Grant agreement no.: 27092

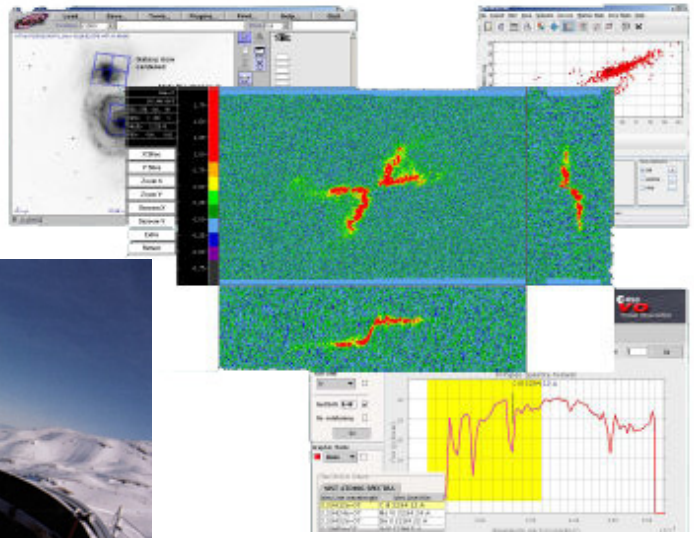
Workflows Preservation

José Enrique Ruiz, Lourdes Verdes-Montenegro, Susana Sánchez,
Juan de Dios Santander-Vela and the Wf4Ever Team
IAA-CSIC

January 18th 2012
7th Workflow Working Group Meeting - AS OV France

Who am I ?

Instituto Astrofísica de Andalucía - CSIC



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

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

What is Wf4Ever ?

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. Universisty of Oxford (OXF, UK)
6. Instituto de Astrofísica de Andalucía (IAA, Spain)
7. Leiden University Medical Centre (LUMC, NL)

What is Wf4Ever ?

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

Partners

- One SME
- Six public organizations

Technological Core Competencies

- Digital Libraries
- Workflow Management
- Semantic Web
- Integrity & Authenticity
- Provenance
- Information Quality

Case Studies

- Astronomy (IAA)
- Genome-wide Analysis and Biobanking (LUMC)

Goals

Archival, classification, and indexing of scientific workflows and their associated materials in scalable semantic repositories, providing advanced access and recommendation capabilities

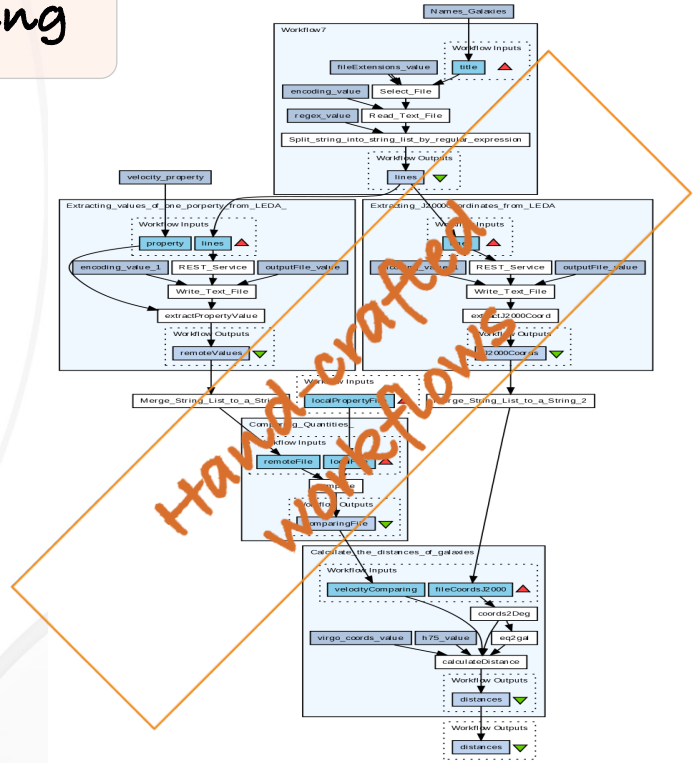
Creation of scientific communities to collaboratively share, reuse and evolve workflows and their parts, stimulating the development of new scientific knowledge

What are our Scientific workflows ?

Combination of data and processes into a configurable and structured set of steps that implement semi-automated computational solutions in problem solving

Types of workflows in Astronomy

- Personal script-based recipes
  Python, IDL, Software..
- Multi-archive VO recipes
- Internal group developments
  GRID, Clusters..
- Processing pipelines
  Provide Data, Computing Infrastructure, Tools..



Scientifically exploitable results vs. scientific insight
Easily accessible and reproducible (Shared)

Wfs on steroids !

Why workflow preservation is important ?

Astronomy research is entirely digital
Time has come to go "Beyond the PDF"

Preserved experiments

- Methodology "in action"
- All data are exposed
- Reproducible
- Repeatable — Trust assessment
- Re-usable
- Re-purposeable
- Participatory
- Collaborative
- Formative — Social aspect

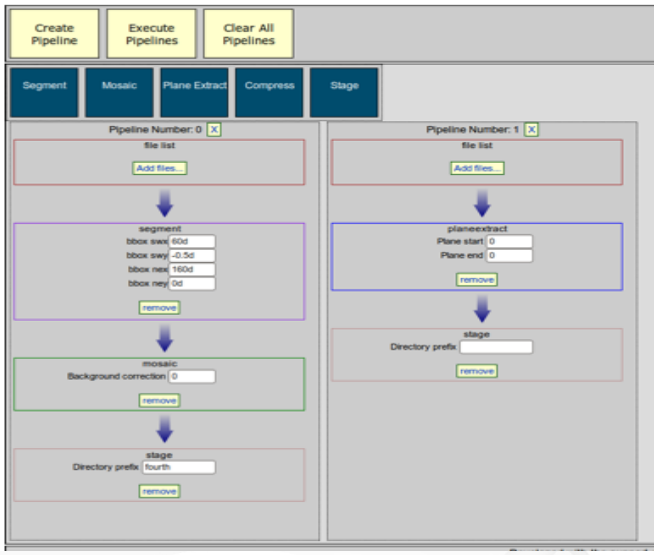
Discoverable !

Cyber-SKA

Provide infrastructure that will be required to address the needs of future radio telescopes such as the Square Kilometre Array

Web based workflow builder

- Image segmentation
- Image mosaicking (Montage)
- Spatial reprojection
- Plane extraction from data cubes



IceCore

University of Helsinki

Web portal for executing workflows – University of Helsinki

Common interface for Wfs distributed in different engine servers

Related Initiatives

Montage

- FITS Image Mosaicking
- Toolkit for Desktops, Clusters and Grids

Astro-WISE

- Distributed data storage and computing infrastructure
- Track process provenance of final data products
- Calibration and analysis of images

Helio-VO

- Solar physics Virtual Observatory
- Enable workflow execution via Taverna Server

Workflows VO France

- Provide use cases mainly oriented VO
- AÏDA Workflow System implements FITS validation with CharDM

Taverna

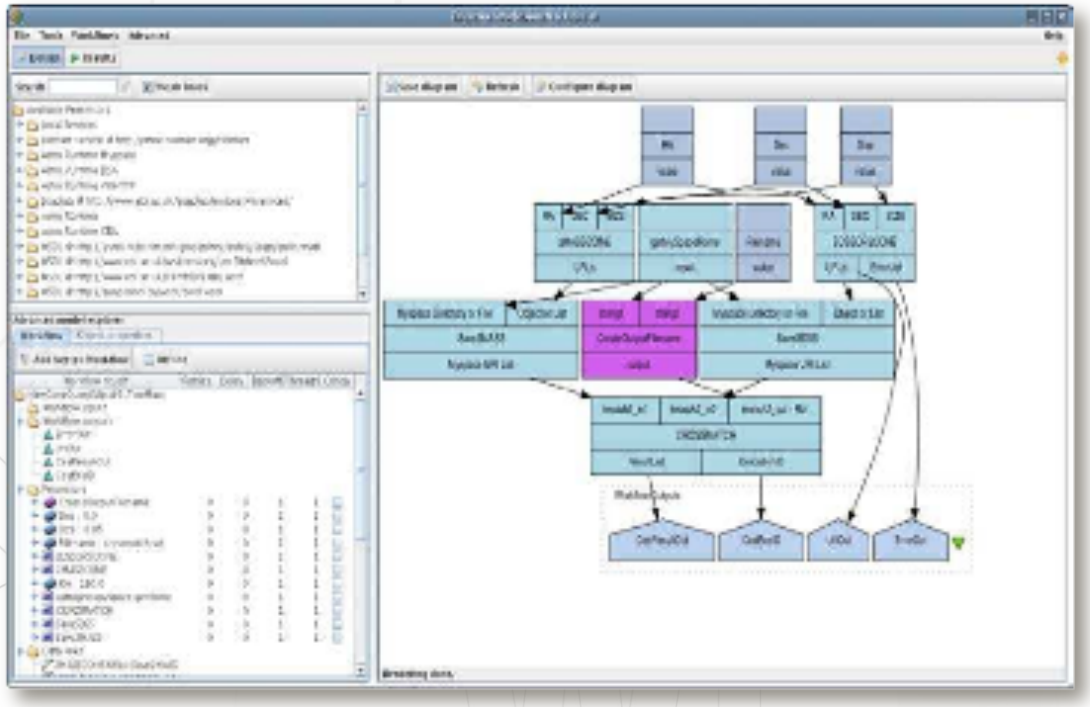
- Strongly typed bioinformatics
- Taverna Engine
- Taverna Server
- Taverna Workbench

Kepler

- Generic Science
- Workflow System

Triana

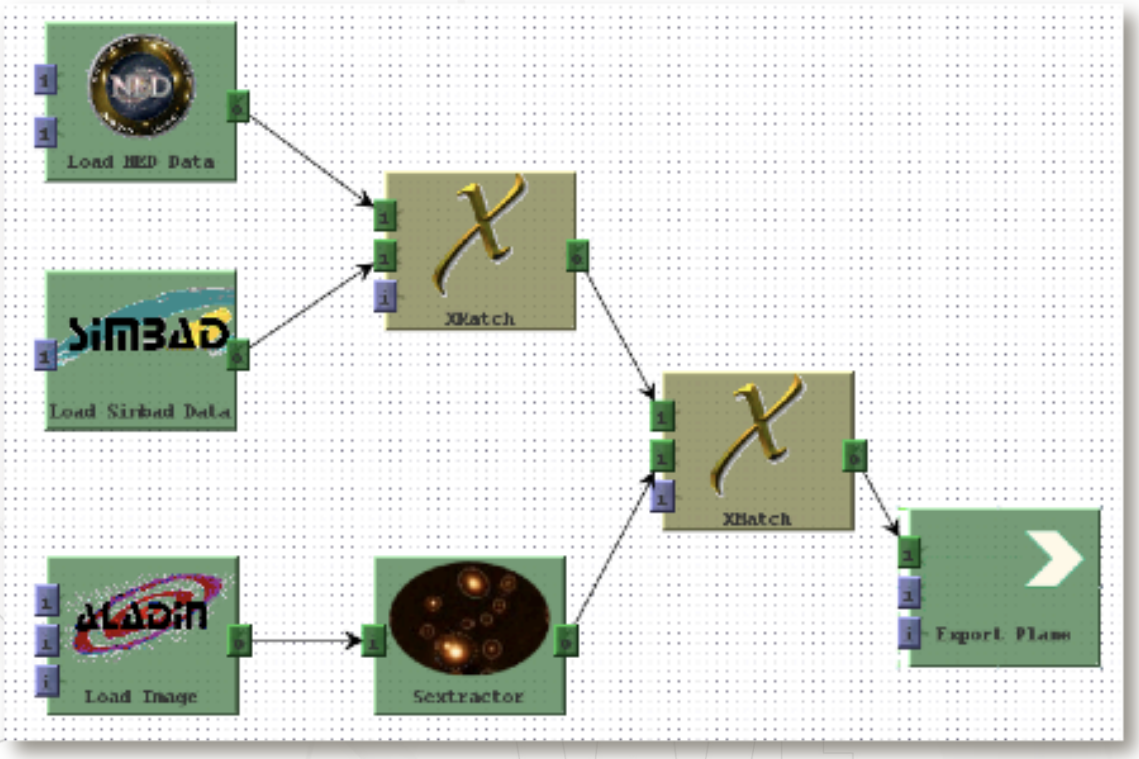
- Local execution
- Clusters RMI
- GRID
- Web Services



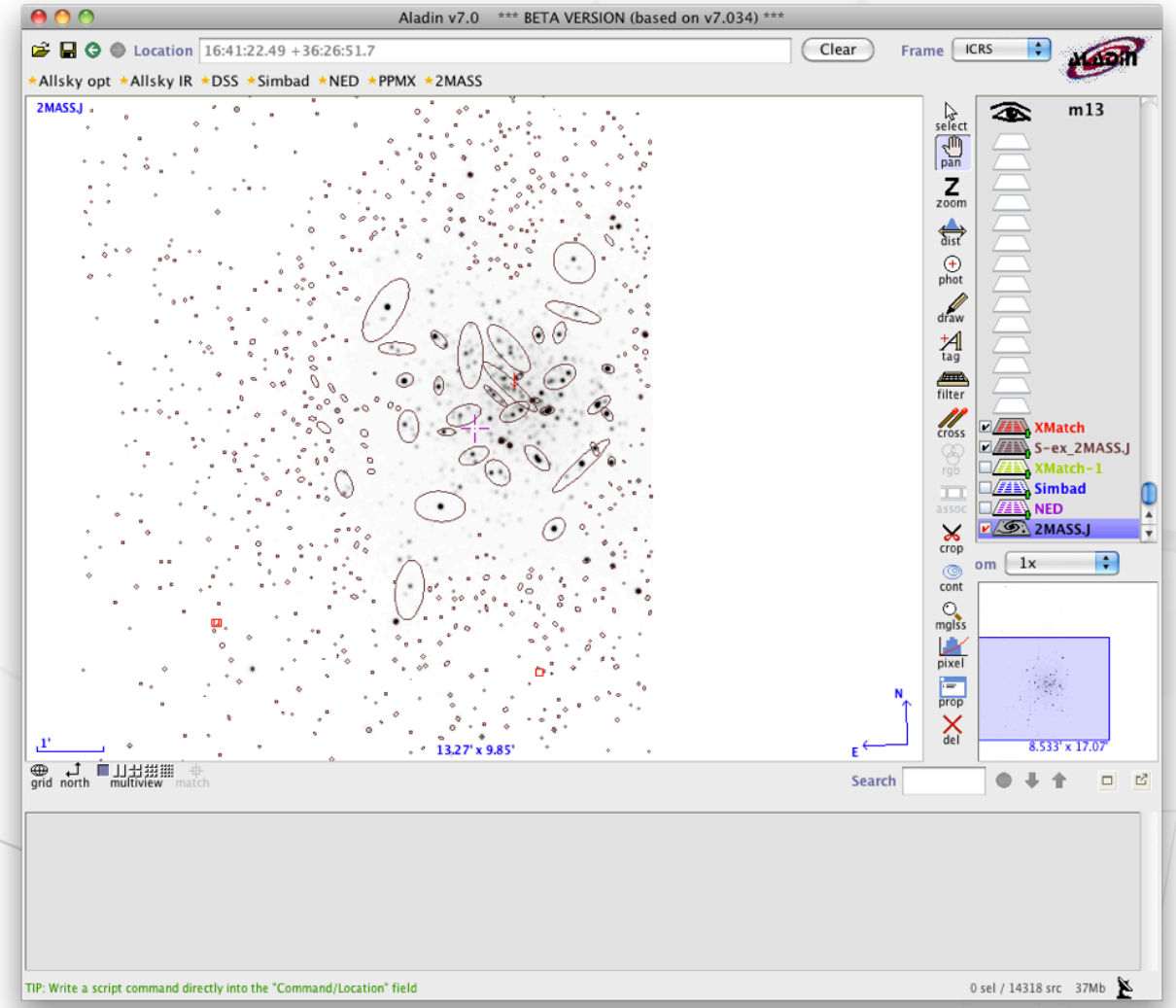
Aladin JLOW Plugin

Aladin plugin API permits graphical replacement of Aladin tools

```
#AJS
get NED m13 14'
get Simbad m13 14'
sync
xmatch NED Simbad 4
sync
get Aladin(2MASS,J) m13
sync
get SExtractor(2MASS.J)
sync
xmatch "XMatch*1" "SExtractor*"
sync
export -votable "XMatch*2"
```



Aladin JLOW Plugin

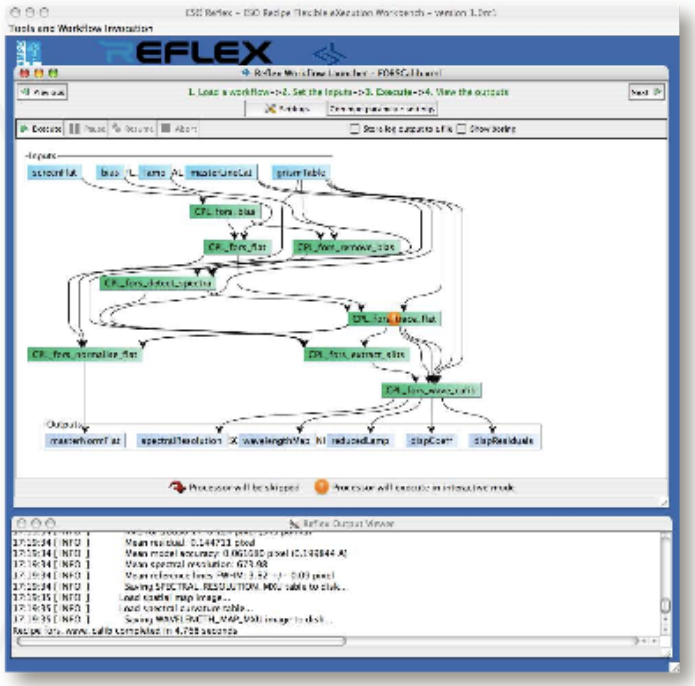


ESO Reflex

Finland's in-kind contribution to ESO

- Prototype/feasibility study
- Initially based on Taverna 1

Current implementation based on Kepler



AstroTaverna

AstroGrid Development

Prototype, marrying of VO Desktop & Taverna 1

Library of Taverna functions to access VO Desktop's API

Status

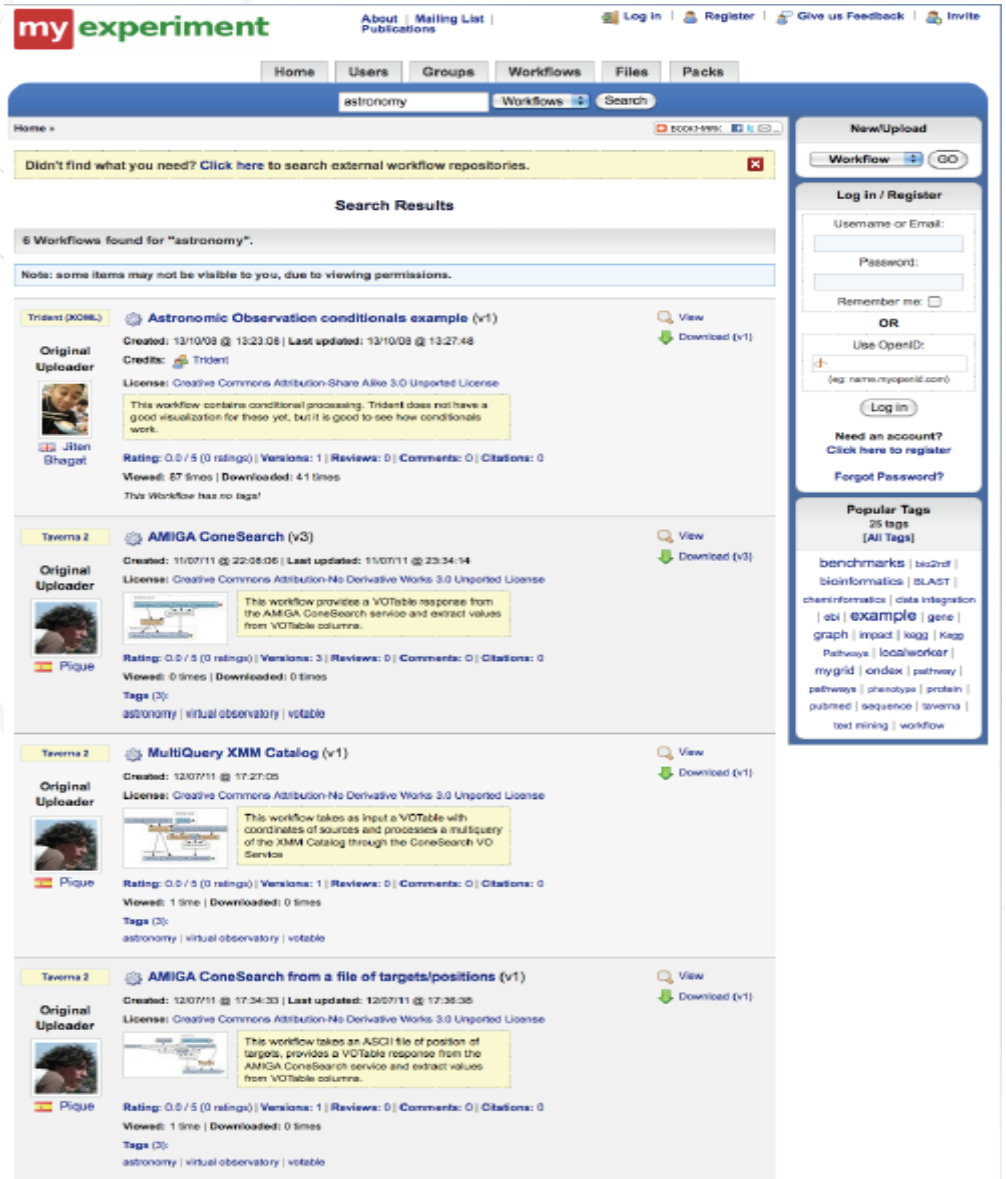
Wrapper libraries only for Taverna 1

The recipes store

Oxford e-Research Centre

- Find workflows
- Share workflows and files
- Find people
- Build communities
- Publish packages
- Tag workflows
- Score and rate workflows
- Comment on workflows
- Write reviews

Digital Repositories



Astronomy in MyExperiment

- 10 interested users
- No VO-services-based Wfs
- Some Helio Project Wfs
  - VOTables parsing
  - Internal services
  - Astro-Shims
- BioCatalogue vs. VORegistry

Astro-Wf4Ever specific Wfs

- Catalogue Queries

The upcoming context

Processes should benefit of the same privileges acquired by Data

Digital Libraries of Workflows may boost the use of the existing infrastructure of data (VO)

Users need templates !

Wf4Ever is also a project about

- How to publish
- How to do review by peers
- Improve visibility by reference and attribution

Publishers should play an import role

The next generation of archives

Much wider FoV and spectral coverage

- Huge sized datasets (~ tens TB)
- Big Data science highly dependent on I/O data rates
- Subproducts as virtual data generated on-the-fly

Automated surveys

- Huge amount of tabular data
- Services for Knowledge Discovery in Databases

The upcoming context

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.

Archives speaking self-descriptif web services

- Smaller virtual data subproducts
- Distributed, multi-archive, multi-wavelength astronomy

~~(Data)~~ Workflow preservation

- Interpreted through their execution
  - Complex models are required to describe them
- Severely vulnerable to obsolescence
  - Applications
  - Libraries
  - Operating environment
- Provenance is a complex issue in a cloud of services
- Resources are often beyond control of scientists
- Alleviate decay of external resources via alternates

~~(Data)~~ Workflow preservation

- Versioning of the whole or its components
- Restricted access on data and processes
- Permissions, licenses, platform, costs, etc.
- Semantic discovery of Wfs, processes, web services
- Metrics for quality: use stats, logs uptime, etc.
- Integrity evaluation
- Completeness checking
- Ensure trustworthiness and authenticity
- Workflows for workflow curation

Preserve, Retrieve, Reconstruct, Replay

- Retrieve
  - Functionality of the Wf or its modules
  - What are the inputs and outputs
  - Metadata, authority, keywords
- Reconstruct
  - Understand dependencies and components
  - Technical specificities
- Replay
  - Check the success of the preservation method

- Referenced and acknowledged

Characterization

Semantics and Modeling

Execution Tools

Long-term IDs

RO. The Research Object

All components related to the research lifecycle of an experiment should be available.

Preserved and easily retrievable

- Proposals
- Data
- Processes
- Publications

LINKED

Astronomy WP in Wf4Ever

Development and Implementation of Golden Exemplars

- Local catalogue curation based on VO Archives
- Sources extraction and crossmatching from 2D images
- Modeling and analysis of 3D velocity cubes of galaxies

Create a community of users

- Development of Prototypes and Tools
- Dissemination

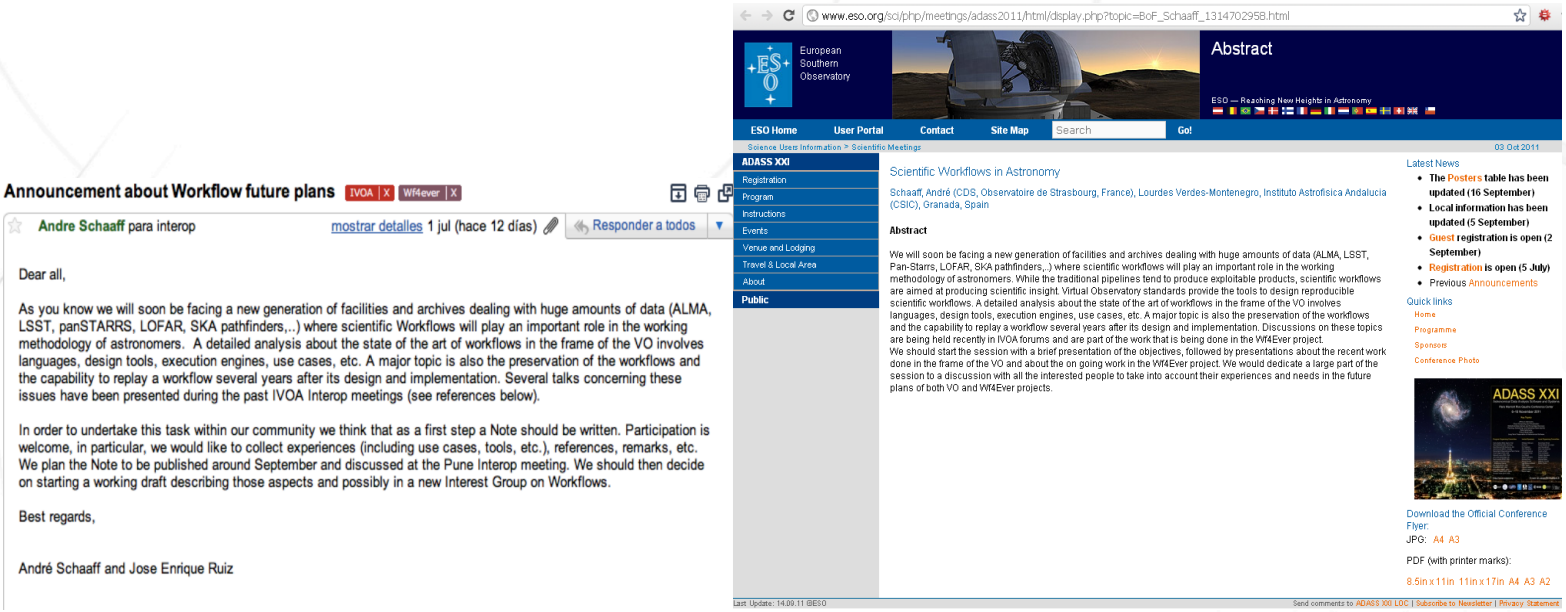
Integrate existing astronomy software with Wf4Ever Tools

- SAMP and WebSAMP

Provide interoperable models, ontologies and vocabularies for the characterization of workflows, processes and RO components

Astronomy WP in Wf4Ever

- Characterization of the Astronomy domain in Wf
- Detailed study of standards and web services in IVOA
- Exploration of similar initiatives for the curation of digital objects
- Sociological study and working methodology of astronomers
- Extraction of user and technical requirements
- Extraction of Taverna user requirements for Astronomy
- Implementation of first Golden Exemplar
- Early contacts in IVOA for the creation of a community of users



Announcement about Workflow future plans IVOA | X Wf4ever | X

Andre Schaaff para interop — mostrar detalles 1 jul (hace 12 días) — Responder a todos

Dear all,

As you know we will soon be facing a new generation of facilities and archives dealing with huge amounts of data (ALMA, LSST, panSTARRS, LOFAR, SKA pathfinders,..) where scientific Workflows will play an important role in the working methodology of astronomers. A detailed analysis about the state of the art of workflows in the frame of the VO involves languages, design tools, execution engines, use cases, etc. A major topic is also the preservation of the workflows and the capability to replay a workflow several years after its design and implementation. Several talks concerning these issues have been presented during the past IVOA Interop meetings (see references below).

In order to undertake this task within our community we think that as a first step a Note should be written. Participation is welcome, in particular, we would like to collect experiences (including use cases, tools, etc.), references, remarks, etc. We plan the Note to be published around September and discussed at the Pune Interop meeting. We should then decide on starting a working draft describing those aspects and possibly in a new Interest Group on Workflows.

Best regards,

André Schaaff and Jose Enrique Ruiz

www.eso.org/sci/php/meetings/adass2011/html/display.php?topic=BoF_Schaaff_1314702958.html

European Southern Observatory

Abstract

ESO — Reaching New Heights in Astronomy

ESO Home | User Portal | Contact | Site Map | Search | Go!

Science Users Information > Scientific Meetings — 03 Oct 2011

ADASS XXI

Registration
Program
Instructions
Events
Venue and Lodging
Travel & Local Area
About
Public

Scientific Workflows in Astronomy

Schaaff, André (CDS, Observatoire de Strasbourg, France), Lourdes Verdes-Montenegro, Instituto Astrofísica Andalucía (CSIC), Granada, Spain

Abstract

We will soon be facing a new generation of facilities and archives dealing with huge amounts of data (ALMA, LSST, Pan-Starrs, LOFAR, SKA pathfinders, ..) where scientific workflows will play an important role in the working methodology of astronomers. While the traditional pipelines tend to produce exploitable products, scientific workflows are aimed at producing scientific insight. Virtual Observatory standards provide the tools to design reproducible scientific workflows. A detailed analysis about the state of the art of workflows in the frame of the VO involves languages, design tools, execution engines, use cases, etc. A major topic is also the preservation of the workflows and the capability to replay a workflow several years after its design and implementation. Discussions on these topics are being held recently in IVOA forums and are part of the work that is being done in the Wf4Ever project. We should start the session with a brief presentation of the objectives, followed by presentations about the recent work done in the frame of the VO and about the on going work in the Wf4Ever project. We would dedicate a large part of the session to a discussion with all the interested people to take into account their experiences and needs in the future plans of both VO and Wf4Ever projects.

Latest News

- The Posters table has been updated (16 September)
- Local information has been updated (5 September)
- Guest registration is open (2 September)
- Registration is open (5 July)
- Previous Announcements

Quick links

Home
Programme
Sponsors
Conference Photo

ADASS XXI

Download the Official Conference Flyer:
JPG: A4 A3

PDF (with printer marks):
8.5in x 11in 11in x 17in A4 A3 A2

Last Update: 14.09.11 @ESO — Send comments to ADASS XXI LOC | Subscribe to Newsletter | Privacy Statement

Users' Requirements

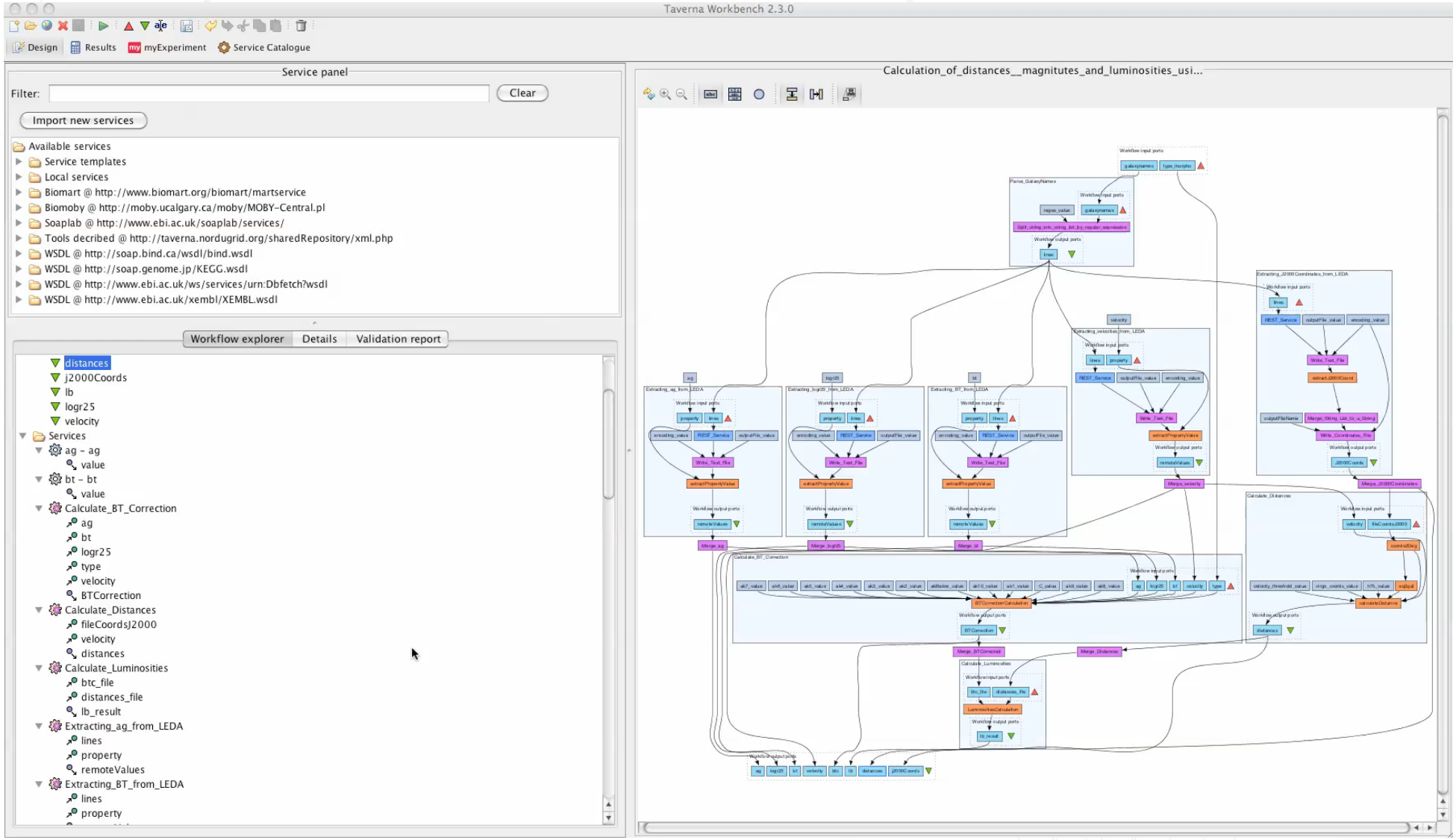
- Functional requirements for Wf4Ever "working" platform
- Focused on improving collaboration and reuse
- Interoperability in exchanging scientific methodology
- Expose experiment in a structured way to be understood by others

RO Modeling

- Model for interlinked components in a Research Object
- Strategies for assessing integrity and authenticity
- Attempts in metrics for Information Quality

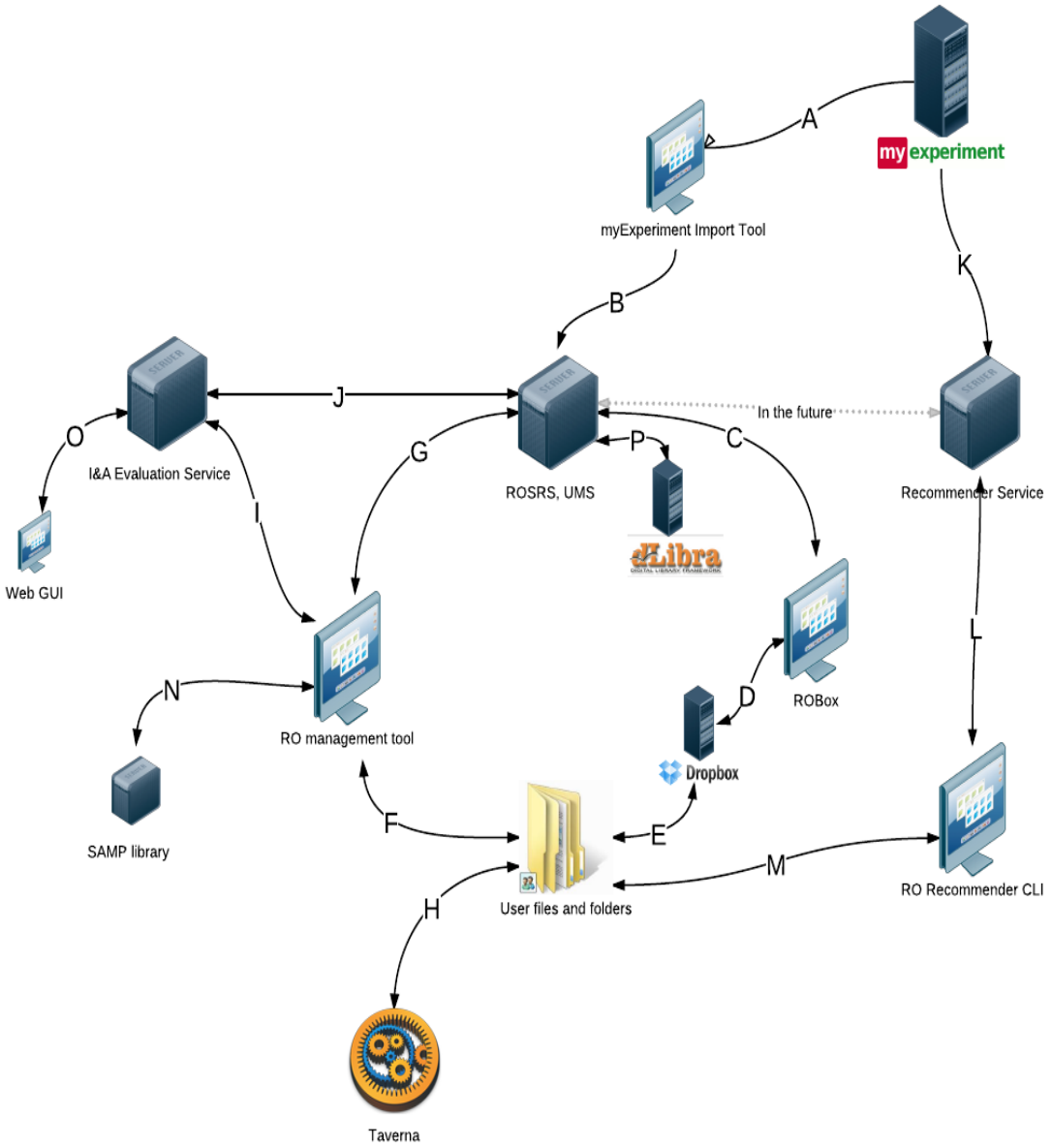
We need to build what we would like to preserve

Wf4Ever Update



Proposed improvements for Taverna

- VO Registry Access perspective
- STILTS VOTable Library Integration
- SAMP (Connectivity with VO Software)
- Python based Beanshells - Done
- Simple standard functions for Astronomy
- ODBC Connector to DB



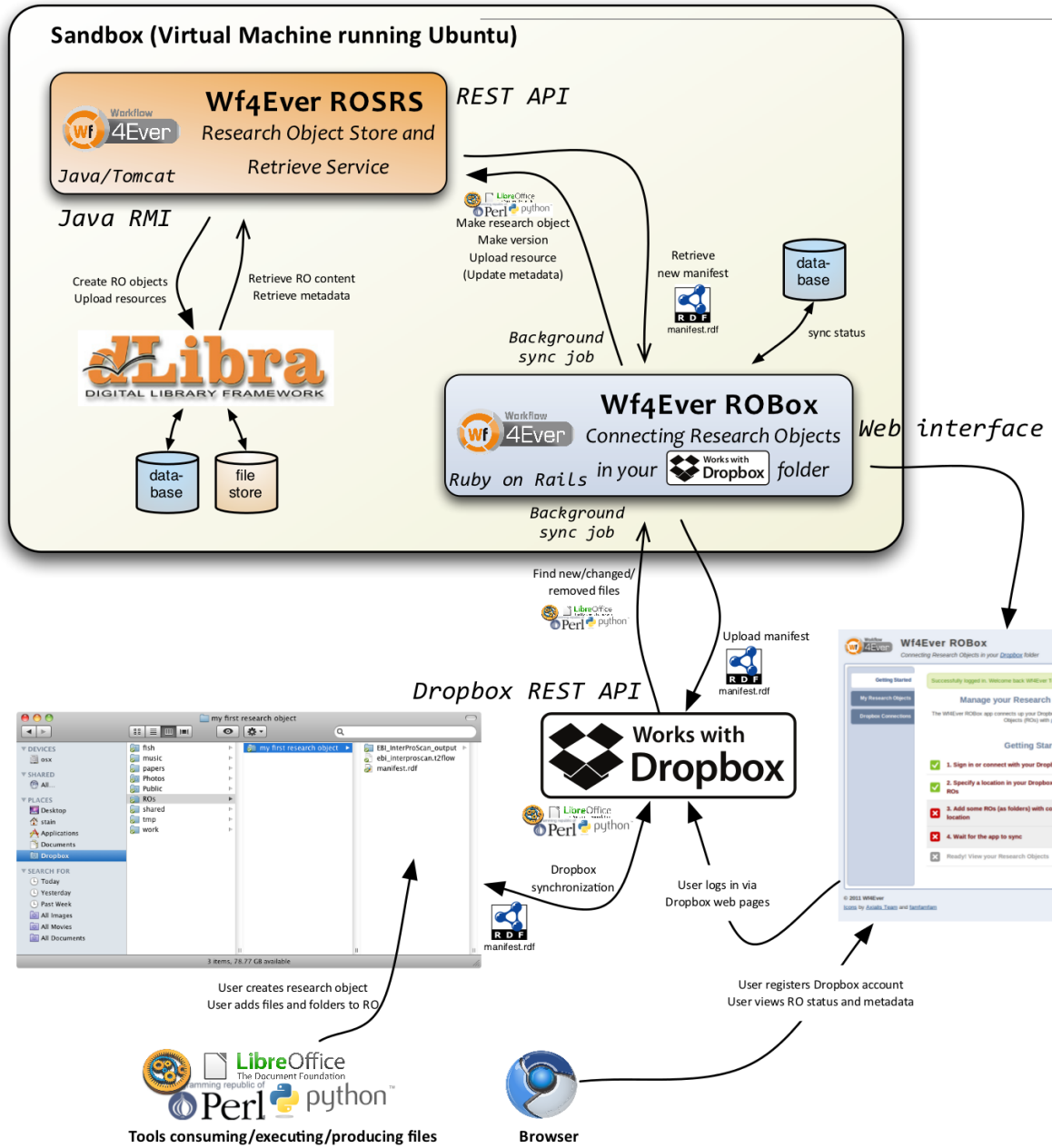
Architecture

- Search & Retrieval Service
- Recommender Service
- I&A Evaluation Service
- Notification Service

User-Tools Prototypes

- RO Command Line Tool
- RO Annotator
- RO Box

Architecture of Prototype 1



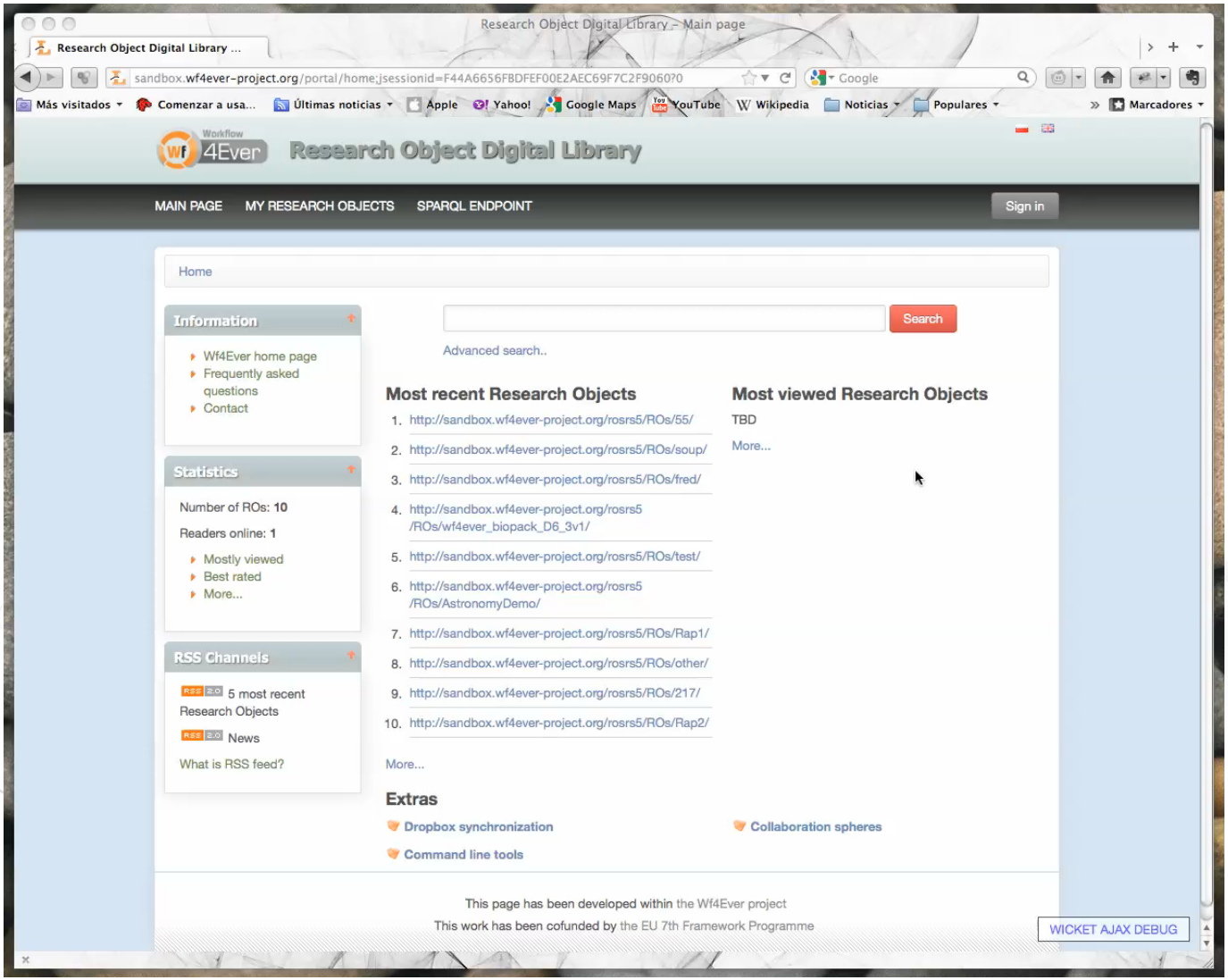
ROBox

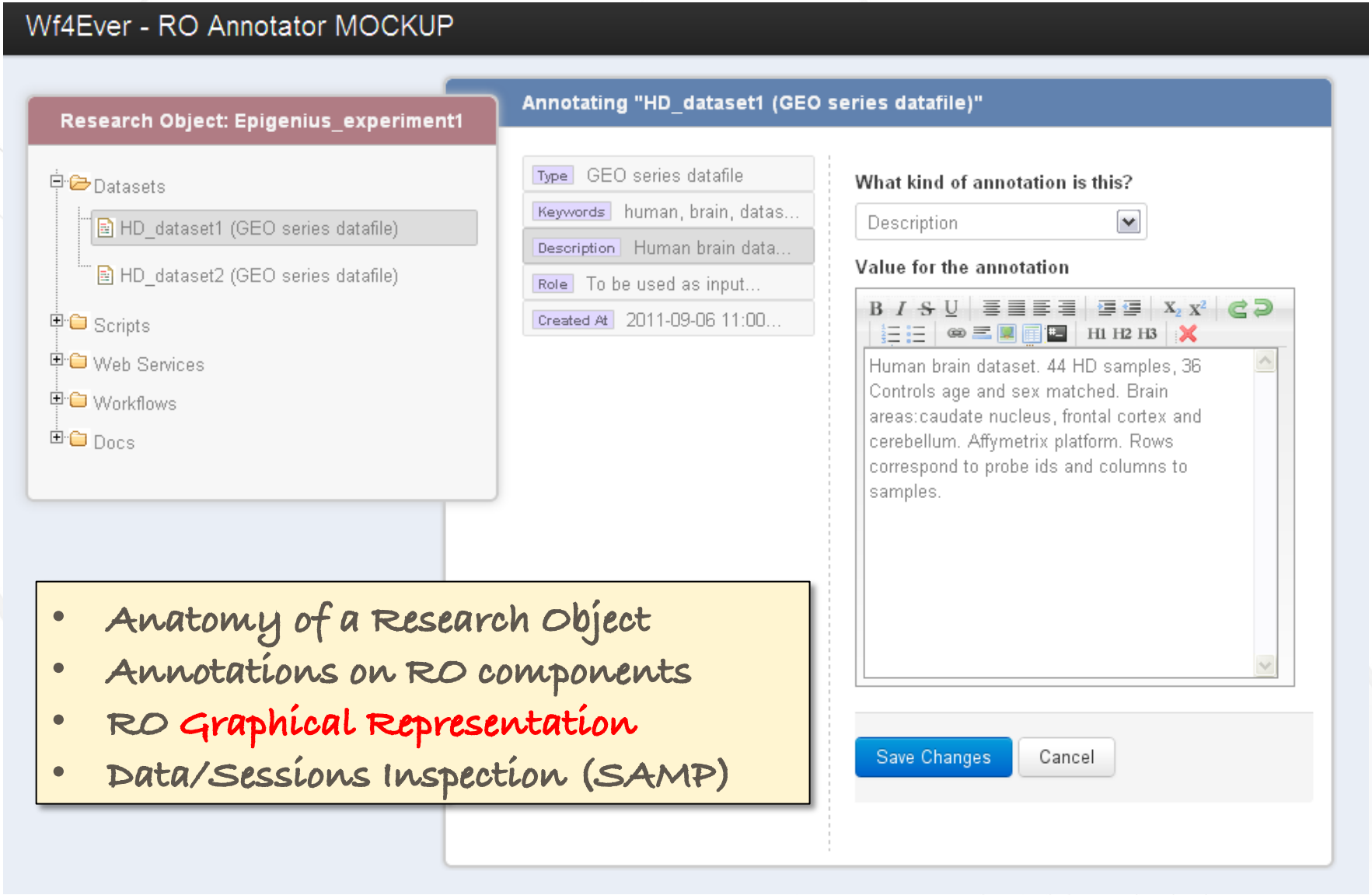
Seamless contribution to a working collaborative platform

A shared folder in Dropbox becomes a Working RO

Automatic generation of metadata

RO Digital Library and RO Import





- Anatomy of a Research Object
- Annotations on RO components
- RO Graphical Representation
- Data/Sessions Inspection (SAMP)

Wf4Ever Update

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

http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDALuminosities/

- RO
  - gathering_of_physical_properties_from_hyperleda_database_77
  - comparison_and_update_values_475535.
  - scripts.zip
  - data.zip
  - references.zip
  - CONTENT.txt
  - D5.3v1_1.0.pdf
  - calculating_the_total_luminosity_of_a_galaxy_using_properties_
  - calculating_the_total_luminosity_of_a_galaxy_using_properties_
  - README.txt
  - RECIPES.txt
  - sql.zip

Add folder | Add resource | Delete resource

Download metadata | Publish

Item info

Resource URI: http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDALuminosities/

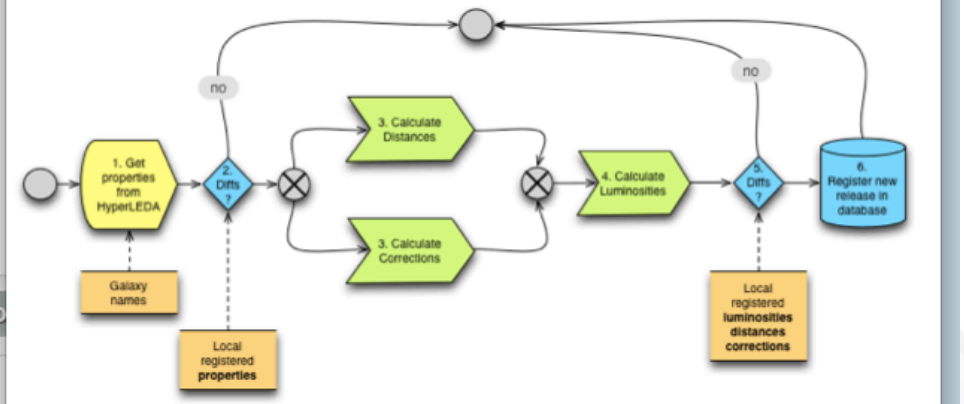
Download URI: http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDALuminosities/

Created by: Jose Enrique Ruiz

Created on: 2012.01.08 17:09:14 CET

Schema.png

Anterior | Siguiente | Zoom | Trasladar | Texto | Seleccionar | Anotación | Barra lateral



Annotating: http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLED

Annotations

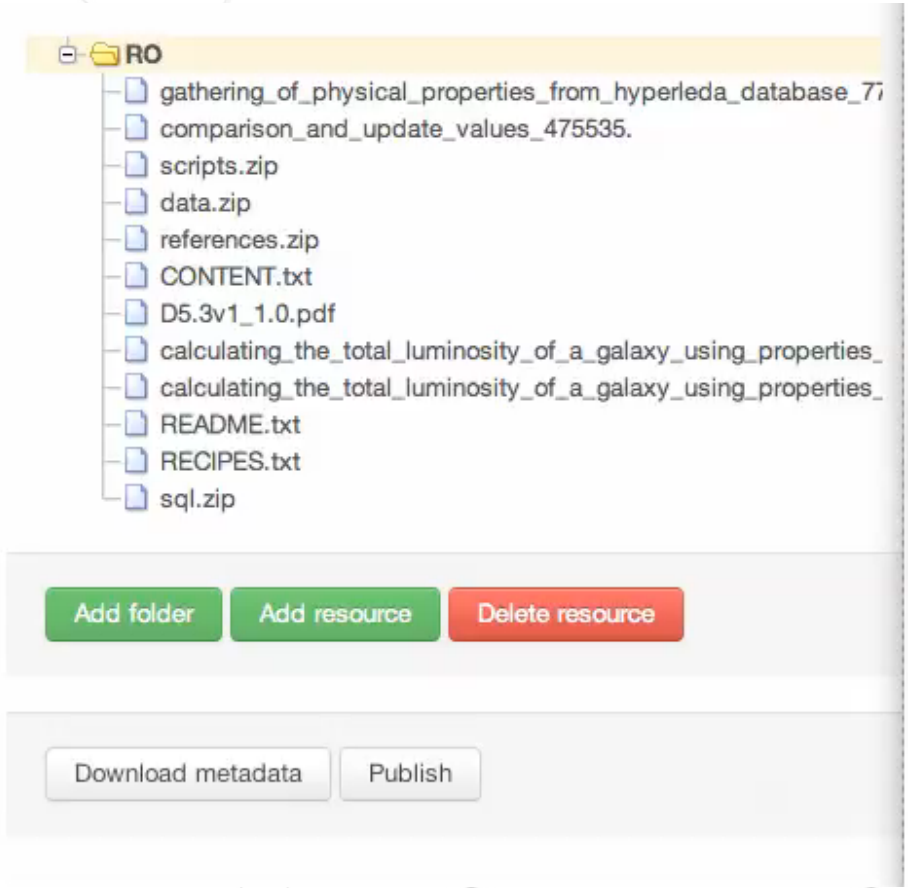
2012.01.08 17:12:35 CET	Jose Enrique Ruiz

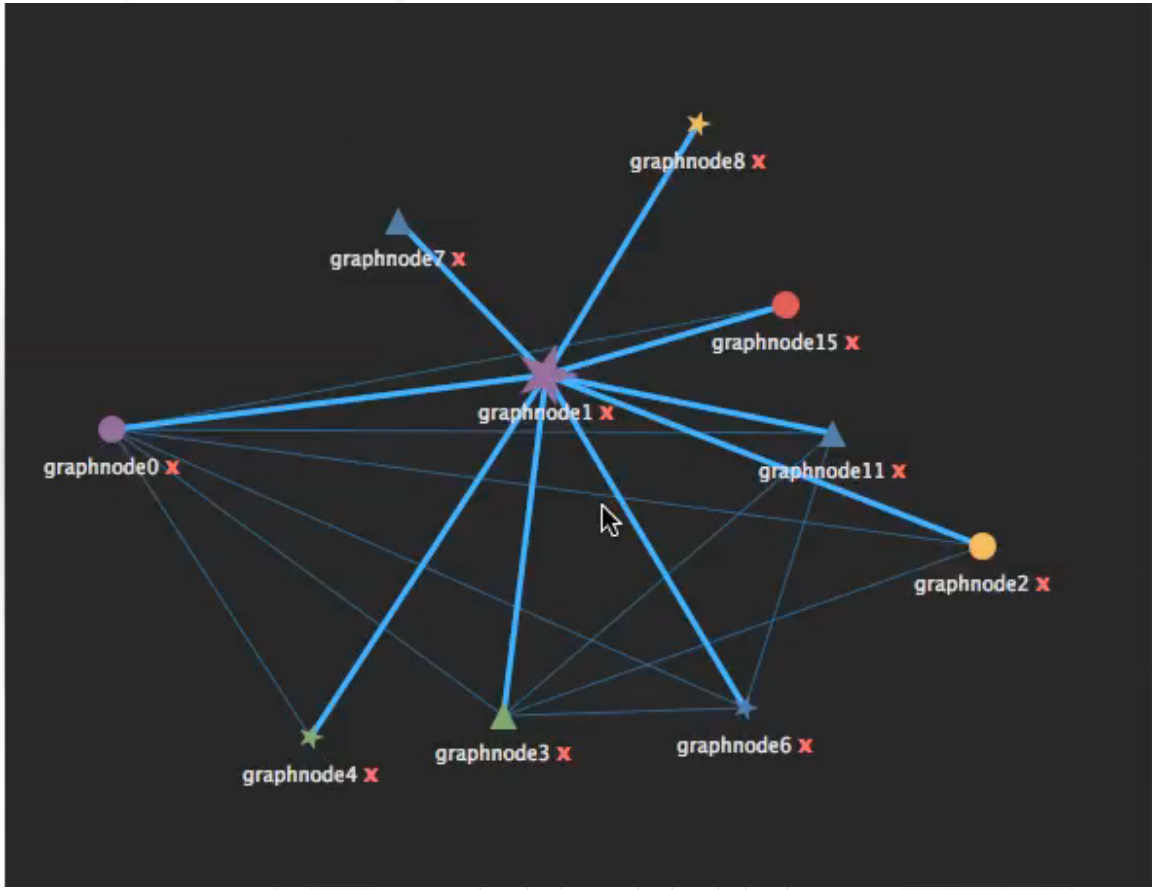
Add annotation | Delete annotation

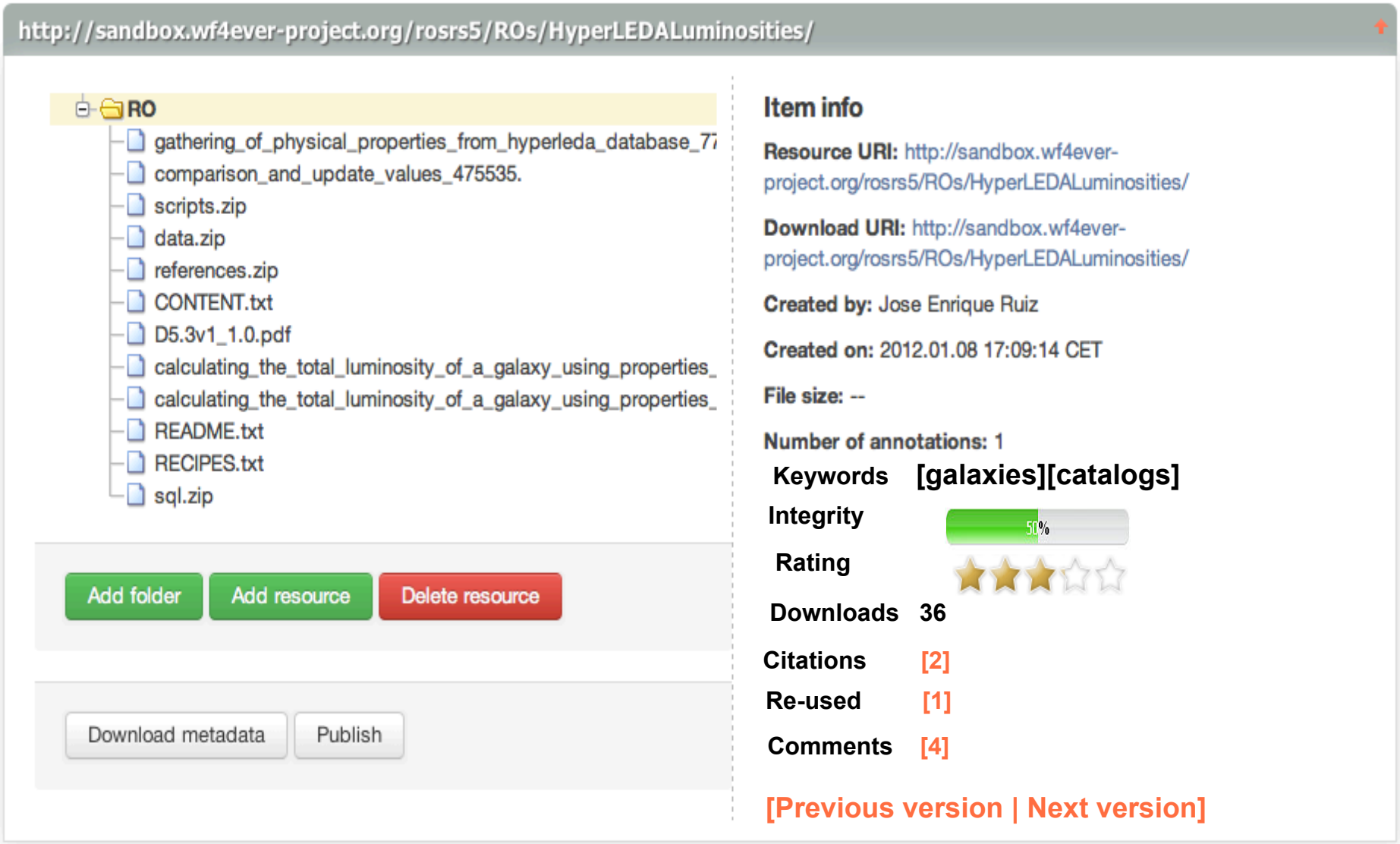
Statements

Add statement | Delete statement

RO Visualization



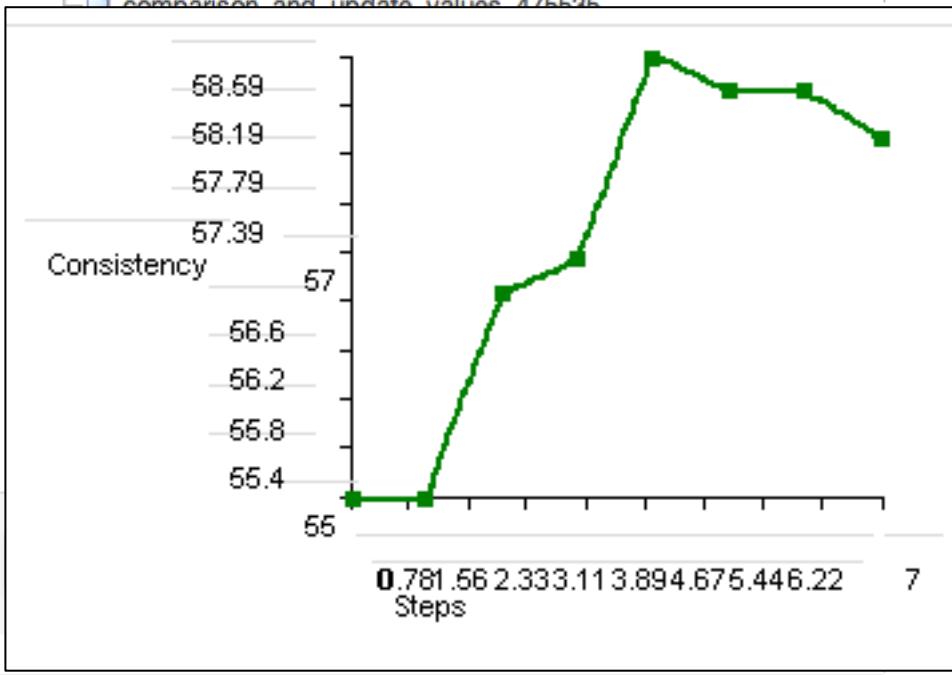



http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDALuminosities/
RO
gathering_of_physical_properties_from_hyperleda_database_77
comparison_and_update_values_475535.
scripts.zip
data.zip
references.zip
CONTENT.txt
D5.3v1_1.0.pdf
calculating_the_total_luminosity_of_a_galaxy_using_properties_
calculating_the_total_luminosity_of_a_galaxy_using_properties_
README.txt
RECIPES.txt
sql.zip
Add folder
Add resource
Delete resource
Download metadata
Publish
Item info
Resource URI: http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDALuminosities/
Download URI: http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDALuminosities/
Created by: Jose Enrique Ruiz
Created on: 2012.01.08 17:09:14 CET
File size: --
Number of annotations: 1
Keywords [galaxies][catalogs]
Integrity 50%
Rating
Downloads 36
Citations [2]
Re-used [1]
Comments [4]
[Previous version | Next version]

http://sandbox.wf4ever-project.org/rosrs5/ROs/HyperLEDALuminosities/

RO

gathering_of_physical_properties_from_hyperleda_database_77



1. Alice adds Recipe.txt
2. John adds Results2.txt
3. Alice adds Bibliography.pdf
4. John removes Scrtip.py
5. John edits annotation on Recipe.txt
6. Unknown adds Dropme.txt

Item info

Resource URI: http://sandbox.wf4ever-

Download metadata | Publish

Re-used [1]

Comments [4]

[Previous version | Next version]

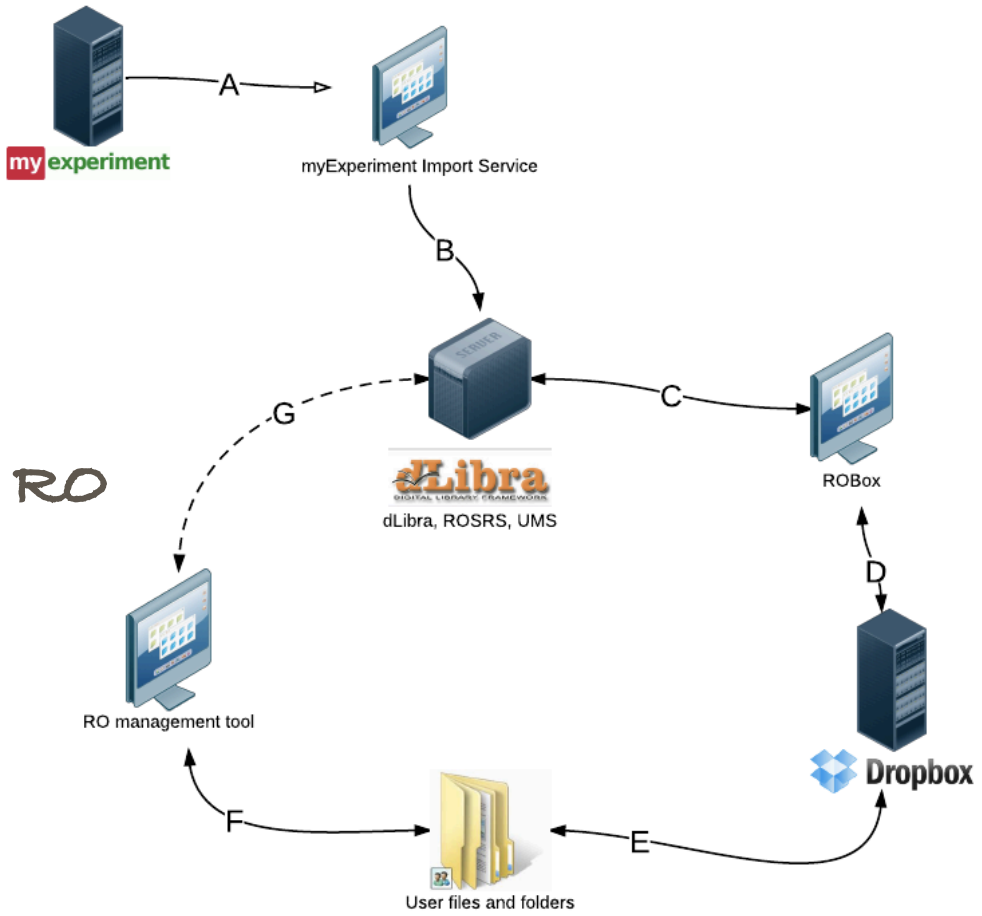
Notification Service for Authors

What should be notified ?

- Fails
- Downloads
- Annotations
- Linked/Similarity
- Modifications on Working RO
- Acknowledgements

Notification Management Tool

Avoid spam



Astronomy WP in Wf4Ever

Astronomy WP

- Development and Implementation of "Extraction of Sources"
- Development and Implementation of "Modelling of 3D Data"
- Explore experiments subject to be migrated to Wf/RO methodology
- Contribute to IVOA in Semantics for Processes

Other WPs

Continue Providing Feedback

- RO Model, Architecture, Integrity & Authenticity, Information Quality, etc.
- Software integration and improved functionalities (SAMP, Taverna, etc.)
- Prototypes for management and visualization of RO

Community engagement

- Approach Astro-Informaticians
- Continue pushing in the IVOA Community
- Tackle collaboration with Publishers

VO France Workflows Group

Distributed data analysis in the VO

- Panchromatic, multi-archive, multi-facility
- Executes in the VO Infrastructure
- Orchestration of simple services

Present processing pipelines

- Produce exploitable data
- Provenance modeling
- VO compliant data

Workflows VO Characterization

- Inputs
- Outputs
- Processes
- Descriptions
- Metadata
- Etc..

Data processing from the VO

- Provide custom re-processing to VO users
- Virtual data generation through UWS in VOSpace

Related activities in the VO

IVOA Working Groups

- **Data Modeling**
  Characterization, Provenance..
- **Semantics**
  Ontologies, Vocabularies for Processes
- **Data Access Layer**
  TAP, self-descriptive Protocols..
- **Grid and Web Services**
  UWS, VOSpace, SSO..
- **Applications**
  SAMP
- **IG. KDD**
  Knowledge Discovery and Data Mining
- **IG. Data Curation and Preservation**
  Persistent Identifiers, Curation of VO Resources..
  Wf4Ever Project, US VAO semantic linking of proposals, publications, data

IVOA Note

Scientific Workflows in the VO

André Schaaff & Jose Enrique Ruiz

workflow@ivoa.net

Questions

More info

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

http://www.wf4ever-project.org

workflow@ivoa.net