Analysis of the Interstellar Medium of Isolated GAlaxies

AMIGA

Digital Science

Towards the Executable Paper

José Enrique Ruiz on behalf of the Wf4Ever and CANUBE Teams

IAA Seminars
IAA - CSIC, THURSDAY 31st OCTOBER 2013

SEVENTH FRAMEWORK PROGRAMME
European Commission
Information Society and Media

Wf4Ever
Advanced Workflow Preservation Technologies for Enhanced Science

2011 – 2013 EU FP7

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 (Poland)
5. University of Oxford and OeRC (OXF, UK)

6. Instituto Astrofísica Andalucía (IAA-CSIC, Sp
7. Leiden University Medical Centre (LUMC

Reproducible Science

IAA – CSIC contribution through AMIGA Group

- **User Functional Requirements**
  - BioGenomics /BioInformaticians
  - Astronomers /AstroInformaticians
  - Publishers /Librarians
  - Computer Scientists

- **Software Development**
  - AstroTaverna Plugin
  - AstroTaverna Starter Pack and Workflows

- **Community Engagement and Collaborations**
  - Spanish Virtual Observatory
  - International Virtual Observatory Alliance
  - Action Spécifique Observatoires Virtuels France
  - Observatoire de Paris-Meudon
  - EU FP7 Projects : Er-Flow and VAMDC
  - SAO NASA /ADS Digital Library

Reproducible Science

CANube
Ciencia Abierta en la Nube

Mars – Dec 2013

Open Science Project granted by the Second Call for Proposals of the **Bio-TIC Campus of International Excellence** of the University of Granada.

- Universidad de Granada
- Instituto Astrofísica Andalucía - CSIC
- Campus CEI-BioTic

- Red del Sur
- Fidesol
- Intelify
- Grupo Trevenque

Open Science

Digital Science - Towards the Executable Paper
Digital Astronomy

Astronomy research lifecycle is **entirely digital**

- Observation proposals
- Data reduction pipelines
- Analysis of science ready data
- Catalogs of objects and data archives
- 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 is not efficiently used

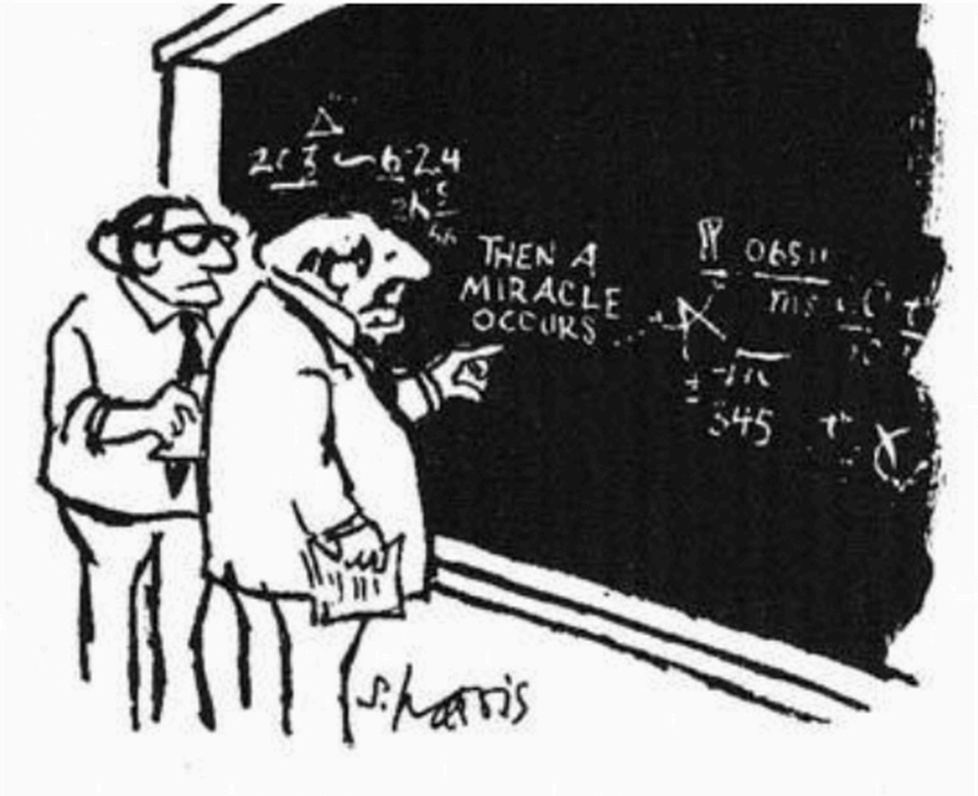
A normalized preservation of methodology is needed

Tools

Reproducibility

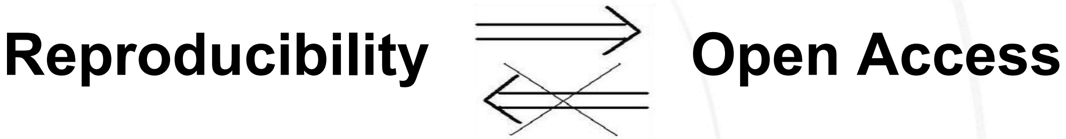
"... up to 70% of research from academic labs **cannot be reproduced**, representing an enormous waste of money and effort."

- Elizabeth Iorns, Science Exchange



"I think you should be more explicit in step two"

Reproducibility is achieved when access is granted for all resources



Reproducibility

Open Access

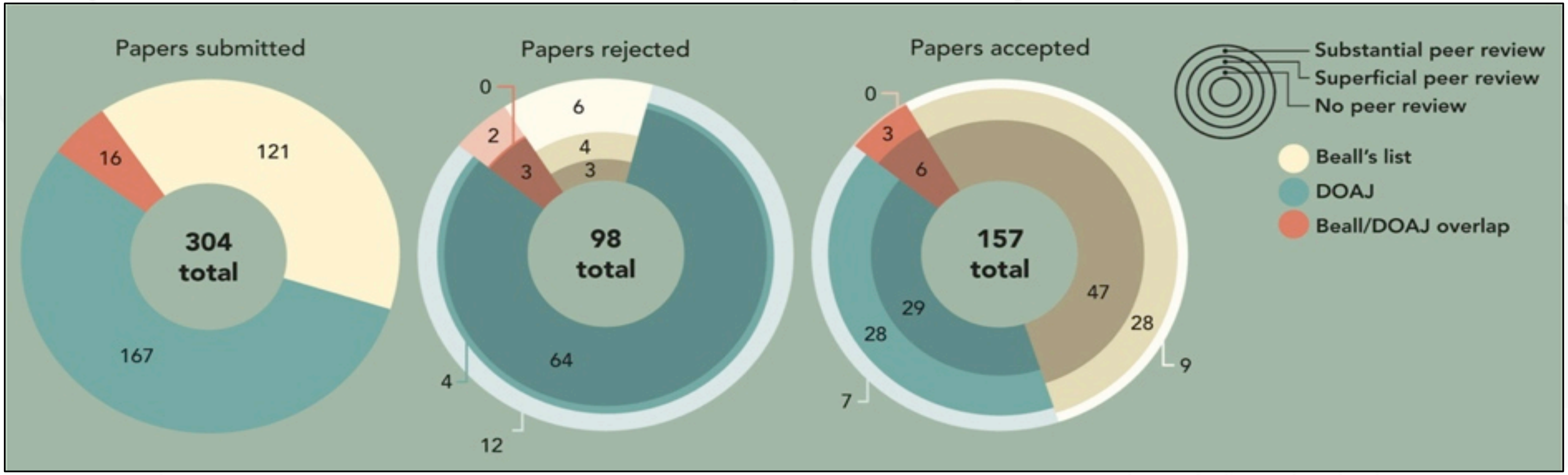
Clamorous fake methods and results published in 157 out of 304 Journals

Who's Afraid of Peer Review?

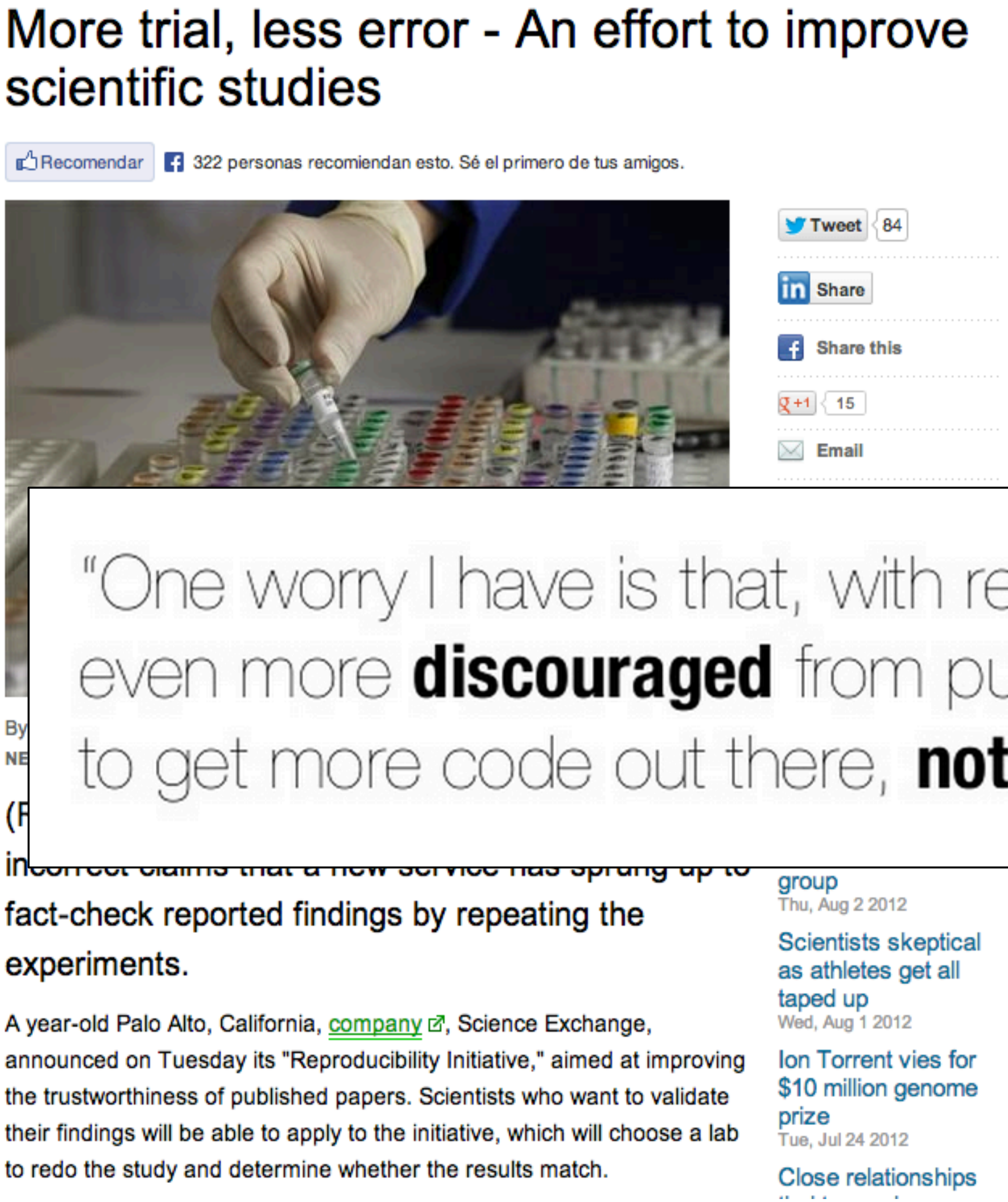
John Bohannon

A spoof paper concocted by *Science* reveals little or no scrutiny at many open-access journals.

Science 4 October 2013:
Vol. 342 no. 6154 pp. 60-65
DOI: 10.1126/science.342.6154.60



Open Science



More trial, less error - An effort to improve scientific studies

By
NE

(R

incorrect claims that a new service has sprung up to fact-check reported findings by repeating the experiments.

A year-old Palo Alto, California, company, Science Exchange, announced on Tuesday its "Reproducibility Initiative," aimed at improving the trustworthiness of published papers. Scientists who want to validate their findings will be able to apply to the initiative, which will choose a lab to redo the study and determine whether the results match.

group
Thu, Aug 2 2012

Scientists skeptical as athletes get all taped up
Wed, Aug 1 2012

Ion Torrent vies for $10 million genome prize
Tue, Jul 24 2012

Close relationships

nature International weekly journal of science
Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Au
Archive > Volume 501 > Issue 7468 > News > Article

NATURE | NEWS

Mozilla plan seeks to debug scientific code

Software experiment raises prospect of extra peer review.

Erika Check Hayden

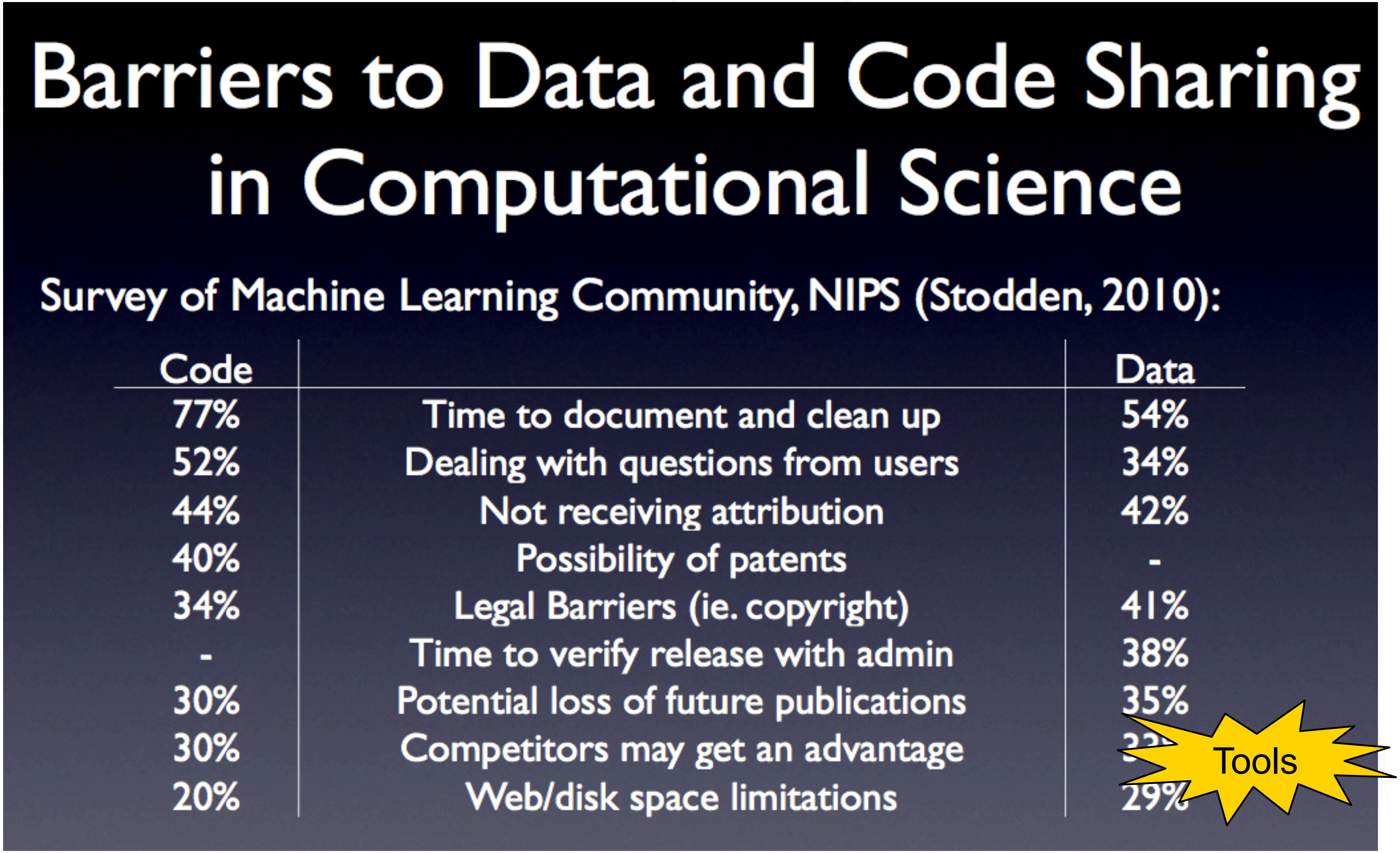
24 September 2013

entered the debate, aiming to discover whether a review process could improve the quality of researcher-built software that is used in myriad fields today, ranging from ecology and biology to social science. In an experiment being run by the Mozilla Science Lab, software engineers have reviewed selected pieces of code from published papers in computational biology. "Scientific code does not have that comprehensive, off-the-shelf nature that we want to be associated with the way science is published and presented, and this is our attempt to poke at that issue," says Mozilla Science Lab director Kaitlin Thaney.

- Cancer institute tackles sloppy data
- Publish your computer code: it is good enough
- Computational science: ...Error

More related stories

> "One worry I have is that, with reviews like this, scientists will be even more **discouraged** from publishing their code [...] We need to get more code out there, **not improve how it looks**."



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%	Possibility of patents	-
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%

Open Science

Repeatable

The methodology is clearly exposed

I could repeat the experiment

Reproducible

Clear methodology and available resources

I could reproduce the results

Reusable

I know how it could be useful for my needs

I could use all or some parts as it is

I could modify and adapt it even for other purposes

Citations ?
30" BUILT-IN ELECTRIC OVEN SELF CLEAN CONVECTION
Part No. W10334312 Rev. A
$399
NEW LOWER PRICES, SAME GREAT QUALITY.
IKEA
IKEA-USA.COM

Visibility, Efficiency and Reuse

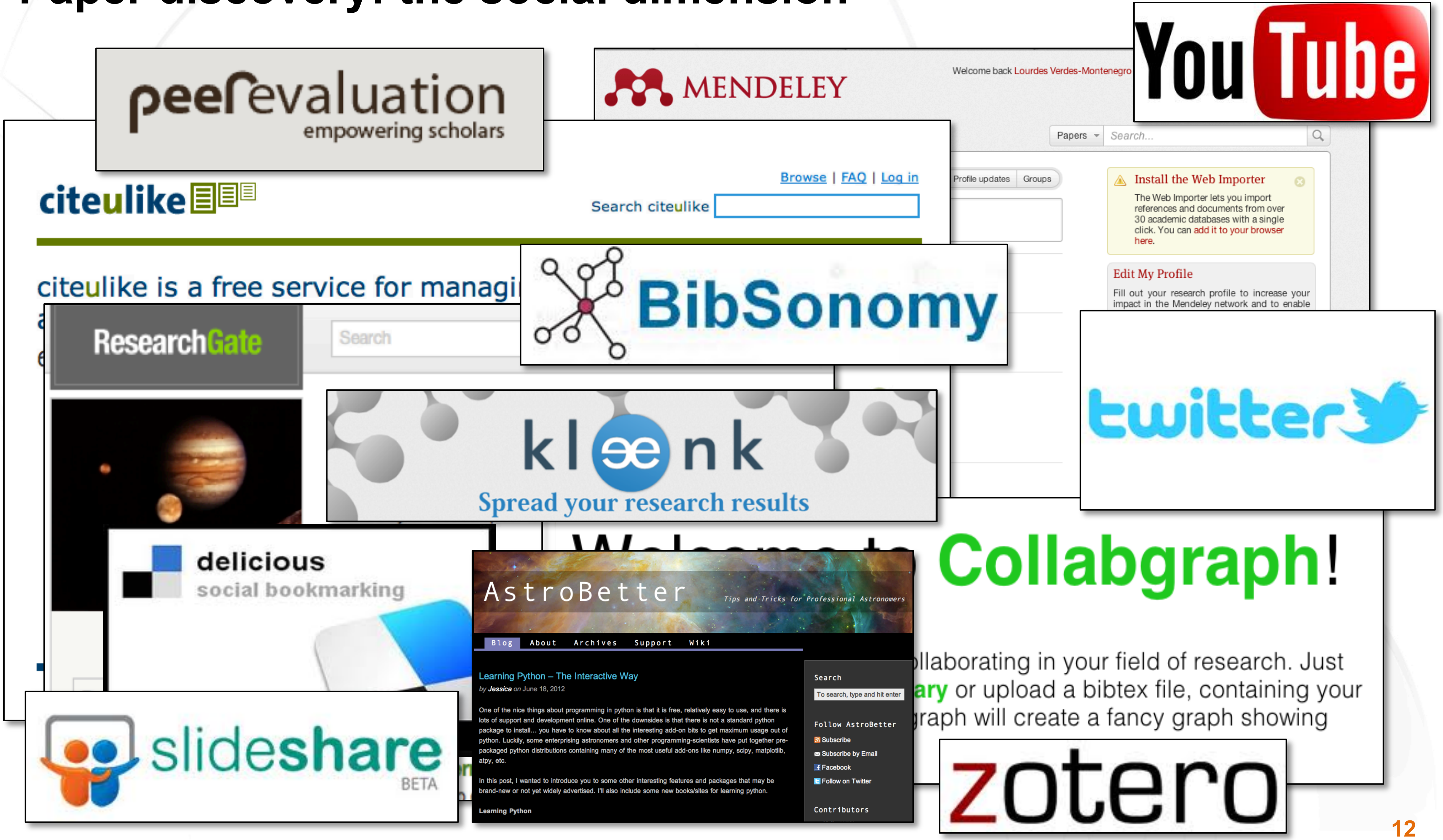
Optimize return on investments made on big facilities

- 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 encoded in software**

More data in archives do not imply more knowledge

- Expose **complete scientific record**, not the story
- Allow easy **discovery** of methods and tools

Paper discovery: the social dimension



Install the Web Importer

The Web Importer lets you import references and documents from over 30 academic databases with a single click. You can add it to your browser here.

Edit My Profile

Fill out your research profile to increase your impact in the Mendeley network and to enable

citeulike is a free service for managi

Learning Python – The Interactive Way

by ***Jessica*** *on* June 18, 2012

One of the nice things about programming in python is that it is free, relatively easy to use, and there is lots of support and development online. One of the downsides is that there is not a standard python package to install... you have to know about all the interesting add-on bits to get maximum usage out of python. Luckily, some enterprising astronomers and other programming-scientists have put together pre-packaged python distributions containing many of the most useful add-ons like numpy, scipy, matplotlib, atpy, etc.

In this post, I wanted to introduce you to some other interesting features and packages that may be brand-new or not yet widely advertised. I'll also include some new books/sites for learning python.

Learning Python

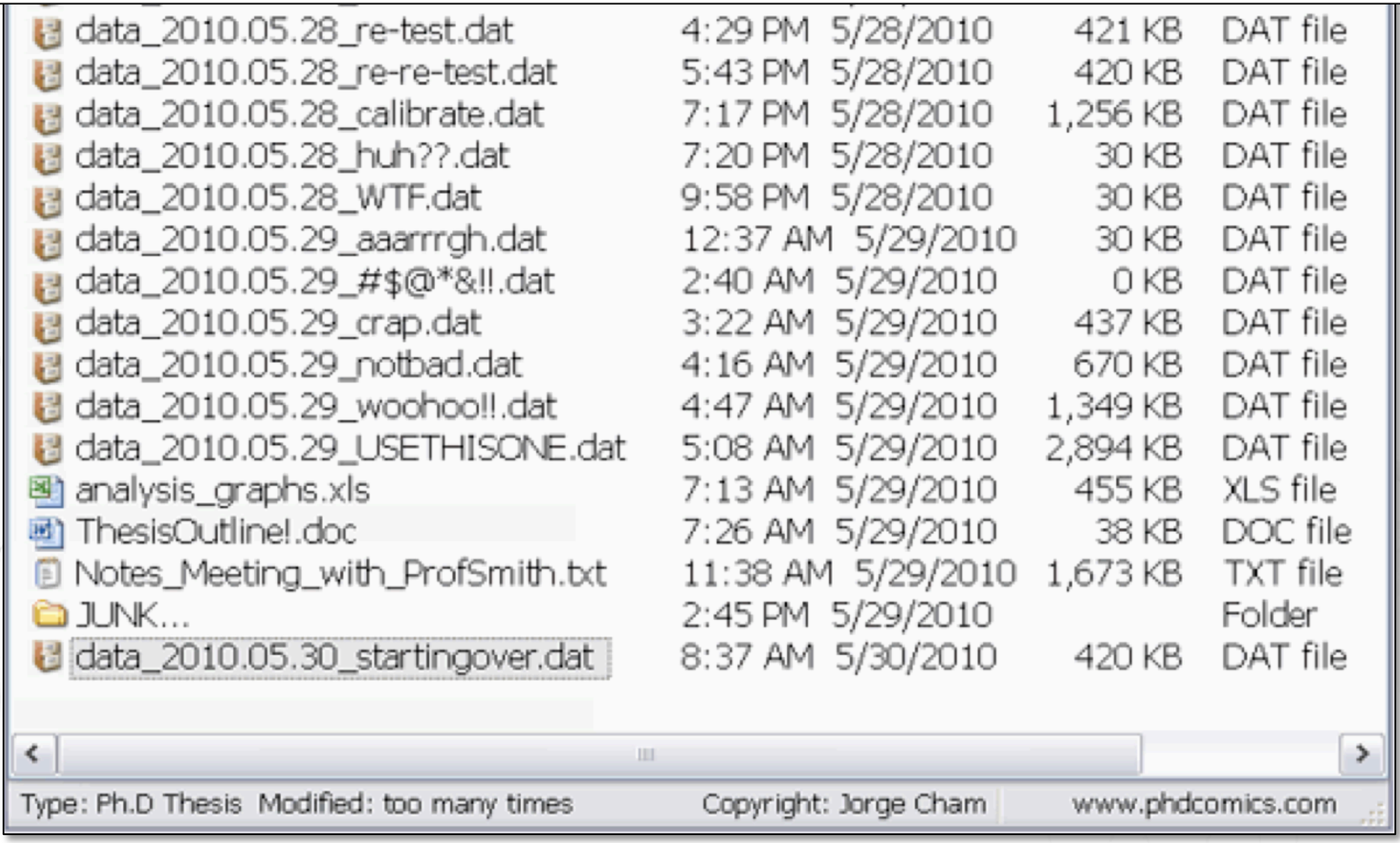
Welcome to Collabgraph!

llaborating in your field of research. Just
ary or upload a bibtex file, containing your
raph will create a fancy graph showing

Time has come to go beyond the PDF

Digital Astronomy in the Local Desktop

Going beyond automation
Organization

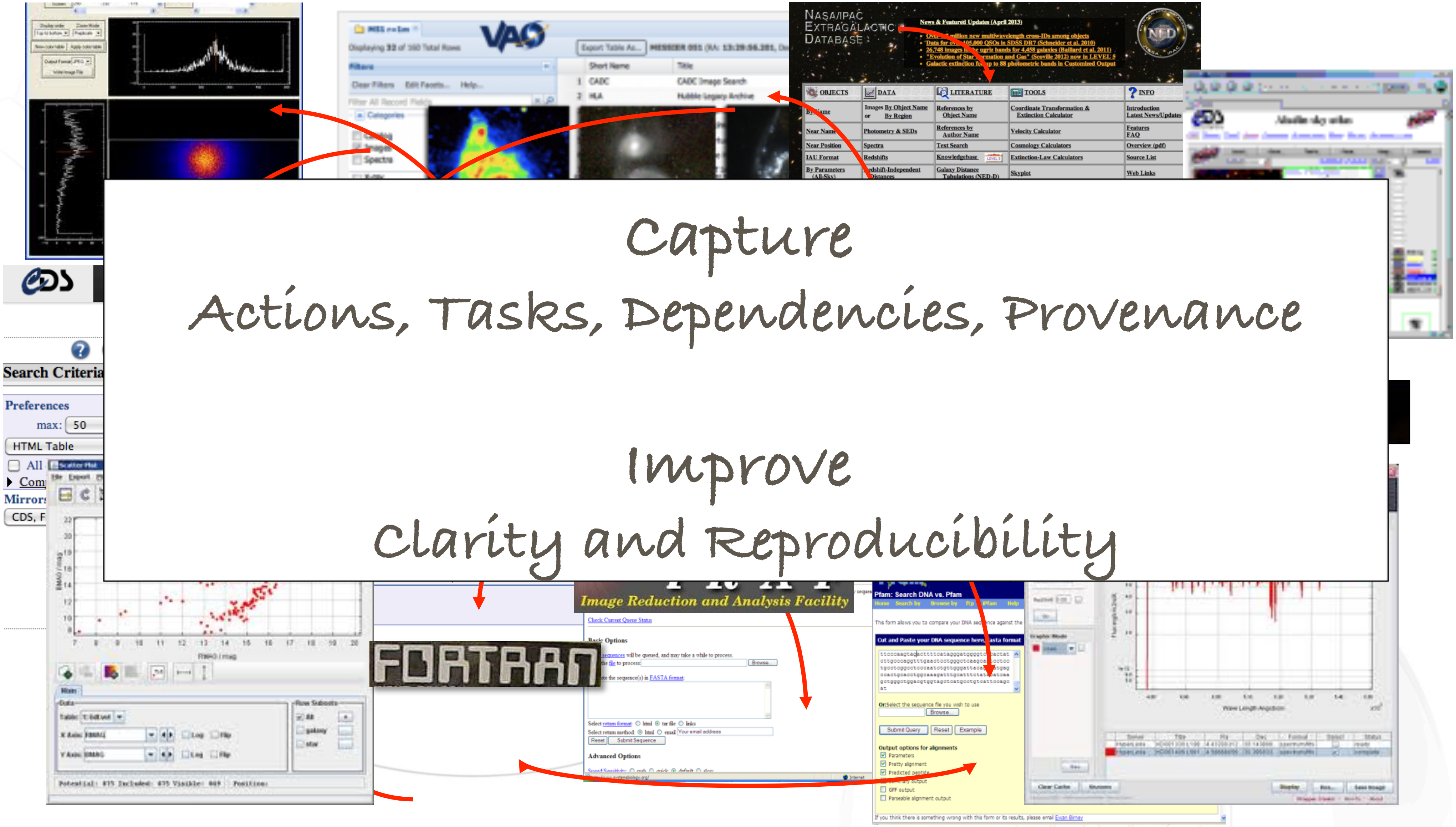


data_2010.05.28_re-test.dat	4:29 PM	5/28/2010	421 KB	DAT file
data_2010.05.28_re-re-test.dat	5:43 PM	5/28/2010	420 KB	DAT file
data_2010.05.28_calibrate.dat	7:17 PM	5/28/2010	1,256 KB	DAT file
data_2010.05.28_huh??.dat	7:20 PM	5/28/2010	30 KB	DAT file
data_2010.05.28_WTF.dat	9:58 PM	5/28/2010	30 KB	DAT file
data_2010.05.29_aaarrrgh.dat	12:37 AM	5/29/2010	30 KB	DAT file
data_2010.05.29_#$@*&!!.dat	2:40 AM	5/29/2010	0 KB	DAT file
data_2010.05.29_crap.dat	3:22 AM	5/29/2010	437 KB	DAT file
data_2010.05.29_notbad.dat	4:16 AM	5/29/2010	670 KB	DAT file
data_2010.05.29_woohoo!!.dat	4:47 AM	5/29/2010	1,349 KB	DAT file
data_2010.05.29_USETHISONE.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

Type: Ph.D Thesis Modified: too many times Copyright: Jorge Cham www.phdcomics.com

Digital Science - Towards the Executable Paper

Digital Astronomy in the Local Desktop


NASA/IPAC EXTRAGALACTIC DATABASE
OBJECTS
DATA
LITERATURE
TOOLS
INFO
VAO
Search Criteria
Preferences
max: 50
HTML Table
Mirrors
FORTRAN
Image Reduction and Analysis Facility
Pfam: Search DNA vs. Pfam

Capture

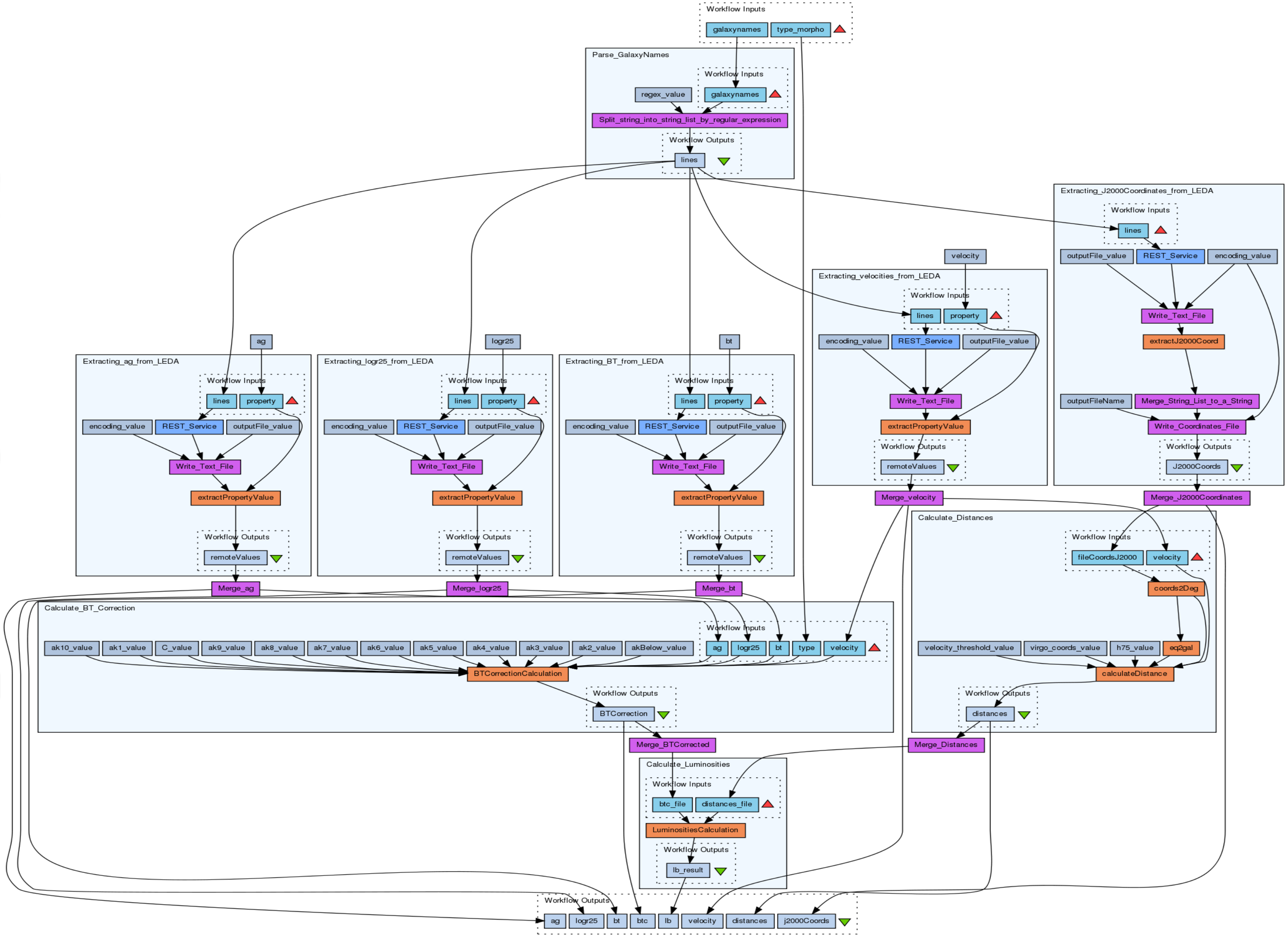
Actions, Tasks, Dependencies, Provenance

Improve

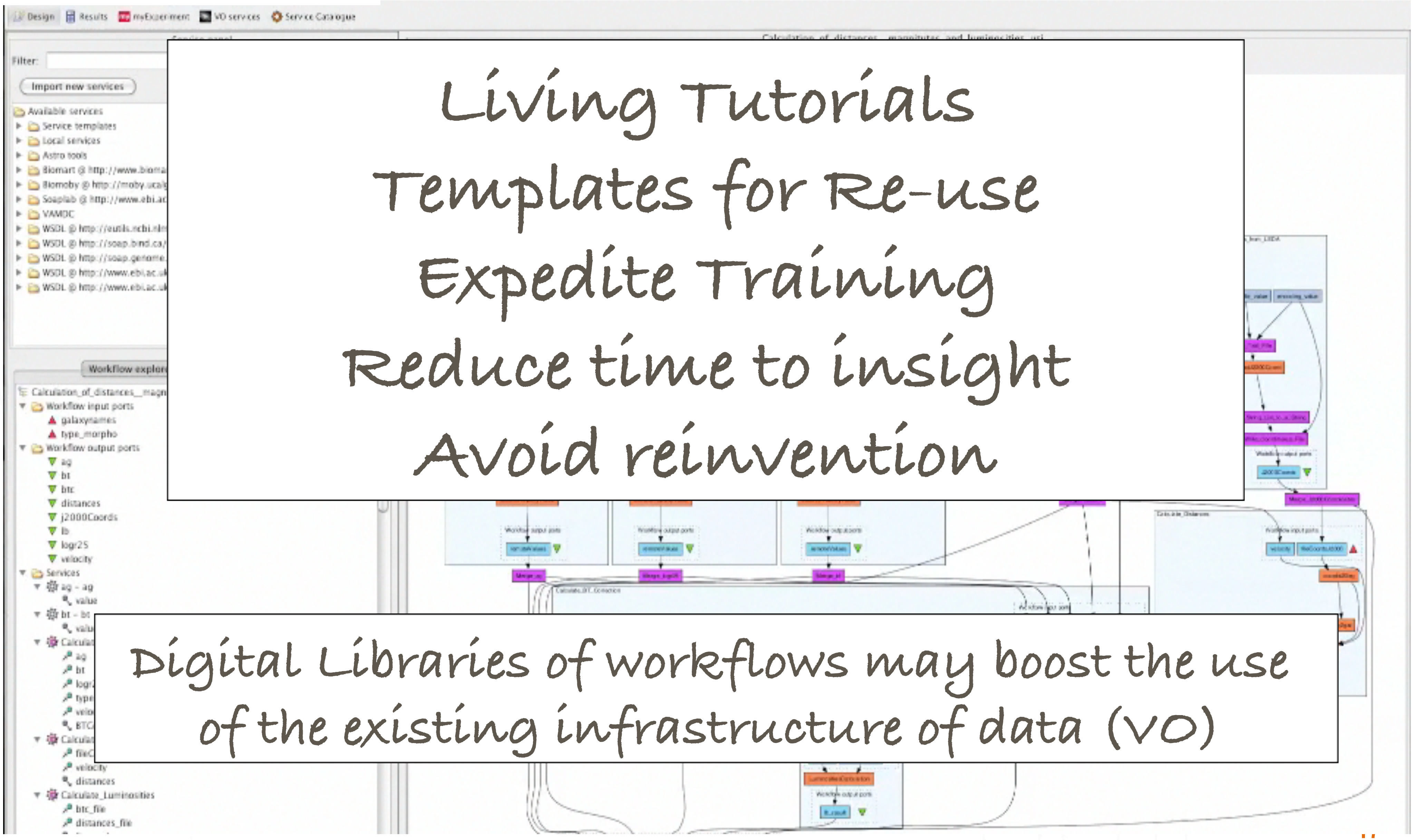
Clarity and Reproducibility

Digital Science - Towards the Executable Paper

Scientific Workflows



Scientific Workflows



Living Tutorials

Templates for Re-use

Expedite Training

Reduce time to insight

Avoid reinvention

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

Scientific Workflows

Related Initiatives

- ER-Flow
- VAMDC
- HELIO
- Cyber-SKA
- IceCore
- Montage
- Astro-WISE
- AstroGrid

Software

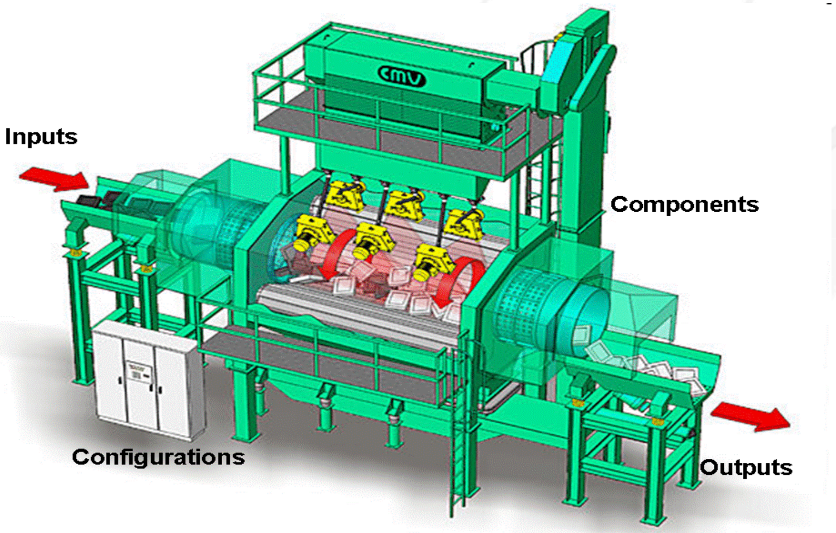
- Taverna
- Kepler
- Pegasus
- Triana
- ESO Reflex

IVOA

- AstroGrid
- Grid&WS WG
- VO France Wf WG

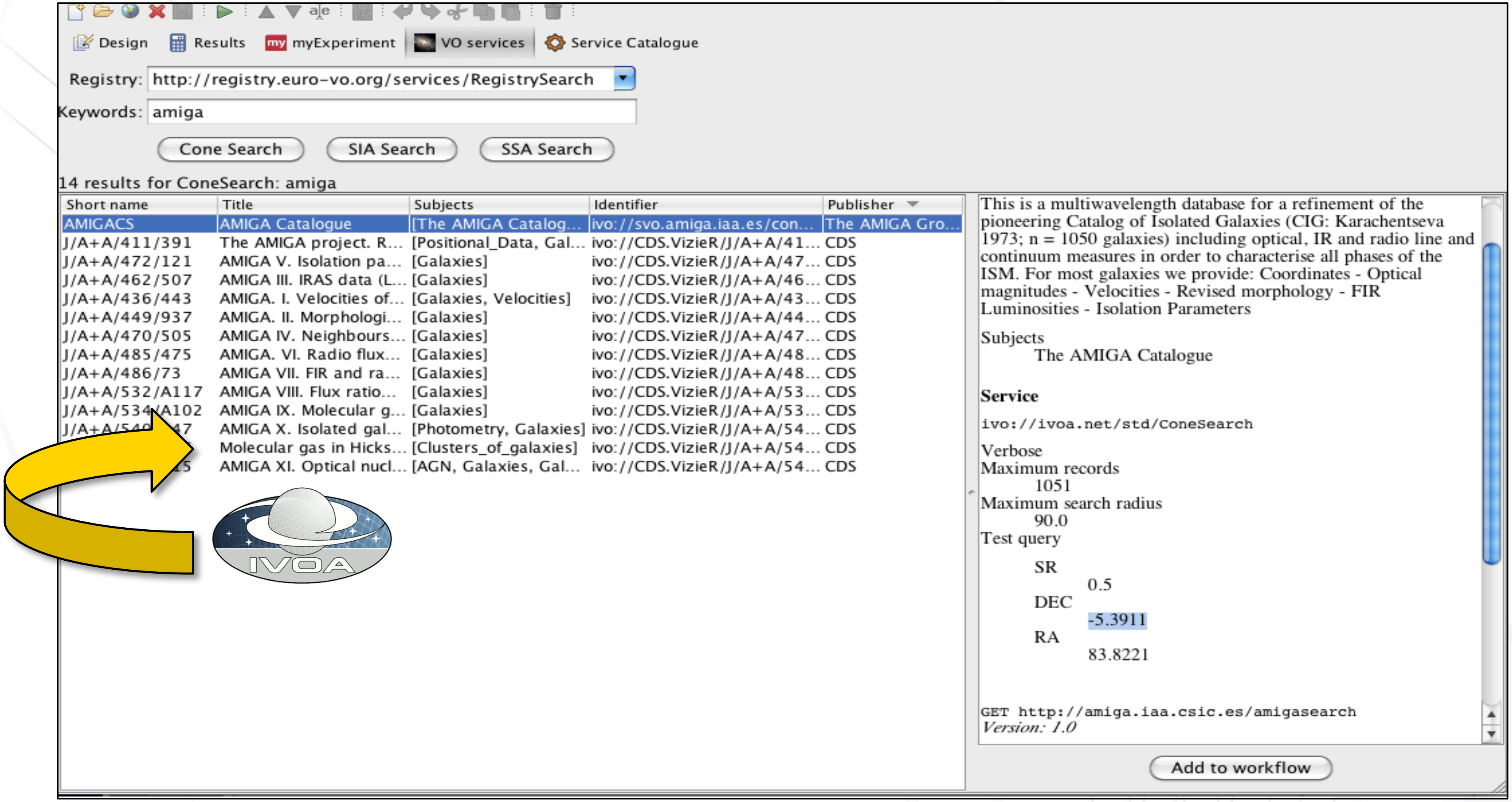
Self descriptive WS

- PDL
- SimDAL, S3



Interoperability Standards

AstroTaverna: Create, annotate and run a workflow



Registry: http://registry.euro-vo.org/services/RegistrySearch

Keywords: amiga

14 results for ConeSearch: amiga

Short name	Title	Subjects	Identifier	Publisher
AMIGACS	AMIGA Catalogue	[The AMIGA Catalog...	ivo://svo.amiga.iaa.es/con...	The AMIGA Gro...
J/A+A/411/391	The AMIGA project. R...	[Positional_Data, Gal...	ivo://CDS.VizieR/J/A+A/41...	CDS
J/A+A/472/121	AMIGA V. Isolation pa...	[Galaxies]	ivo://CDS.VizieR/J/A+A/47...	CDS
J/A+A/462/507	AMIGA III. IRAS data (L...	[Galaxies]	ivo://CDS.VizieR/J/A+A/46...	CDS
J/A+A/436/443	AMIGA. I. Velocities of...	[Galaxies, Velocities]	ivo://CDS.VizieR/J/A+A/43...	CDS
J/A+A/449/937	AMIGA. II. Morphologi...	[Galaxies]	ivo://CDS.VizieR/J/A+A/44...	CDS
J/A+A/470/505	AMIGA IV. Neighbours...	[Galaxies]	ivo://CDS.VizieR/J/A+A/47...	CDS
J/A+A/485/475	AMIGA. VI. Radio flux...	[Galaxies]	ivo://CDS.VizieR/J/A+A/48...	CDS
J/A+A/486/73	AMIGA VII. FIR and ra...	[Galaxies]	ivo://CDS.VizieR/J/A+A/48...	CDS
J/A+A/532/A117	AMIGA VIII. Flux ratio...	[Galaxies]	ivo://CDS.VizieR/J/A+A/53...	CDS
J/A+A/534/A102	AMIGA IX. Molecular g...	[Galaxies]	ivo://CDS.VizieR/J/A+A/53...	CDS
J/A+A/54[illegible]47	AMIGA X. Isolated gal...	[Photometry, Galaxies]	ivo://CDS.VizieR/J/A+A/54...	CDS
[illegible]	Molecular gas in Hicks...	[Clusters_of_galaxies]	ivo://CDS.VizieR/J/A+A/54...	CDS
[illegible]5	AMIGA XI. Optical nucl...	[AGN, Galaxies, Gal...	ivo://CDS.VizieR/J/A+A/54...	CDS

This is a multiwavelength database for a refinement of the pioneering Catalog of Isolated Galaxies (CIG: Karachentseva 1973; n = 1050 galaxies) including optical, IR and radio line and continuum measures in order to characterise all phases of the ISM. For most galaxies we provide: Coordinates - Optical magnitudes - Velocities - Revised morphology - FIR Luminosities - Isolation Parameters

Subjects
The AMIGA Catalogue

Service

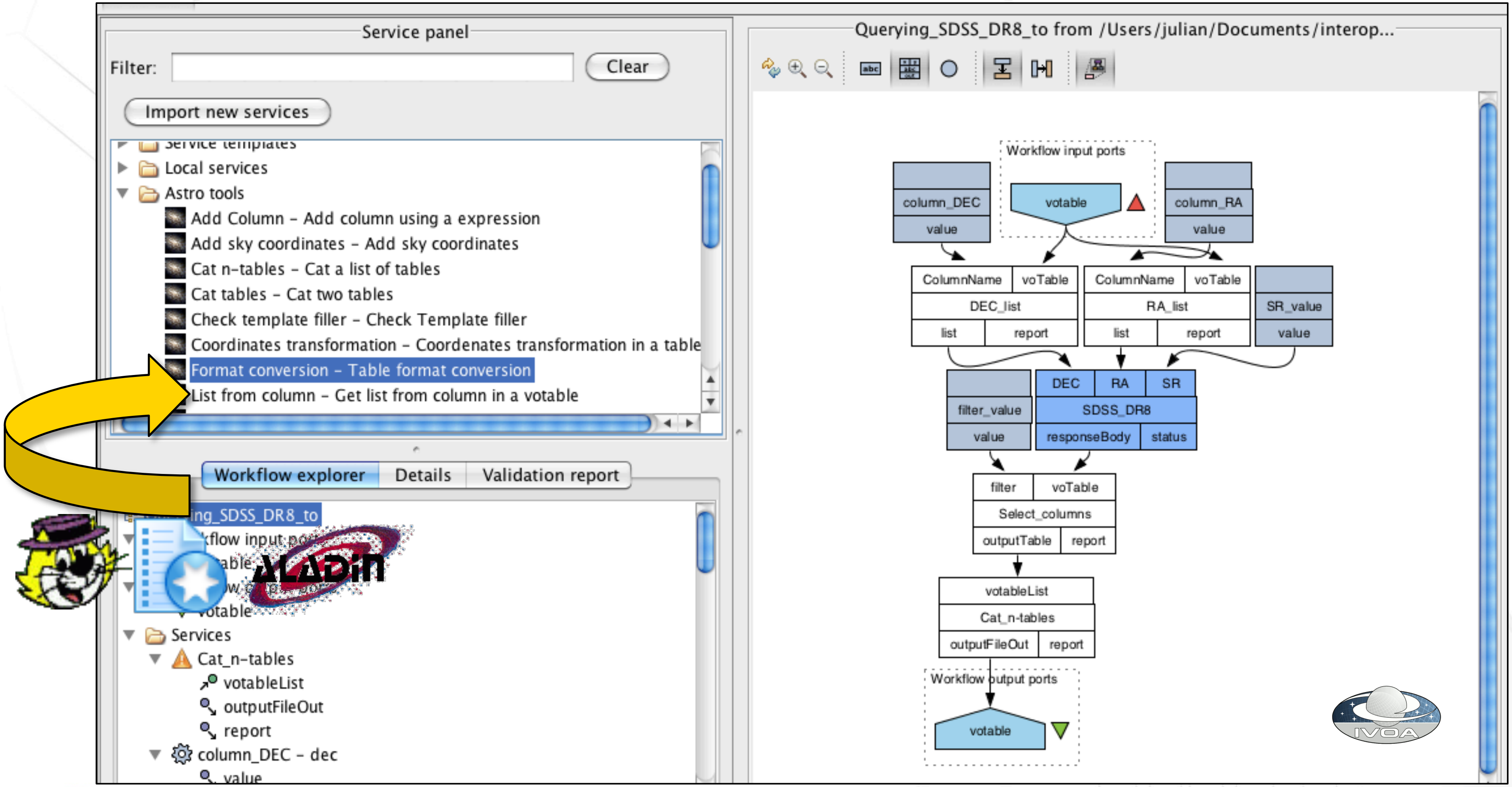
ivo://ivoa.net/std/ConeSearch

Verbose
Maximum records
1051
Maximum search radius
90.0
Test query
SR
0.5
DEC
-5.3911
RA
83.8221

GET http://amiga.iaa.csic.es/amigasearch
Version: 1.0

http://amiga.iaa.es/p/290-astrotaverna.htm

AstroTaverna: Create, annotate and run a workflow



Service panel

Filter: Clear

Import new services

- Service templates
- Local services
- Astro tools
  - Add Column – Add column using a expression
  - Add sky coordinates – Add sky coordinates
  - Cat n-tables – Cat a list of tables
  - Cat tables – Cat two tables
  - Check template filler – Check Template filler
  - Coordinates transformation – Coordenates transformation in a table
  - Format conversion – Table format conversion
  - List from column – Get list from column in a votable

Workflow explorer | Details | Validation report

- ...ing_SDSS_DR8_to
  - ...kflow input po...
    - ...able
  - ...w...
    - votable
  - Services
    - Cat_n-tables
      - votableList
      - outputFileOut
      - report
    - column_DEC – dec
      - value

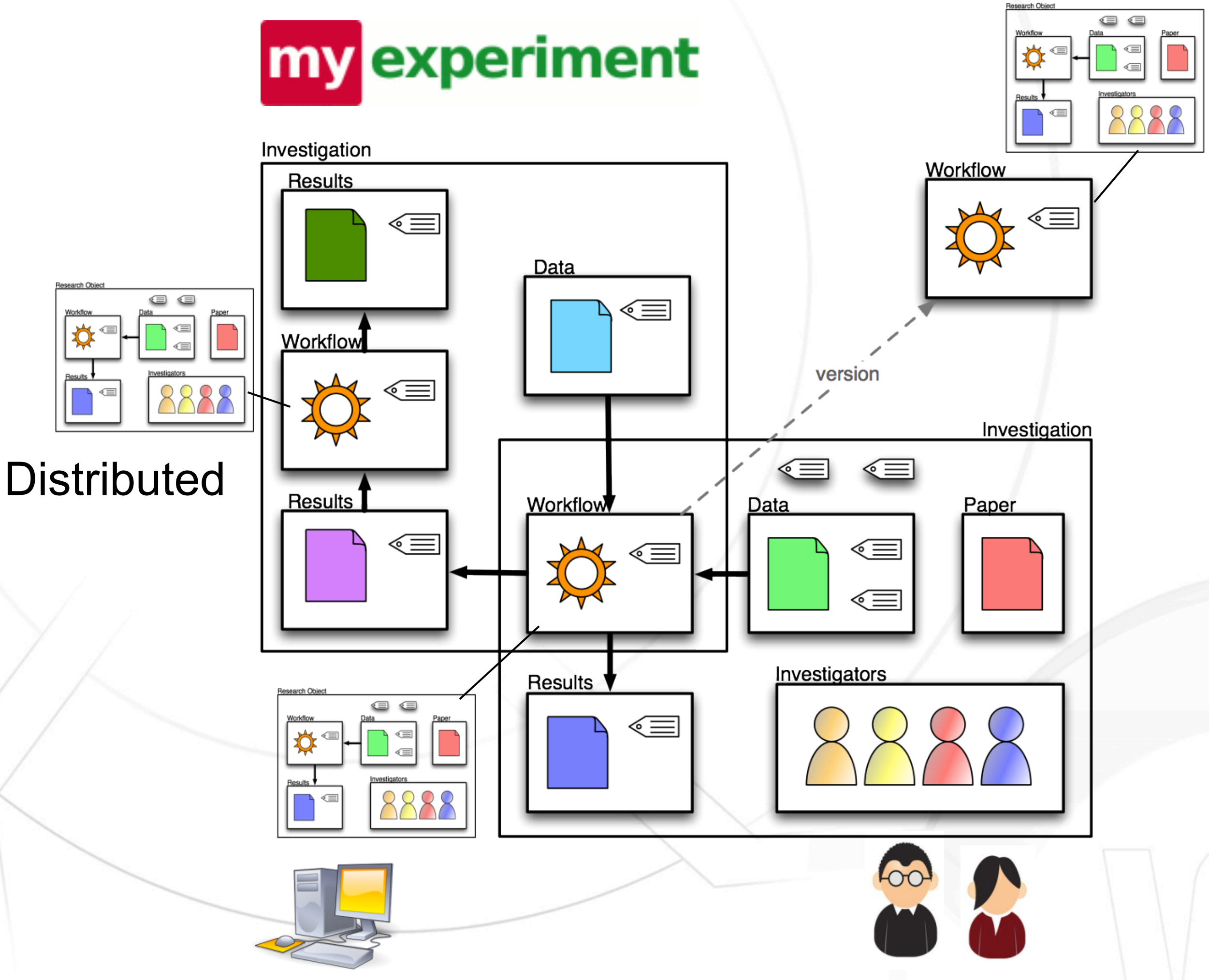
Querying_SDSS_DR8_to from /Users/julian/Documents/interop...

http://amiga.iaa.es/p/290-astrotaverna.htm

Digital Science - Towards the Executable Paper

Research Objects

Expose experimental context in a structured way in order to be understood



Distributed

Technical Objects

Social Objects

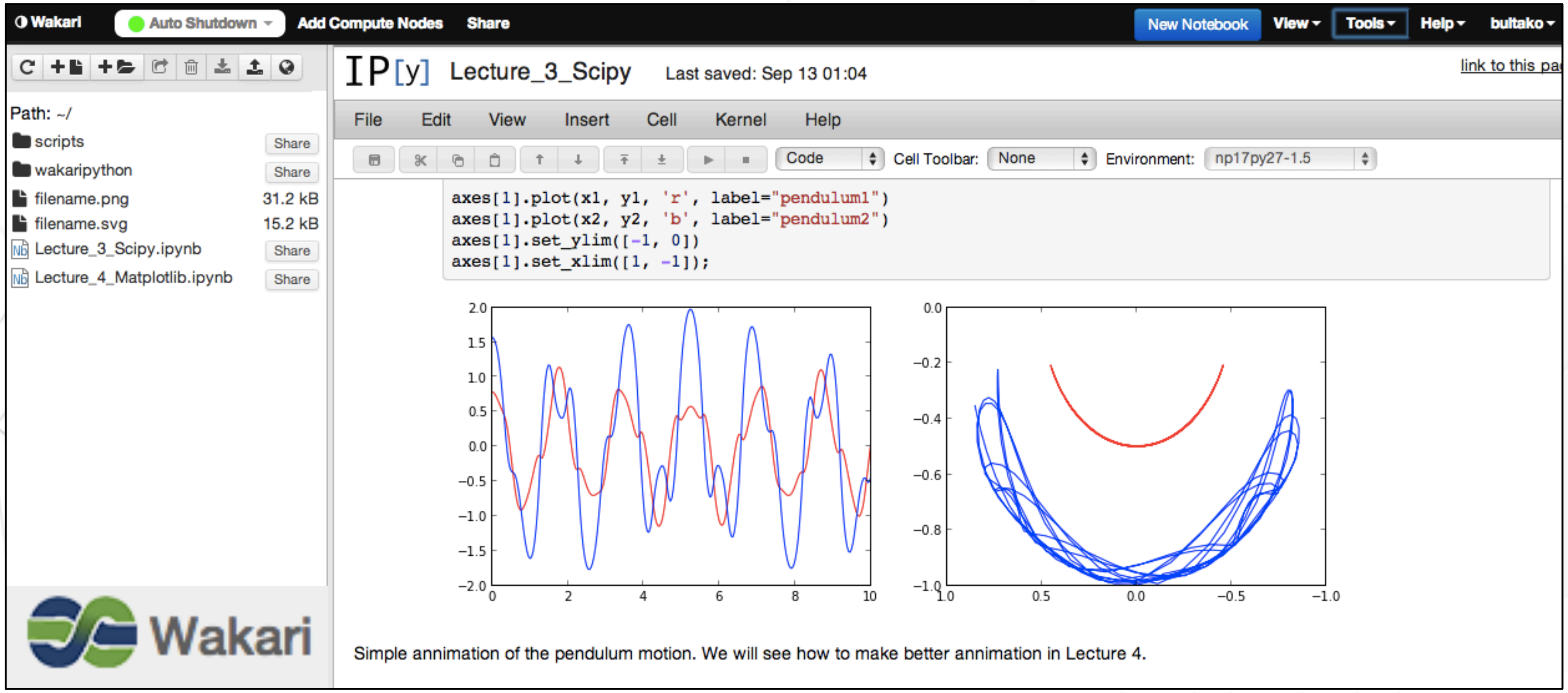
The IPython Notebook

Hands On !

Digital Science - Towards the Executable Paper
Research Objects

IPython Notebook solutions

- Web-browser as the working desktop
- Python code, plots and data, living with rich-text documentation
- Cloud-based adaptive scalable computing environment
- Fully shareable, re-usable and executable wikis
- Social platform and Git versioning

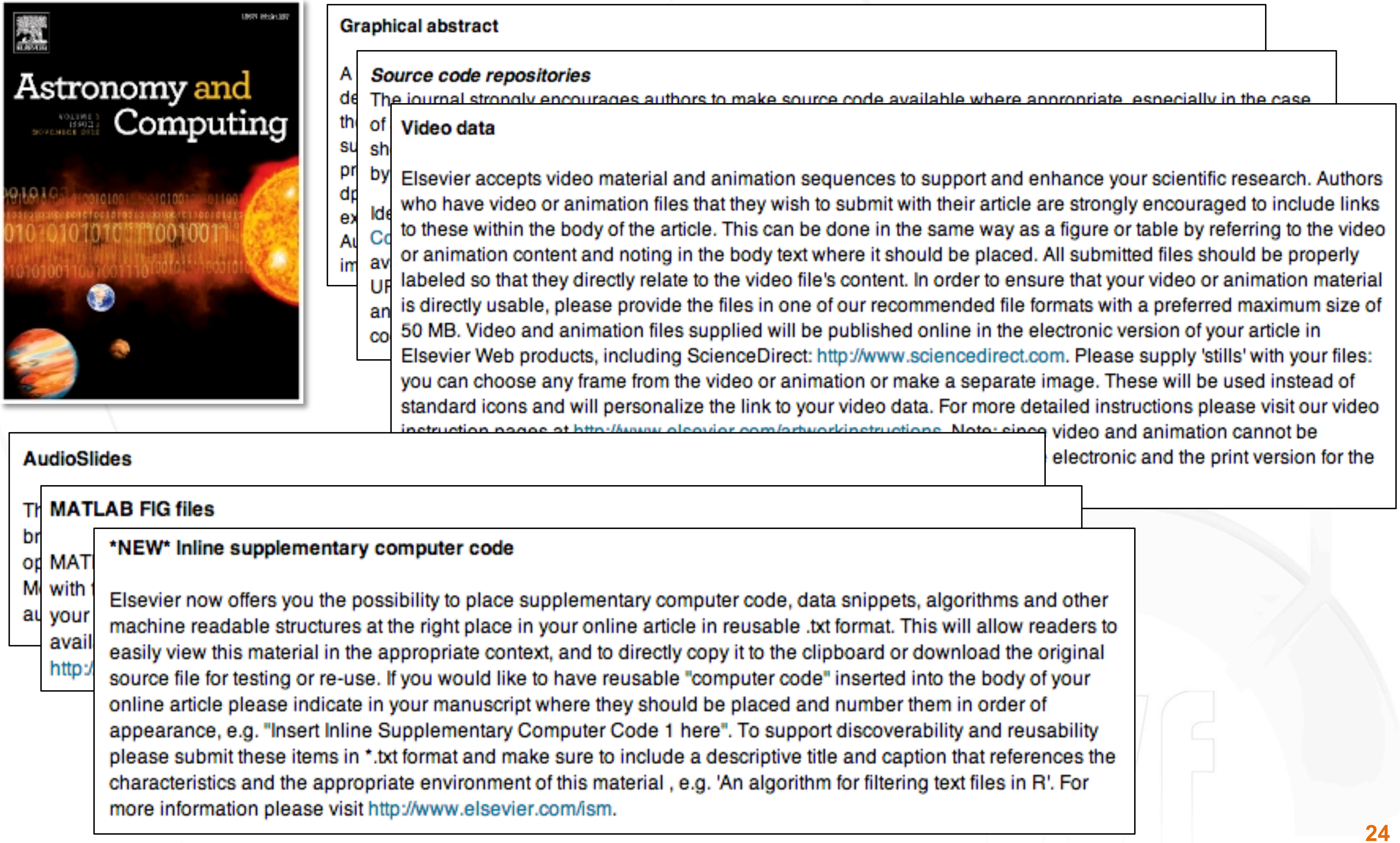


```
axes[1].plot(x1, y1, 'r', label="pendulum1")
axes[1].plot(x2, y2, 'b', label="pendulum2")
axes[1].set_ylim([-1, 0])
axes[1].set_xlim([1, -1]);
```

Simple annimation of the pendulum motion. We will see how to make better annimation in Lecture 4.

ELSEVIER article of the future

Digital Science - Towards the Executable Paper
Research Objects



Graphical abstract

A
de
th
su
pr
dp
ex
Au
im

Source code repositories

The journal strongly encourages authors to make source code available where appropriate, especially in the case
of
sh
by
Ide
Co
av
UF
an
co

Video data

Elsevier accepts video material and animation sequences to support and enhance your scientific research. Authors who have video or animation files that they wish to submit with their article are strongly encouraged to include links to these within the body of the article. This can be done in the same way as a figure or table by referring to the video or animation content and noting in the body text where it should be placed. All submitted files should be properly labeled so that they directly relate to the video file's content. In order to ensure that your video or animation material is directly usable, please provide the files in one of our recommended file formats with a preferred maximum size of 50 MB. Video and animation files supplied will be published online in the electronic version of your article in Elsevier Web products, including ScienceDirect: http://www.sciencedirect.com. Please supply 'stills' with your files: you can choose any frame from the video or animation or make a separate image. These will be used instead of standard icons and will personalize the link to your video data. For more detailed instructions please visit our video instruction pages at http://www.elsevier.com/artworkinstructions. Note: since video and animation cannot be
electronic and the print version for the

AudioSlides

Th
br
op
M
au

MATLAB FIG files

MAT
with
your
avail
http:/

***NEW* Inline supplementary computer code**

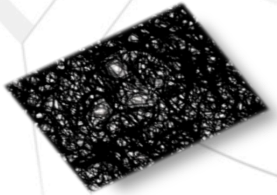
Elsevier now offers you the possibility to place supplementary computer code, data snippets, algorithms and other machine readable structures at the right place in your online article in reusable .txt format. This will allow readers to easily view this material in the appropriate context, and to directly copy it to the clipboard or download the original source file for testing or re-use. If you would like to have reusable "computer code" inserted into the body of your online article please indicate in your manuscript where they should be placed and number them in order of appearance, e.g. "Insert Inline Supplementary Computer Code 1 here". To support discoverability and reusability please submit these items in *.txt format and make sure to include a descriptive title and caption that references the characteristics and the appropriate environment of this material , e.g. 'An algorithm for filtering text files in R'. For more information please visit http://www.elsevier.com/ism.

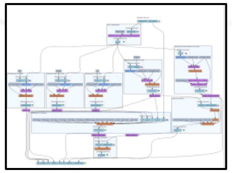
Digital Science - Towards the Executable Paper
Research Objects

ADSLabs

ADO Linked Components

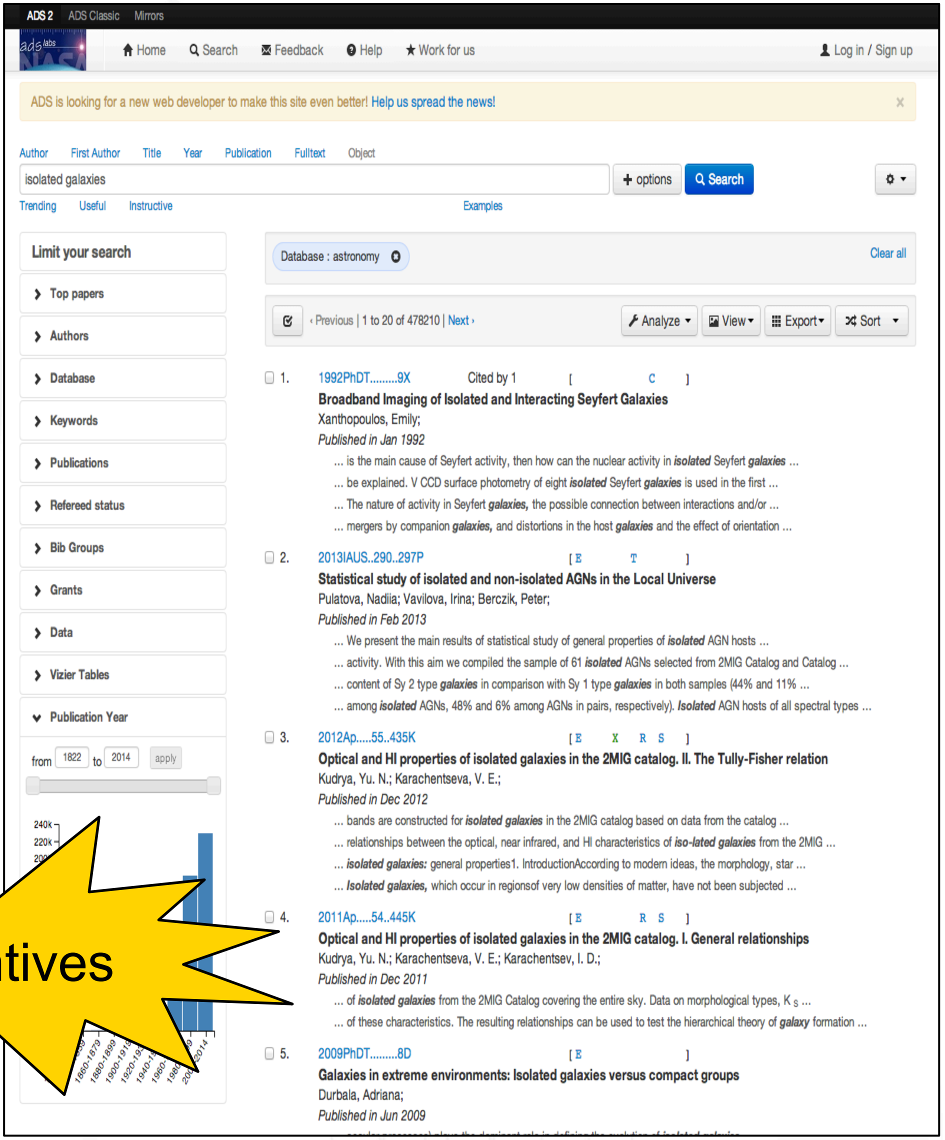
- » Authors
- » Publications
- » Journals
- » Objects SIMBAD
- » Tabular data behind the plots CDS
- » Observing time Proposals
- » Used facilities, surveys or missions
- » ASCL reference of used software





Incentives

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



ADS is looking for a new web developer to make this site even better! Help us spread the news!

isolated galaxies

Database : astronomy

‹ Previous | 1 to 20 of 478210 | Next ›

□ 1. 1992PhDT.........9X Cited by 1 [C]
Broadband Imaging of Isolated and Interacting Seyfert Galaxies
Xanthopoulos, Emily;
Published in Jan 1992
... is the main cause of Seyfert activity, then how can the nuclear activity in ***isolated*** Seyfert ***galaxies*** ...
... be explained. V CCD surface photometry of eight ***isolated*** Seyfert ***galaxies*** is used in the first ...
... The nature of activity in Seyfert ***galaxies,*** the possible connection between interactions and/or ...
... mergers by companion ***galaxies,*** and distortions in the host ***galaxies*** and the effect of orientation ...

□ 2. 2013IAUS..290..297P [E T]
Statistical study of isolated and non-isolated AGNs in the Local Universe
Pulatova, Nadiia; Vavilova, Irina; Berczik, Peter;
Published in Feb 2013
... We present the main results of statistical study of general properties of ***isolated*** AGN hosts ...
... activity. With this aim we compiled the sample of 61 ***isolated*** AGNs selected from 2MIG Catalog and Catalog ...
... content of Sy 2 type ***galaxies*** in comparison with Sy 1 type ***galaxies*** in both samples (44% and 11% ...
... among ***isolated*** AGNs, 48% and 6% among AGNs in pairs, respectively). ***Isolated*** AGN hosts of all spectral types ...

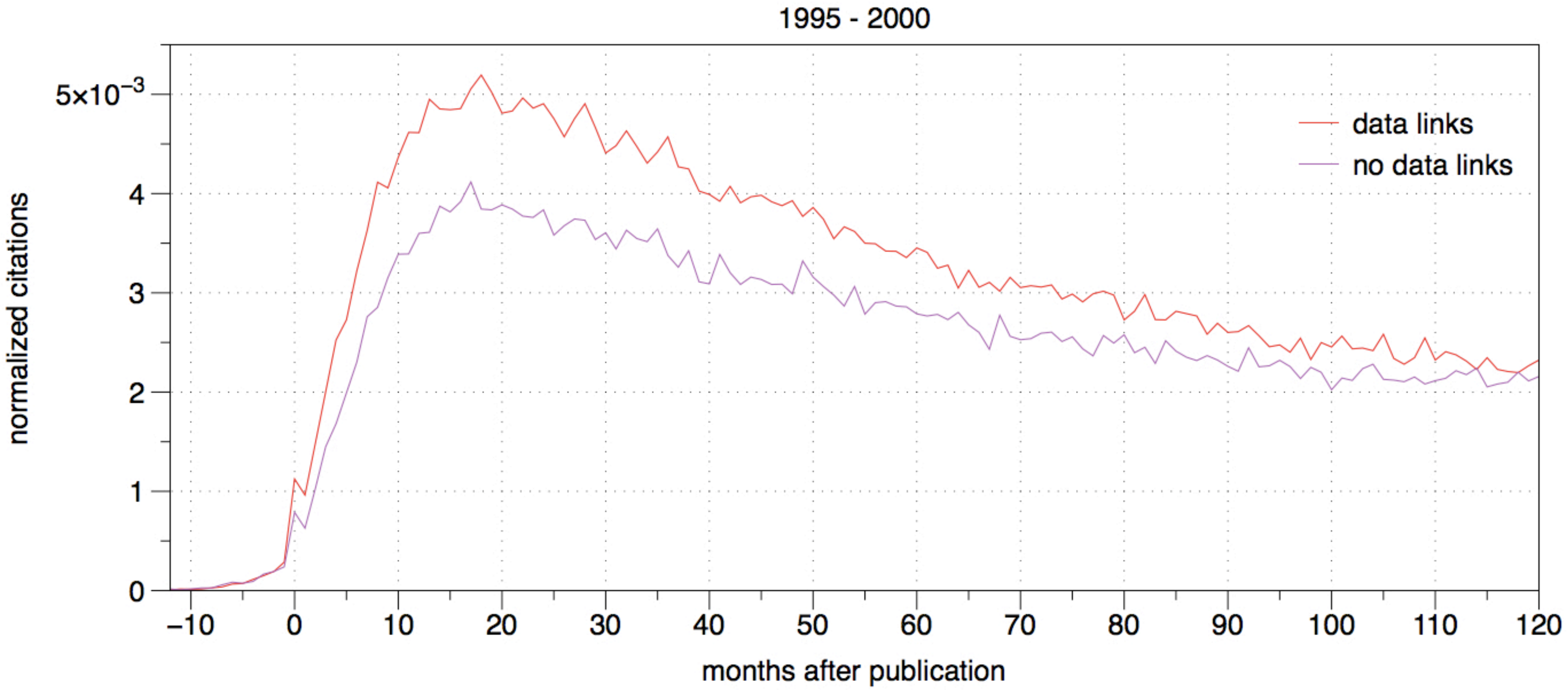
□ 3. 2012Ap.....55..435K [E X R S]
Optical and HI properties of isolated galaxies in the 2MIG catalog. II. The Tully-Fisher relation
Kudrya, Yu. N.; Karachentseva, V. E.;
Published in Dec 2012
... bands are constructed for ***isolated galaxies*** in the 2MIG catalog based on data from the catalog ...
... relationships between the optical, near infrared, and HI characteristics of ***iso-lated galaxies*** from the 2MIG ...
... ***isolated galaxies:*** general properties1. IntroductionAccording to modern ideas, the morphology, star ...
... ***Isolated galaxies,*** which occur in regionsof very low densities of matter, have not been subjected ...

□ 4. 2011Ap.....54..445K [E R S]
Optical and HI properties of isolated galaxies in the 2MIG catalog. I. General relationships
Kudrya, Yu. N.; Karachentseva, V. E.; Karachentsev, I. D.;
Published in Dec 2011
... of ***isolated galaxies*** from the 2MIG Catalog covering the entire sky. Data on morphological types, K_S ...
... of these characteristics. The resulting relationships can be used to test the hierarchical theory of ***galaxy*** formation ...

□ 5. 2009PhDT.........8D [E]
Galaxies in extreme environments: Isolated galaxies versus compact groups
Durbala, Adriana;
Published in Jun 2009

The Incentive

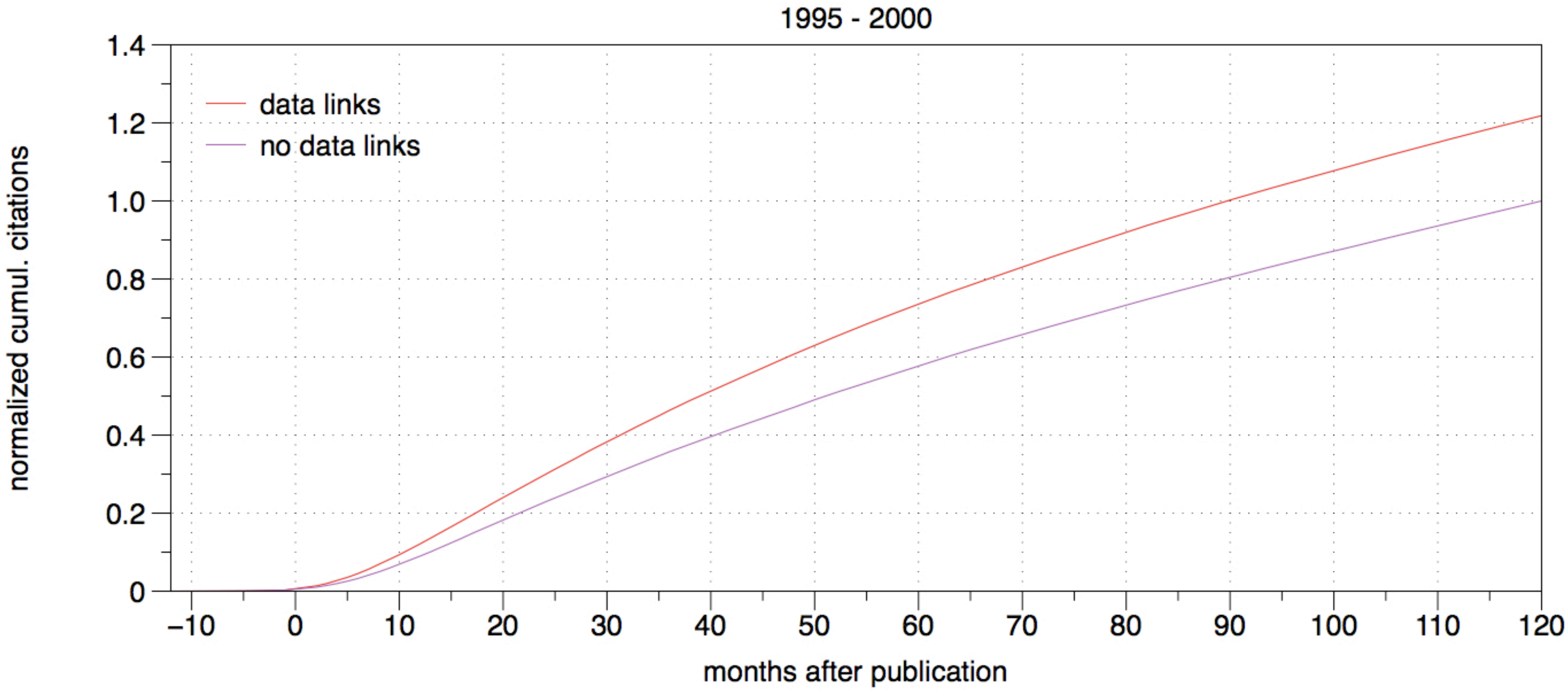
Papers with data links are cited more than those without



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

The 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.

Conclusions

- » **AMIGA** group invested in reproducible science projects
- » **Reproducibility** is at the very heart of the scientific method
- » Open data – access to all resources involved must be granted
- » **Re-use** needed in highly specialized science to achieve efficiency
- » Improving visibility is key in order to avoid reinvention
- » Social dimension of science stressed in the discovery process
- » Time has come to go **Beyond the PDF**
- » Capture provenance and structure in the local desktop
- » Scientific workflows go beyond automation – provide clarity and structure
- » **Research Object**
  - › Modular distributed aggregation of digital resources
  - › Executable, re-usable, documented, socially curated and inspected..
- » Other initiatives
  - › IPython notebooks-based solutions
  - › Elsevier Paper of the Future
  - › ADSLabs, ..

Conclusions

How NOT to be a good Astronomer in XXI Century

- In marketing just advertise your results – do not say how to reproduce them
- 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
- Do not provide data results – including the plots is just fine
- Practise the "data mine-ing" – input data and/or results are mine
- Practise the "data flirting" – please call me, if you want to know more
- Always cite the same authors and papers or those that cite you
- Do not reference other resources than published papers – never provide URL links
- Do not search info on Internet with other tools than ADS or arXiv
- Do not contact others if you re-use – duplicate and reinvent for your own

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

http://www.wf4ever-project.org

http://canu.be

jer@iaa.es