Deploying astronomical workflows in heterogeneous distributed computing infrastructures

Susana Sánchez Expósito

Instituto de Astrofísica de Andalucía - CSIC

Pablo Martín, José Enrique Ruiz, Elena Ortega, Lourdes Verdes-Montenegro, Antonio Ruíz-Falcó, Julián Garrido, Juan De Dios Santander-Vela, Raül Sirvent, Enric Tejedor, Rosa M. Badia and Jesús Lorenzana

IWSG 2013, Zurich

AMIGA4GAS: AMIGA for GTC, ALMA and SKA pathfinders

- 4th project of AMIGA group at IAA-CSIC

  (AYA2011-30491-C02)

Study of the intrinsic properties of the galaxies

Multi- wavelength data of ~1000 galaxies

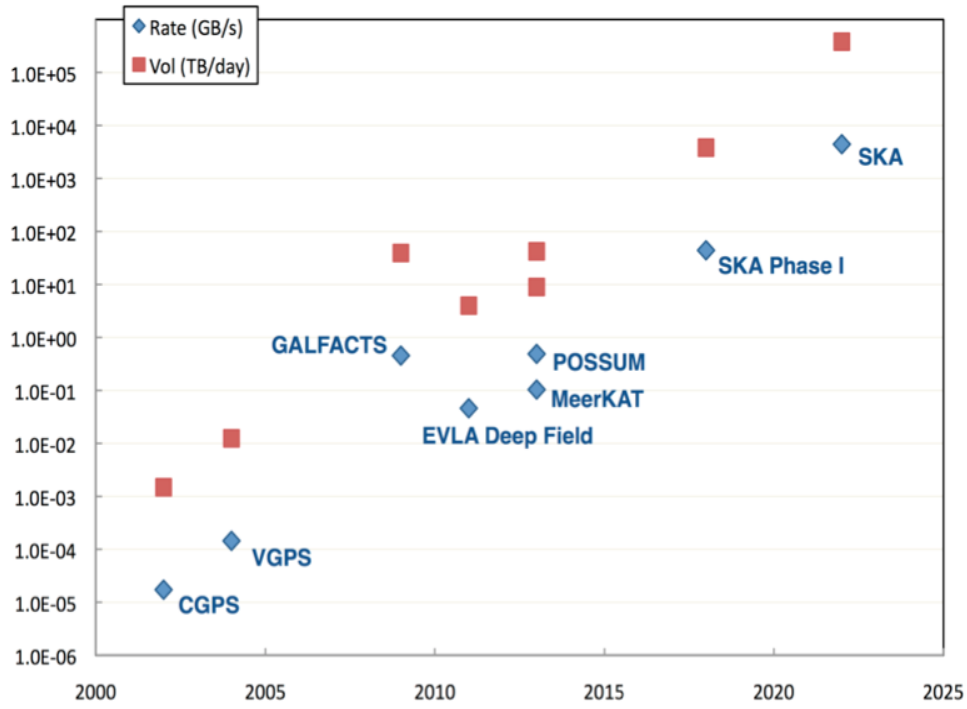
Access to different archives

Virtual Observatory Activities

- Data consumers and data publishers
- Contributing to data interoperability standards

"Exa-scale Era"

- Coordinated with the FCSCL
  - Technical challenge: Delivery and efficient exploitation of very large data volumes
- Close collaboration with the BSC



"CyberSKA: An On-line Collaborative Portal for Data-intensive Radio Astronomy". Cameron Kiddle, et al. GCE 2011

How to deal with such large data volumes?

→Web-services-based workflows for analysis of multidimensional data on federated DCIs

- Why web services?
  - Flexibility: a web service can be called from different tools
  - Modularity: each analysis web service can be used in different experiments
  - Transparency: users are unaware of software issues (SaaS benefits)
  - IVOA (International Virtual Observatory Alliance) standards for astronomical data services

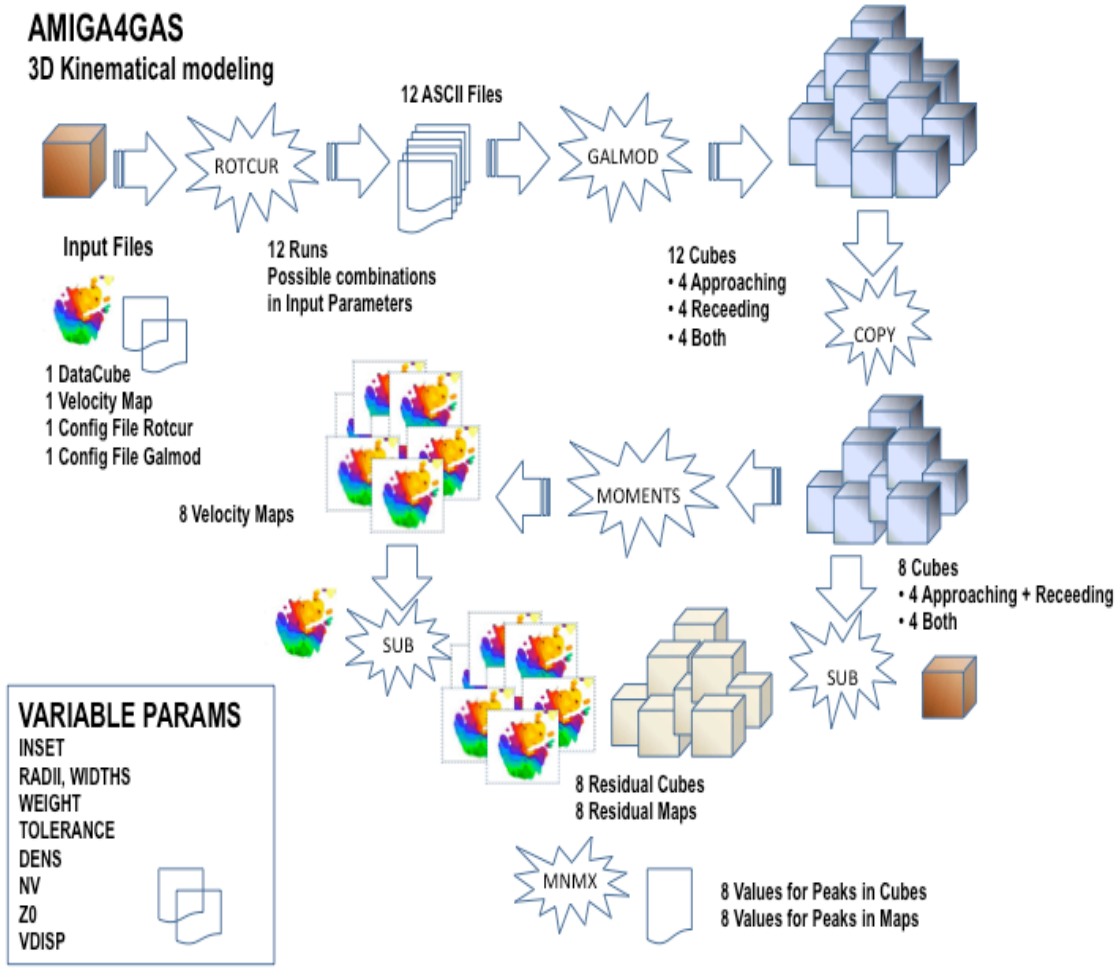
- Why workflows?
  - User level: Orchestration of the different applications involved in an experiment
  - Infrastructure level (not web-services-based): Parallelize applications

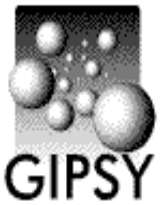
- Multidimensional Data
  - Data from interferometric instruments (SKA)

- Why federated DCIs?
  - Facilitate the access to the infrastructures (Authentication/Authorization)
  - Optimize the use of resources

Web-services-based workflows for analysis of multidimensional data

1) Use Case: Kinematical analysis of galaxies using multidimensional data





Groningen Image Processing System[1]

Software for analysis of multi-D data

- Designed originally for the Westerbork telescope
- Complex installation
- Non parallelized software

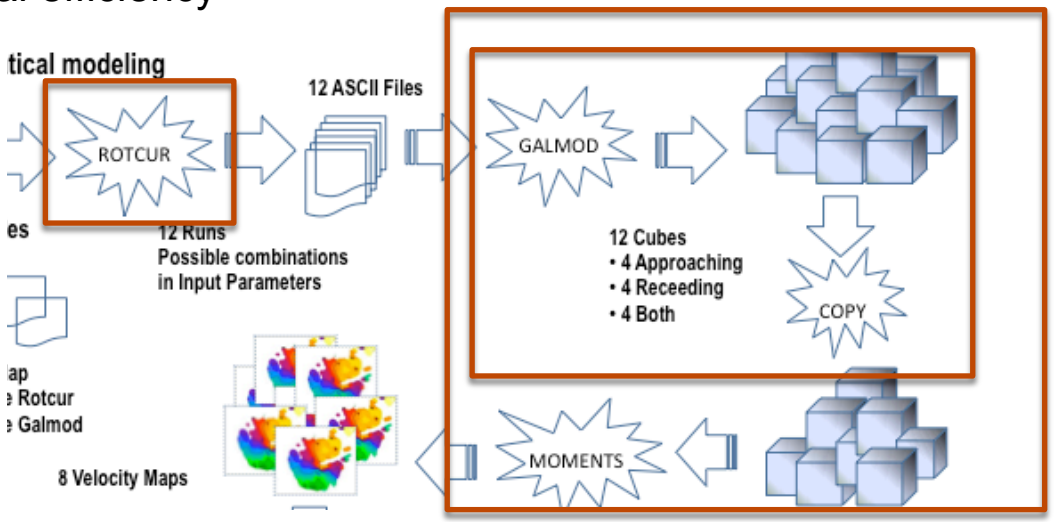
[1] http://www.astro.rug.nl/~gipsy

Web-services-based workflows for analysis of multidimensional data

- Install the software on the infrastructures:
  - IBERGRID (phys.vo.ibergrid.eu)
  - SGE Cluster at FCSCL
  - Best option Cloud Computing?

Recommendation for selecting the infrastructure

- Design the analysis web services (nesting level)
  - Computational efficiency
  - Reusability

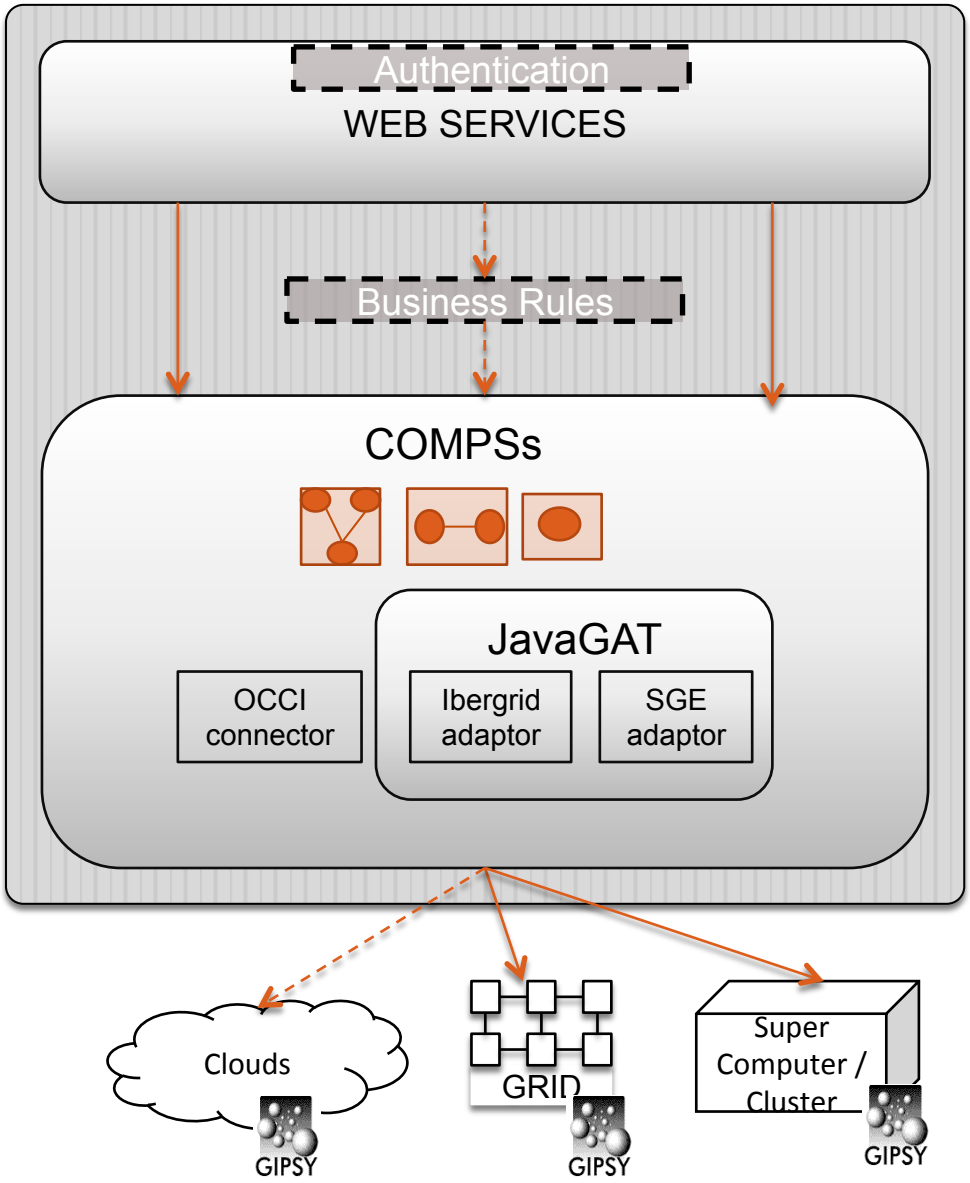


Guide to provide efficient and reusable analysis web services

- Apply new standards (PDL [1]) to characterize analysis service
  - Input parameters checking for the correct execution of the workflow

[1] http://www.ivoa.net/documents/PDL/20130510/index.html

Federated layer



Authentication:

- LDAP's clusters & Grid certificates & Cloud accounts

Business Rules:

- Scheduling among DCIs
- Latency, energy efficiency, time to complete, probability of completion or costs

COMPSs [1]:

- Job launcher
- Infrastructure level workflows

[1] http://www.bsc.es/es/publications/comp-superscalar-bringing-grid-superscalar-and-gcm-together

Our targets vs. Science Gateway targets

EGI-InSPIRE

SCIENCE GATEWAY PRIMER

Document link	https://documents.egi.eu/document/1463
Last modified	12/03/2013
Version	0.9.2
Document Status	Public - draft

Outcomes of the project

1) Specific (ready-to-use) outcomes

- Set of web services for analysing interferometric multidimensional data running on heterogeneous DCIs

2) General outcomes

- How to schedule among DCIs:
  - Business rules
  - Statistics of performance of the applications on different infrastructures
- How to get astrophysics Software as a Service running on DCI
  - Installation and deployment issues
  - Web service design: efficiency and reusability
- Characterization of analysis services
  - IVOA – International Virtual Observatory Alliance
  - PDL – Parameter Description Language

Questions?

sse@iaa.es

http://amiga.iaa.es/p/263-federated-computing.htm

http://www.fcsc.es