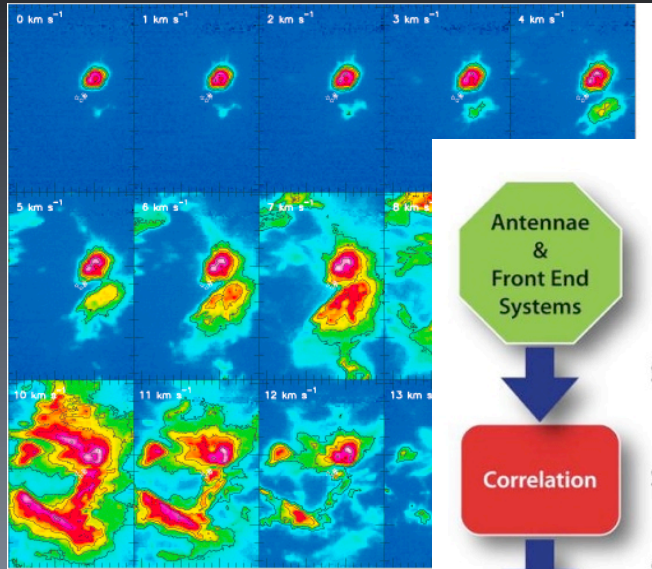
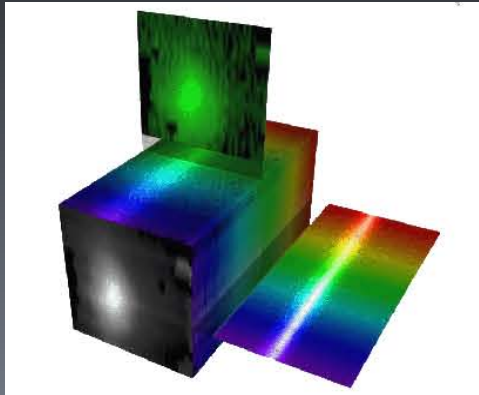


SKA: Square Kilometre Array



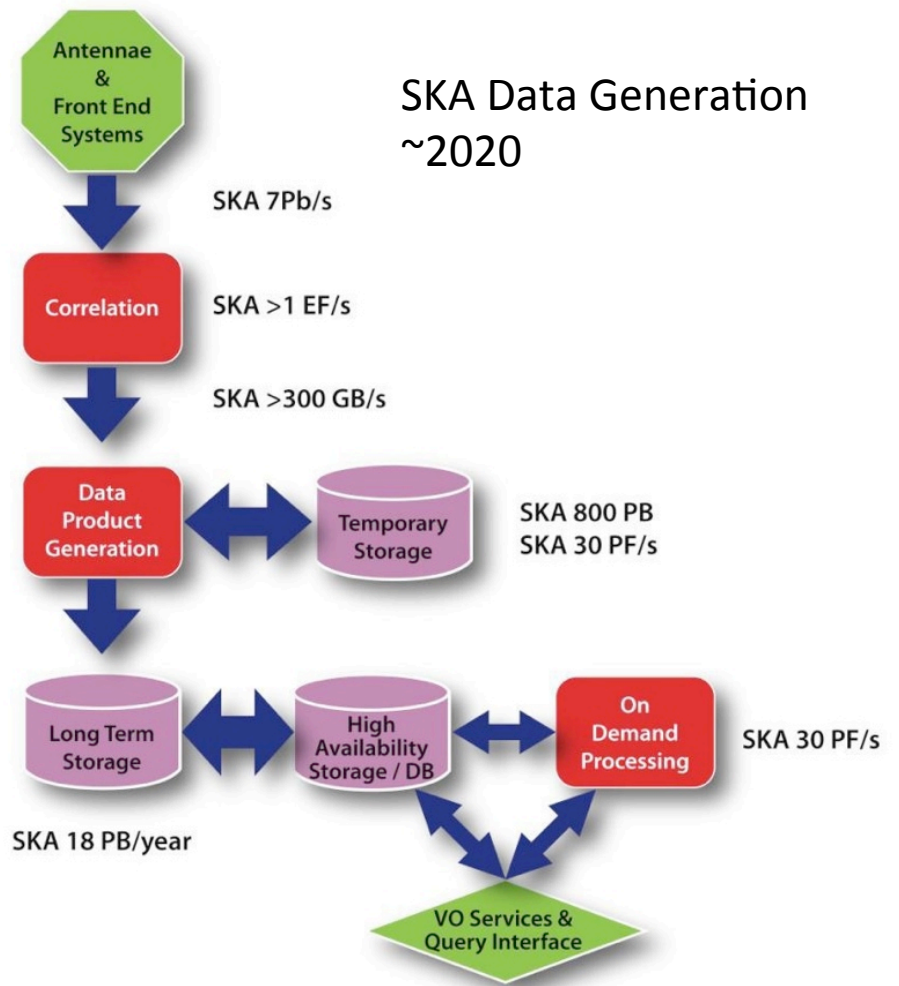
- 1000 -1500 antennas x15m in 5km
  - 1000 -1500 antennas x15m up to 3000 km
  - Candidate sites: South Africa or Australia
- When will it be ready?
- SKA1 = 10% collecting area 2016 -2019
  - SKA2= 100% collecting area 2018 -2023

# About the data



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

## SKA Data Generation ~2020



→ Deploy web-services-based workflows for analysis of multidimensional data on heterogeneous e-infrastructures

## Web-services-based users oriented workflows

- Automate the analysis process
- Orchestrate the specific software tools involved
- Get modularity (Re-use)

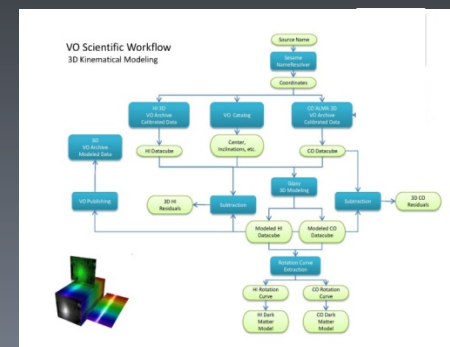
## Multidimensional data

→ Virtual Observatory

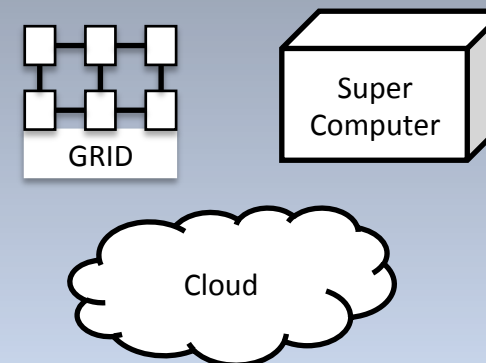
- Standards for publishing and accessing astronomical data
- Service Oriented Architecture (SoA)

## Heterogeneous e-infrastructures

- Optimize the use of resources
- Facilitate the access to the infrastructures
  - → FEDERATION?



VIRTUAL OSERVATORY



# Workflows Management System and web services as interface between users and infrastructures



**AstroTaverna** Workflow 4Ever

Taverna Workbench 2.3-SNAPSHOT Developer

Registry: <http://registry.euro-vo.org/services/RegistrySearch>  
 Keywords: amiga  
 Cone Search SIA Search SSA Search

13 results for ConeSearch: amiga

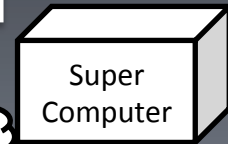
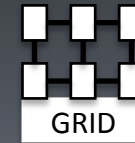
Short name	Title	Subjects	Identifier	Publisher
AMIGACS	AMIGA Catalogue	[The AMIGA Catalogue]	ivo://svo.amiga...	The AMIGA ...
J/A+A/411/391	The AMIGA proje...	[Positional Data, Galaxies]	ivo://CDS.Viz...	CDS
J/A+A/472/121	AMIGA V. Isolatio...	[Galaxies]	ivo://CDS.Viz...	CDS
J/A+A/462/507	AMIGA III. IRAS ...	[Galaxies]	ivo://CDS.Viz...	CDS
J/A+A/436/443	AMIGA. I. Velociti...	[Galaxies, Velocities]	ivo://CDS.Viz...	CDS
J/A+A/449/937	AMIGA. II. Morph...	[Galaxies]	ivo://CDS.Viz...	CDS
J/A+A/470/505	AMIGA IV. Neighb...	[Galaxies]	ivo://CDS.Viz...	CDS
J/A+A/485/475	AMIGA. VI. Radio...	[Galaxies]	ivo://CDS.Viz...	CDS
J/A+A/486/73	AMIGA VII. FIR a...	[Galaxies]	ivo://CDS.Viz...	CDS
J/A+A/532/A117	AMIGA VIII. Flux ...	[Galaxies]	ivo://CDS.Viz...	CDS
J/A+A/534/A102	AMIGA IX. Molecu...	[Galaxies]	ivo://CDS.Viz...	CDS
J/A+A/540/A47	AMIGA X. Isolate...	[Photometry, Galaxies]	ivo://CDS.Viz...	CDS
J/A+A/540/A96	Molecular gas in H...	[Clusters_of_galaxies]	ivo://CDS.Viz...	CDS

**J/A+A/411/391: The AMIGA project. Revised positions for CIG galaxies (Leon+ 2003)**

We present revised positions for the 1051 galaxies belonging to the Karachentseva Catalog of Isolated Galaxies (CIG, Cat. 82>). New positions were calculated by applying SExtractor to the Digitized Sky Survey CIG fields with a spatial resolution of 1.2". We visually checked the results and for 118 galaxies had to recompute the assigned positions due to complex morphologies (e.g. distorted isophotes, undefined nuclei, knotty galaxies) or the presence of bright stars. We found differences between older and newer positions of up to 38" with a mean value of 2.96" relative to SIMBAD and up to 38" and 2.42" respectively relative to UZC (Cat. PASP/111/438>). Based on star positions from the APM catalog (Cat. 267>) we determined that the DSS astrometry of five CIG fields has a mean offset in  $(\{\alpha\}, \{\delta\})$  of  $(-0.90", 0.93")$  with a dispersion of 0.4". These results have been confirmed using the 2MASS All-Sky Catalog of Point Sources (Cat. 246>). The intrinsic errors of our method combined with the astrometric ones are of the order of 0.5".

Publisher  
CDS  
Documentation  
<http://cdsarc.u-strasbg.fr/cgi-bin/Cat?J/A+A/411/391>

Add J/A+A/411/391 to workflow

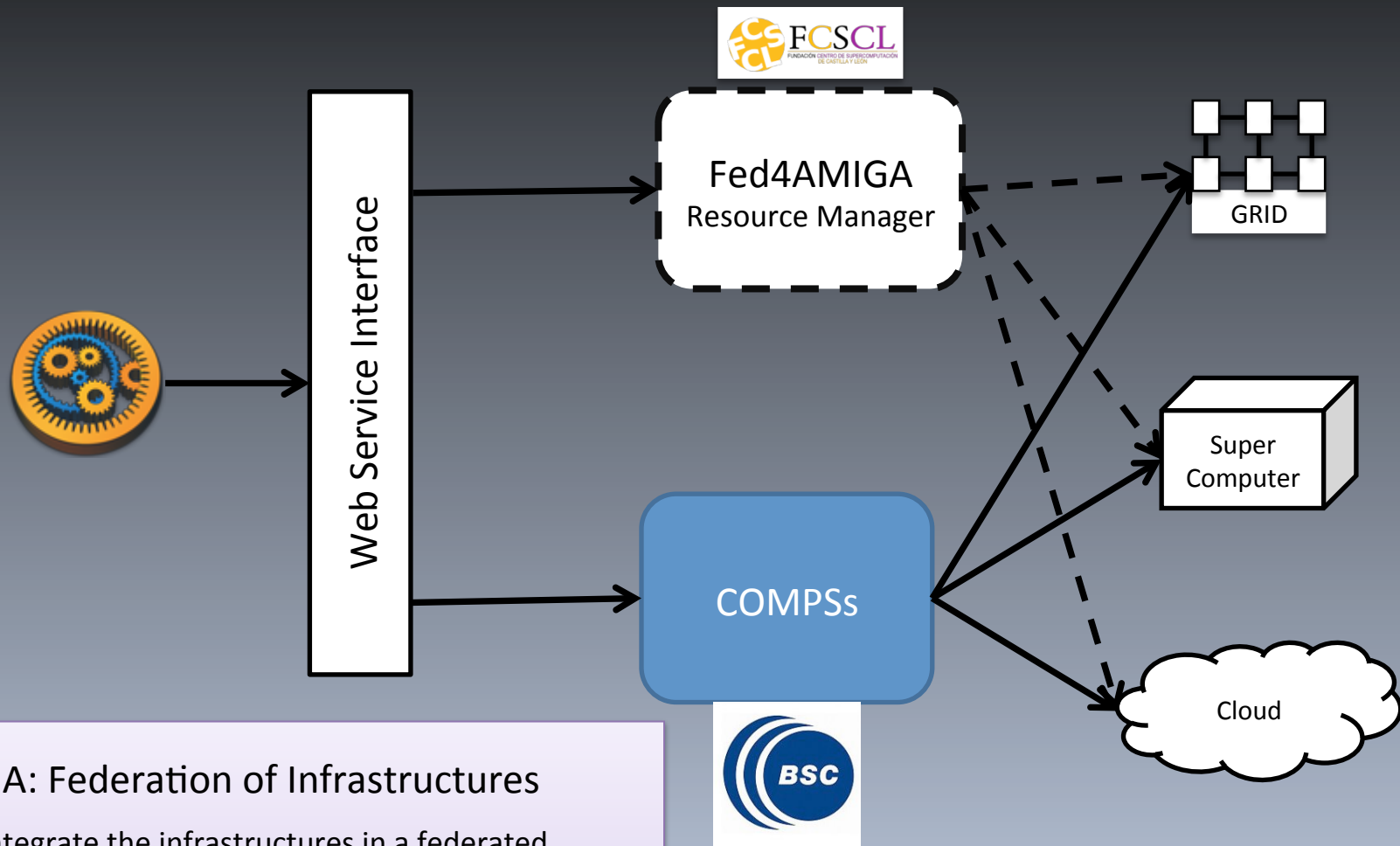


Ad

- S
- N

- Get flexibility: a web service can be called from different tools
- Modularity allows Re-use and Re-purpose
- Reproducibility of the experiments

# Federation of heterogeneous infrastructures

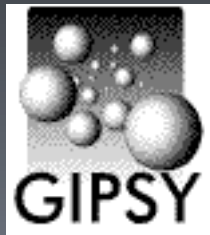


## Fed4AMIGA: Federation of Infrastructures

- How to integrate the infrastructures in a federated system?
- How to authenticate the users?
- How to implement business rules to decide in which infrastructure the task should run?

## Getting started:

→ USE CASE: 3D Kinematical modeling

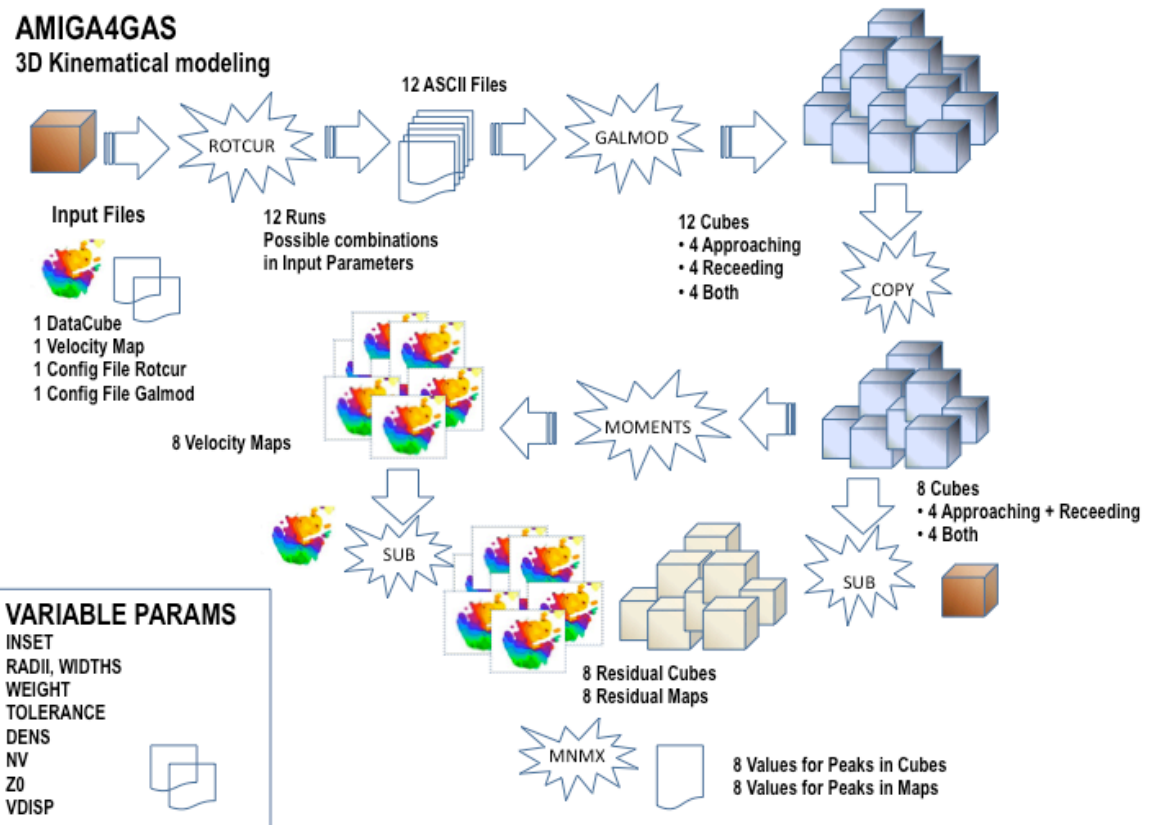


Groningen Image  
Processing System

- Software for analysis of interferometric data
- Specially powerful for kinematical modelling
- A complex software which has been growing during 20 years

### AMIGA4GAS

3D Kinematical modeling





## Getting started:

- GIPSY ON GRID



- How deploy it on the Grid nodes?
- It is better encapsulate it in a Virtual Machine?
- Fedcloud Task Force of EGI



- We start using the [vo.physics.ibergrid.eu](http://vo.physics.ibergrid.eu)



- GIPSY ON CLOUD AND ON SUPERCOMPUTING



- Searching the suitable infrastructure for GIPSY tasks:
  - Tests with different inputs (datacube sizes and parameters) and virtual machine configurations