



The Astronomy challenge: How can workflow preservation help?

Susana Sánchez, Jose Enrique Ruíz, Lourdes Verdes-Montenegro,
Julian Garrido, Juan de Dios Santander-Vela and Wf4Ever team.

Instituto de Astrofísica de Andalucía – CSIC

SHIWA Summer School. 3 July 2012

- » Introduction to AMIGA group
- » The astronomy challenge and context of Wf4Ever project
- » How Wf4ever tools can help the astronomers
- » Our astronomy use case

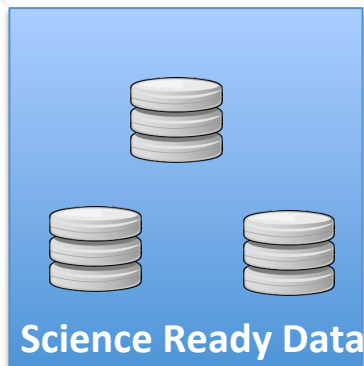
AMIGA

Analysis of the interstellar
Medium of Isolated GALaxies

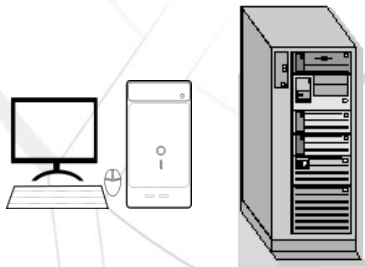
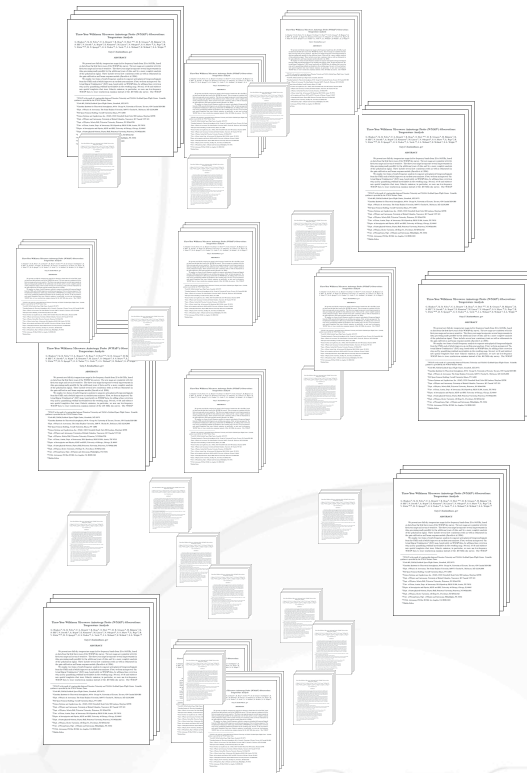
An international collaboration coordinated from the IAA-CSIC
P.I. Lourdes Verdes-Montenegro
<http://amiga.iaa.es>

- Statistical baseline of isolated galaxies to compare with the behavior of galaxies in denser environments
- Multi- λ study of ~ 1000 galaxies:
 - Need of intensive and complex analysis of multidimensional data





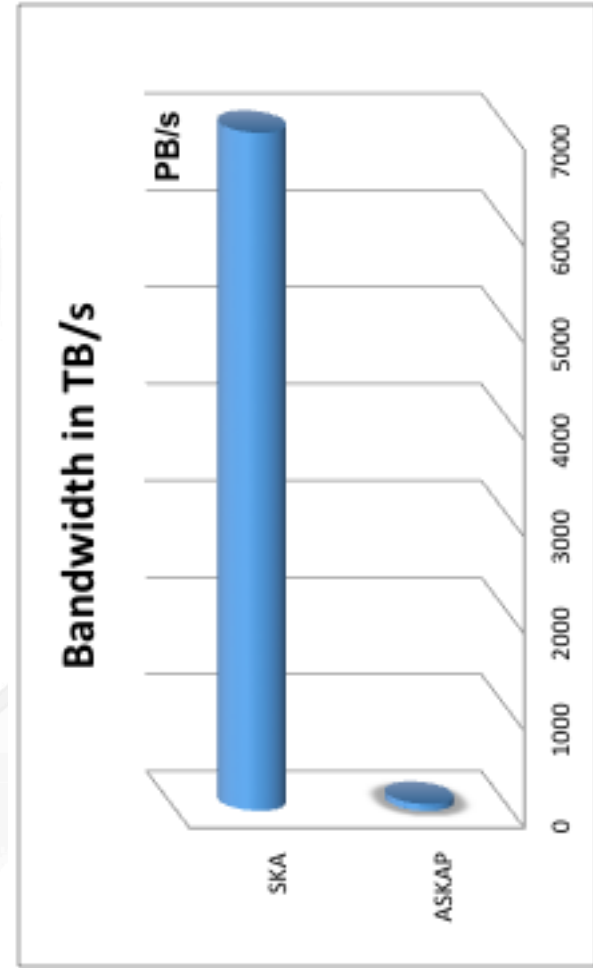
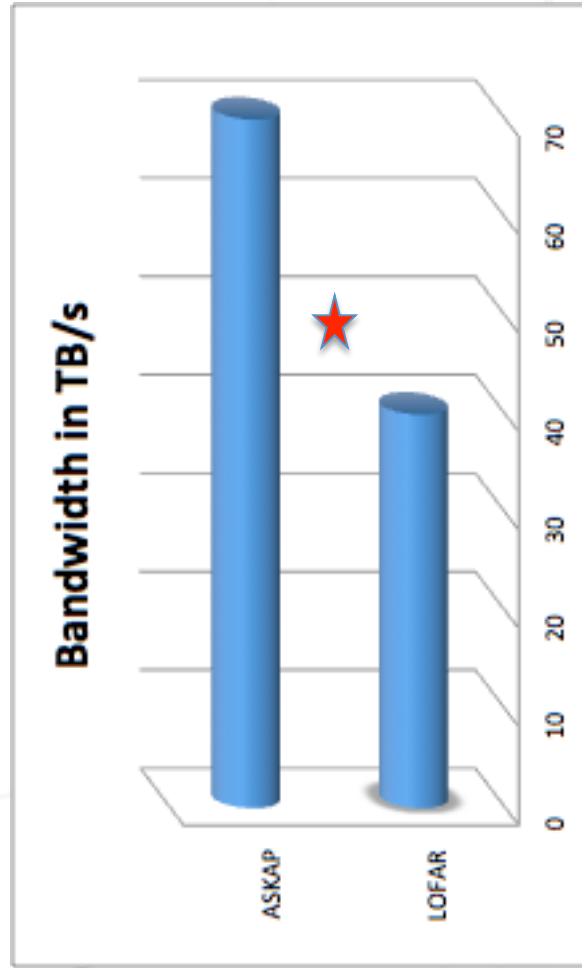
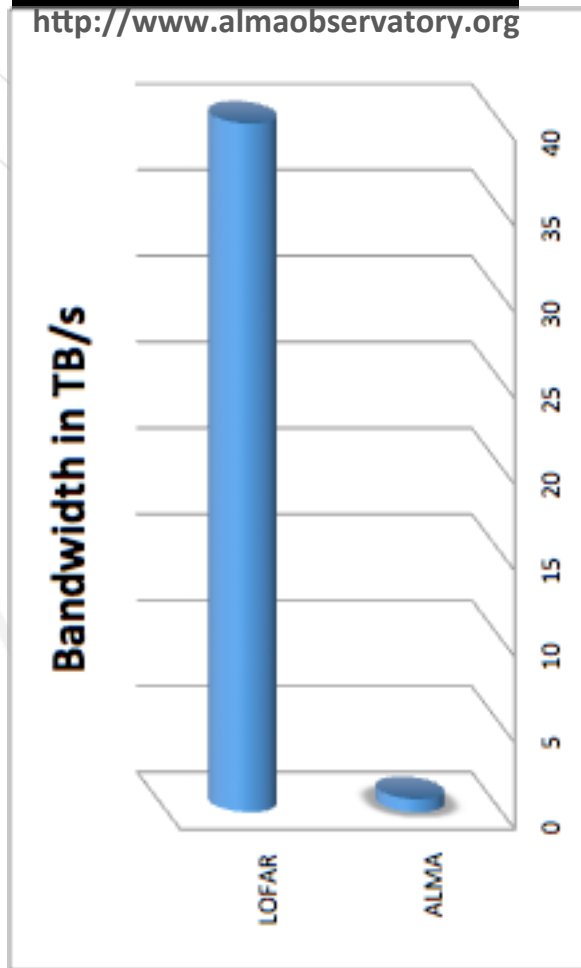
Analysis



The astronomy challenge and context of Wf4Ever project



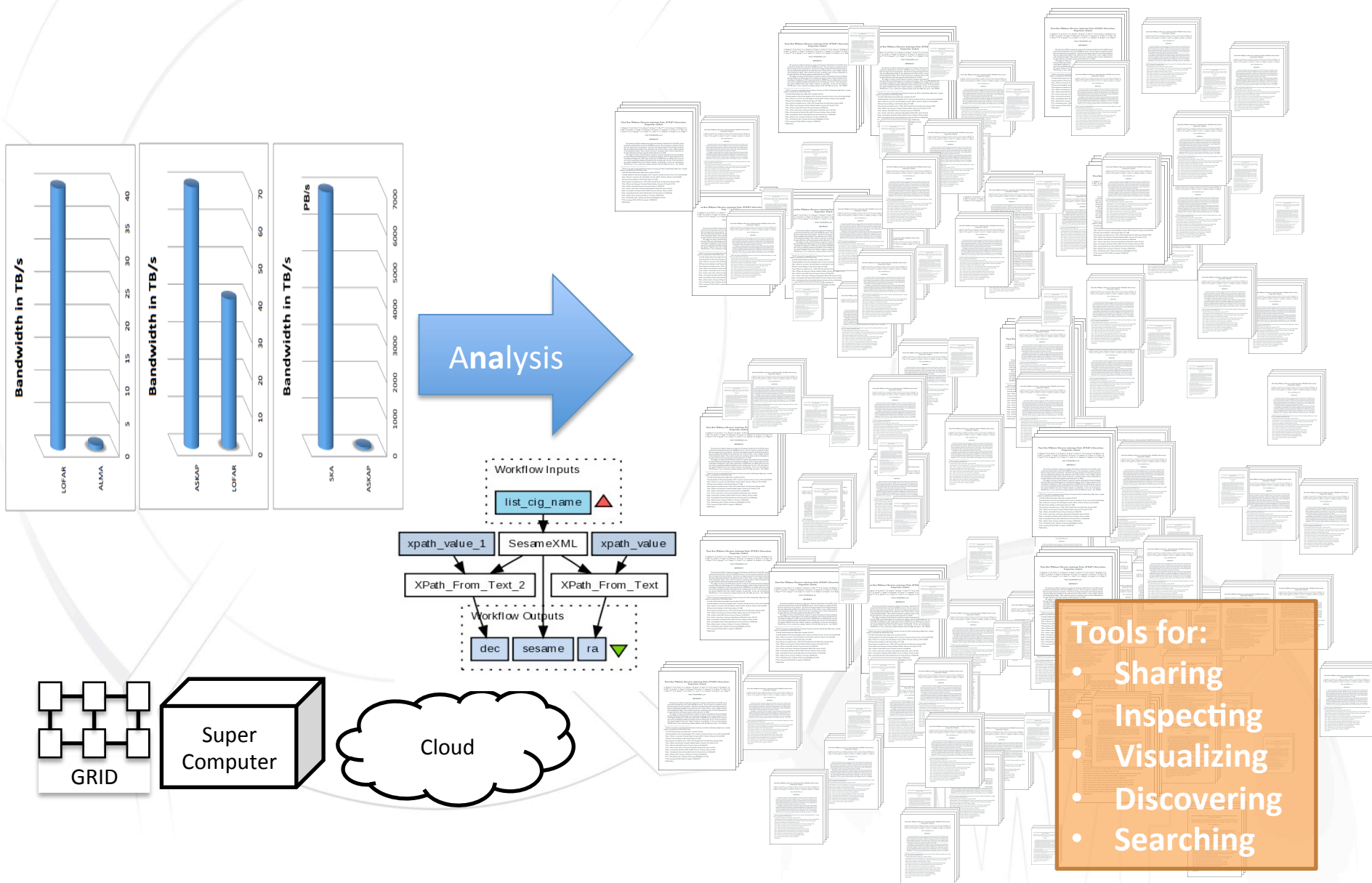
<http://www.almaobservatory.org>



★ LHC – Tier 1

The astronomy challenge and context of Wf4Ever project

A disruptive change in the methodology is needed



» Keywords: efficient use of data and reliable science

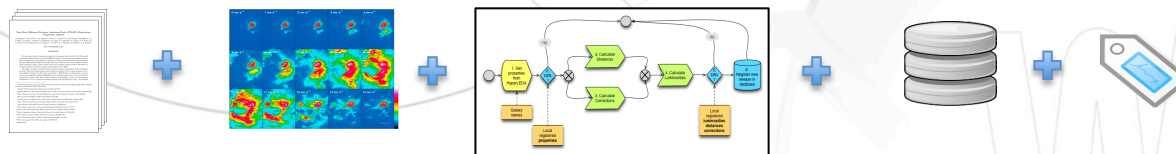
» Scientific Workflows:

- Enable automation and expose the flow of scientific methods
- Encourage best practices in packing the experiment
- Provide a way to share the experiment

» But more is needed:

- Reusability, fundamental for incremental scientific development
- Reproducibility, key for reliable science

→ Preserve the data and the scientific method



Workflow preservation

Wf4Ever - Preservation of scientific workflows in data-intensive science

EU funded FP7 STREP Project
December 2010 – December 2013



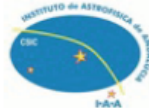
- 1. Intelligent Software Components (ISOCO, Spain)**
- 2. University of Manchester (UNIMAN, UK)**
- 3. Universidad Politécnica de Madrid (UPM, Spain)**
- 4. Poznan Supercomputing and Networking Centre (PSNC, Poland)**
- 5. University of Oxford (OXF, UK)**
- 6. Instituto de Astrofísica de Andalucía (IAA, Spain)**
- 7. Leiden University Medical Centre (LUMC, NL)**

isoco
enabling the networked economy



The University of Manchester

MANCHESTER
1824




UPM



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)





Research Object Digital Library

Sign out

Location: /home/susana/D

MAIN PAGE MY RESEARCH OBJECTS SPARQL ENDPOINT

Home / Research Object: <http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDA%20Luminosities/>

Interactive Conceptual Physical

HyperLEDA Luminosities/

- Web Services
- Datasets
 - agNew.txt
 - lbOld.txt
 - j2000Coords.txt
 - lbNew.txt
 - diff_lb.txt
 - lb.sql
 - NamesLEDA.txt
 - logr25New.txt
 - velocitiesNew.txt
 - distancesNew.txt
 - morphoNew.txt
 - btcNew.txt
 - btNew.txt
- Scripts
- Workflows
 - comparison_and_update_values_475535.
 - calculating_the_total_luminosity_of_a_galaxy_using_properties_from_text_1
 - gathering_galaxy_properties_using_hyperleda_129473.
- Workflow Runs
- Documents
 - GoldenTrace.txt

Selected resource info

URI: <http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDA%20Luminosities/>

Created by: Piotr Holubowicz

Created on: 12 April 2012 08:56

Number of annotations: 2

Relations:

- is referenced by D5.3v1_1.0.pdf
- source D2.1.pdf, D5.1.pdf, 3 more...
- (inverse of) relation README.txt, CONTENT.txt, RECIPES.txt
- aggregates diff_lb.txt, D1.2.pdf, 39 more...

Keywords [galaxies][catalogs]

Completeness 50%

Rating ★★★★★

Downloads 36

Citations [2]

Re-used [1]

Comments [4]

[\[Previous version\]](#) | [Next version](#)

Home / My Research Objects

Title	Created
<input type="checkbox"/> http://sandbox.wf4ever-project.org/rosrs5/ROs/Internal%20Extinction%20calculation/	Monday 12:52

Choose file x

Resource type A local file
 A web resource

File to upload

Resource type(s)

- Documents
- Datasets
- Workflows
- Web Services
- Workflow Runs
- Scripts

my experiment About | Mailing List | Publications | Log in | Register

[Home](#) | [Users](#) | [Groups](#) | [Workflows](#) | [Files](#) | [Packs](#) | [Services](#) | [Tools](#)

All

Home > Packs > Propagation of properties extracted from the HyperLEDA catalog in the calculation of luminosities of galaxies

Pack: Propagation of properties extracted from the HyperLEDA catalog in the calculation of luminosities of galaxies

Created at: 24/11/11 @ 19:12:05 | Last updated: 01/12/11 @ 22:13:34
[Tags \(5\)](#) | [Featured in Packs \(0\)](#) | [Favourited By \(0\)](#) | [Comments \(0\)](#)

Title: Propagation of properties extracted from the HyperLEDA catalog in the calculation of luminosities of galaxies

Description

The scientific experiment represented by this research object pertains to the multi-wavelength study for a sample of the most isolated galaxies in the local universe. This study characterizes each galaxy of this sample through both the measurement of basic astrophysical properties:

- The equatorial coordinates in J2000 epoch
- The velocities in km/s (v)
- The dust extinction coefficient (ag)
- The axis ratio of the isophote 25 mag/arcsec2 (logr25)
- The apparent total B magnitude (BT)
- The morphological type (t)


and the calculation of the more complex properties:

- The distance in Mega parsecs (D)
- The corrected apparent B magnitude (bct)
- The optical luminosity in B-band (LB)

Specifically, this research object is focused on the calculation of the intrinsic luminosity in the Johnson B-band, in order to achieve it the measurement or calculation of all those astrophysical properties is needed.

All the data involved in the characterization of the sample are stored in a local relational MySQL database. To maintain up to date this database and to register all the updates properly is part of this scientific experiment.

Creator



Jose Enrique Ruiz

12 items in this pack

Tags (5)

Creator tags

astronomy | curation | database storing | galaxies | hyperleda | virtual observatory

[Log in to add Tags](#)

Tags from Items (5)

astronomy | curation | database storing | galaxies | hyperleda

Shared with Groups (1)

INTEROPERABILITY

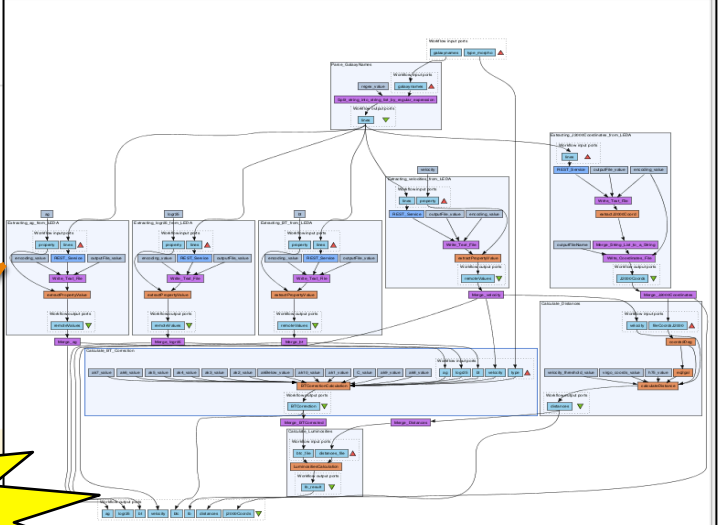
Workflow
Wf 4Ever **Research Object Digital Library**

MAIN PAGE MY RESEARCH OBJECTS SPARQL ENDPOINT Sign out

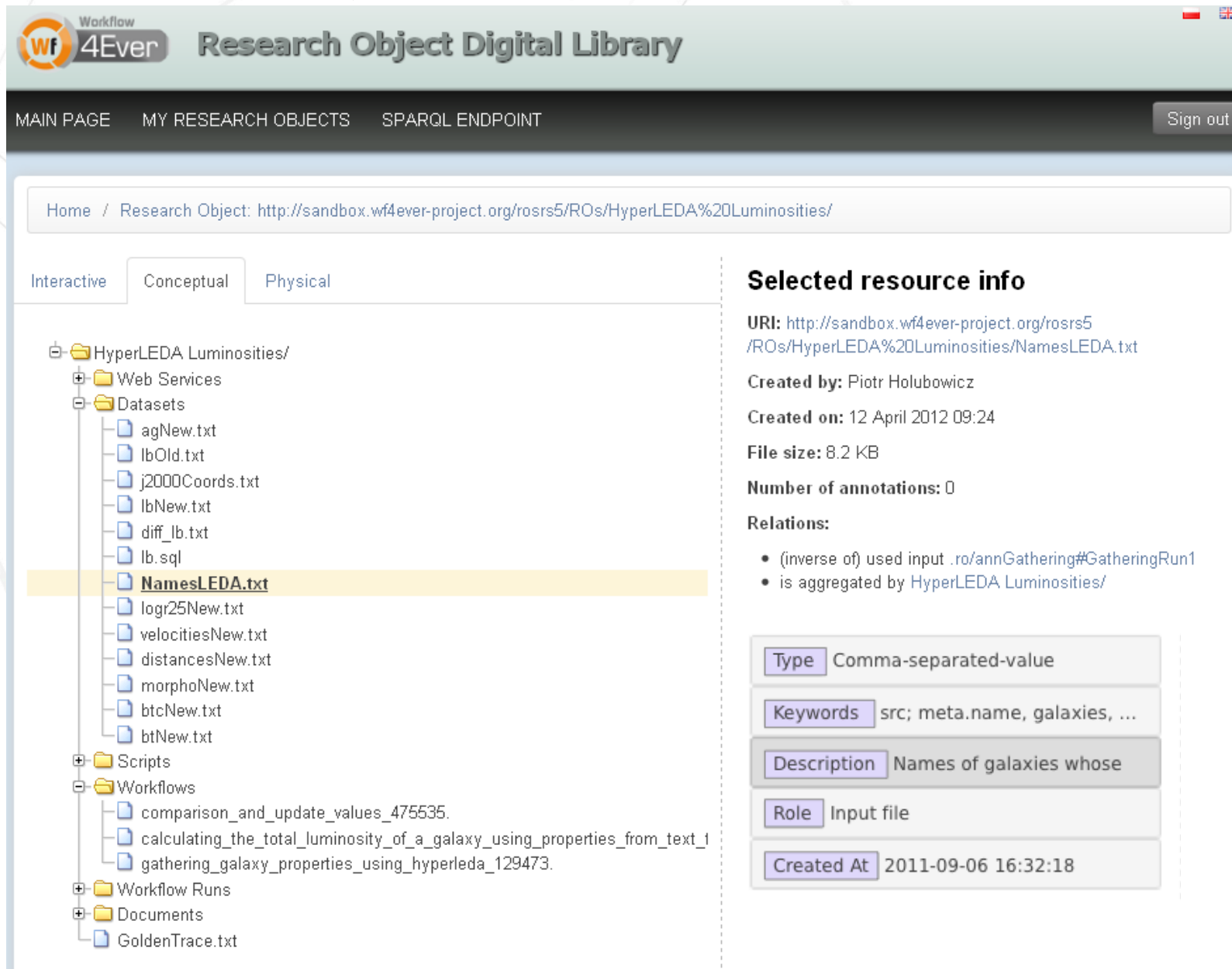
Home / Research Object: <http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDA%20Luminosities/>

Interactive Conceptual Physical

- HyperLEDA Luminosities/
 - Web Services
 - Datasets
 - agNew.txt
 - lbOld.txt
 - j2000Coords.txt
 - lbNew.txt
 - diff_lb.txt
 - lb.sql
 - NamesLEDA.txt**
 - logr25New.txt
 - velocitiesNew.txt
 - distancesNew.txt
 - morphoNew.txt
 - btNew.txt
 - Scripts
 - Workflows
 - comparison_and_update_values_475535.
 - calculating_the_total_luminosity_of_a_galaxy_using_properties_from_text_1
 - gathering_galaxy_properties_using_hyperleda_129473.
 - Workflow Runs
 - Documents
 - GoldenTrace.txt



Extract !



Workflow 4Ever Research Object Digital Library

MAIN PAGE MY RESEARCH OBJECTS SPARQL ENDPOINT Sign out

Home / Research Object: <http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDA%20Luminosities/>

Interactive Conceptual Physical

- HyperLEDA Luminosities/
 - Web Services
 - Datasets
 - agNew.txt
 - lbOld.txt
 - j2000Coords.txt
 - lbNew.txt
 - diff_lb.txt
 - lb.sql
 - NamesLEDA.txt**
 - logr25New.txt
 - velocitiesNew.txt
 - distancesNew.txt
 - morphoNew.txt
 - btcNew.txt
 - btNew.txt
 - Scripts
 - Workflows
 - comparison_and_update_values_475535.
 - calculating_the_total_luminosity_of_a_galaxy_using_properties_from_text_1
 - gathering_galaxy_properties_using_hyperleda_129473.
 - Workflow Runs
 - Documents
 - GoldenTrace.txt

Selected resource info

URI: <http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDA%20Luminosities/NamesLEDA.txt>

Created by: Piotr Holubowicz

Created on: 12 April 2012 09:24

File size: 8.2 KB

Number of annotations: 0

Relations:

- (inverse of) used input `./ro/annGathering#GatheringRun1`
- is aggregated by HyperLEDA Luminosities/

Type: Comma-separated-value

Keywords: src; meta.name, galaxies, ...

Description: Names of galaxies whose

Role: Input file

Created At: 2011-09-06 16:32:18

Home / Research Object: <http://sandbox.wf4ever-project.org/rosrs5/ROs>

Interactive Conceptual Physical

HyperLEDA Luminosities/

Web Services

Datasets

- agNew.txt
- lbOld.txt
- j2000Coords.txt
- lbNew.txt
- diff_lb.txt
- lb.sql
- NamesLEDA.txt**
- logr25New.txt
- velocitiesNew.txt
- distancesNew.txt
- morphoNew.txt
- btcNew.txt
- btNew.txt

Scripts

Workflows

- comparison_and_update_values_475535.
- calculating_the_total_luminosity_of_a_galaxy_using_properties_129473.
- gathering_galaxy_properties_using_hyperleda_129473.

Workflow Runs

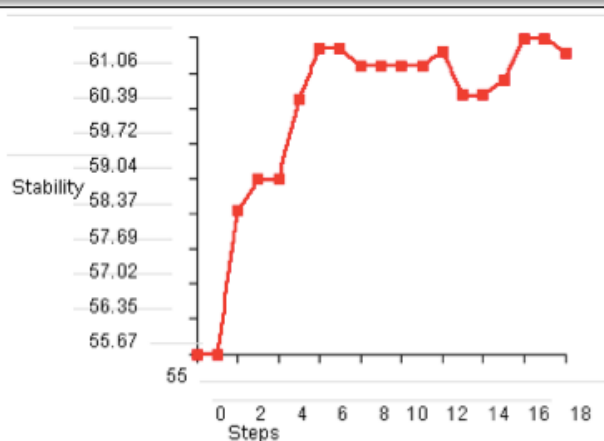
Documents

GoldenTrace.txt

A good understanding of the experiment is key for the reusability and repurposability



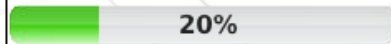



Stability: Changes made by different kind of users on the RO, can improve it or make it worse



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

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

	Runnable	Service up, software working, etc.
	Repeatable	The output can be reproduced
	Shareable	Enough annotation.
	Publishable	Service up, software working and all well commented

Decay: The health of the RO: state of the services (up or down), of the applications (updated or deprecated), permissions to access the input data

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

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

Rating 

Downloads 36

Citations [2]

Re-used [1]



Comments [4]

galaxies

Keywords are looked for in annotations describing the RO (not its resources).

Main view | Search results

Search results

1. [Internal Extinction calculation/](#)
by Susana Sanchez , Monday 12:52 Score: 54% 
2. [HyperLEDA Luminosities/ Propagation of properties extracted from the HyperLEDA catalog in the calculation of luminosities of galaxies](#)
by Piotr Holubowicz , 12 April 2012 08:56 Score: 34% 
3. [AstronomyPack/](#)
by ,

Search by
keywords

Past ratings
Ratings of similar users

Recommender | Search results

Recommended Research Objects

1. [Luminosities/](#)
by Susana Sanchez , Monday 17:14
2. [Internal Extinction calculation/](#)
by Susana Sanchez , Monday 12:52
3. [16/](#)
by Piotr Holubowicz , 14 June 2012 16:04

REUSABILITY AND REPURPOSABILITY

- Annotations for the **description** of the whole and its components
- **Visualization** of the relationships between the components
- **Versioning** of the whole or its components

SHARING

- Restricted **access** on data and processes
- Ensuring **authorship** and allow to RO to be **cited**

REPRODUCIBILITY

- **Completeness** checking
- **Decay monitoring** and notification, **reacting** to decay
- **Execution** and **interoperability** support

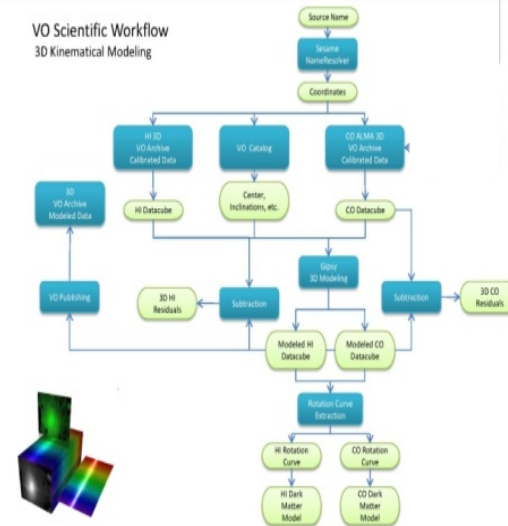
DISCOVERY

- **Semantic** discovery of ROs, processes, web services
- **Recommendation** capabilities.

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

	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

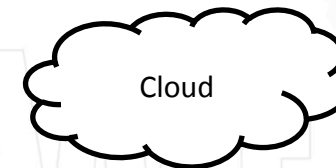
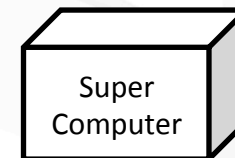
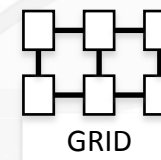
ASKAP Cubes Prof. Kevin Vinsen



VIRTUAL OBSERVATORY



- Standards for publishing and accessing astronomical data (Service Oriented Architecture)
- Data provider → analysis service provider



Susana Sánchez: sse@iaa.es

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

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

