

# The Reality of Reproducibility of *in silico* Science



Prof Carole Goble FREng FBCS CITP  
JCDL Washington DC, June 2012

X REUNIÓN CIENTÍFICA  
DE LA SOCIEDAD ESPAÑOLA  
DE ASTRONOMÍA

VALENCIA  
9/13 JULIO



# Digital Science

## Reproducibility and Visibility in Astronomy

José Enrique Ruiz, Lourdes Verdes-Montenegro, Susana Sánchez,  
Julian Garrido, Juan de Dios Santander and the Wf4Ever Team

SESIÓN INSTRUMENTACIÓN Y COMPUTACIÓN  
VALENCIA, VIERNES 13 JULIO 2012



# Digital Science - Reproducibility and Visibility in Astronomy

## Astronomy Research Lifecycle

Astronomy research lifecycle is **entirely digital**

- » Observation proposals 
- » Data reduction pipelines
- » Analysis of science ready data
- » Catalogs of objects and data
- » Publish process
  - › Final data results 
  - › Experiment in DL  
ADS/arXiv

**Reproducible research is still not possible in a digital world**

**A rich infrastructure of data (VO) is not efficiently used**



**A normalized preservation of methodology is needed**

**Tools**

# Digital Science - Reproducibility and Visibility in Astronomy

## The next generation of archives

### Much wider FoV and spectral coverage

- » Large volumes for an observed datacube
- » Subproducts are **Virtual Data** generated on-the-fly

	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

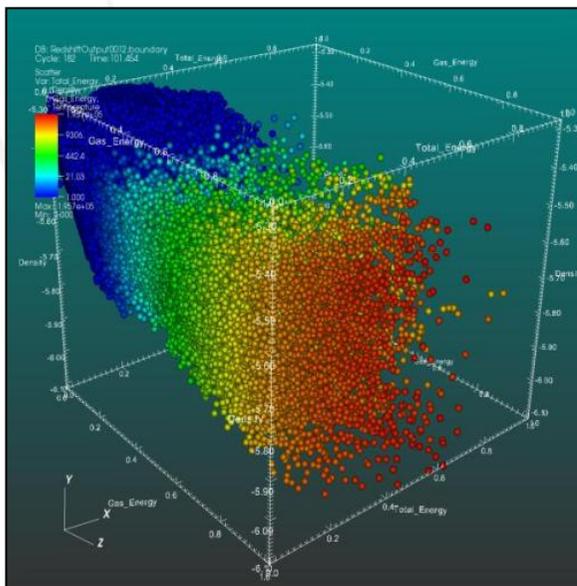
ASKAP Cubes  
Prof. Kevin Vinsen

# Digital Science - Reproducibility and Visibility in Astronomy

## The next generation of archives

### Automated surveys

- » Huge amounts of tabular data
- » Services for KDD



Extraction of scientifically relevant information from a multidimensional parameter space

- » Exploration services
- » Anomaly detection
- » Cross-matching data
- » Dimensionality reduction

# Digital Science - Reproducibility and Visibility in Astronomy

## The next generation of archives

---

### » A cloud of Web Services

Archives should evolve from data providers into

- » Virtual data providers
- » Software tasks providers

### » Archives speaking Web Services

Astronomy of multi archives/facilities/wavelength

Interconnected and interoperable archives

- » Software Tasks
- » Data



Preservation

**Process should benefit of the same privileges acquired by data**

Preserving the method ensures replication of final results at any moment

# Digital Science - Reproducibility and Visibility in Astronomy

## Efficiency and Reuse

### Optimize return on investments made on big facilities

- » Avoid duplication of efforts and reinvention
- » How to discover and not duplicate ?
- » How to re-use and not duplicate ?
- » How to make use of best practices ?
- » How to use the rich infrastructure of data ?
- » **Intellectual contributions are encoded in softw**

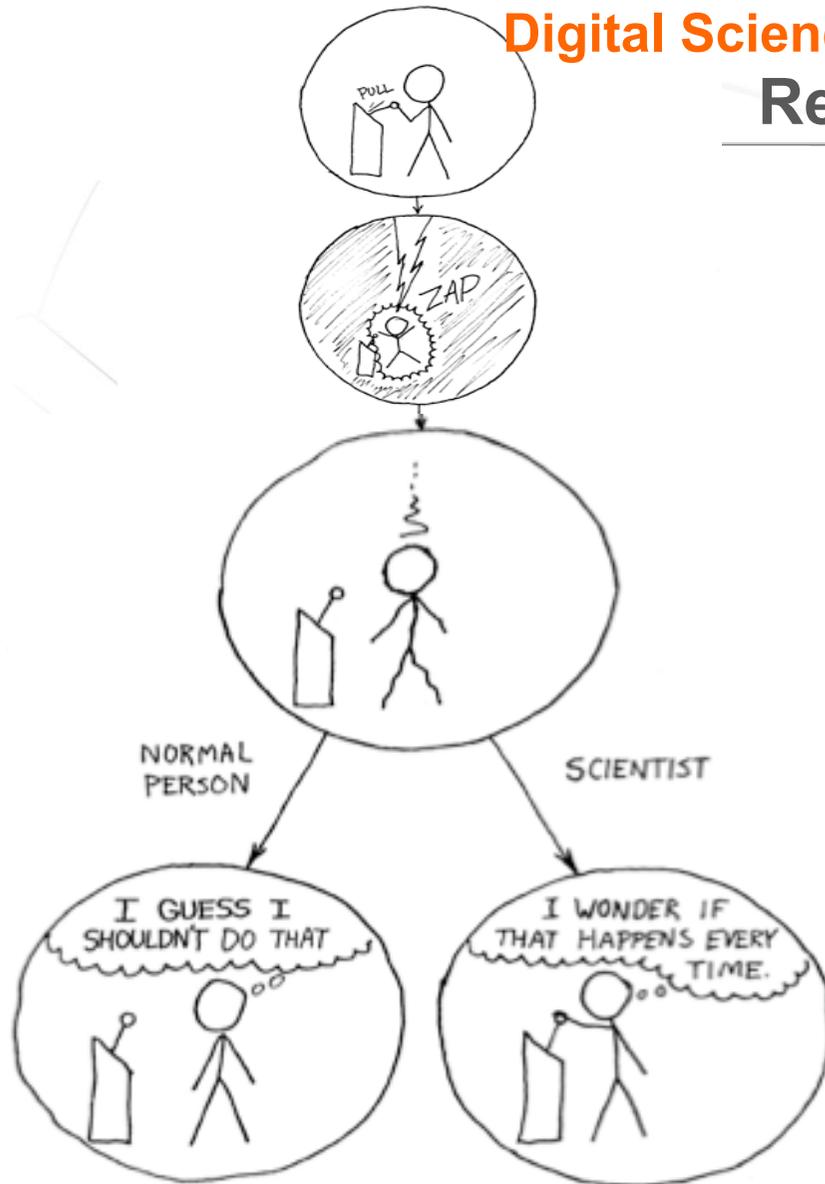
### More data in archives does not imply more knowledge

- » Time has come to go beyond the PDF
- » Expose complete scientific record, not the story
- » Allow easy discovery of methods and tools



# Digital Science - Reproducibility and Visibility in Astronomy

## Reproducibility and The Scientific Method



<http://xkcd.com/242/>

### Benefits

- » Publishing knowledge, **not advertising**
- » The author, the referee and the re-user
- » Reputation, prestige and respect
- » **Higher quality of publications**
  - › Authors will be more careful
  - › Many eyes to check results

### Challenges

- » Hard and time consuming
- » Need incentives – not rewarded now

### Initiatives

- » **Elsevier Executable Papers Challenge**
- » Open Data / Open Science

# Barriers to Data and Code Sharing in Computational Science

Survey of Machine Learning Community, NIPS (Stodden, 2010):

Code		Data
77%	Time to document and clean up	54%
52%	Dealing with questions from users	34%
44%	Not receiving attribution	42%
40%		-
34%	<b>I don't know how</b>	41%
-	Time to verify release with admin	38%
30%	Potential loss of future publications	35%
30%	Competitors may get an advantage	33%
20%	Web/disk space limitations	29%

# Digital Science - Reproducibility and Visibility in Astronomy

## Discovery, Visibility and Credit

**nature** International weekly journal of science

Search   [Advanced search](#)

[Home](#) | [News & Comment](#) | [Research](#) | [Careers & Jobs](#) | [Current Issue](#) | [Archive](#) | [Audio & Video](#) | [For Authors](#)

[Archive](#) > [Specials & supplements archive](#) > [Science Metrics](#)

### SPECIALS

[▶ See all specials](#)

[Journal home](#) | [Subscribe](#) | 

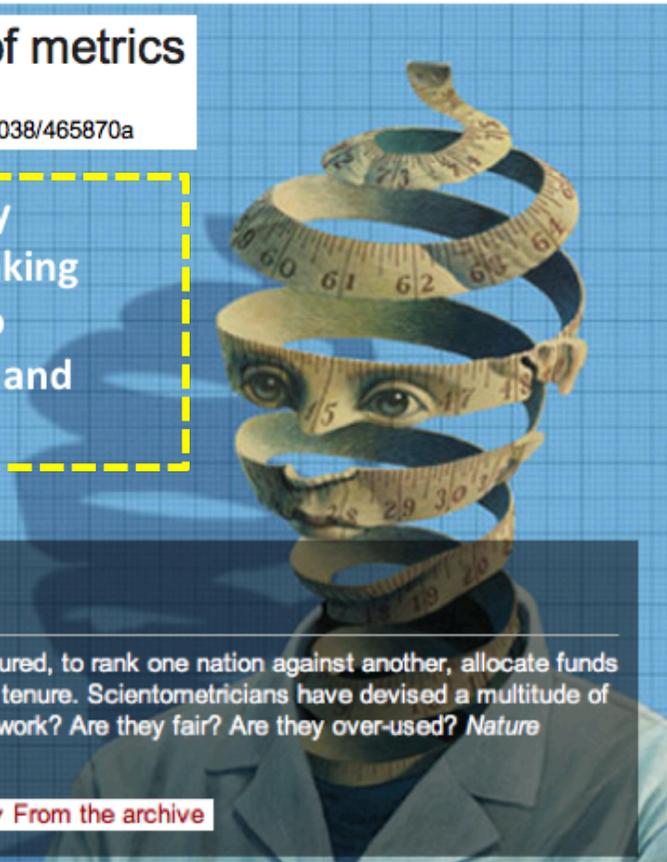
[Current issue](#) | [E-alert sign up](#)

[For authors](#) | [RSS feed](#)

## How to improve the use of metrics

Nature **465**, 870–872 (17 June 2010) | doi:10.1038/465870a

... “Science is being killed by numerical ranking,”[...] Ranking systems lures scientists into pursuing high rankings first and good science second.



### SCIENCE METRICS

The value of scientific output is often measured, to rank one nation against another, allocate funds between universities, or even grant or deny tenure. Scientometricians have devised a multitude of 'metrics' to help in these rankings. Do they work? Are they fair? Are they over-used? *Nature* investigates.

[▼ Editorial](#) | [▼ Features](#) | [▼ Opinion](#) | [▼ From the archive](#)

### Top content

Emailed	Downloaded	Blogged
1. <b>Public health: The toxic truth about sugar</b> <i>Nature</i>   01 February 2012		
2. <b>The case for open computer programs</b> <i>Nature</i>   22 February 2012		
3. <b>The great beyond</b> <i>Nature</i>   29 February 2012		
4. <b>The darker side of stem cells</b> <i>Nature</i>   29 February 2012		
5. <b>Cancer: Solving an age-old problem</b> <i>Nature</i>   29 February 2012		
<a href="#">View all ▶</a>		

# Digital Science - Reproducibility and Visibility in Astronomy

## Discovery, Visibility and Credit

**nature** International weekly journal of science

Search   [Advanced search](#)

[Home](#) | [News & Comment](#) | [Research](#) | [Careers & Jobs](#) | [Current Issue](#) | [Archive](#) | [Audio & Video](#) | [For Authors](#)

[Archive](#) > [Specials & supplements archive](#) > [Science Metrics](#)

### SPECIALS

[▶ See all specials](#)

[Journal home](#) | [Subscribe](#) | 

[Current issue](#) | [E-alert sign up](#)

[For authors](#) | [RSS feed](#)

## How to improve the use of metrics

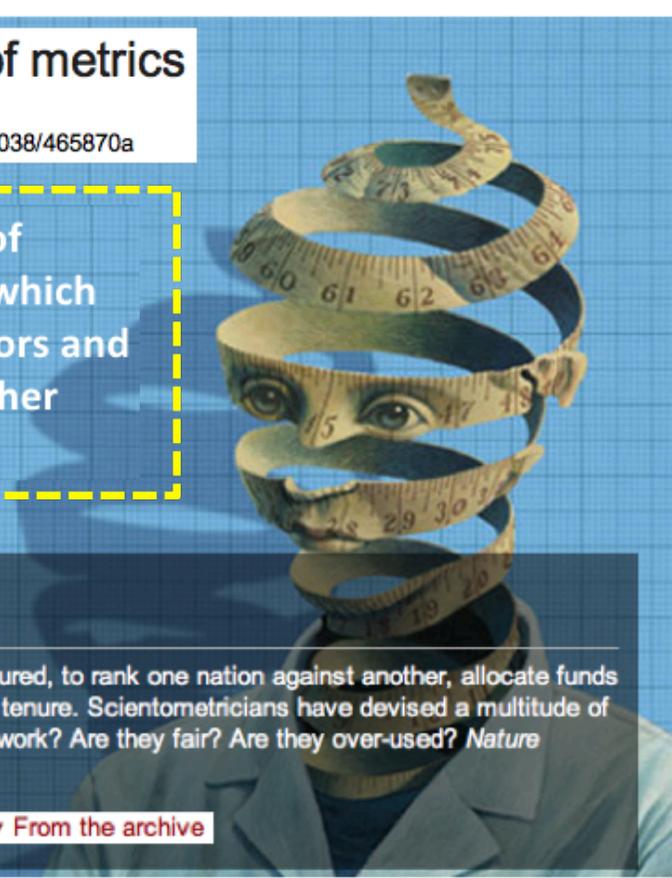
Nature 465, 870–872 (17 June 2010) | doi:10.1038/465870a

Research reverts to a kind of 'academic prostitution', in which work is done to please editors and referees rather than to further knowledge.

### SCIENCE METRICS

The value of scientific output is often measured, to rank one nation against another, allocate funds between universities, or even grant or deny tenure. Scientometricians have devised a multitude of 'metrics' to help in these rankings. Do they work? Are they fair? Are they over-used? *Nature* investigates.

[▼ Editorial](#) | [▼ Features](#) | [▼ Opinion](#) | [▼ From the archive](#)



### Top content

Emailed	Downloaded	Blogged
1. <b>Public health: The toxic truth about sugar</b> <i>Nature</i>   01 February 2012		
2. <b>The case for open computer programs</b> <i>Nature</i>   22 February 2012		
3. <b>The great beyond</b> <i>Nature</i>   29 February 2012		
4. <b>The darker side of stem cells</b> <i>Nature</i>   29 February 2012		
5. <b>Cancer: Solving an age-old problem</b> <i>Nature</i>   29 February 2012		

[View all ▶](#)

# Digital Science - Reproducibility and Visibility in Astronomy

## Discovery, Visibility and Credit

**nature** International weekly journal of science

Search   [Advanced search](#)

[Home](#) | [News & Comment](#) | [Research](#) | [Careers & Jobs](#) | [Current Issue](#) | [Archive](#) | [Audio & Video](#) | [For Authors](#)

[Archive](#) > [Specials & supplements archive](#) > [Science Metrics](#)

### SPECIALS

[▶ See all specials](#)

[Journal home](#) | [Subscribe](#) | 

[Current issue](#) | [E-alert sign up](#)

[For authors](#) | [RSS feed](#)

## How to improve the use of metrics

Nature 465, 870–872 (17 June 2010) | doi:10.1038/465870a

... an author's h-index can reflect longevity as much as quality — and can never go down with age, even if a researcher drops out of science altogether.

### SCIENCE METRICS

The value of scientific output is often measured, to rank one nation against another, allocate funds between universities, or even grant or deny tenure. Scientometricians have devised a multitude of 'metrics' to help in these rankings. Do they work? Are they fair? Are they over-used? *Nature* investigates.

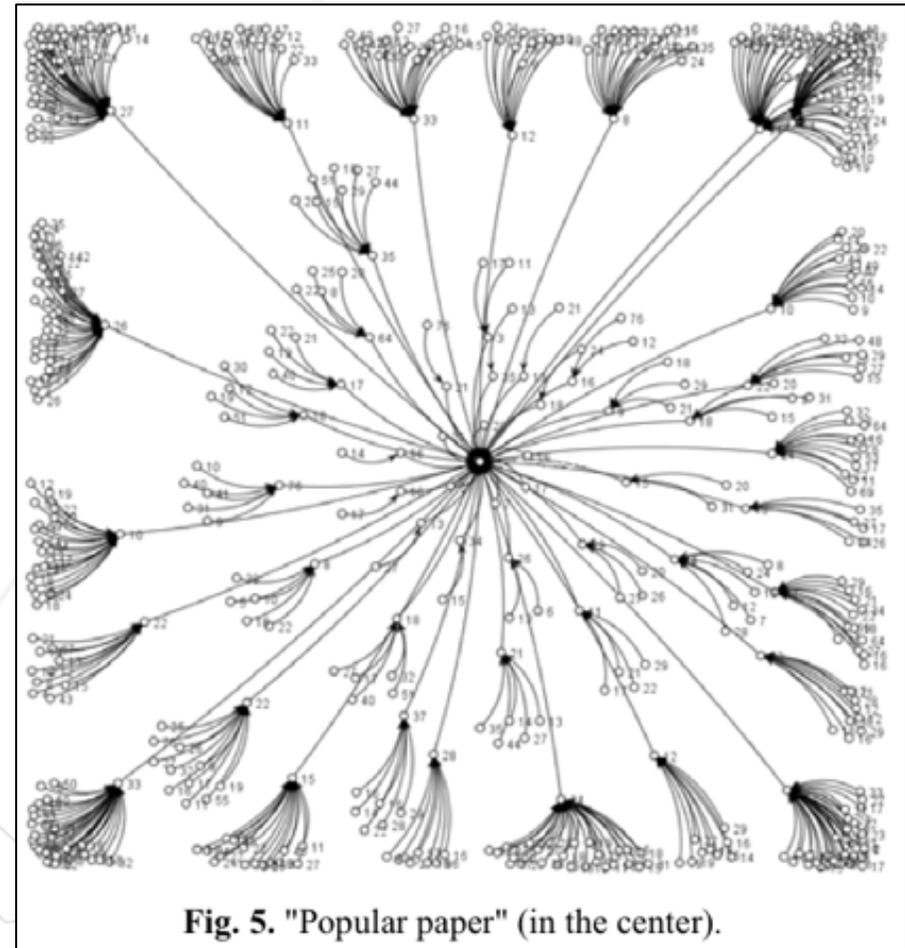
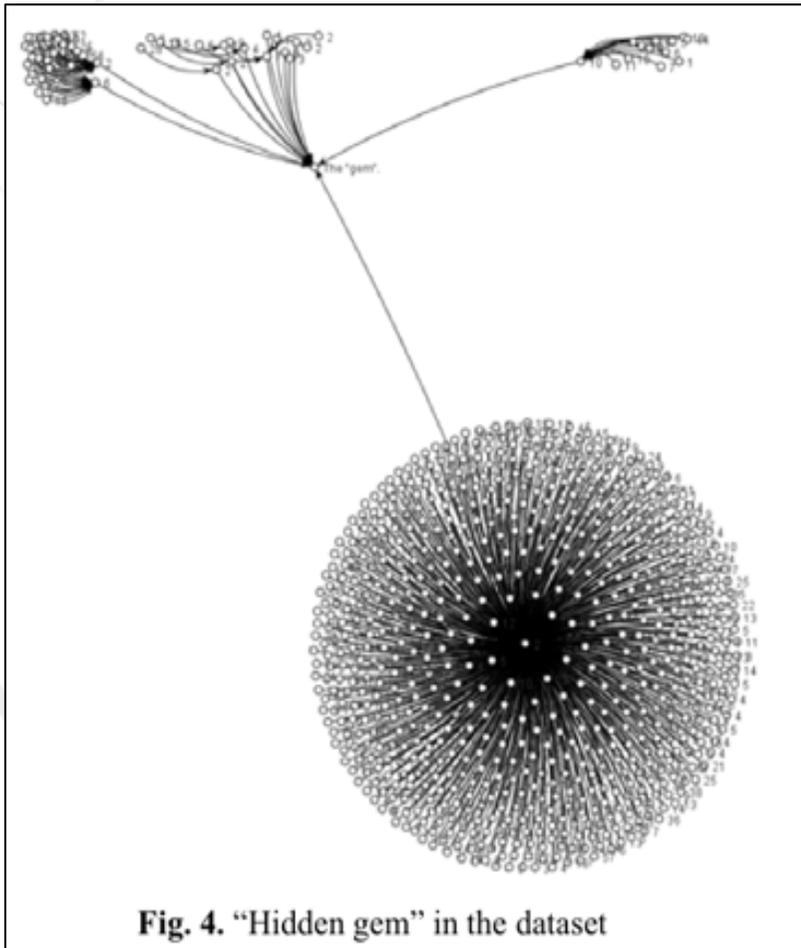
[▼ Editorial](#) | [▼ Features](#) | [▼ Opinion](#) | [▼ From the archive](#)



### Top content

Emailed	Downloaded	Blogged
1. <b>Public health: The toxic truth about sugar</b> <i>Nature</i>   01 February 2012		
2. <b>The case for open computer programs</b> <i>Nature</i>   22 February 2012		
3. <b>The great beyond</b> <i>Nature</i>   29 February 2012		
4. <b>The darker side of stem cells</b> <i>Nature</i>   29 February 2012		
5. <b>Cancer: Solving an age-old problem</b> <i>Nature</i>   29 February 2012		
<a href="#">View all</a> ▶		

## Exploring and understanding scientific metrics in citation



# Digital Science - Reproducibility and Visibility in Astronomy

## Discovery, Visibility and Credit

### Paper discovery: the social dimension

The image is a collage of various digital science and social media platforms. At the top left is the Peeref logo with the tagline 'empowering scholars'. Next to it is the Mendeley logo with the text 'Welcome back Lourdes Verdes-Montenegro'. To the right is the YouTube logo. Below these are the Citeulike logo and a search bar for 'Search citeulike'. Further down is the BibSonomy logo. To the right of BibSonomy is a screenshot of a Mendeley profile page with a 'Papers' dropdown and a search bar. Below BibSonomy is the Klonk logo with the tagline 'Spread your research results'. To the right of Klonk is a Twitter logo with the hashtag '#SEA2012'. Below Klonk is the ResearchGate logo with a search bar. To the right of ResearchGate is a screenshot of the AstroBetter website, which has a navigation menu (Blog, About, Archives, Support, Wiki) and a search bar. Below ResearchGate is the Delicious logo with the tagline 'social bookmarking'. To the right of Delicious is a screenshot of the Slideshare logo with the text 'BETA'. At the bottom left is the Slideshare logo. At the bottom right is the Zotero logo. The word 'Collabgraph!' is written in large green letters in the center-right area. Below it is a paragraph of text: 'Collaborating in your field of research. Just [redacted] or upload a bibtex file, containing your [redacted] graph will create a fancy graph showing [redacted]'. The word 'Collabgraph!' is also written in large green letters in the center-right area.

# Digital Science - Reproducibility and Visibility in Astronomy

## The Wf4Ever Project

### 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)

**iSOOCO**  
enabling the networked economy



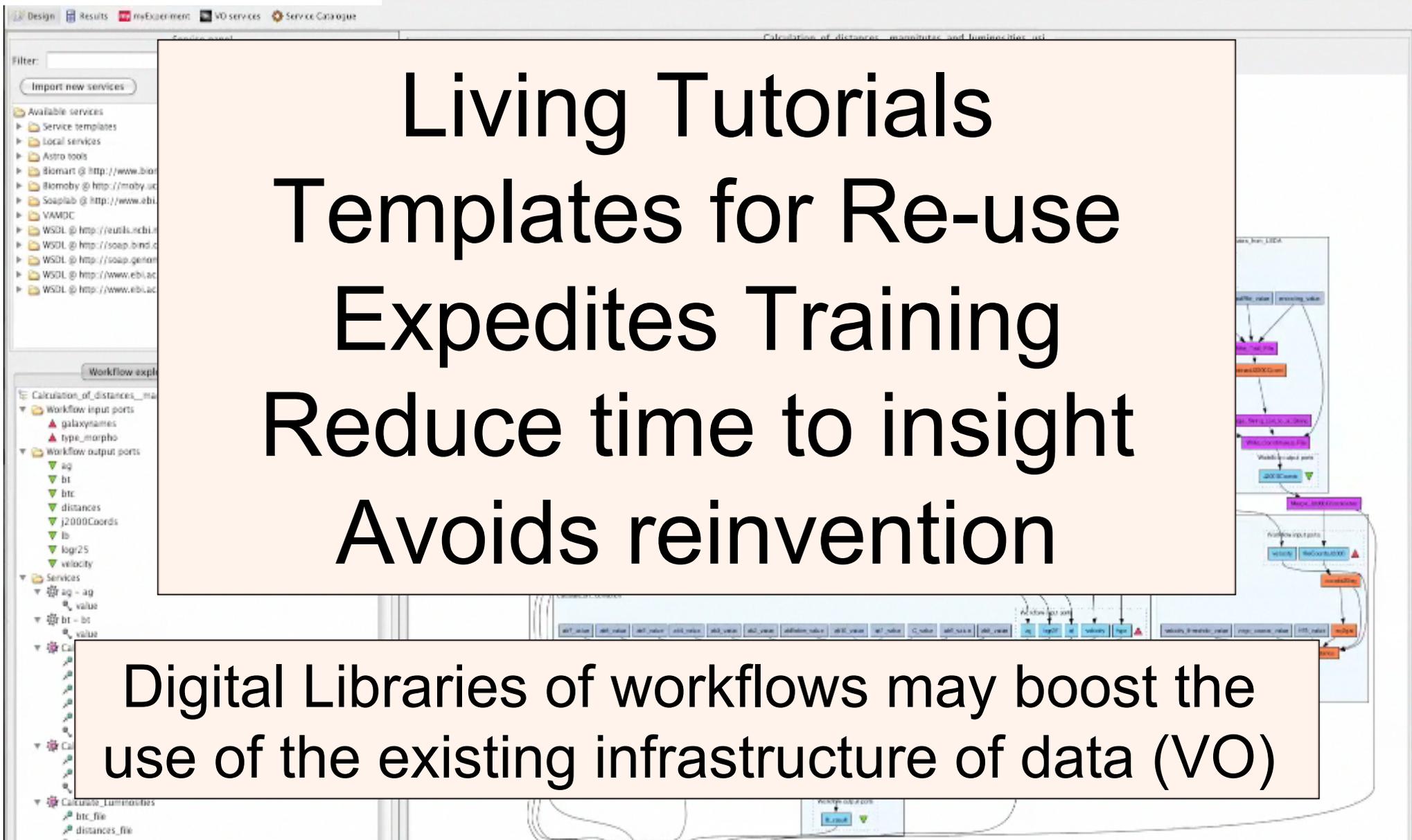
The University  
of Manchester

MANCHESTER  
1824



UPM



The background image is a screenshot of a scientific workflow management software. On the left, there is a 'Service Catalogue' panel with a search filter and a list of available services, including 'Biomart', 'Blomoby', 'Soaplab', 'VAMDC', and several 'WSDL' services. Below this is a 'Workflow explorer' panel showing a tree view of a workflow named 'Calculation\_of\_distances\_ma'. The main area of the screenshot shows a complex workflow diagram with various nodes and connections. A large, semi-transparent orange box is overlaid in the center, containing the text 'Living Tutorials Templates for Re-use Expedites Training Reduce time to insight Avoids reinvention'. At the bottom, another semi-transparent orange box contains the text 'Digital Libraries of workflows may boost the use of the existing infrastructure of data (VO)'.

Living Tutorials  
Templates for Re-use  
Expedites Training  
Reduce time to insight  
Avoids reinvention

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

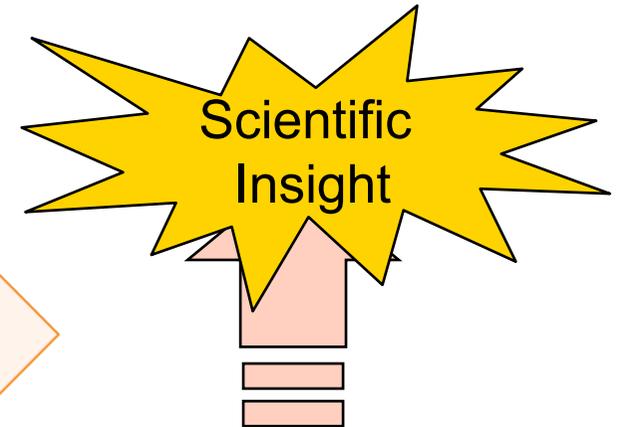
# Digital Science - Reproducibility and Visibility in Astronomy

## Scientific Workflows

**Survey** in the domain of astrophysical workflows

- › Personal script-based recipes
  - Python, IDL, Software..
- › Multi-archive VO recipes
  - Euro-VO, IVOA..
- › Internal group developments
  - GRID, Clusters, Specific knowledge
- › Processing pipelines
  - Facilities provide data, compute infrastructure, tools..

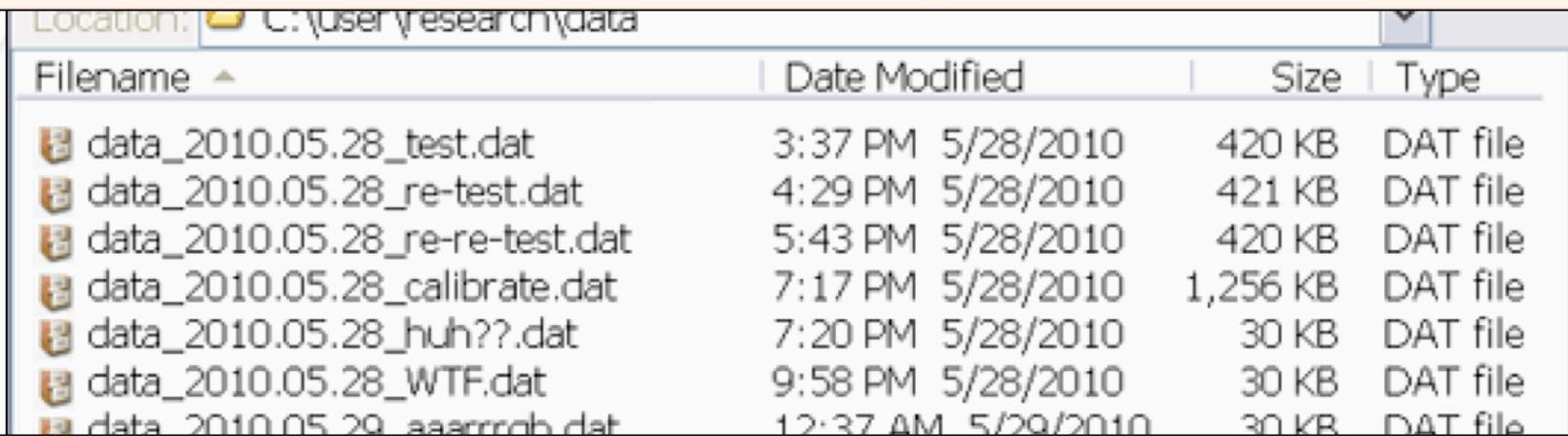
*Hand-crafted workflows*



Accessible  
Shareable  
Reusable  
Adaptable  
Understandable

- » **Clarity** (workflows) for re-use and re-purpuse **vs. automation** (pipelines)
- » A black box is not **re-usable**, cannot be broken into parts
- » **Reproducibility vs. industrial** paper publishing

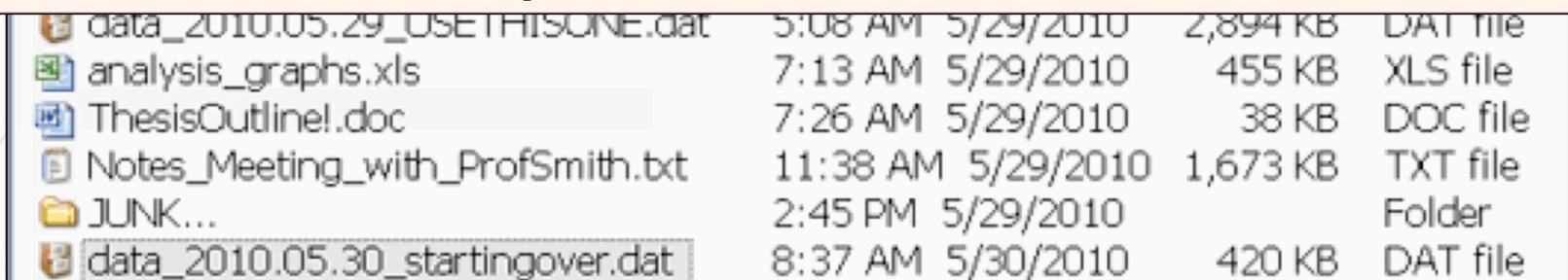
Organization is more sexy than automation



Location: C:\user\research\data

Filename	Date Modified	Size	Type
data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
data_2010.05.28_huh??.dat	7:20 PM 5/28/2010	30 KB	DAT file
data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file
data_2010.05.29_aaarrrob.dat	12:37 AM 5/29/2010	30 KB	DAT file

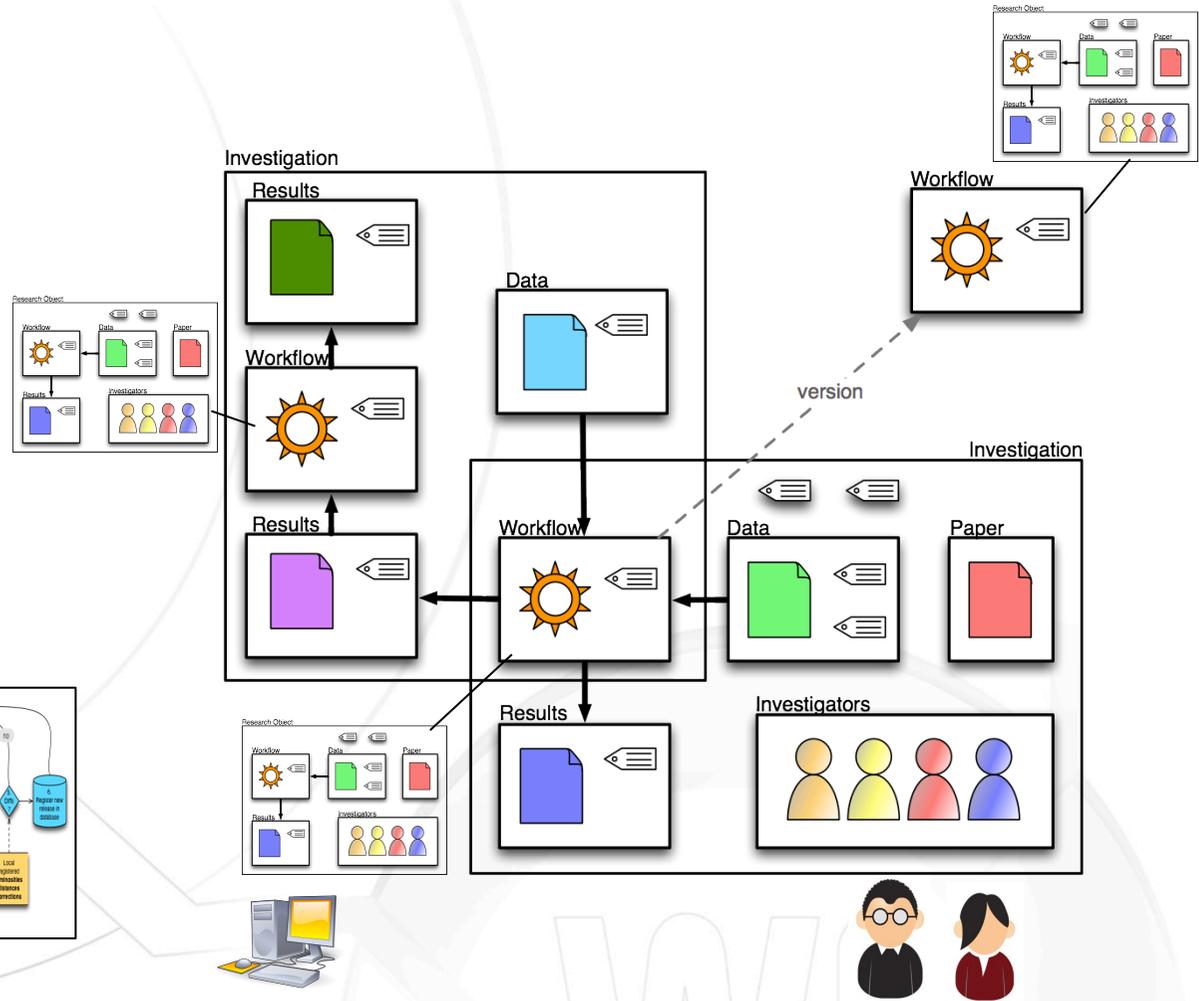
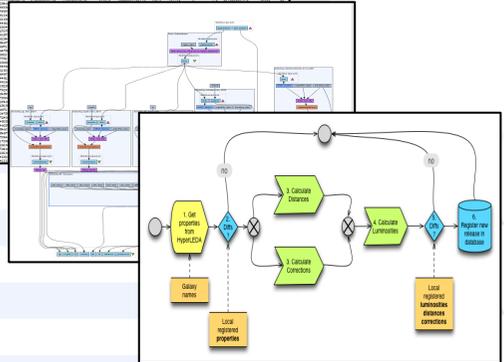
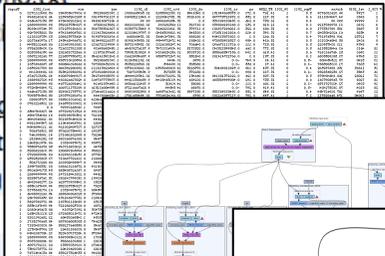
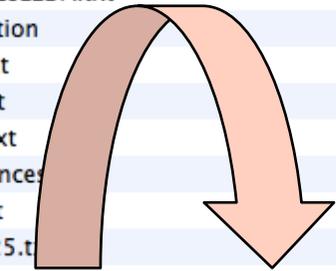
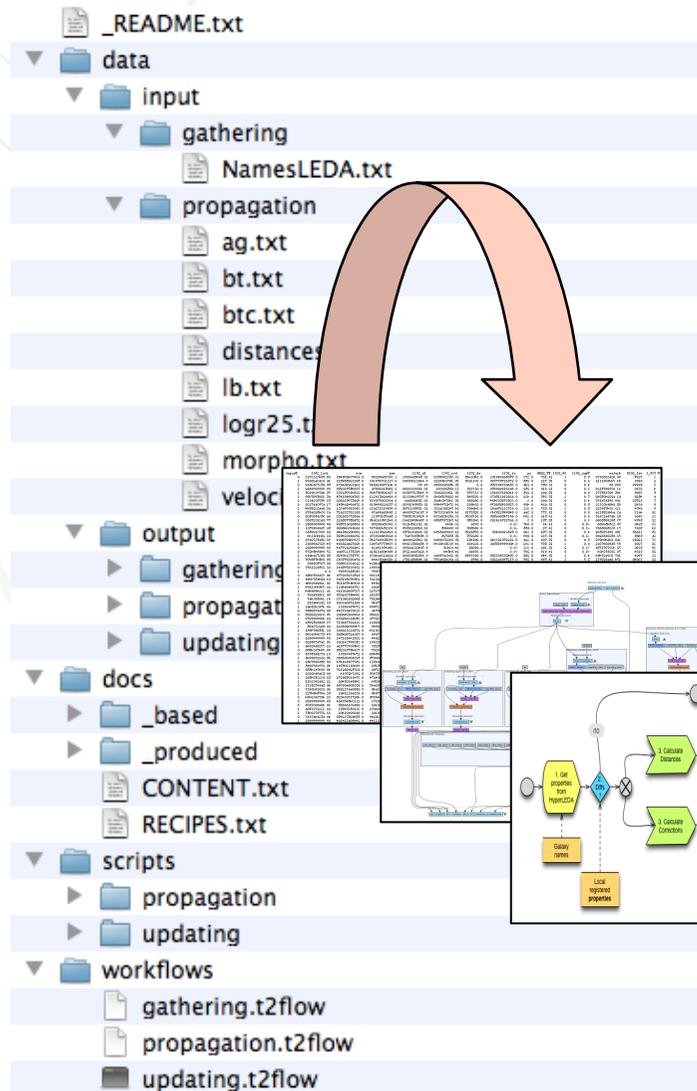
Assistive building  
Completeness evaluation



data_2010.05.29_USETHIRSONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
JUNK...	2:45 PM 5/29/2010		Folder
data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file

# Digital Science - Reproducibility and Visibility in Astronomy Research Objects

Expose experiment in a structured way in order to be **understood**



**Technical Objects**  
Distributed

**Social Objects**

## Similar initiatives in Astronomy

- » **Semantic curation** of digital objects
  - › CDS Centre Données Strasbourg
  - › US Virtual Astronomical Observatory
  - › SAO/NASA ADSLabs
- » **Workflow users platforms**
  - › Cyber-SKA
  - › IceCore
  - › Montage
  - › Astro-WISE
  - › Helio-VO
- » **Semantically auto descriptive WS**
  - › Workflows VO-France

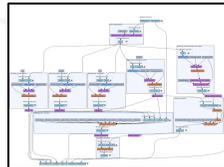
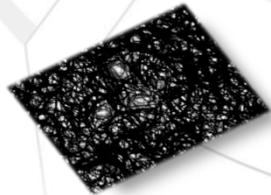


# Digital Science - Reproducibility and Visibility in Astronomy Research Objects

## ADSLabs Initiative

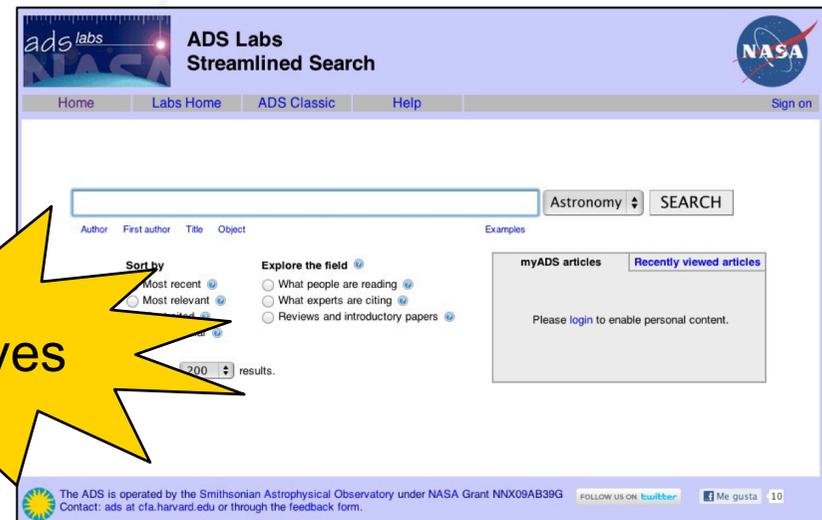
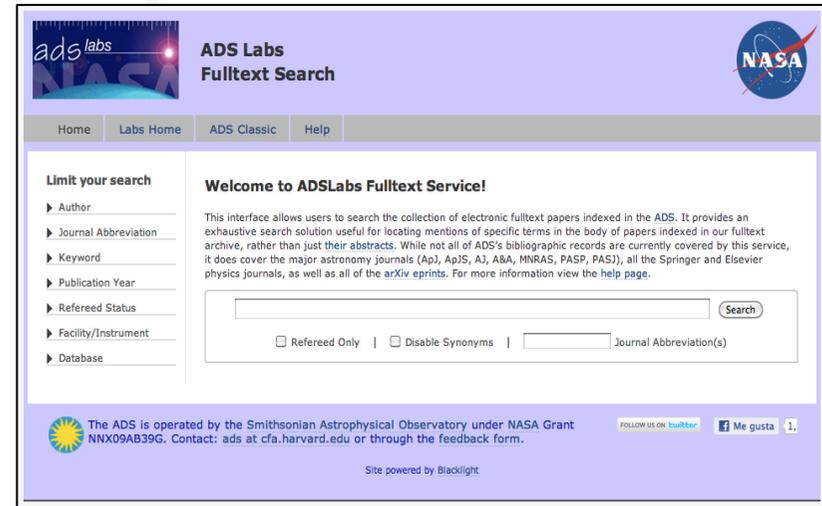
### ADO Linked Components

- » Authors
- » Publications
- » Journals
- » Objects SIMBAD
- » Tabular data behind the plots CDS
- » ASCL reference of used software
- » Observing time Proposals
- » Used facilities, surveys or missions



Incentives

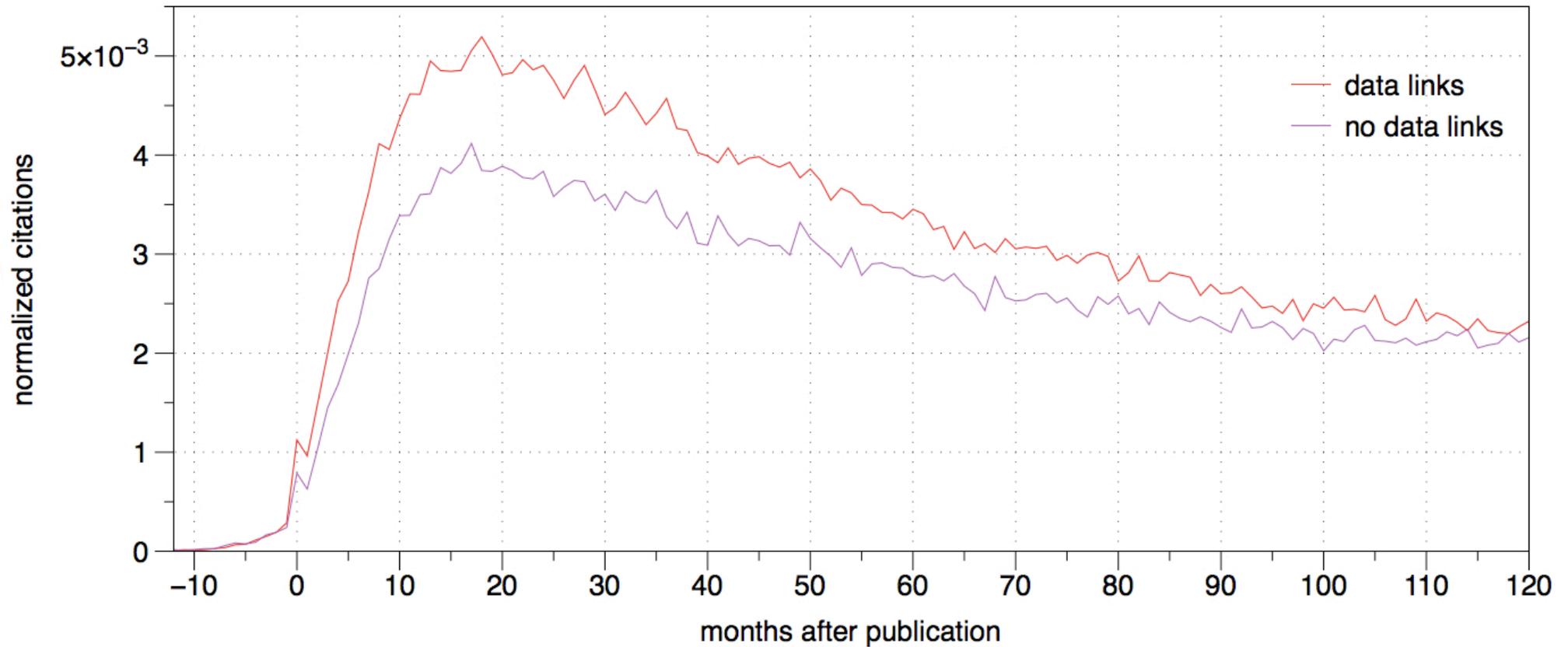
 <http://labs.adsabs.harvard.edu/>



## The Incentive

Papers with data links are cited more than those without

1995 - 2000



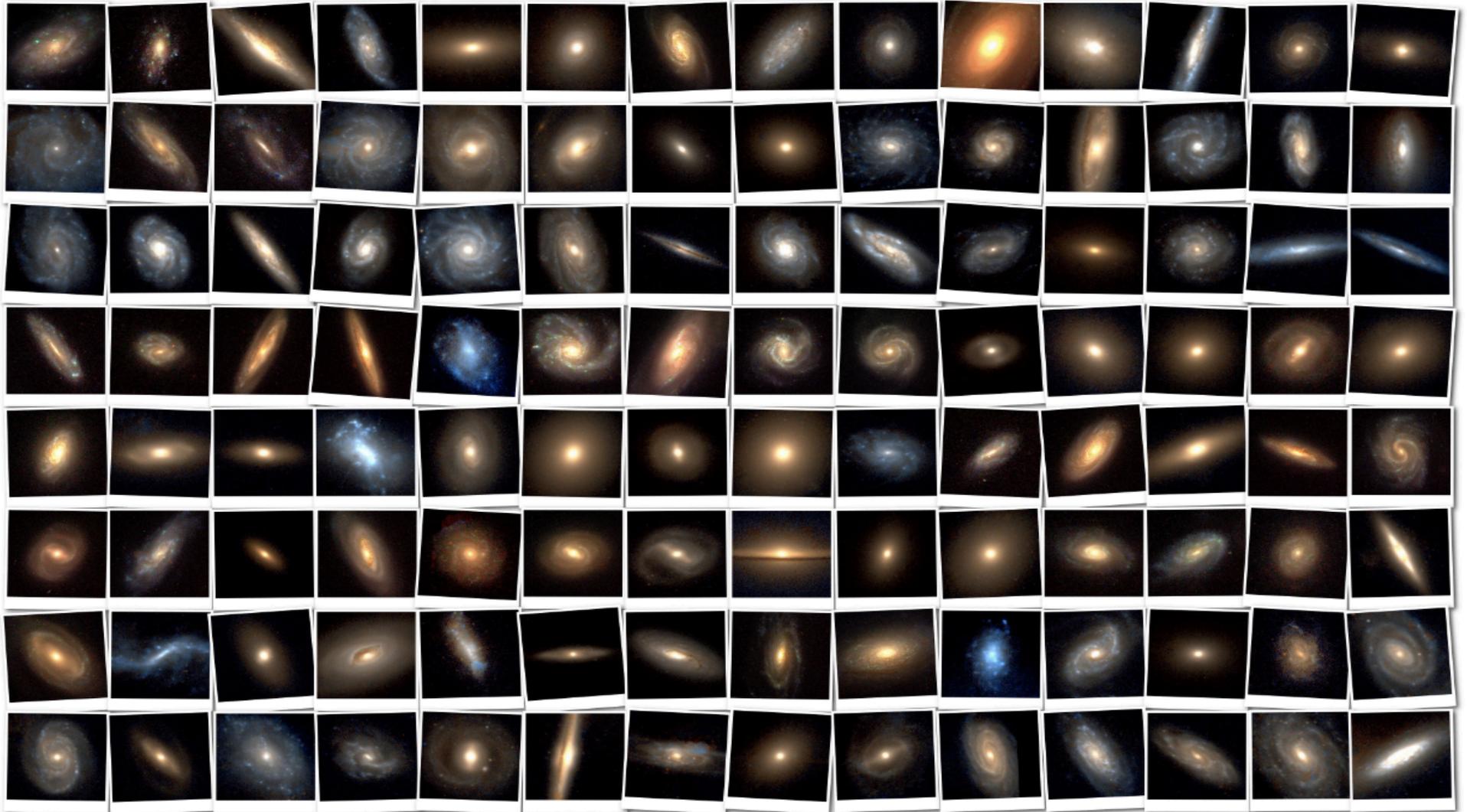
Effect of E-printing on Citation Rates in Astronomy and Physics  
2006. Edwin A. Henneken et al.



# Digital Science - Reproducibility and Visibility in Astronomy

## Astronomical Research Objects in Action

### Curation by inspecting propagation of changes in quantities



Credit: Zsolt Frei and James E. Gunn. The Galaxy Catalog

# Digital Science - Reproducibility and Visibility in Astronomy

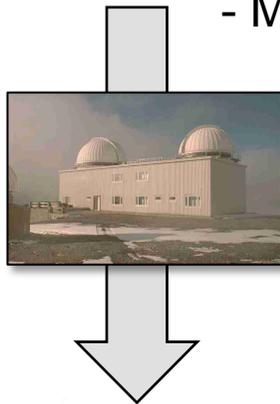
## Astronomical Research Objects in Action

### AMIGA Catalog

Panchromatic properties for a sample of the most isolated nearby galaxies

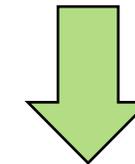
How is the User DB affected ?

- Changes in External DB
- Modifications in Calculations

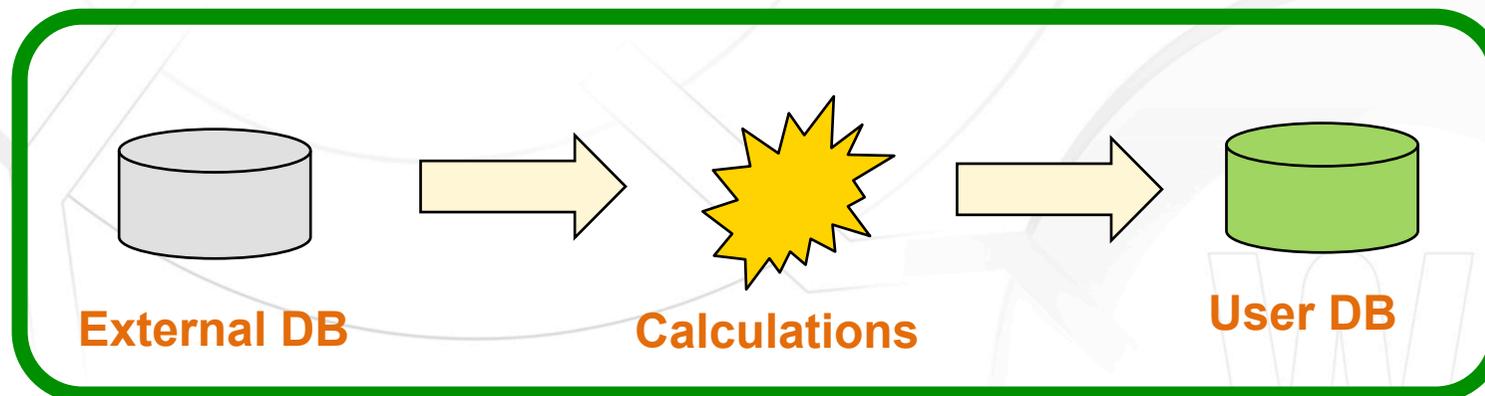


**Evaluate variations with time**

- Modified External Quantities
- Affected User DB Quantities



**Update**

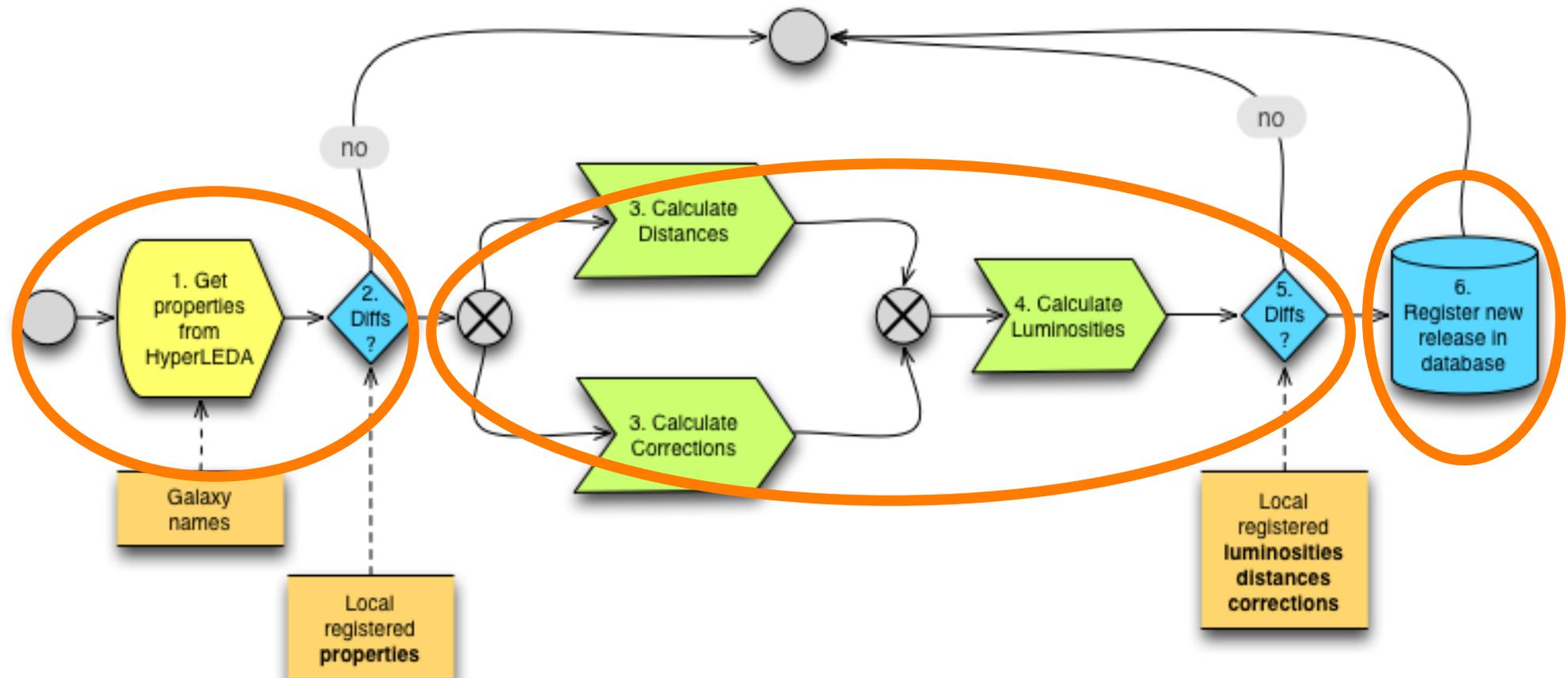


# Digital Science - Reproducibility and Visibility in Astronomy

## Astronomical Research Objects in Action

### Curation by inspecting propagation of changes in quantities

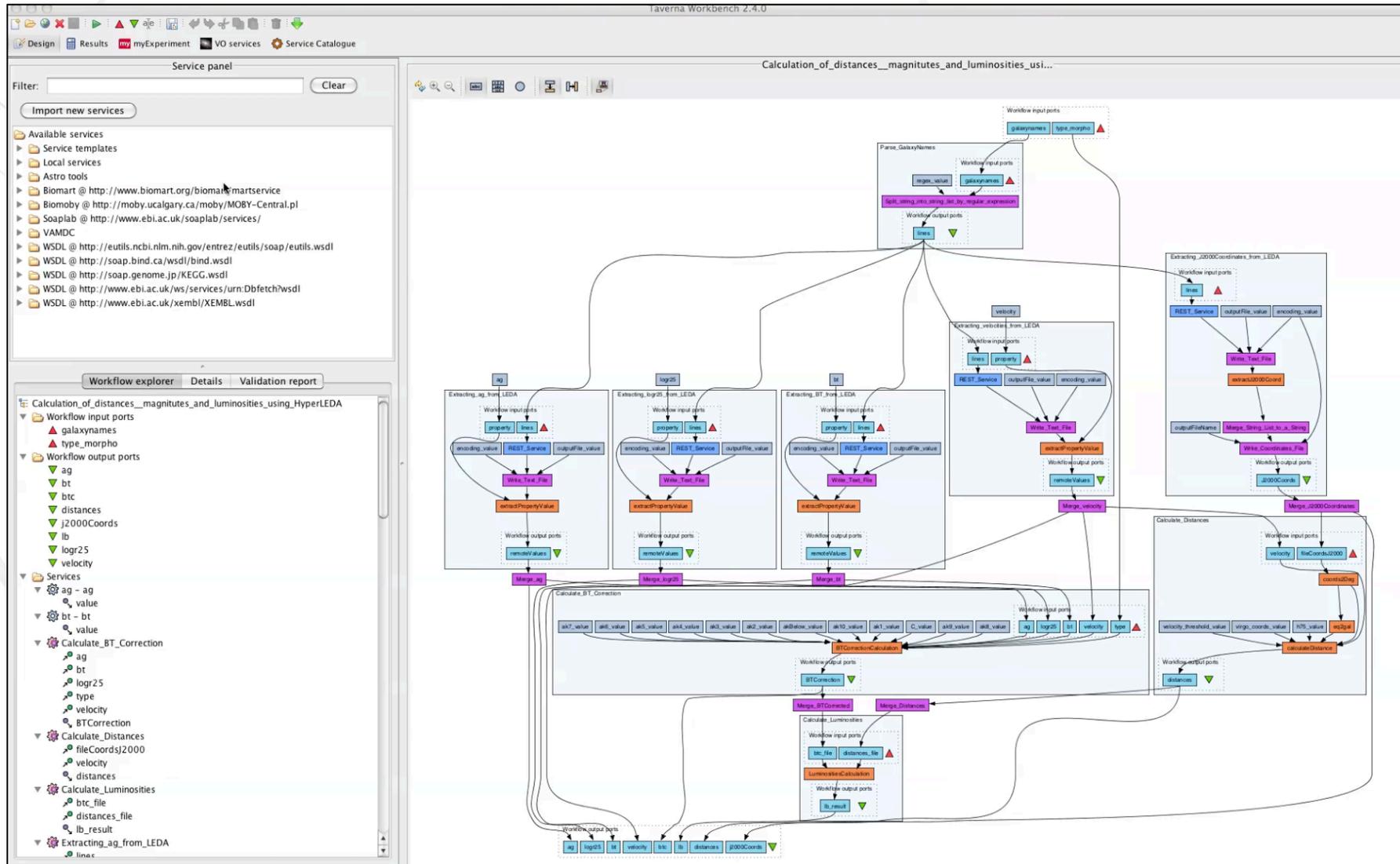
### Multi-workflow Research Object



# Digital Science - Reproducibility and Visibility in Astronomy

## Astronomical Research Objects in Action

### Create, annotate and run a workflow



# Digital Science - Reproducibility and Visibility in Astronomy

## Astronomical Research Objects in Action

### Populate the Research Object and annotate

Wf4Ever - RO Annotator MOCKUP

#### Research Object: Distance Estimation

- Datasets
  - Galaxy\_Names.csv
  - Apparent\_Magnitudes.csv
- Scripts
- Web Services
- Workflows

#### Annotating "Galaxy\_Names.csv"

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

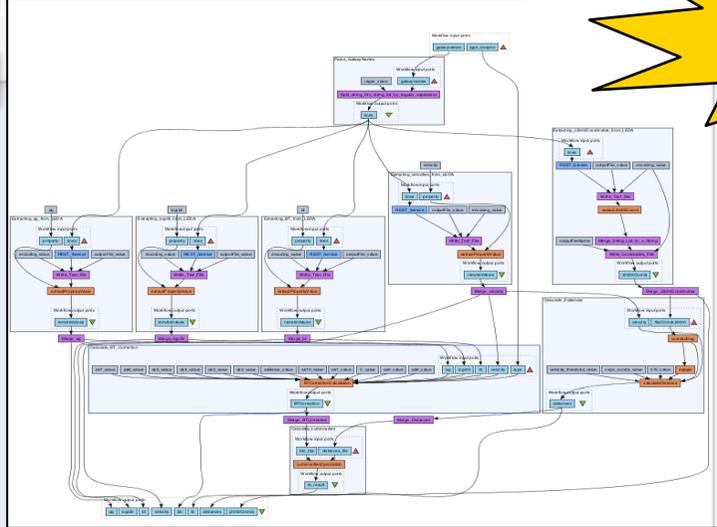
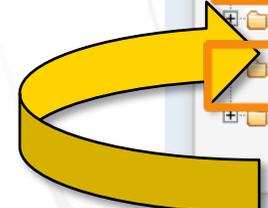
What kind of annotation is this?  
Description:

Value for the annotation

Names of galaxies whose distance is to be estimated. Each line represents a different galaxy. Optional information on the galaxy is added as comma-separated values, in this format:

- Galaxy name
- Morphology type (NED)
- NED distance

Save Changes Cancel



# Digital Science - Reproducibility and Visibility in Astronomy

## Astronomical Research Objects in Action

### Add documents and references

Wf4Ever - RO Annotator MOCKUP

#### Research Object: Distance Estimation

- Datasets
  - Galaxy\_Names.csv
  - Apparent\_Magnitudes.csv
- Scripts
- Web Services
- Workflows
- Docs
  - 2012A&A...536A.108V.AMIGA XIII: Workflow-based distance assessment...

#### Annotating "Galaxy\_Names.csv"

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

Modified At: 2012-02-07 08:44:32

What kind of annotation is this?

Description:

Value for the annotation

Names of galaxies whose distance is to be estimated. Each line represents a different galaxy. Optional information on the galaxy is added as comma-separated values, in this order:

- Galaxy name
- Morphology type (NED)
- NED distance
- Estimation Method

Save Changes Cancel

User-agent: \*  
Allow: /

```
# October 11 2010 Whyte & Mackay whisky are running a promotion where 250 bottles of 30 year old whisky, each worth £150, are hidden in bottles of Whyte & Mackay Special whisky (learn more at http://bit.ly/whiskyhuntvideo )  
  
# The bottles are hidden in stores across the UK but we wanted to hide one in our new-look website as well - so if you're reading this congratulations on being a winner  
  
# Drop an email to richard at themasterblender dot com with the subject I Read Robots.txt Files and if you're one of the first to reply, are of a legal drinking age and your local licensing laws allow it we'll send you a bottle of 30 year old Whyte & Mackay. If we can't send you that, we'll send you something else.  
  
# And of course make sure you are following us on Twitter and Facebook - @the_nose @whyteandmackay facebook.com/whyteandmackay see what we do next. We do like our stunts!  
  
# If you weren't first, there's still 200 bottles hidden in stores across the UK  
October 11 2010
```

PDF



# Digital Science - Reproducibility and Visibility in Astronomy

## Astronomical Research Objects in Action

### Add schema of the experiment

Wf4Ever - RO Annotator MOCKUP

#### Research Object: Distance Estimation

- Datasets
  - Galaxy\_Names.csv
  - Apparent\_Magnitudes.csv
- Scripts
- Web Services
- Workflows
- Docs

#### Annotating "Galaxy\_Names.csv"

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

Modified At: 2012-02-07 08:44:32

What kind of annotation is this?

Description:

Value for the annotation

Names of galaxies whose distance is to be estimated. Each line represents a different galaxy. Optional information on the galaxy is added as comma-separated values, in this order:

- Galaxy name
- Morphology type (NED)
- NED distance
- Estimation Method

```
graph TD; Start(( )) --> T1[1. Get properties from HyperLEDA]; T1 --> D1{2. Diff?}; D1 -- no --> T2[3. Calculate Distances]; D1 -- yes --> T3[3. Calculate Corrections]; T2 --> T4[4. Calculate Luminosities]; T3 --> T4; T4 --> D2{5. Diff?}; D2 -- no --> T1; D2 -- yes --> T6[6. Register new release in database]; D2 --> End(( ));
```

The workflow diagram illustrates the process of distance estimation. It begins with a start node leading to a task '1. Get properties from HyperLEDA', which receives 'Galaxy names' as input. This task leads to a decision diamond '2. Diff?'. If the answer is 'no', the flow proceeds to '3. Calculate Distances'. If 'yes', it goes to '3. Calculate Corrections'. Both paths merge and lead to '4. Calculate Luminosities', which receives 'Local registered luminosities distances corrections' as input. This task leads to another decision diamond '5. Diff?'. If 'no', the flow loops back to '1. Get properties from HyperLEDA'. If 'yes', it leads to '6. Register new release in database', which then leads to an end node.

# Digital Science - Reproducibility and Visibility in Astronomy

## Astronomical Research Objects in Action

### Publication for later discovery

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

**Item info**

**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** 1

**Re-used** 4

**Comments** 2

<< Previous version | Next version >>



# Digital Science - Reproducibility and Visibility in Astronomy

## Astronomical Research Objects in Action

### Curation by inspecting propagation of changes in quantities

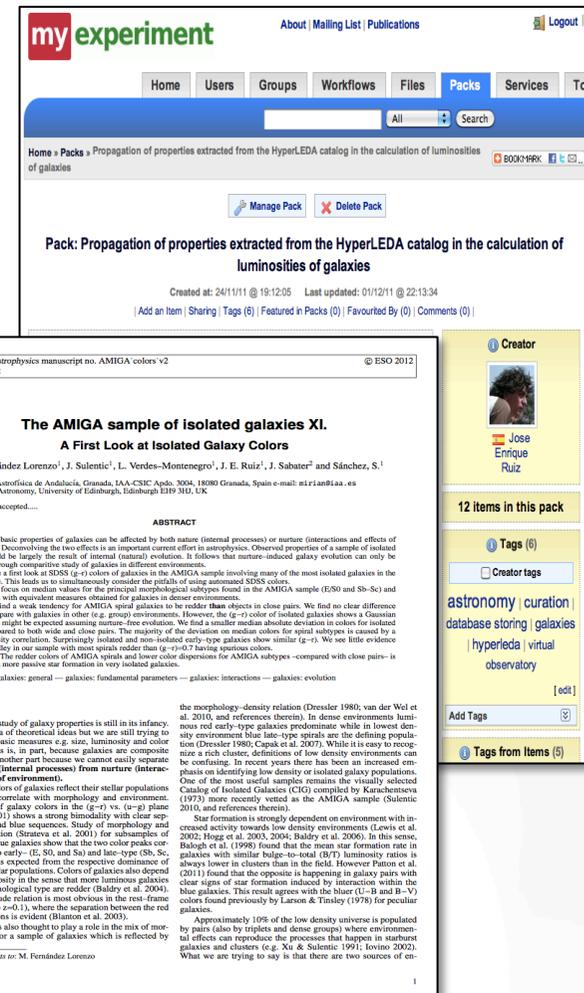
- » Taverna 2.3 
- » MyExperiment Pack
  - » <http://www.myexperiment.org/packs/231>

### Related Publication

## The AMIGA sample of isolated galaxies XI.

## A First Look at Isolated Galaxy Colors

2012 A&A 540, A.47



The screenshot shows the MyExperiment website interface. At the top, there is a navigation bar with 'Home', 'Users', 'Groups', 'Workflows', 'Files', 'Packs', 'Services', and 'Logout'. Below this is a search bar and a breadcrumb trail: 'Home > Packs > Propagation of properties extracted from the HyperLEDA catalog in the calculation of luminosities of galaxies'. The main content area displays the pack title, creation and update dates, and options to 'Manage Pack' or 'Delete Pack'. A sidebar on the right shows the pack creator, Jose Enrique Ruiz, and a list of tags including 'astronomy', 'curation', 'database storing', 'galaxies', 'hyperleda', and 'virtual observatory'. The main text area contains the abstract and introduction of the publication 'The AMIGA sample of isolated galaxies XI. A First Look at Isolated Galaxy Colors' by M. Fernández Lorenzo et al.

### How NOT to be a good e-astronomer

- » Search the **beautiful plot** for high impact instead of real scientific results
- » Write a **obscure paper**, do not say clearly how to reproduce the results
- » Do things **quickly** and forget about them once you've submitted the paper
- » Be untidy, **spread your code and data** in a variety of formats, folders and disks
- » Practise the "**data mine-ing**" – input data are mine
- » Practise the "**data flirting**" – call me if you would like to have more
- » Do not provide data results, **including the plots is just fine**
- » Always **cite the same** authors and papers or those that cite you
- » Do not cite other resources than papers, **neither provide their URL links**
- » Do not search info on **Internet** with other tools than ADS or arXiv
- » **Work alone** and email/phone one friend if you have any doubt

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

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

 [jer@iaa.es](mailto:jer@iaa.es)

 [bultako](#)