



## **WF4Ever: Supporting Reuse and Reproducibility in Experimental Science**

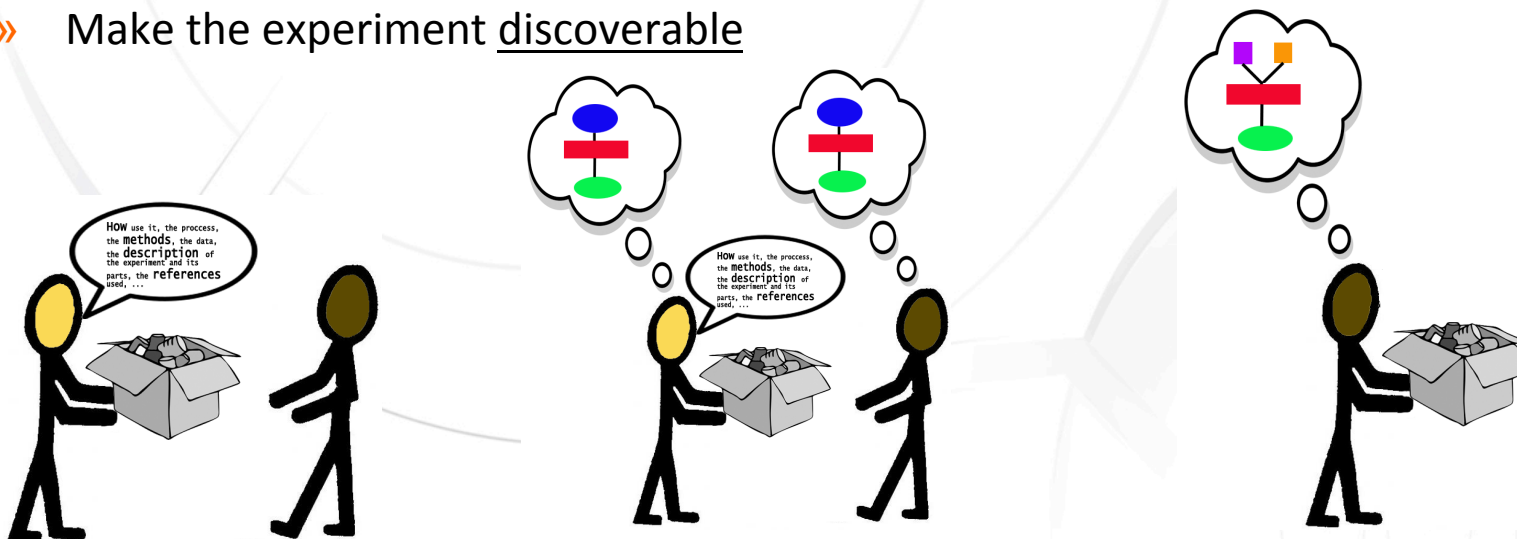
Susana Sánchez, Jose Enrique Ruíz, Lourdes Verdes-Montenegro,  
Julian Garrido, Juan de Dios Santander-Vela and Wf4Ever team  
Instituto de Astrofísica de Andalucía – CSIC

**EGI Technical Forum. 21 September 2012**

- » The challenge: Reuse and reproduce scientific experiments
  - Reusability, fundamental for incremental scientific development
  - Reproducibility, key for reliable science
- » How Wf4Ever is addressing these challenges
- » Impact in Astrophysics
- » Conclusions

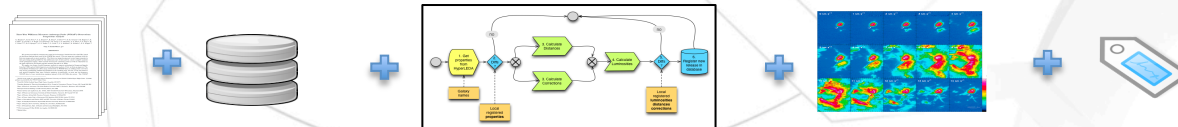
An experiment is reproducible when someone else, working independently, can get similar results following the same methods and using the same inputs  
Someone can reuse (part of) an experiment to create new experiments (\*)

- » Compile all the elements needed by the experiment
- » Share it with other scientists
- » Describe it: methods, data and all the elements involved.
- » Be independent of the environment
- » Make the experiment discoverable



(\*) Replacing the Paper: The Twelve Rs of the e-Research Record” on <http://blogs.nature.com/eresearch/>

- » Scientific Workflows: part of the solution
  - › Enable automation and expose the flow of scientific methods
  - › Encourage best practices in packing the experiment
  - › Provide a way to share the method
- » But more is needed:
  - › Share/provide the data, annotations, references, etc.
  - › Strategies for avoiding decay
  - › Tools for discovering the experiment



Workflow preservation

**Wf4Ever** - Preservation of scientific workflows in data-intensive science

EU funded FP7 STREP Project  
December 2010 – December 2013



- 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 de Astrofísica de Andalucía (IAA, Spain)**
- 7. Leiden University Medical Centre (LUMC, NL)**

**isoco**  
enabling the networked economy



The University of Manchester

MANCHESTER  
1824

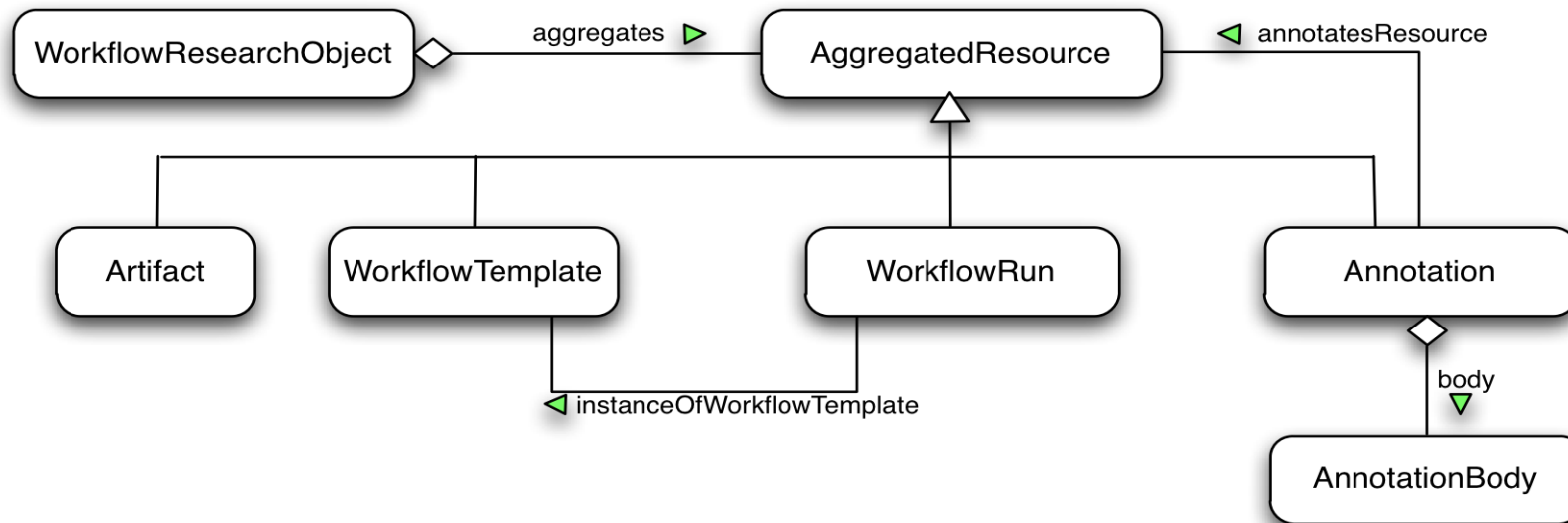


Technological infrastructure for the preservation and efficient retrieval and reuse of scientific workflows in a range of disciplines

- Encapsulate the scientific methodology (the workflows and all the associated information) in an artefact called **Research Object**.
- Archival, classification and indexing of the research object in scalable semantic repositories, providing advanced access and recommendation capabilities based on monitoring and metrics to evaluate similarities, decay, quality, stability, completeness.
- Creation of scientific communities to collaboratively share, reuse and evolve Research Objects stimulating the development of new scientific knowledge
- Use Cases:
  - Astronomy (IAA)
  - Genome-wide Analysis and Biobanking (LUMC)



» Model for Workflow-Centric Research Objects [1]



» Semantic ontologies to implement the model

- > Object Exchange and Reuse (ORE) for specifying aggregation of resources
- > Annotation Ontology (OA) for annotating the resources

[1]Workflow-Centric Research Objects: A First Class Citizen in the Scholarly Discourse, Khalid Belhajjame et al.

» Provide digital libraries with RO preservation functionalities

### Curation of Astrophysical Quantities

Materials and Methods

RO Versions [Prev](#) | [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) | [Next](#)

Conceptual Relations

- Research Object
  - Gathering
    - Digital File
      - Hyperleda.tf2
    - Web Services
      - Hyperleda
      - NED CS
    - Input Data
      - ListCIGS
      - MorphoTypes
    - Results
      - Luminosities
    - Scripts
      - Calculation.py
  - Propagation
  - Comparison
  - Update
  - Bibliography
    - Help
    - Used
    - Produced

Content Annotations Internal Comments Public Comments Access Rights Dependencies Statistics

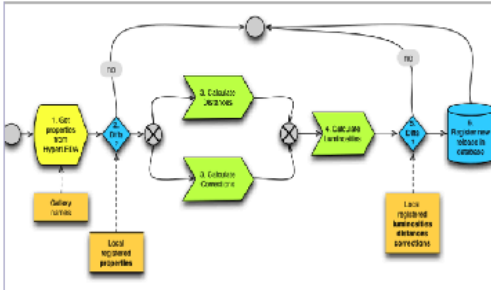
x Astronomy x Curation x Databases

Download  
Edit Info

☆☆☆☆☆  
Citations: 52  
Downloads: 342  
Comments: 24

Created 12/01/2012 by John Doe  
Updated 24/06/2012 by Alice Darn  
Authors: John Doe; Simon Wright; Alice Darn  
Groups: AMIGA Group  
Size: 12 MB 26 files including 3 workflows  
Description and Purpose

Schema



Check State  
Completeness  
Health

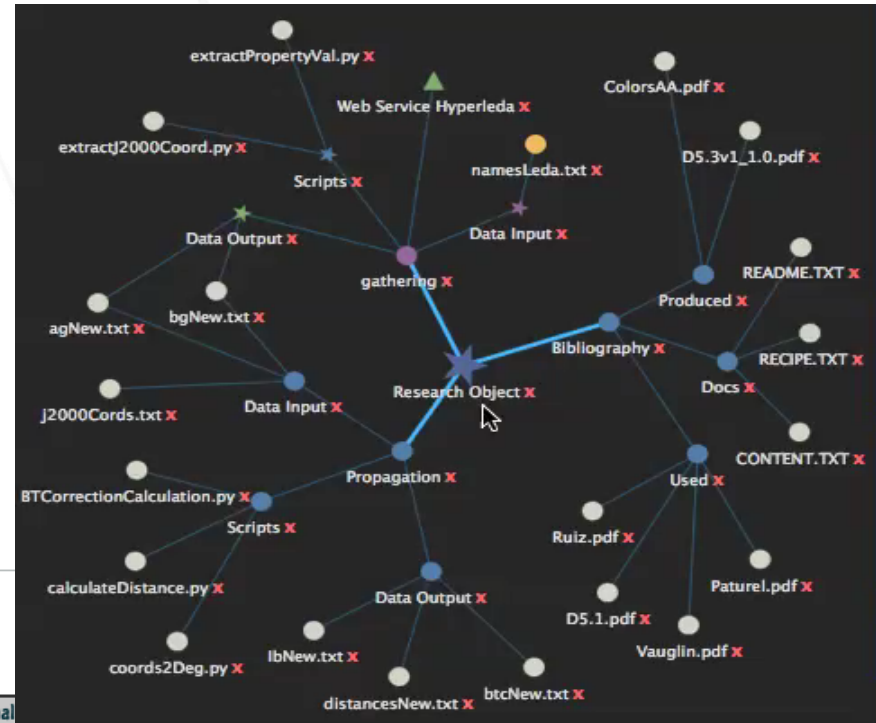




Describe it: methods, data and all the elements involved.

## Mechanisms for

- » Defining relationships between the elements
- » Annotating each element and the whole RO



## Curation of Astrophysical Quantities

Materials and Methods

Conceptual Relations

- Research Object
  - Gathering
    - Digital File
      - Hyperleda.tf2
    - Web Services
      - Hyperleda
      - NED CS
    - Input Data
      - ListCIGS
      - MorphoTypes

Content Annotations Provenance Internal

Role	Type	Annotation	Who
26/01/12	Purpose	Gather info from Hyperleda	John Doe
14/01/12	Description	Annot. coming from Tavema	Tavema
10/01/12	FileType	Workflow	System



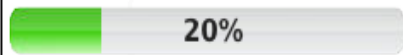
**Runnable**

Service up, software working, etc.



**Repeatable**

The output can be reproduced



**Shareable**

Enough annotation.



**Publishable**

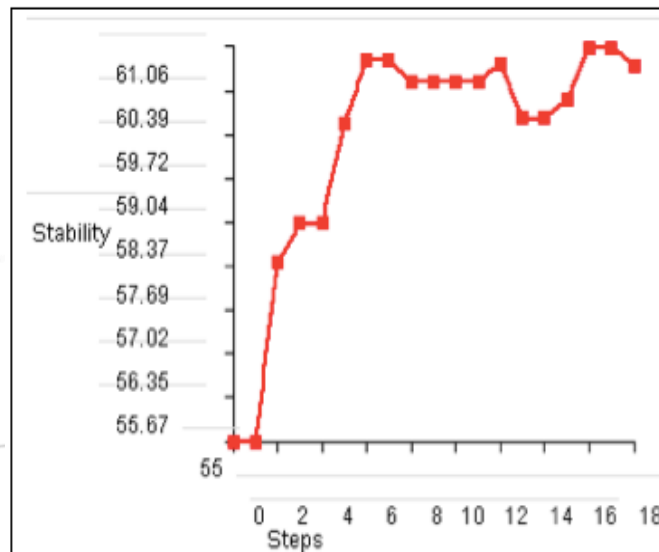
Service up, software working and all well commented

## » Completeness

It contains all the resources needed to be run, published, shared or repeated

## » Stability

Changes made by different kind of users on the RO, can improve it or make it worse



0. Alice creates RO
1. Alice adds Recipe.txt
2. Alice adds Content.txt
3. John adds Results.txt
4. Alice adds Bibliography.pdf
5. Alice annotates Bibliography.pdf
6. John removes Script.py
7. John edits annotation on Recipe.txt
8. Unknown adds Dropme.txt

### Decay Information

Last check was performed 2 days ago and returned one error:

The service SDSS-DR7, needed by the workflow  
calculate\_galaxy\_distances is **down**

Check now

Try to repair

## » Decay

State of the services (up/down), of the applications (updated/deprecated), permissions to access the data

## » Interoperability

- › RO level
- › Component level



## » Tracking

Rating by other users, who used the RO, comments, etc.

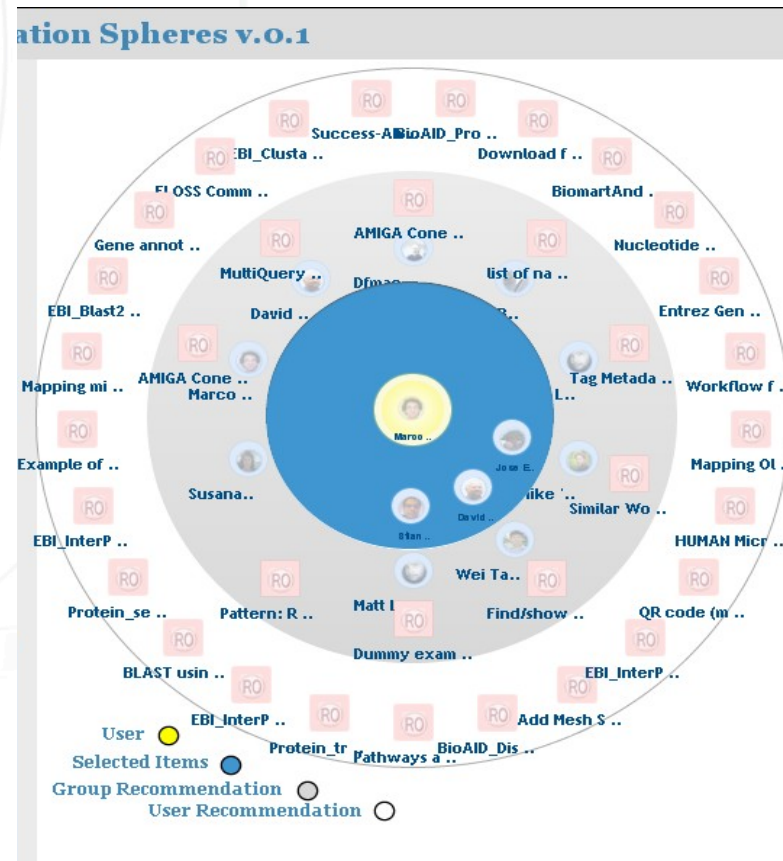
## » Recommender Service

Which exploits semantic description, relations and other metadata to support advanced search mechanisms

## » Collaboration Spheres

Visual mechanism to find similar elements (users, ROs, workflows) to others previously selected

Rating	★★★★☆
Downloads	36
Citations	[2]
Re-used	[1]
Comments	[4]



# AMIGA

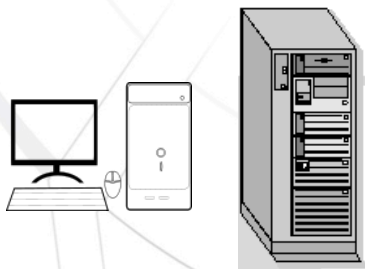
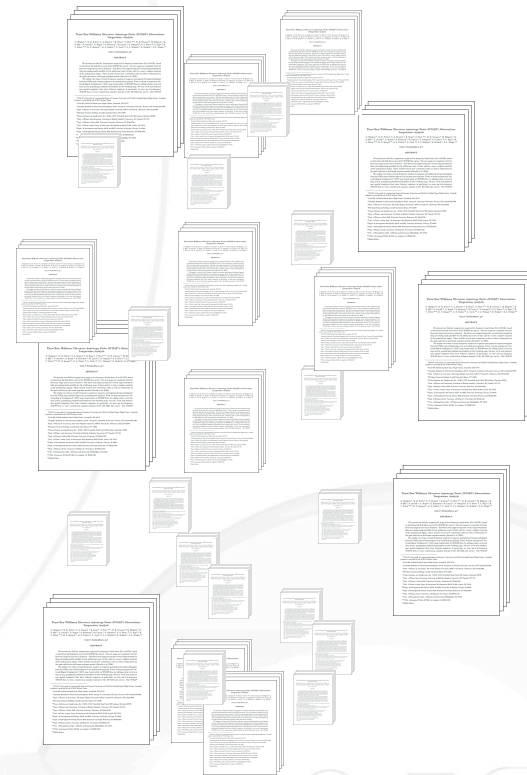
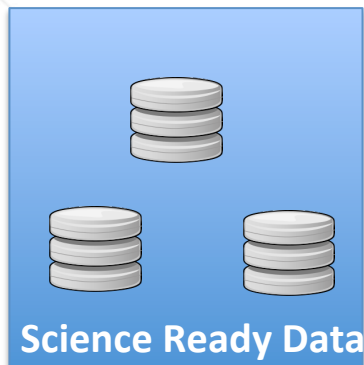
Analysis of the interstellar  
Medium of Isolated Galaxies  
<http://amiga.iaa.es>

- Statistical baseline of isolated galaxies to compare with the behavior of galaxies in denser environments
- Multi- $\lambda$  study of  $\sim 1000$  galaxies  
→ Intensive and complex analysis of multidimensional data

- » Uses case for WF4Ever:
  - › Astrophysical quantities propagation (1D) (\*)
  - › Properties extraction from images (2D)
  - › 3D Data analysis.

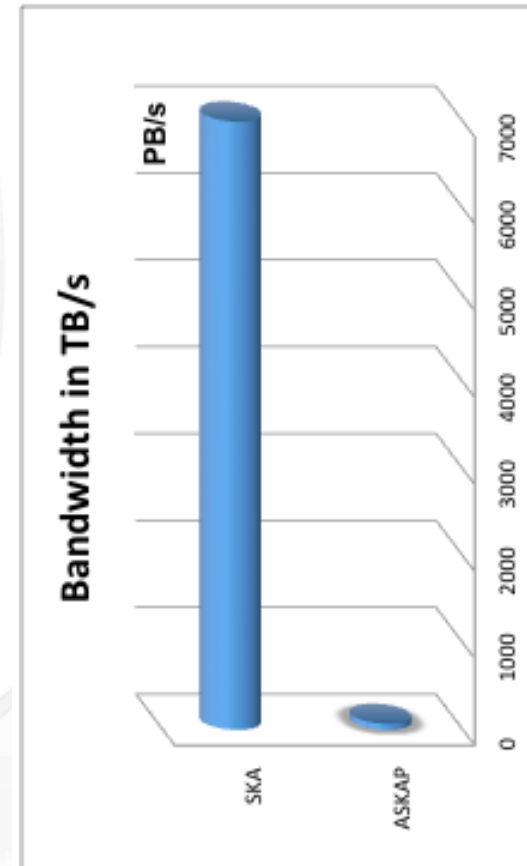
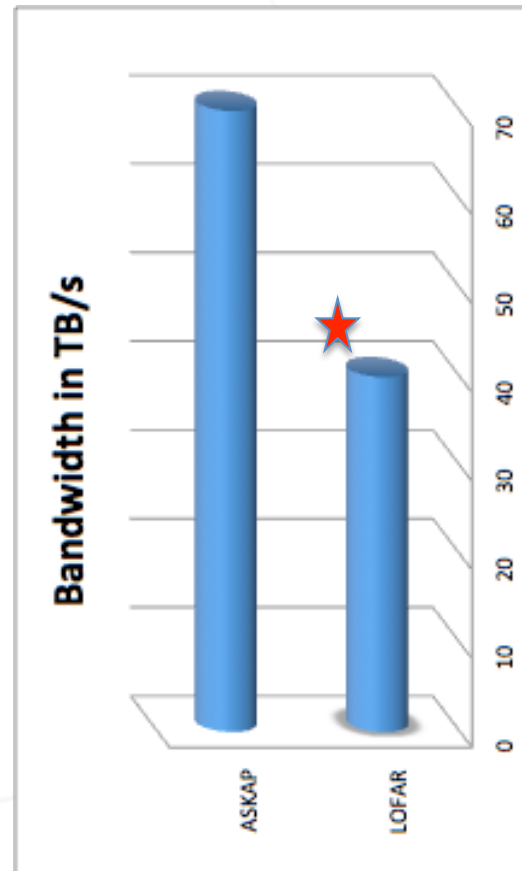
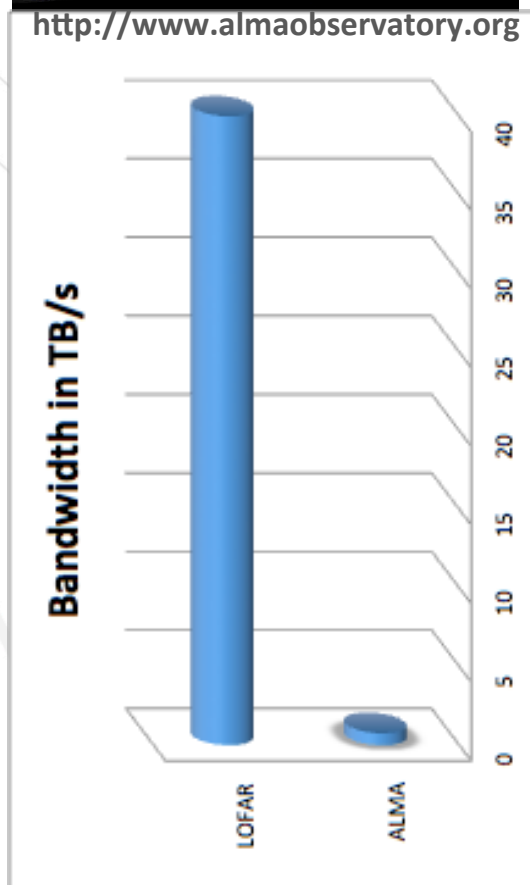


(\*) As part of the work in “The AMIGA sample of isolated galaxies XI. A First Look at Isolated Galaxy Colors”, M. Fernandez Lorenzo et al. Ref. AA/2011/18660

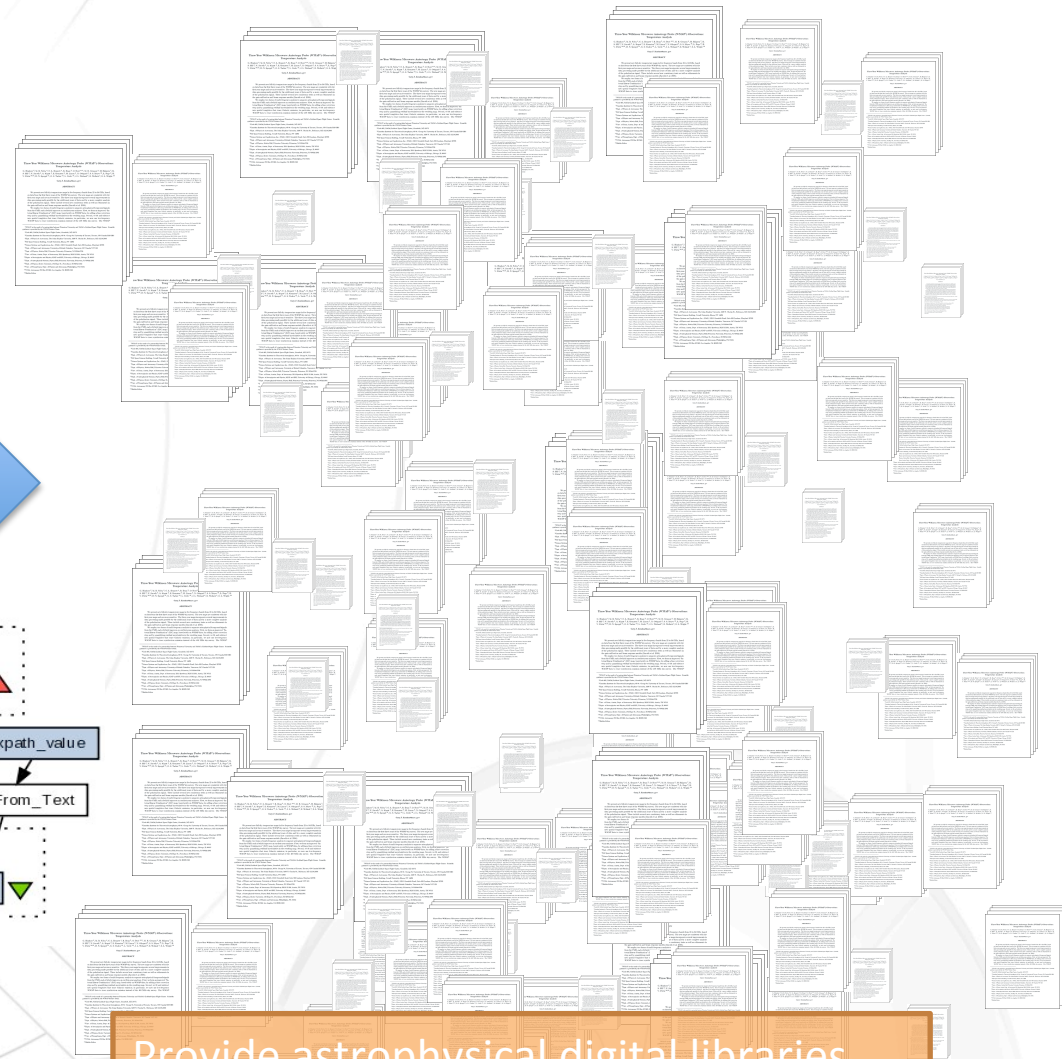
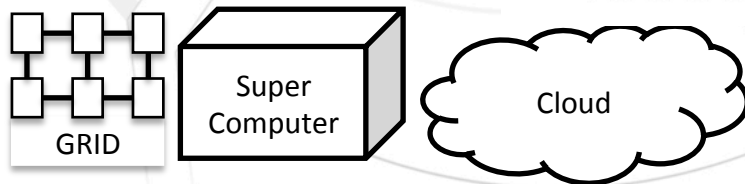
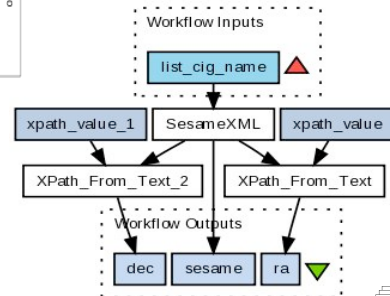
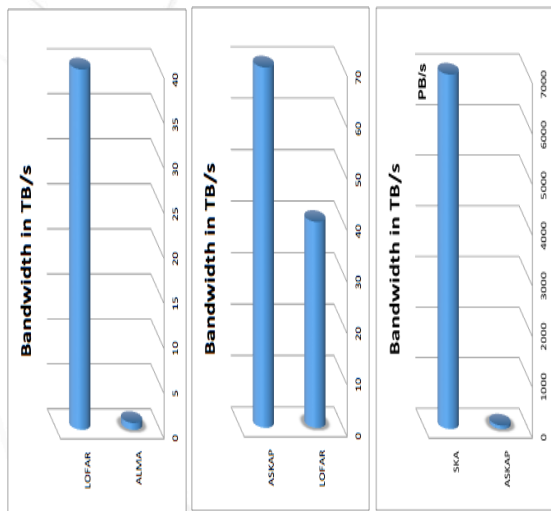




<http://www.almaobservatory.org>



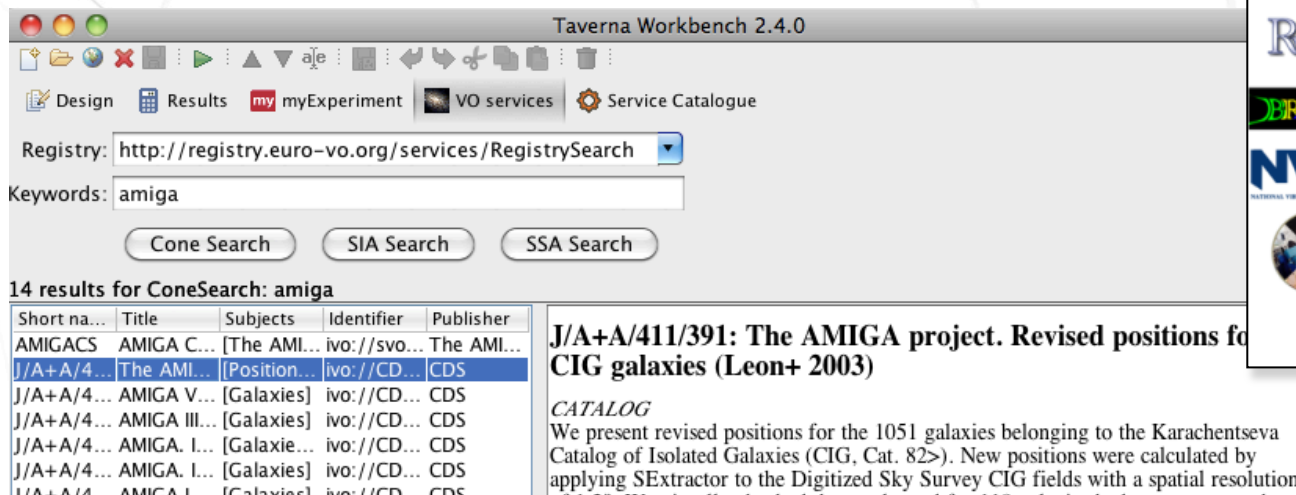
★ LHC – Tier 1



Provide astrophysical digital libraries with standards and models about how store and publish Research Object



- » AstroTaverna plugin <http://wf4ever.github.com/astrotaverna/>
  - » Connection with Virtual Observatory



Taverna Workbench 2.4.0

Registry: <http://registry.euro-vo.org/services/RegistrySearch>

Keywords:

Cone Search   SIA Search   SSA Search

14 results for ConeSearch: amiga

Short na...	Title	Subjects	Identifier	Publisher
AMIGACS	AMIGA C...	[The AMI...	ivo://svo...	The AMI...
J/A+A/4...	The AMI...	[Position...	ivo://CD...	CDS
J/A+A/4...	AMIGA V...	[Galaxies]	ivo://CD...	CDS
J/A+A/4...	AMIGA III...	[Galaxies]	ivo://CD...	CDS
J/A+A/4...	AMIGA. I...	[Galaxie...	ivo://CD...	CDS
J/A+A/4...	AMIGA. I...	[Galaxies]	ivo://CD...	CDS
J/A+A/4...	AMIGA. I...	[Galaxies]	ivo://CD...	CDS

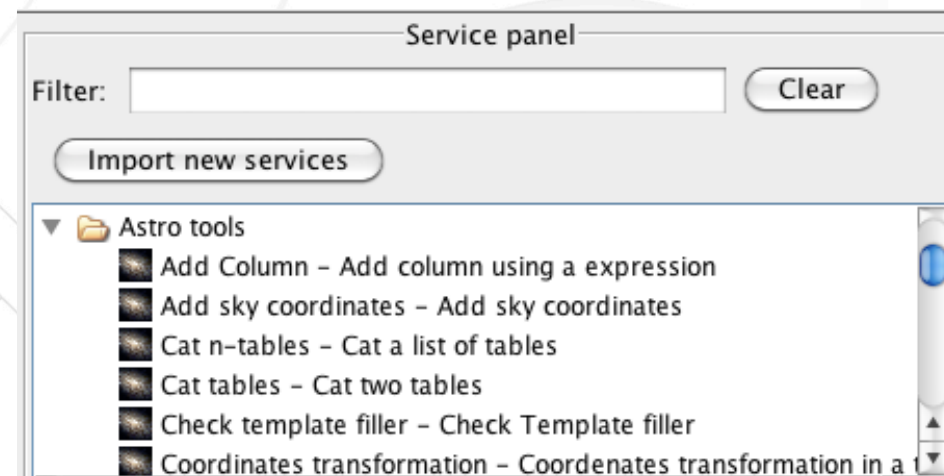
**J/A+A/411/391: The AMIGA project. Revised positions for CIG galaxies (Leon+ 2003)**

*CATALOG*

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



- » Services for managing and visualizing VOTable
- » Astronomical utilities: coordinate transforms.



Service panel

Filter:

- ▼ Astro tools
  - Add Column - Add column using a expression
  - Add sky coordinates - Add sky coordinates
  - Cat n-tables - Cat a list of tables
  - Cat tables - Cat two tables
  - Check template filler - Check Template filler
  - Coordinates transformation - Coordenates transformation in a t

- » Re-using workflows between different workflow languages scientists
- » Data Scientific methodology interoperability to increase the collaboration among users
- » Workflow validation
- » Sharing workflows within and among communities

Susana Sánchez: [sse@iaa.es](mailto:sse@iaa.es)

<http://amiga.iaa.es/p/212-workflows.htm>

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

