Digital Science

Going beyond automation

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

SESIONES CCD
GRANADA, SEPTEMBER 26th 2012

AMIGA
Analysis of the Interstellar Medium of Isolated GAlaxies

INSTITUTO de ASTROFISICA de ANDALUCIA
CSIC
I·A·A

CSIC

Workflow 4Ever

SEVENTH FRAMEWORK PROGRAMME

European Commission
Information Society and Media

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

Efficiency and Reuse

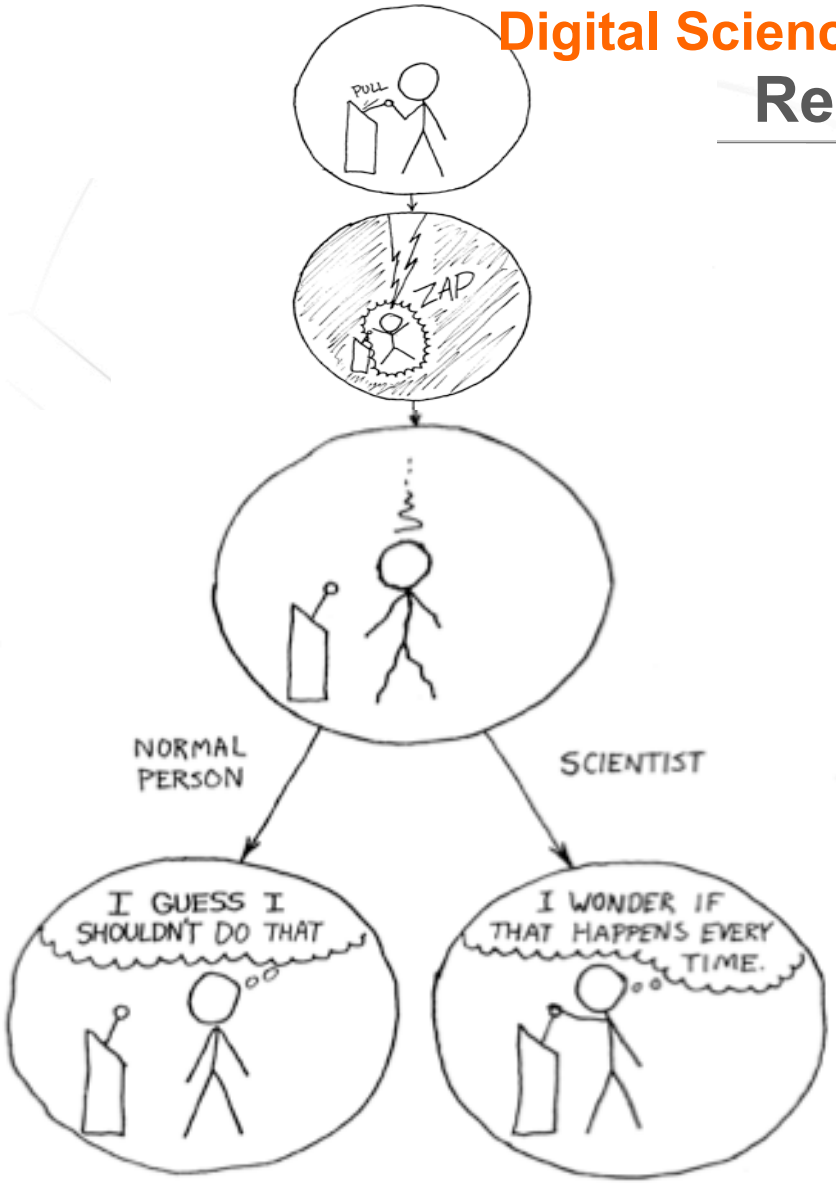
Optimize return on investments made on big facilities

- » 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

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%	I don't know how	-
34%	Legal Barriers (ie. copyright)	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%

Publishing: Discovery, Visibility and Credit

nature International weekly journal of science

Search Go

Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Specials & supplements archive > Science Metrics

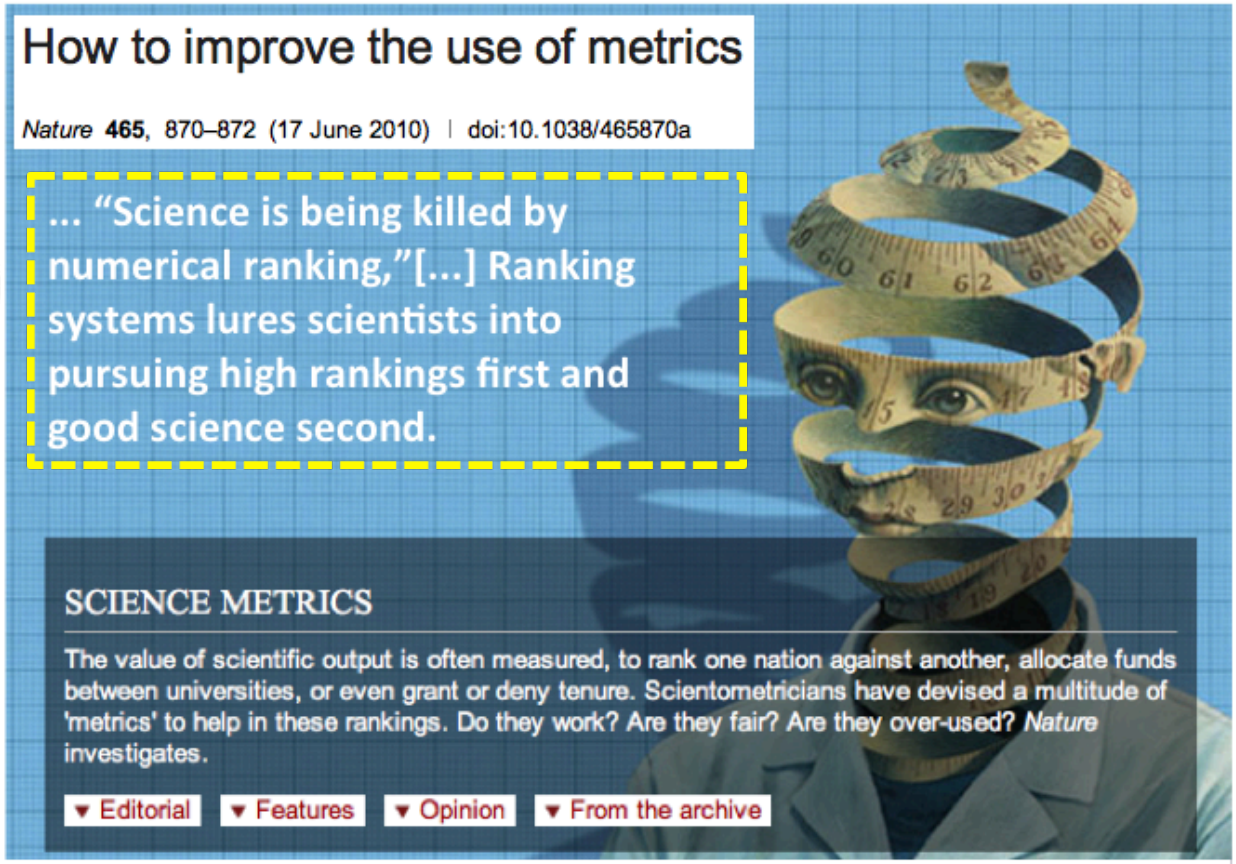
SPECIALS

See all specials

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

Journal home | Subscribe
Current issue | E-alert sign up
For authors | RSS feed

Top content

Emailed | Downloaded | Blogged

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

View all

Publishing: Discovery, Visibility and Credit

nature International weekly journal of science

Search | Go | 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

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

Journal home | Subscribe
Current issue | E-alert sign up
For authors | RSS feed

Top content

Emailed | Downloaded | Blogged

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

View all

Publishing: Discovery, Visibility and Credit

nature International weekly journal of science

Search Go

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

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

Journal home | Subscribe
Current issue | E-alert sign up
For authors | RSS feed

Top content

Emailed | Downloaded | Blogged

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

View all

Exploring and understanding scientific metrics in citation

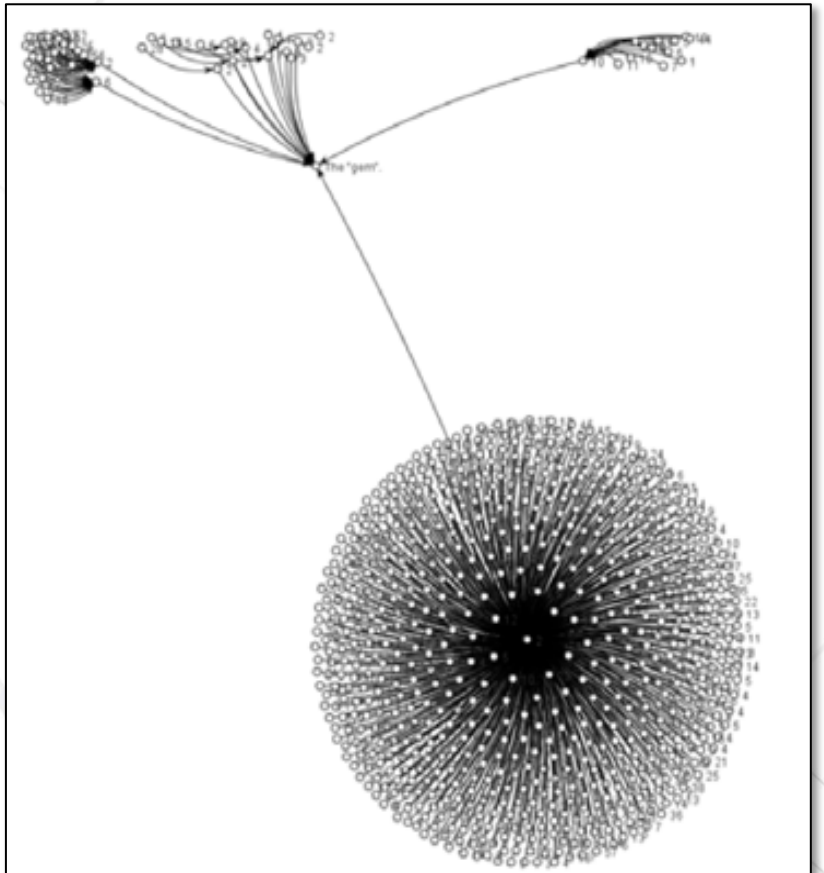


Fig. 4. "Hidden gem" in the dataset

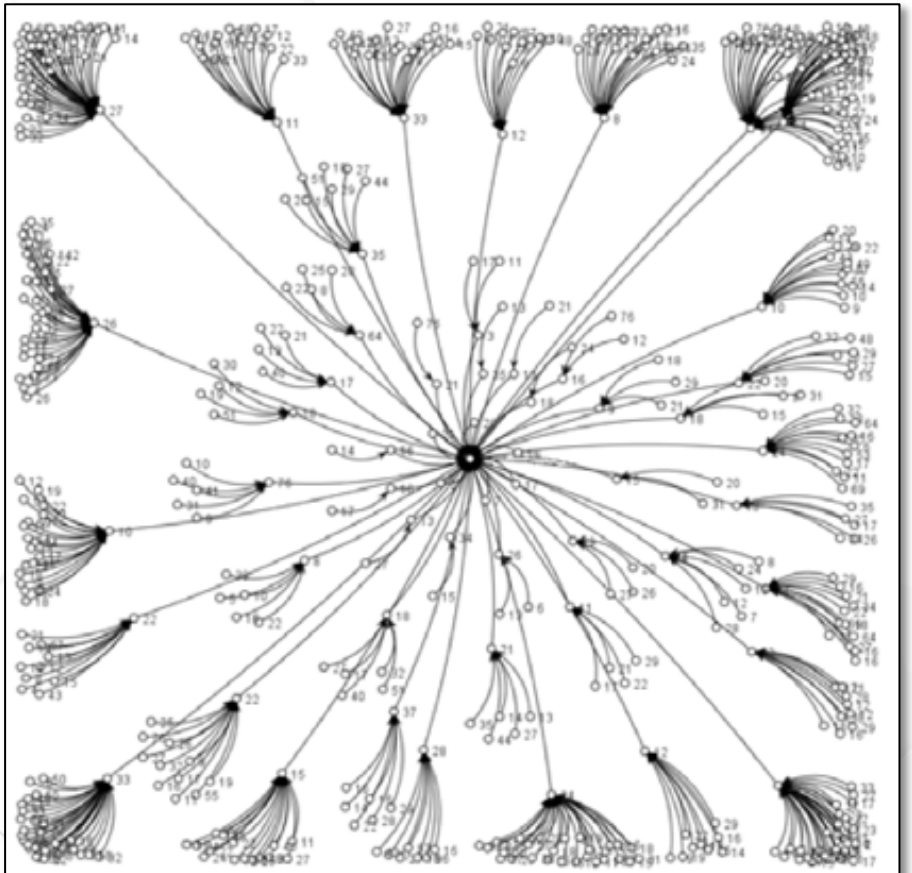


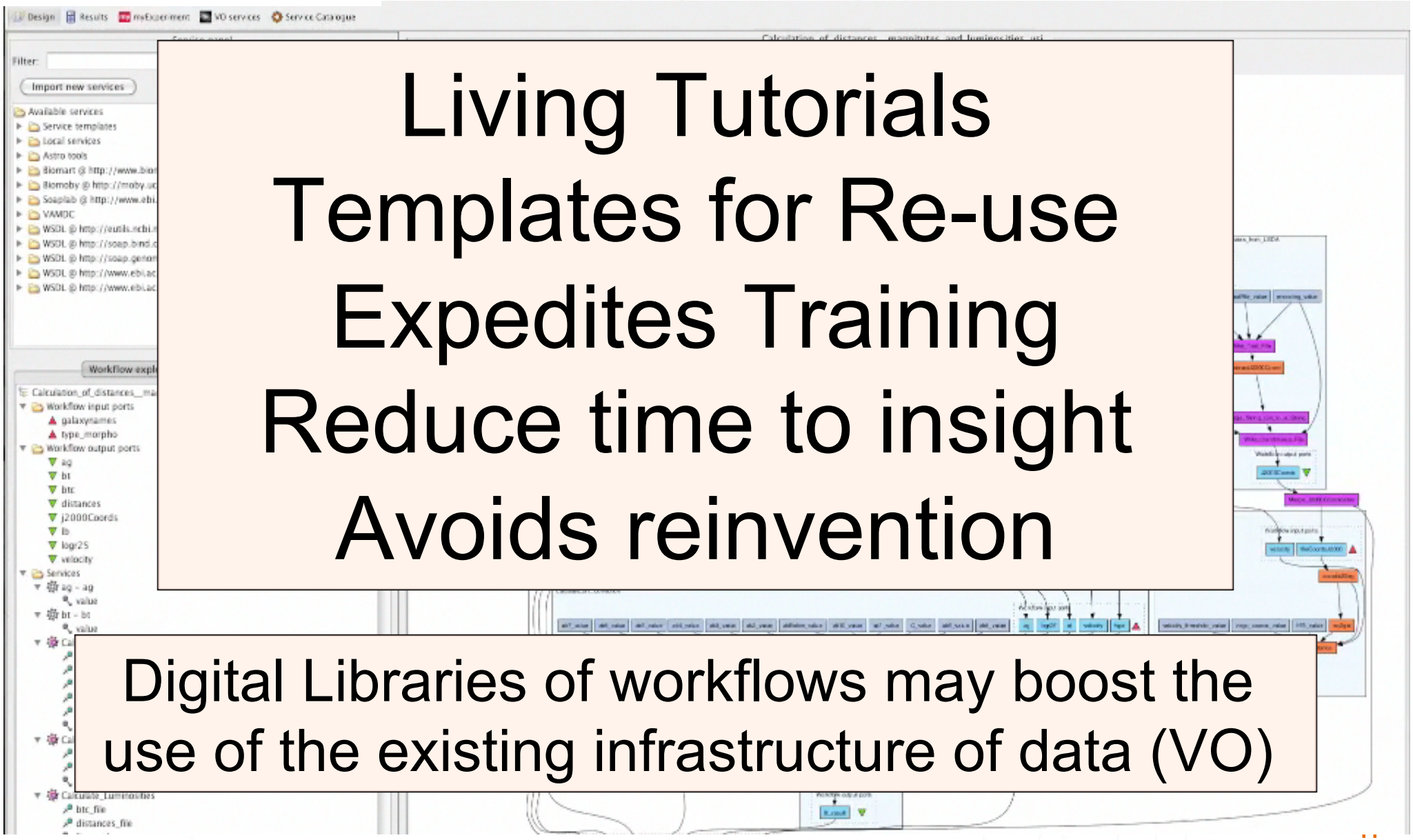
Fig. 5. "Popular paper" (in the center).

LiquidPub

2010 Krapivin et al.

Paper discovery: the social dimension

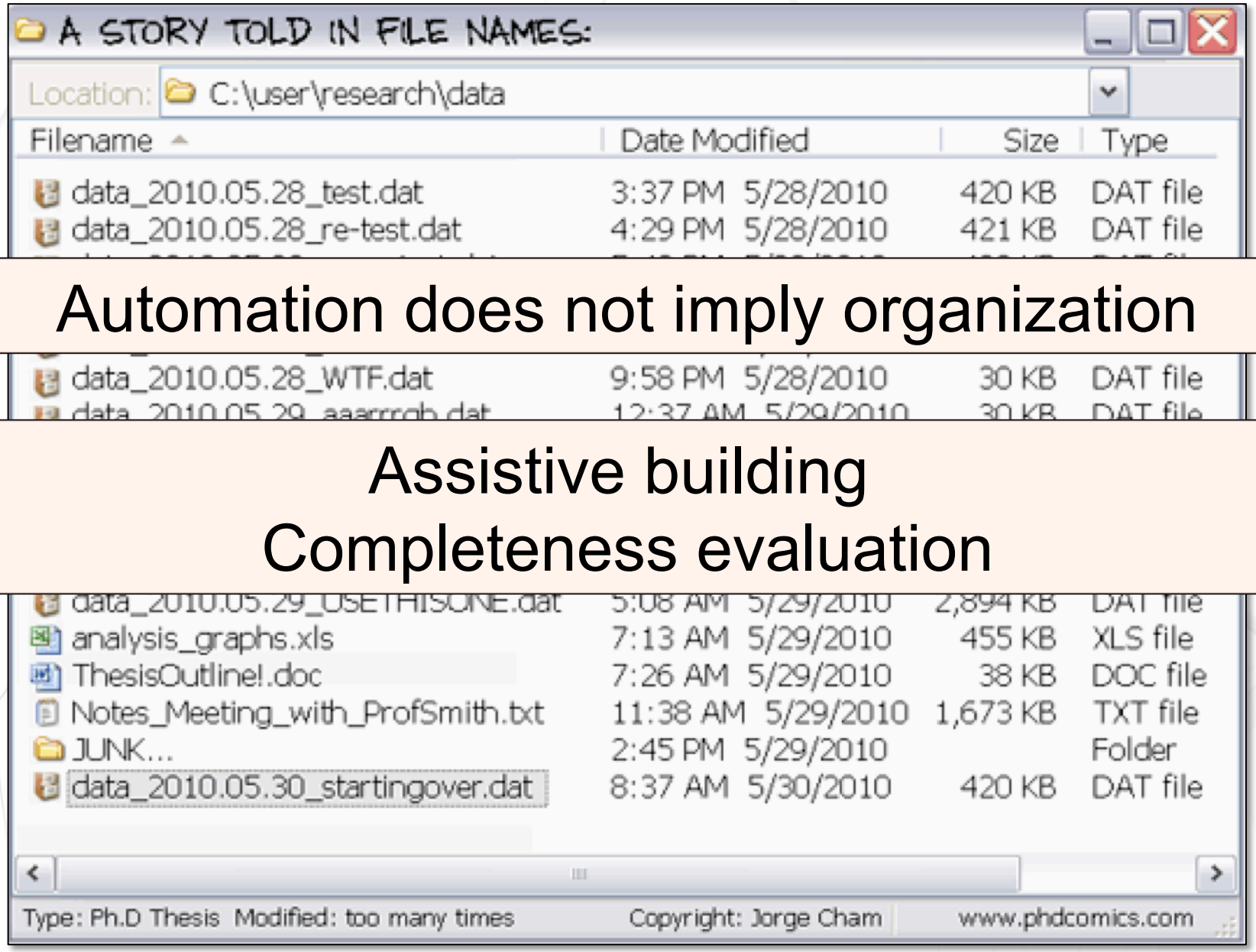
Going beyond automation: Scientific Workflows



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)

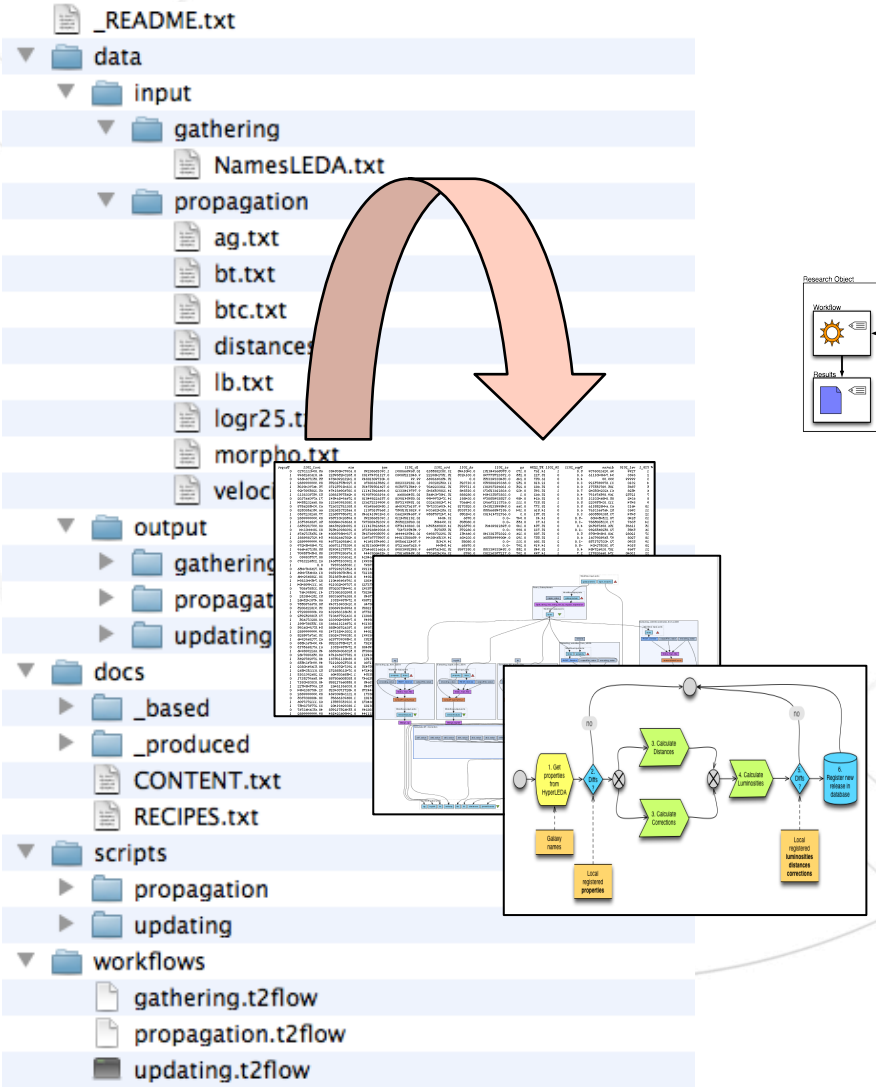
Going beyond automation: Scientific Workflows

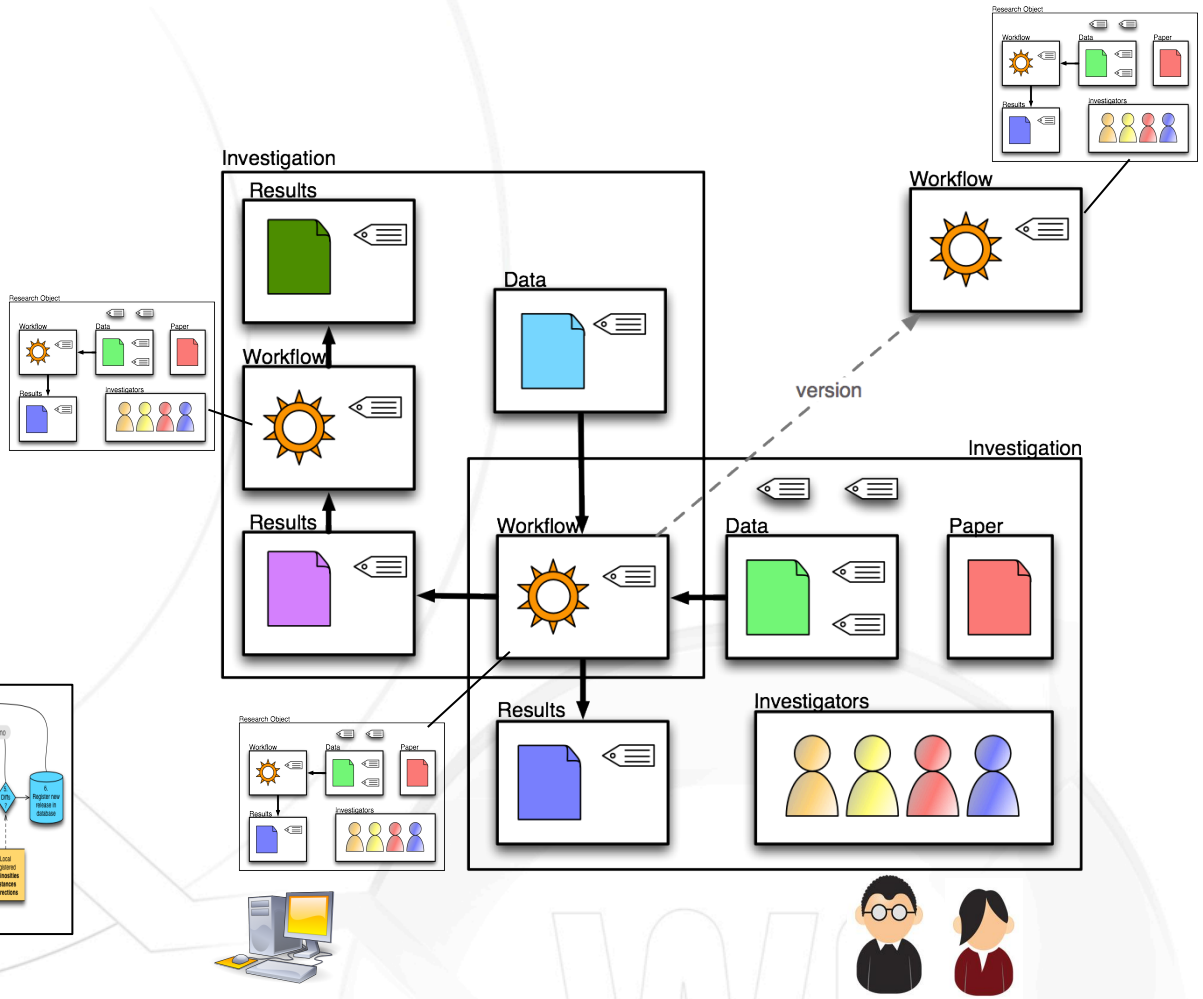


Automation does not imply organization

Assistive building
Completeness evaluation

Expose experiment in a structured way in order to be understood





Technical Objects

Distributed

Social Objects

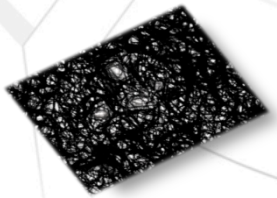
Similar initiatives in Astronomy

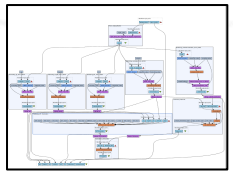
- **Semantic curation** of digital objects
  - CDS Centre Données Strasbourg
  - US Virtual Astronomical Observatory
  - SAO/NASA ADSLabs
- **Workflow users platforms**
  - Cyber-SKA
  - IceCore
  - Montage
  - Astro-WISE
  - Helio-VO
- **Semantically auto descriptive WS**
  - Workflows VO-France

ADSLabs Initiative

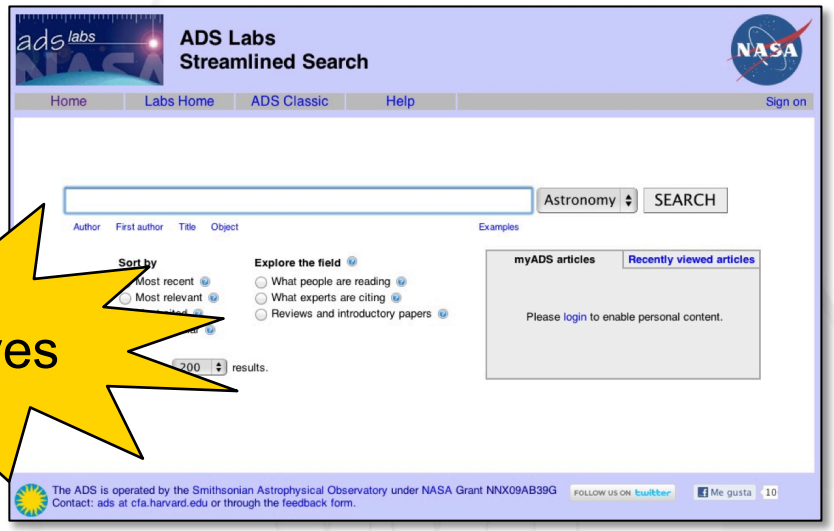
ADO Linked Components

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





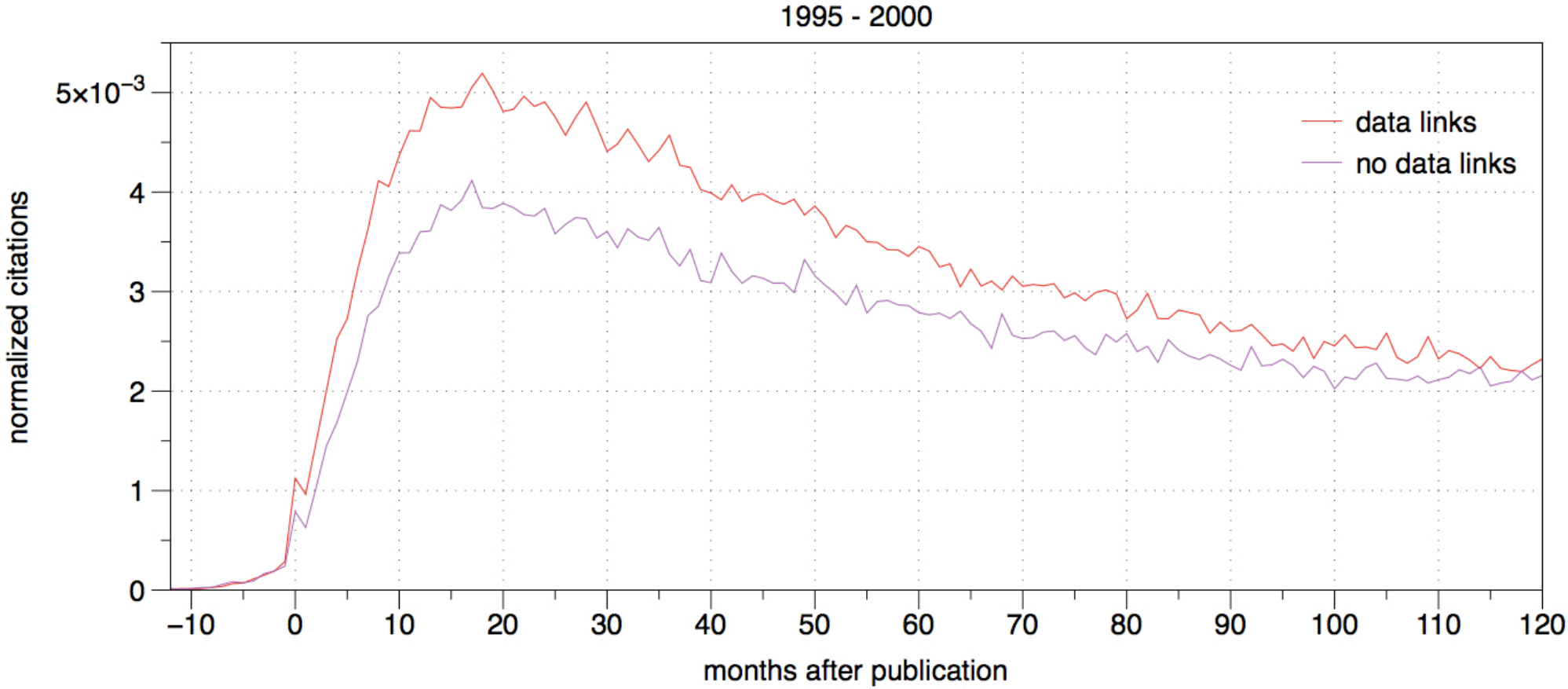
Incentives



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

The Incentive

Papers with data links are cited more than those without



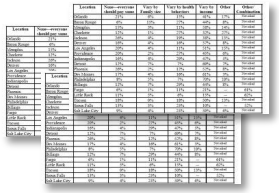
Effect of E-printing on Citation Rates in Astronomy and Physics
2006. Edwin A. Henneken et al.

The Wf4Ever Project

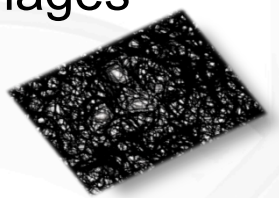
» Development of **AstroTaverna plugins** to access and manage VO data

» Development **Golden Exemplars** of astrophysical Workflows and Research Objects that use the Wf4Ever technological support

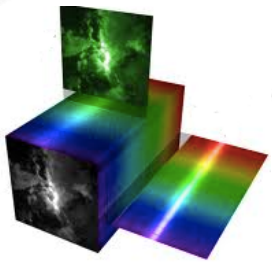
› **Curation** of physical quantities in 1D catalogues
- Data retrieved from external repositories and stored locally
- Only local processes for calculations



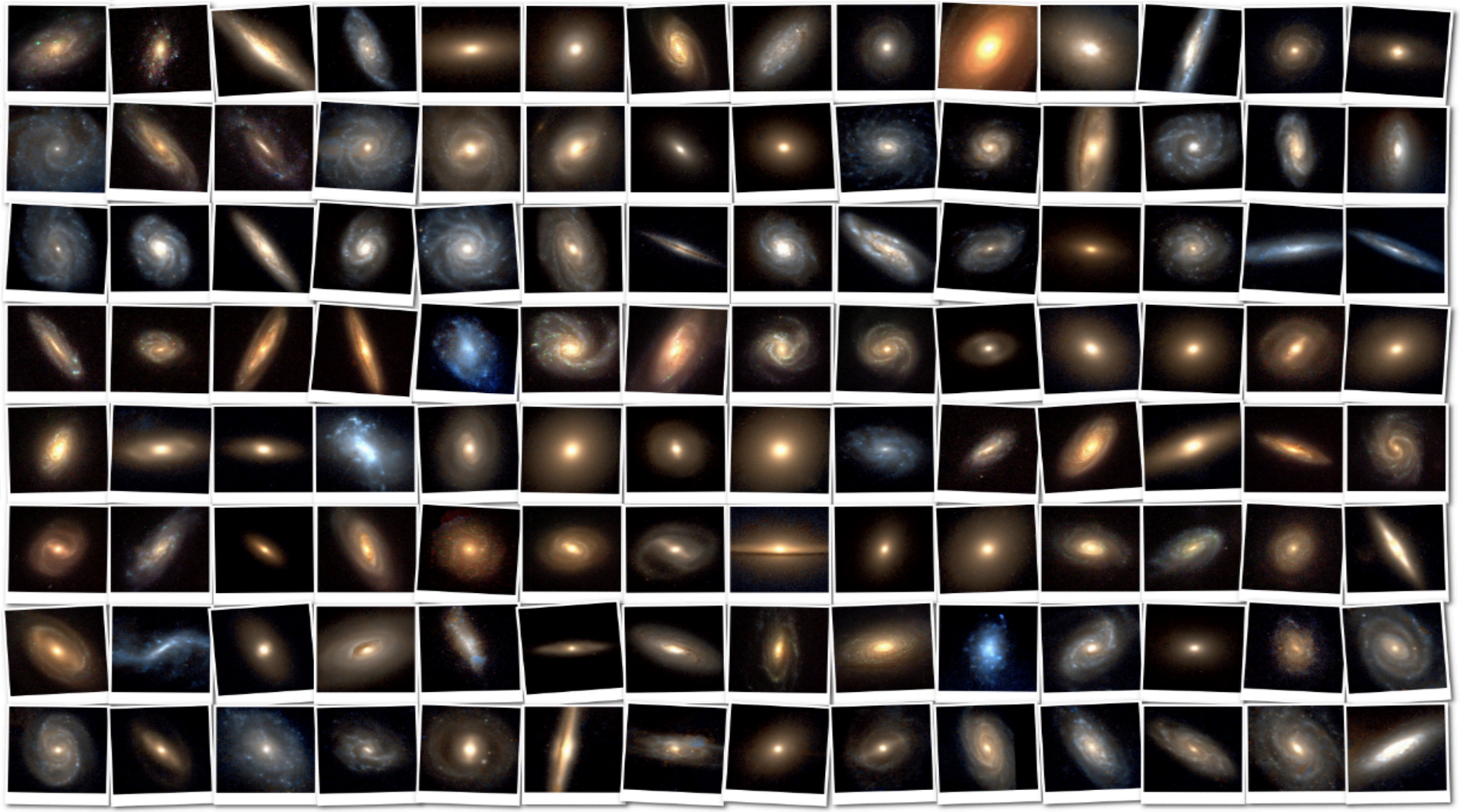
› **Environment and Modelling** from 1D catalogues and 2D images
- Data retrieved from external repositories (SDSS DR7)
- Local software and external web services as processes



› **Modelling and Analysis** of 3D formatted data
- Only external data and processes

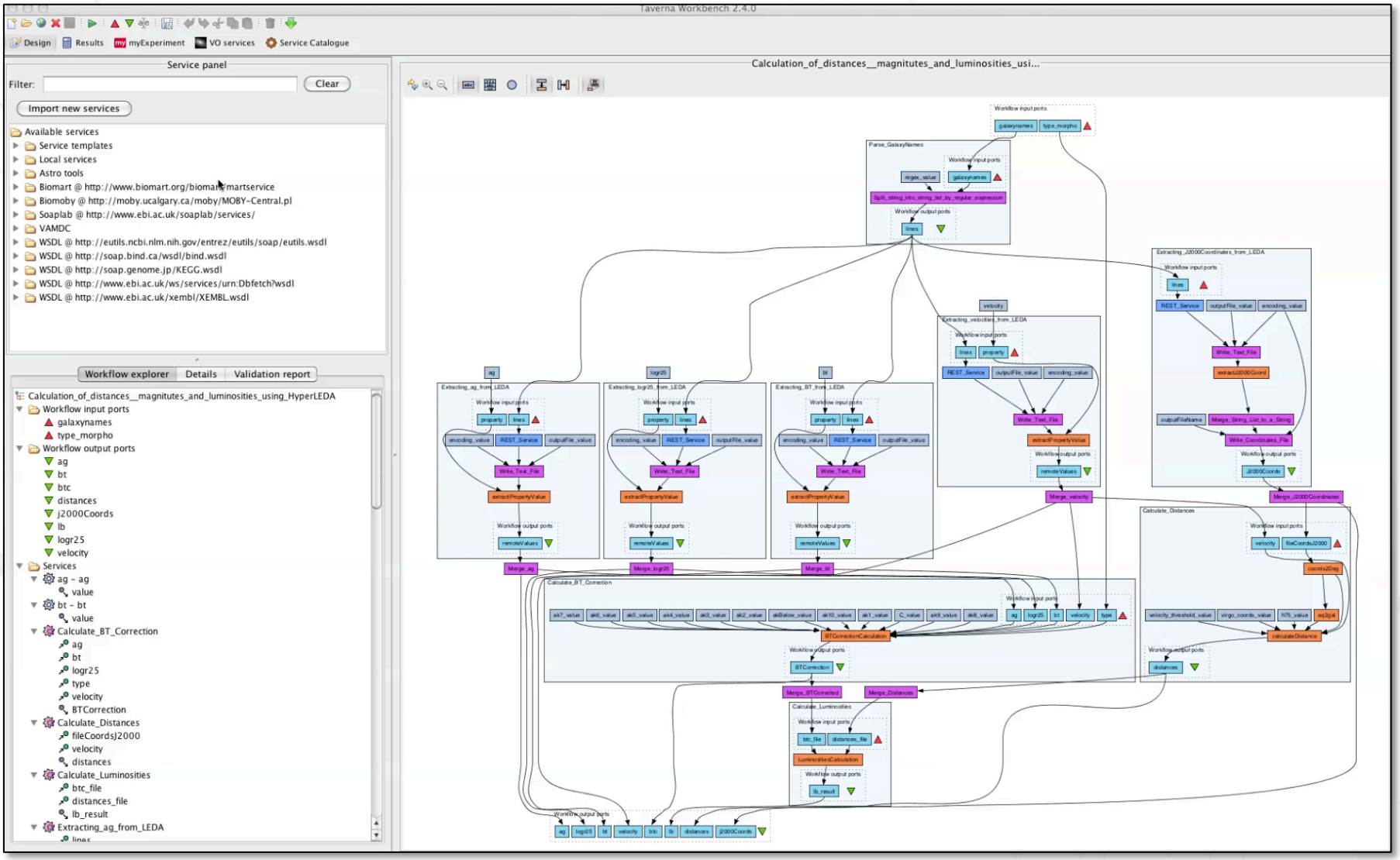


Curation by inspecting propagation of changes in quantities

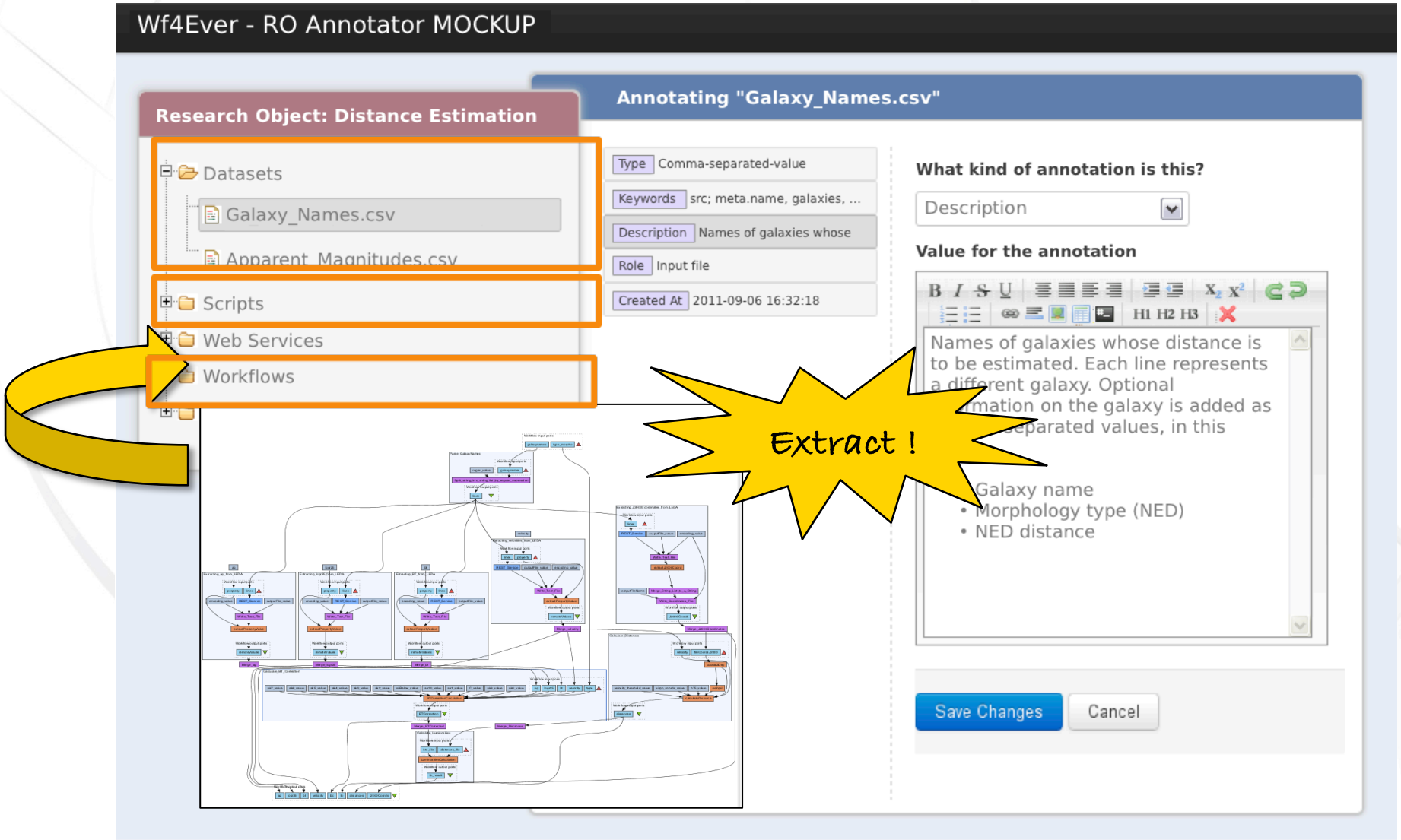


Credit: Zsolt Frei and James E. Gunn. The Galaxy Catalog

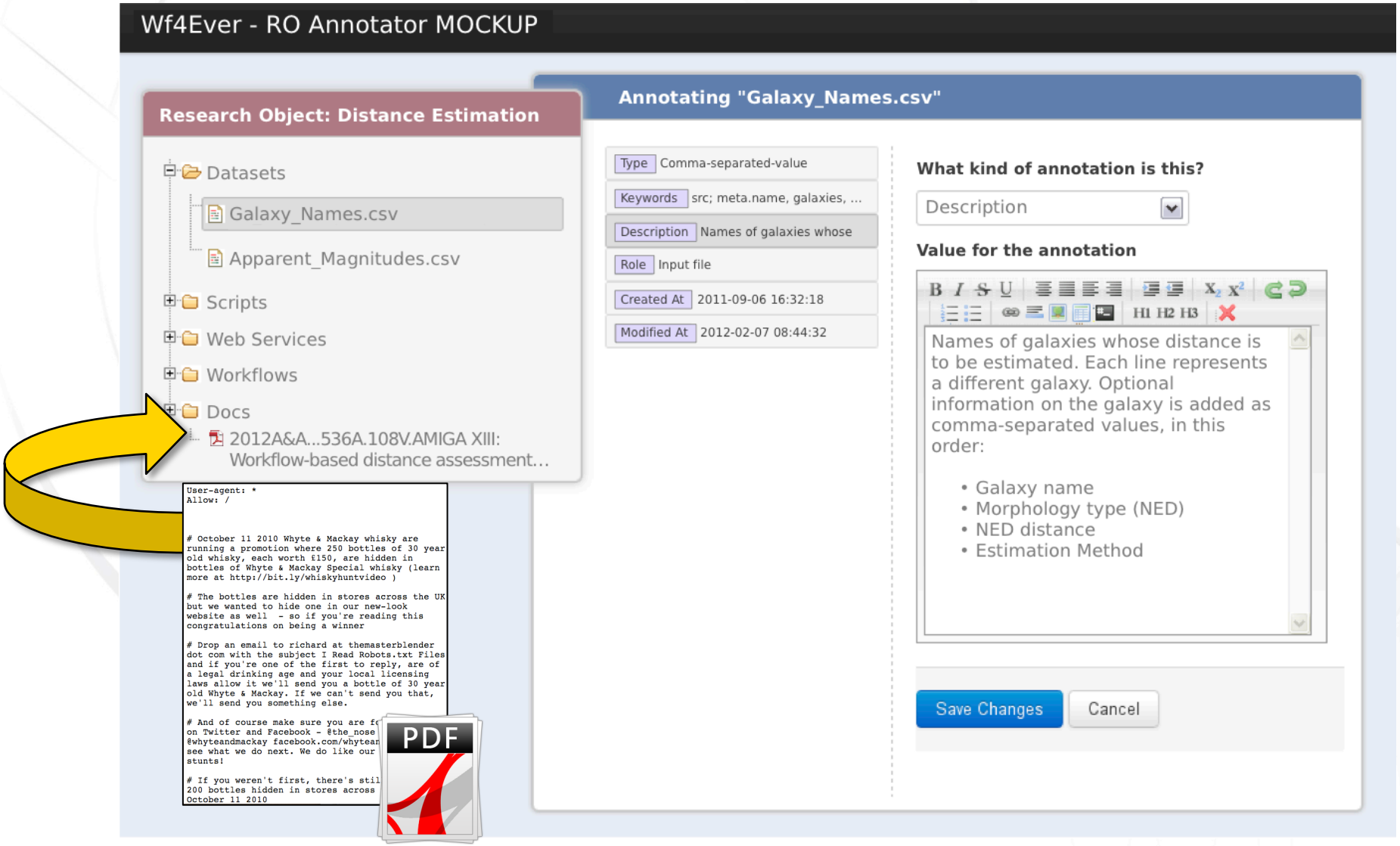
Create, annotate and run a workflow



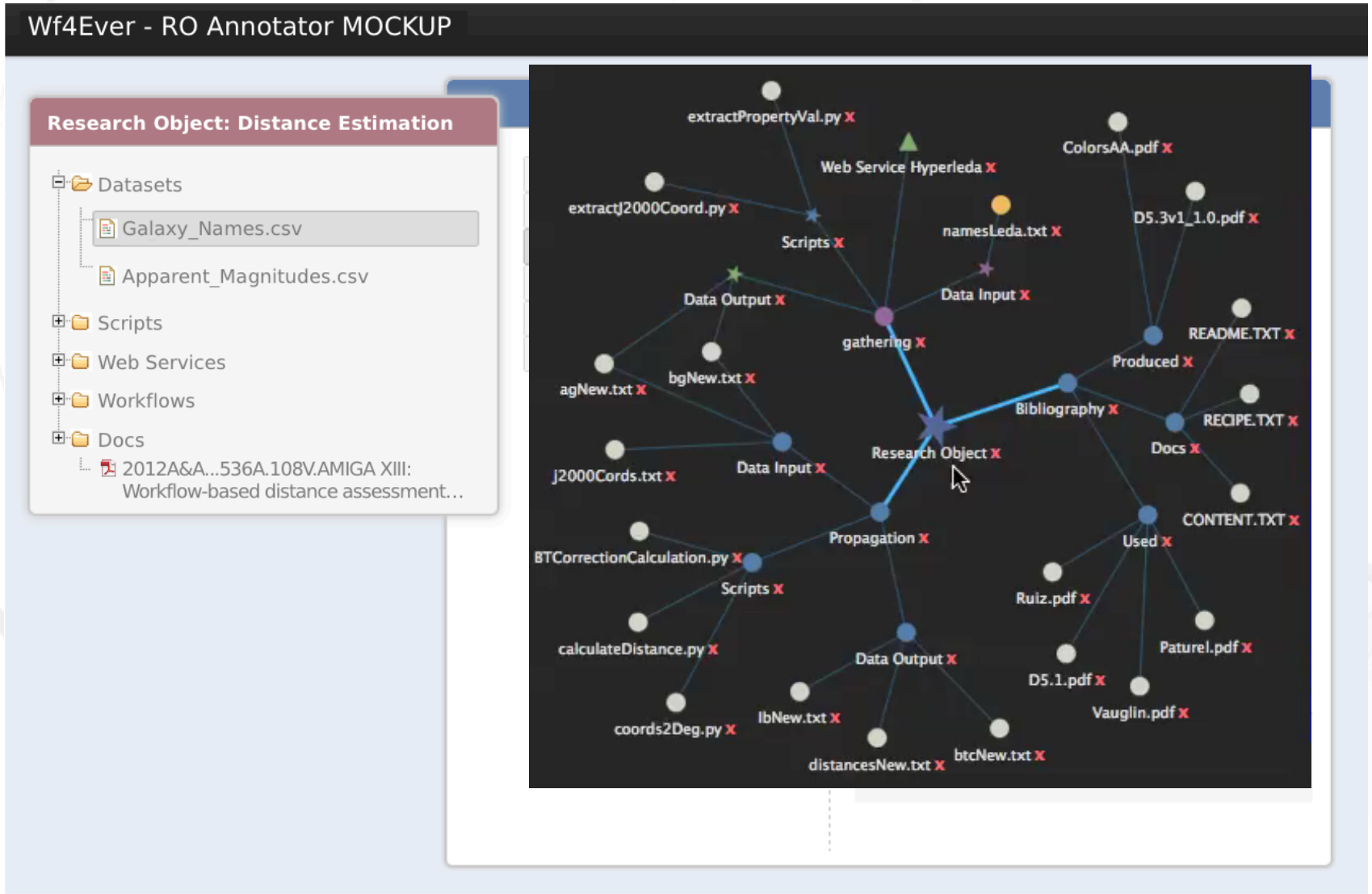
Populate the Research Object and annotate



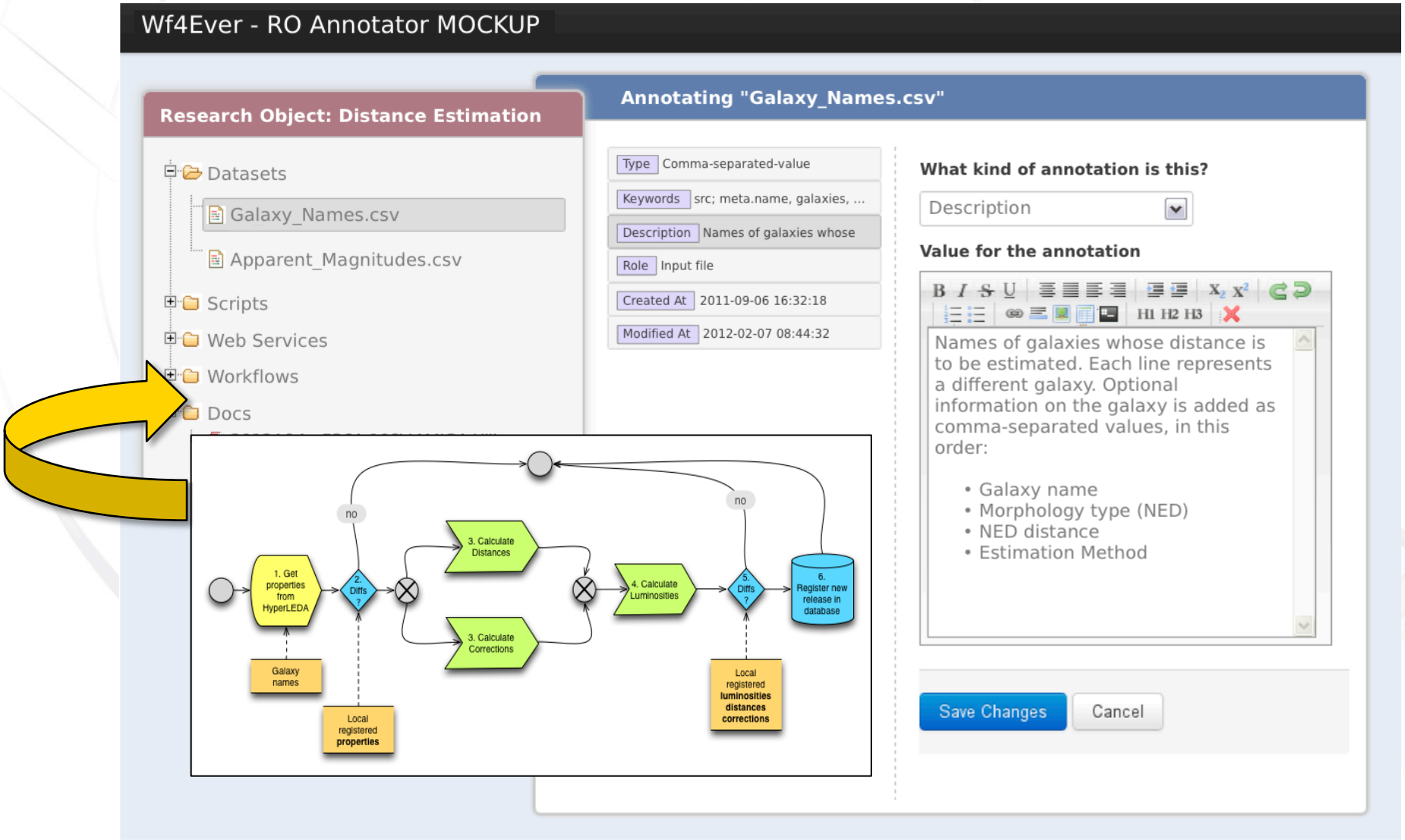
Add documents and references



Create and explore relations among components



Add schema of the experiment



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_f
    - gathering_galaxy_properties_using_hyperleda_129473.
  - Workflow Runs
  - Documents
  - GoldenTrace.txt

Import and re-use !

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

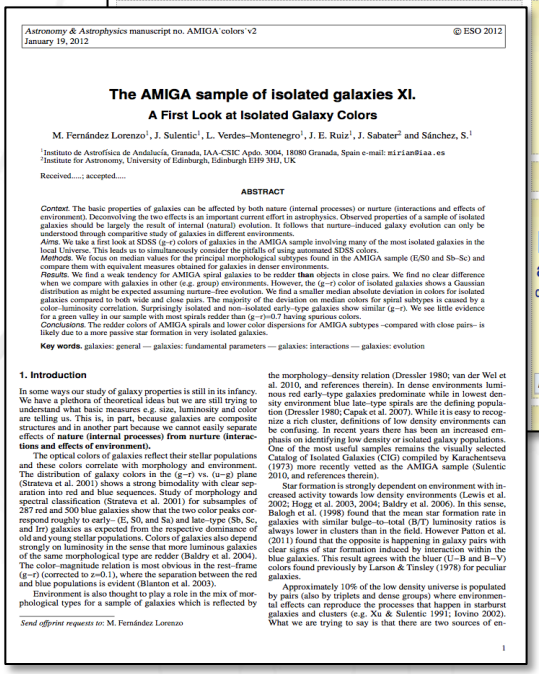
Re-used 4

Comments 2

<< Previous version | Next version >>

Curation by inspecting propagation of changes in quantities

- Taverna 2.3
- MyExperiment Pack
  - *http://www.myexperiment.org/packs/231*



Astronomy & Astrophysics manuscript no. AMIGA'colors'v2 © ESO 2012
January 19, 2012

The AMIGA sample of isolated galaxies XI.
A First Look at Isolated Galaxy Colors

M. Fernández Lorenzo[1], J. Sulentic[1], L. Verdes–Montenegro[1], J. E. Ruiz[1], J. Sabater[2] and Sánchez, S.[1]

[1]Instituto de Astrofísica de Andalucía, Granada, IAA-CSIC Apdo. 3004, 18080 Granada, Spain e-mail: mirian@iaa.es
[2]Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK

Received.....; accepted.....

ABSTRACT

Context. The basic properties of galaxies can be affected by both nature (internal processes) or nurture (interactions and effects of environment). Deconvolving the two effects is an important current effort in astrophysics. Observed properties of a sample of isolated galaxies should be largely the result of internal (natural) evolution. It follows that nurture–induced galaxy evolution can only be understood through comparitive study of galaxies in different environments.
Aims. We take a first look at SDSS (g−r) colors of galaxies in the AMIGA sample involving many of the most isolated galaxies in the local Universe. This leads us to simultaneously consider the pitfalls of using automated SDSS colors.
Methods. We focus on median values for the principal morphological subtypes found in the AMIGA sample (E/S0 and Sb–Sc) and compare them with equivalent measures obtained for galaxies in denser environments.
Results. We find a weak tendency for AMIGA spiral galaxies to be redder than objects in close pairs. We find no clear difference when we compare with galaxies in other (e.g. group) environments. However, the (g−r) color of isolated galaxies shows a Gaussian distribution as might be expected assuming nurture–free evolution. We find a smaller median absolute deviation in colors for isolated galaxies compared to both wide and close pairs. The majority of the deviation on median colors for spiral subtypes is caused by a color–luminosity correlation. Surprisingly isolated and non–isolated early–type galaxies show similar (g−r). We see little evidence for a green valley in our sample with most spirals redder than (g−r)=0.7 having spurious colors.
Conclusions. The redder colors of AMIGA spirals and lower color dispersions for AMIGA subtypes –compared with close pairs– is likely due to a more passive star formation in very isolated galaxies.

Key words. galaxies: general — galaxies: fundamental parameters — galaxies: interactions — galaxies: evolution

1. Introduction

In some ways our study of galaxy properties is still in its infancy. We have a plethora of theoretical ideas but we are still trying to understand what basic measures e.g. size, luminosity and color are telling us. This is, in part, because galaxies are composite structures and in another part because we cannot easily separate effects of **nature (internal processes) from nurture (interactions and effects of environment).**

The optical colors of galaxies reflect their stellar populations and these colors correlate with morphology and environment. The distribution of galaxy colors in the (g−r) vs. (u−g) plane (Strateva et al. 2001) shows a strong bimodality with clear separation into red and blue sequences. Study of morphology and spectral classification (Strateva et al. 2001) for subsamples of 287 red and 500 blue galaxies show that the two color peaks correspond roughly to early– (E, S0, and Sa) and late–type (Sb, Sc, and Irr) galaxies as expected from the respective dominance of old and young stellar populations. Colors of galaxies also depend strongly on luminosity in the sense that more luminous galaxies of the same morphological type are redder (Baldry et al. 2004). The color–magnitude relation is most obvious in the rest–frame (g−r) (corrected to z=0.1), where the separation between the red and blue populations is evident (Blanton et al. 2003).

Environment is also thought to play a role in the mix of morphological types for a sample of galaxies which is reflected by the morphology–density relation (Dressler 1980; van der Wel et al. 2010, and references therein). In dense environments luminous red early–type galaxies predominate while in lowest density environment blue late–type spirals are the defining population (Dressler 1980; Capak et al. 2007). While it is easy to recognize a rich cluster, definitions of low density environments can be confusing. In recent years there has been an increased emphasis on identifying low density or isolated galaxy populations. One of the most useful samples remains the visually selected Catalog of Isolated Galaxies (CIG) compiled by Karachentseva (1973) more recently vetted as the AMIGA sample (Sulentic 2010, and references therein).

Star formation is strongly dependent on environment with increased activity towards low density environments (Lewis et al. 2002; Hogg et al. 2003, 2004; Baldry et al. 2006). In this sense, Balogh et al. (1998) found that the mean star formation rate in galaxies with similar bulge–to–total (B/T) luminosity ratios is always lower in clusters than in the field. However Patton et al. (2011) found that the opposite is happening in galaxy pairs with clear signs of star formation induced by interaction within the blue galaxies. This result agrees with the bluer (U−B and B−V) colors found previously by Larson & Tinsley (1978) for peculiar galaxies.

Approximately 10% of the low density universe is populated by pairs (also by triplets and dense groups) where environmental effects can reproduce the processes that happen in starburst galaxies and clusters (e.g. Xu & Sulentic 1991; Iovino 2002). What we are trying to say is that there are two sources of en-

Send offprint requests to: M. Fernández Lorenzo

1

Related Publication

The AMIGA sample of isolated galaxies XI.

A First Look at Isolated Galaxy Colors

2012 A&A 540, A.47

SOFTWARE

- http://wf4ever.github.com/astrotaverna/
- http://www.taverna.org.uk/

Hands on
AstroTaverna !

VIDEO TUTORIALS

- http://amiga.iaa.es/files/tavernavideos/AstroTavernaIntro.m4v
- http://amiga.iaa.es/files/tavernavideos/NEDImages.mov

How NOT to be a good e-astronomer

- 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" – data are mine
- Practise the "data flirting" – call me if you would like to see 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