e-Science for the SKA

WF4Ever: Supporting Reuse and Reproducibility in Experimental Science

Lourdes Verdes-Montenegro*
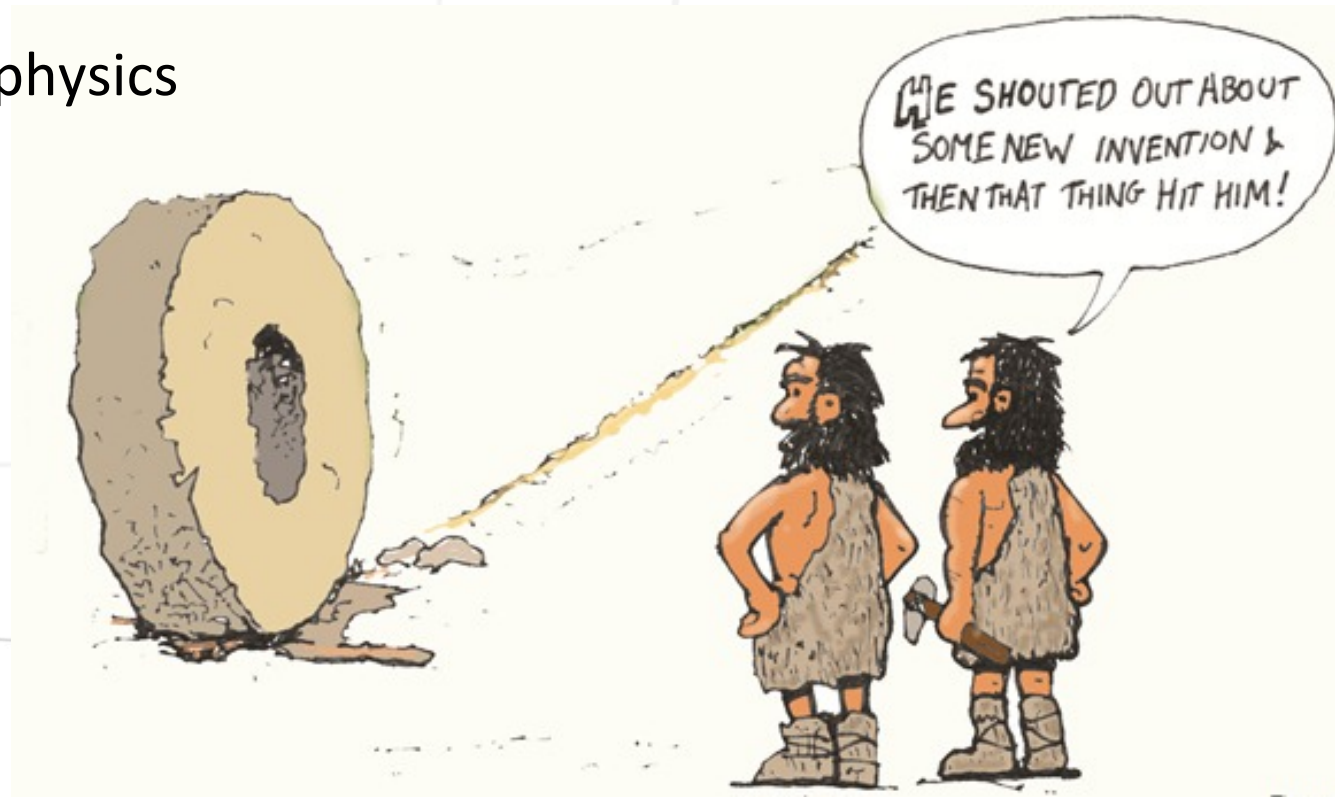
AMIGA and Wf4Ever teams

Instituto de Astrofísica de Andalucía – CSIC

*PI of VIA-SKA, Feasibility study of the Spanish technological participation in SKA

RadioNet Advanced Radio Astronomy, Commissioning Skills and Preparation for the SKA
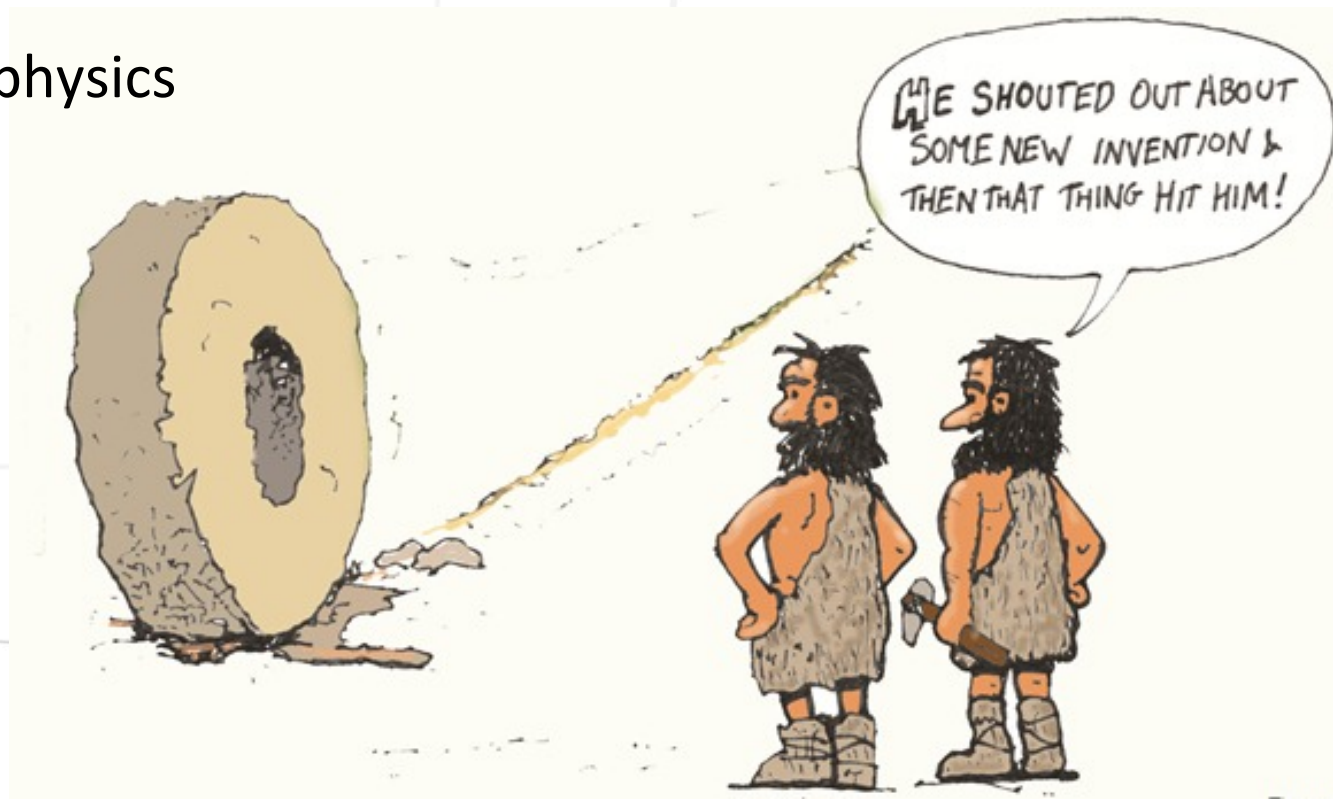
Manchester, November 15th 2012

» The challenge: Reuse and Reproduce scientific experiments

- Reusability, fundamental for incremental scientific development

- Reproducibility, key for reliable science

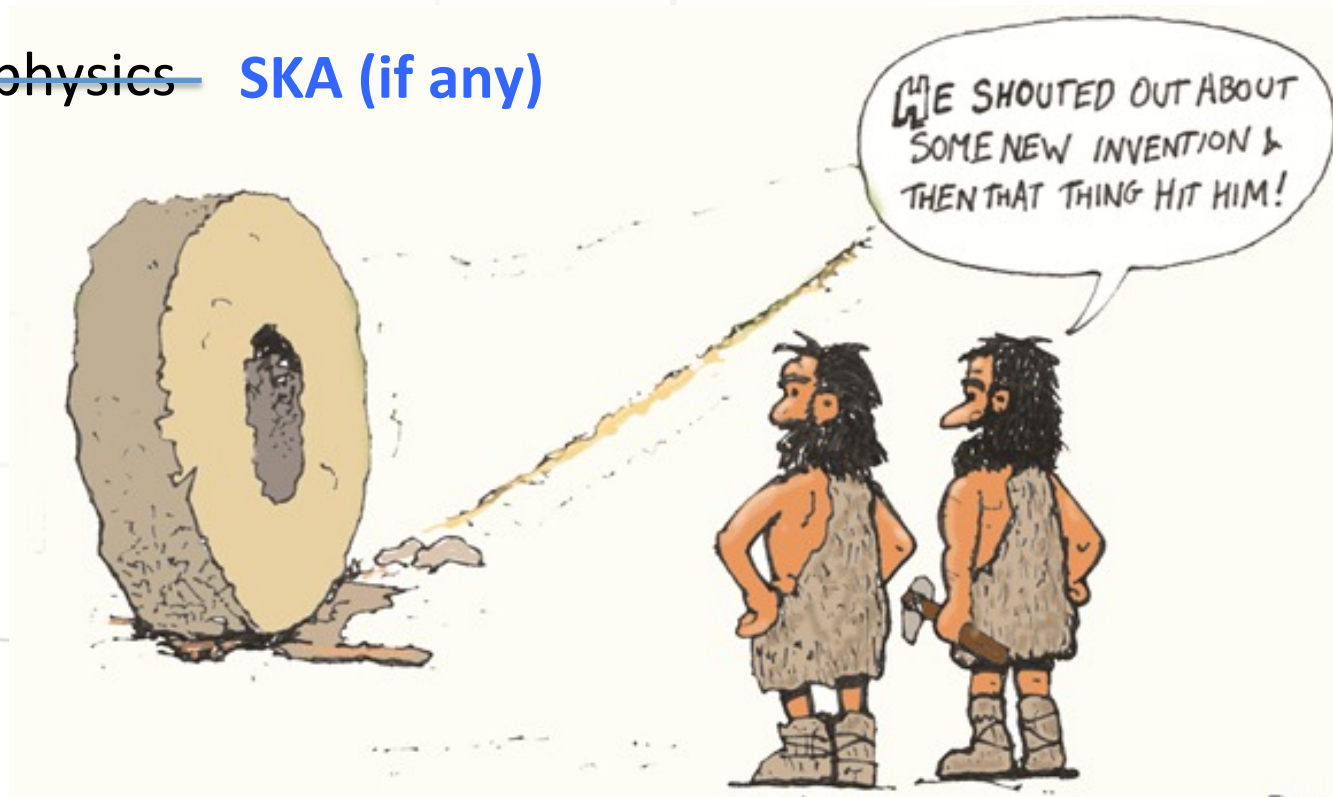» How Wf4Ever is addressing these challenges

» Impact in Astrophysics

And commissioning looks soooo similar to a scientific experiment...

» 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

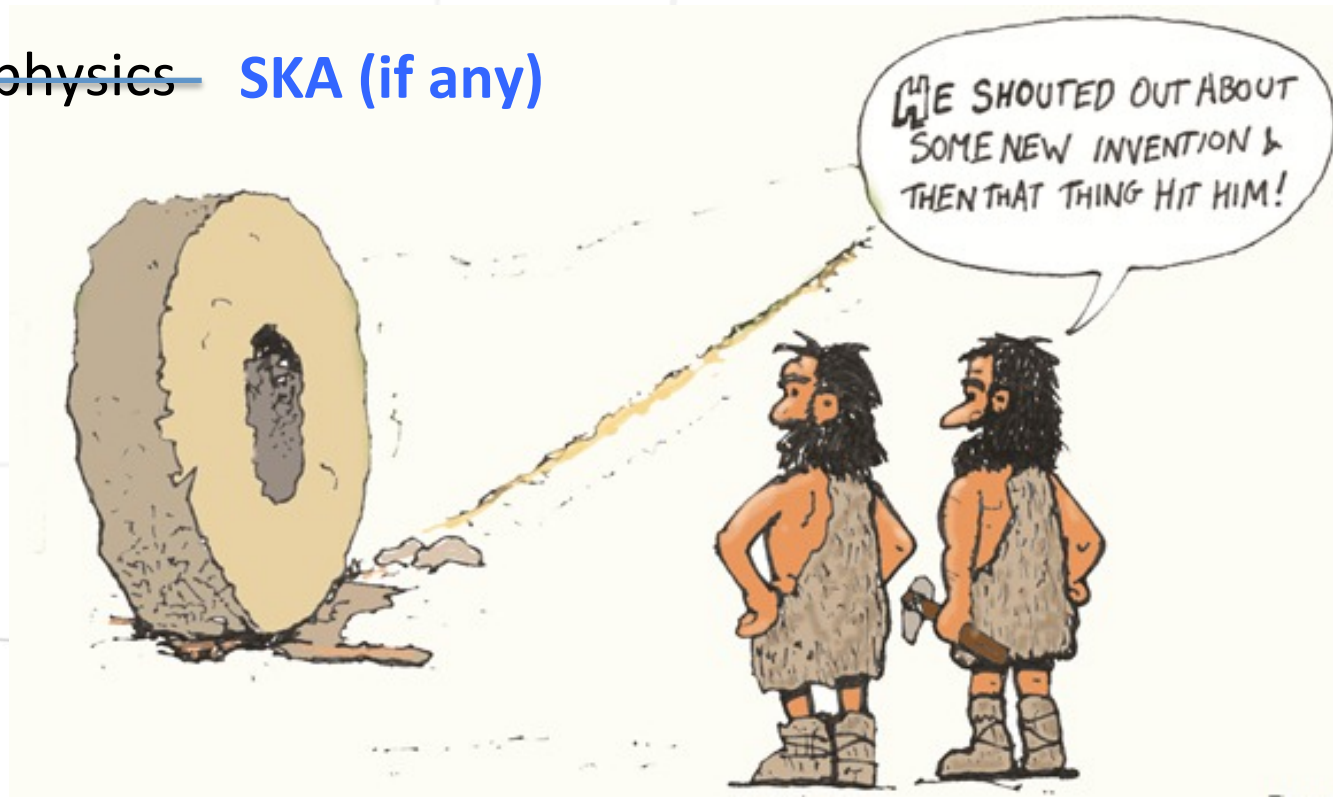# And commissioning looks soooo similar to a scientific experiment...

» 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~~ **SKA (if any)**

HE SHOUTED OUT ABOUT SOME NEW INVENTION & THEN THAT THING HIT HIM!

**And commissioning looks soooo similar to a scientific experiment...**

An experiment is <u>reproducible</u> when someone else, working independently, can get the same results following the same methods and using the same inputs

» How Wf4Ever is addressing these challenges

» Impact in ~~Astrophysics~~ **SKA (if any)**

HE SHOUTED OUT ABOUT SOME NEW INVENTION & THEN THAT THING HIT HIM!

And commissioning looks soooo similar to a scientific experiment...

An experiment is <u>reproducible</u> when someone else, working independently, can get the same results following the same methods and using the same inputs

mmm... commissioning

» How Wf4Ever is addressing these challenges

» Impact in ~~Astrophysics~~ **SKA (if any)**
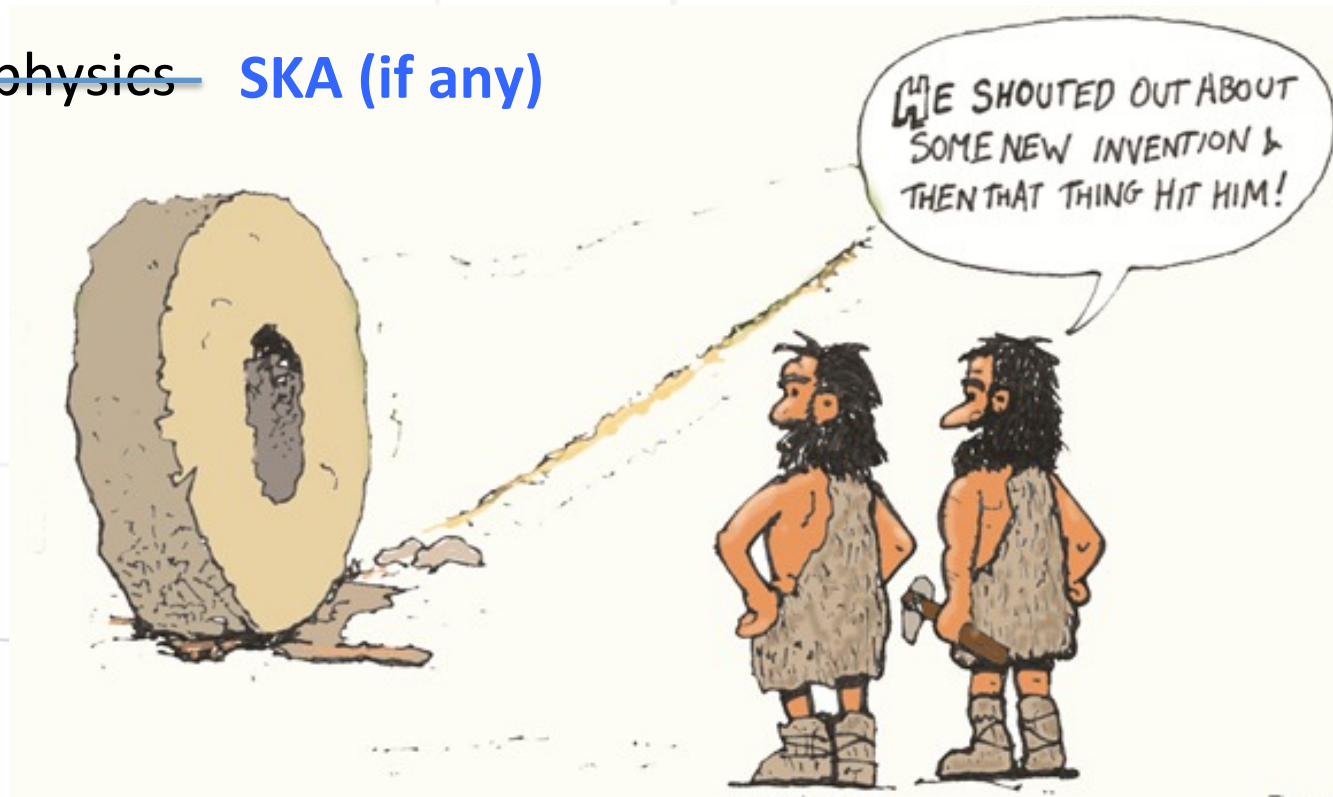
» Compile all needed elements

 » Share it with other … Commissioners

  » Describe it: methods, data,etc.

   Make the experiment discoverable

» Scientific Workflows: part of the solution

 › Automation, encourage best practices, method for sharing

» But more is needed:

 › Share/provide the data, annotations, references, etc.

 › Strategies for avoiding decay

 › Tools for discovering the experiment
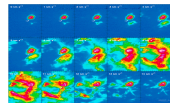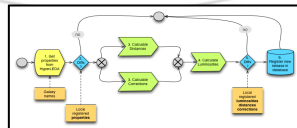
Workflow preservation

Wf4Ever - Preservation of scientific workflows in data-intensive science

EU funded FP7 STREP Project
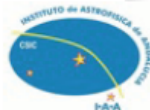December 2010 – December 2013

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)

- Encapsulate the scientific methodology (the workflows and all the associated information/components of the digital experiment) in an artefact called Research Object (beyond the PDF).

- What to do with the Research Object?
  - Archival, classification and indexing in scalable semantic repositories
  - 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

- Use Cases:
  - Astronomy (IAA)
  - Genome-wide Analysis and Biobanking (LUMC)

- Encapsulate the scientific methodology (the workflows and all the associated information/components of the digital experiment) in an artefact called Research Object (beyond the PDF).

- What to do with the Research Object?
  - Archival, classification and indexing in scalable semantic repositories
  - 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

Commissioning is multidisciplinary

- Use Cases:
  - Astronomy (IAA)
  - Genome-wide Analysis and Biobanking (LUMC)

» Model for Workflow-Centric Research Objects [1]
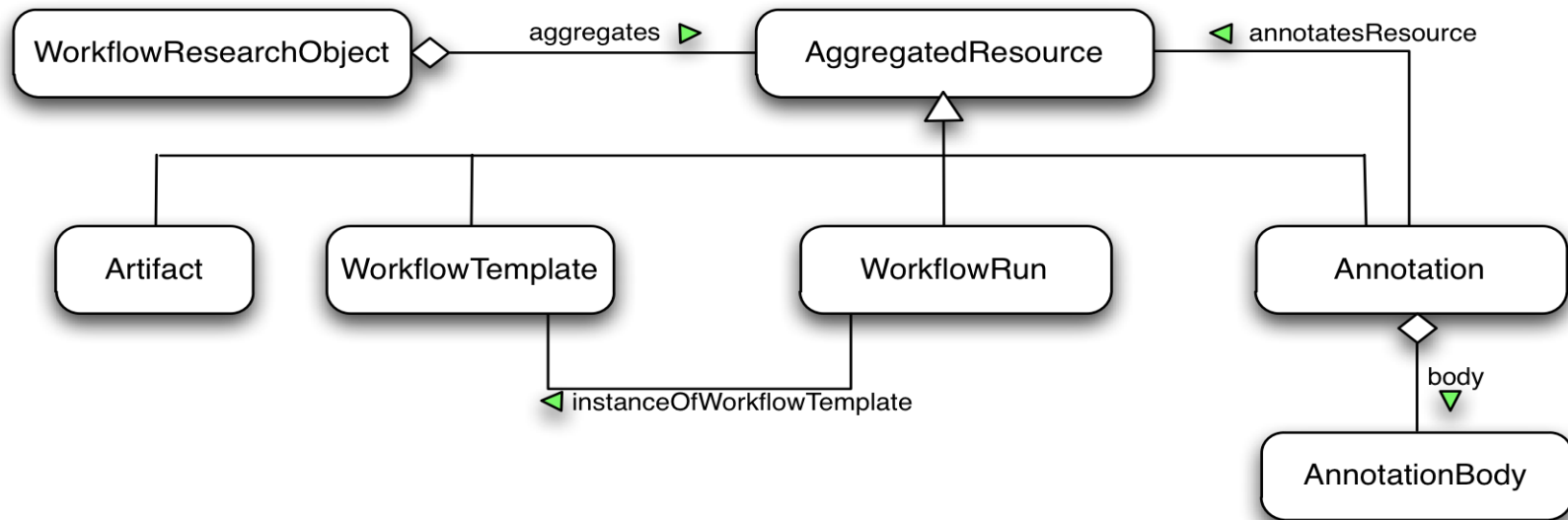
» Semantic ontologies to implement the model

› Object Exchange and Reuse (ORE) for specifying aggregation of resources

› Annotation Ontology (OA) for annotating the resources

Metadata of the processes

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

» Provide digital libraries with RO preservation functionalities

» Provide digital libraries with RO preservation functionalities

Search for trustable reusable material for commissioning (e.g from other disciplines)

Useful for complex processes as commissioning?

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

Metrics and tools for QA Evolution with time

|		
75%	Runnable	Service up, software working, etc.
60%	Repeatable	The output can be reproduced
20%	Shareable	Enough annotation.
50%	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

Stability

61.06
60.39
59.72
59.04
58.37
57.69
57.02
56.35
55.67
55

Steps
0 2 4 6 8 10 12 14 16 18
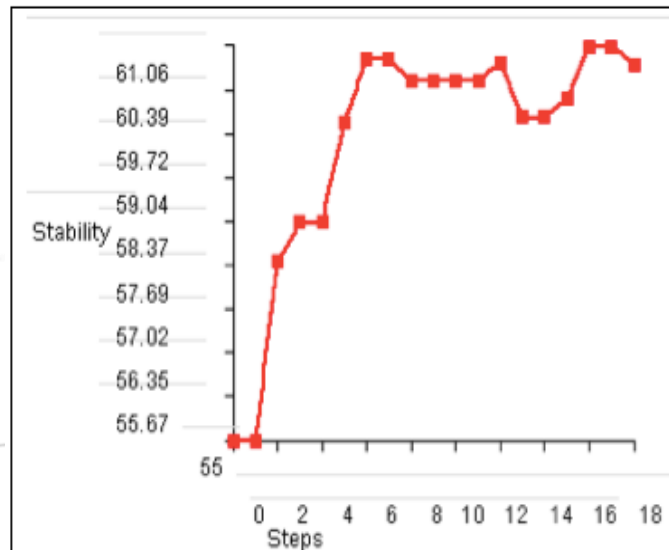
0. Alice creates RO
1. Alice adds Recipe.txt
2. Alice adds Content.txt
3. John adds Results.txt
4. Alice adds Bibliography.pdf
5. Alice annotates Bibliography.pdf
6. John removes Script.py
7. John edits annotation on Recipe.txt
8. Unknown adds Dropme.txt

Metrics and tools for QA Evolution with time

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

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

» 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

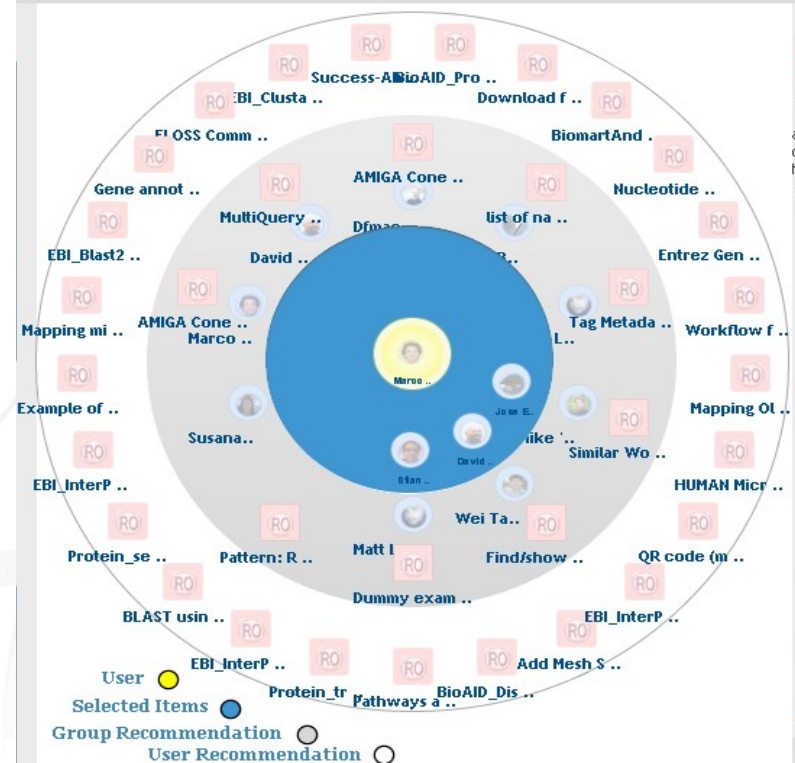Visual mechanism to find similar elements (users, ROs, workflows) to others previously selected

Large international teams to commission an experiment spread in two continents would be the most similar thing to a Social Network

Make the experiment <u>discoverable</u>

» Tracking

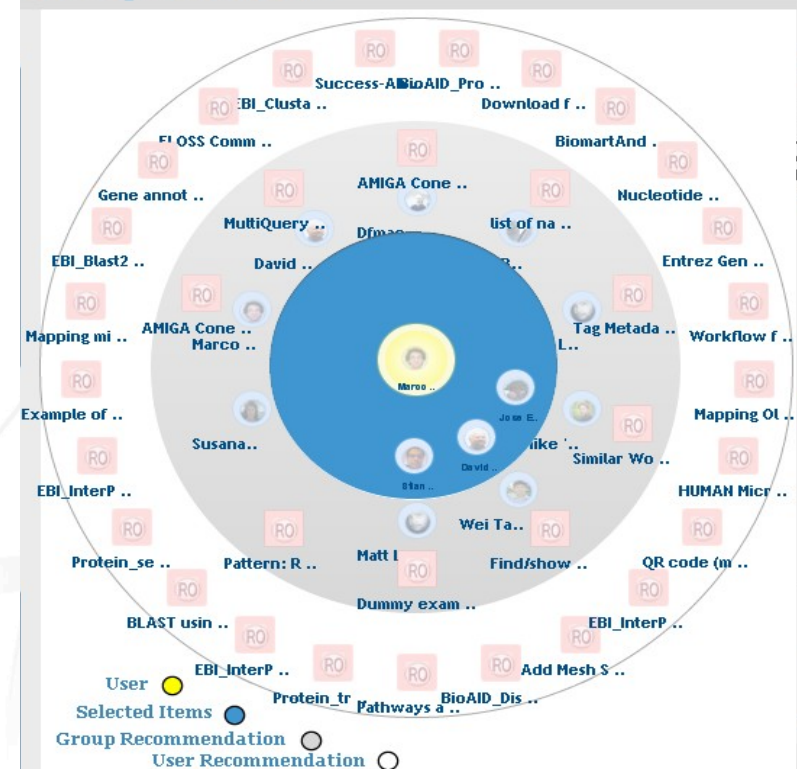Rating by other users, who used the RO, comments, etc.

» 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]

Building complex commissioning tools from existing ones

» AstroTaverna plugin http://wf4ever.github.com/astrotaverna/

› Connection with Virtual Observatory

› Services for managing and visualizing VOTable

› Astronomical utilities: coordinate transforms.

This talk was an example of R-use!

http://amiga.iaa.es/p/212-workflows.htm
http://www.wf4ever-project.org

The R dimensions

Best practices for everything... and gRRRowing

Reusable. The key tenet of Research Objects is to support the sharing and reuse of data, methods and processes.

Repurposeable. Reuse may also involve the reuse of constituent parts of the Research Object.

Repeatable. There should be sufficient information in a Research Object to be able to repeat the study, perhaps years later.

Reproducible. A third party can start with the same inputs and methods and see if a prior result can be confirmed.

Replayable. Studies might involve single investigations that happen in milliseconds or protracted processes that take years.

Referenceable. If research objects are to augment or replace traditional publication methods, then they must be referenceable or citeable.

Revealable. Third parties must be able to audit the steps performed in the research in order to be convinced of the validity of results.

Respectful. Explicit representations of the provenance, lineage and flow of intellectual property.

Replacing the Paper: The Twelve Rs of the e-Research Record" on http://blogs.nature.com/eresearch/

Dave de Roure (Oxford e-Research Center)