The Role of Repositories and Journals in the Astronomy Research Lifecycle

Alberto Accomazzi
NASA Astrophysics Data System
Smithsonian Astrophysical Observatory
http://ads.harvard.edu

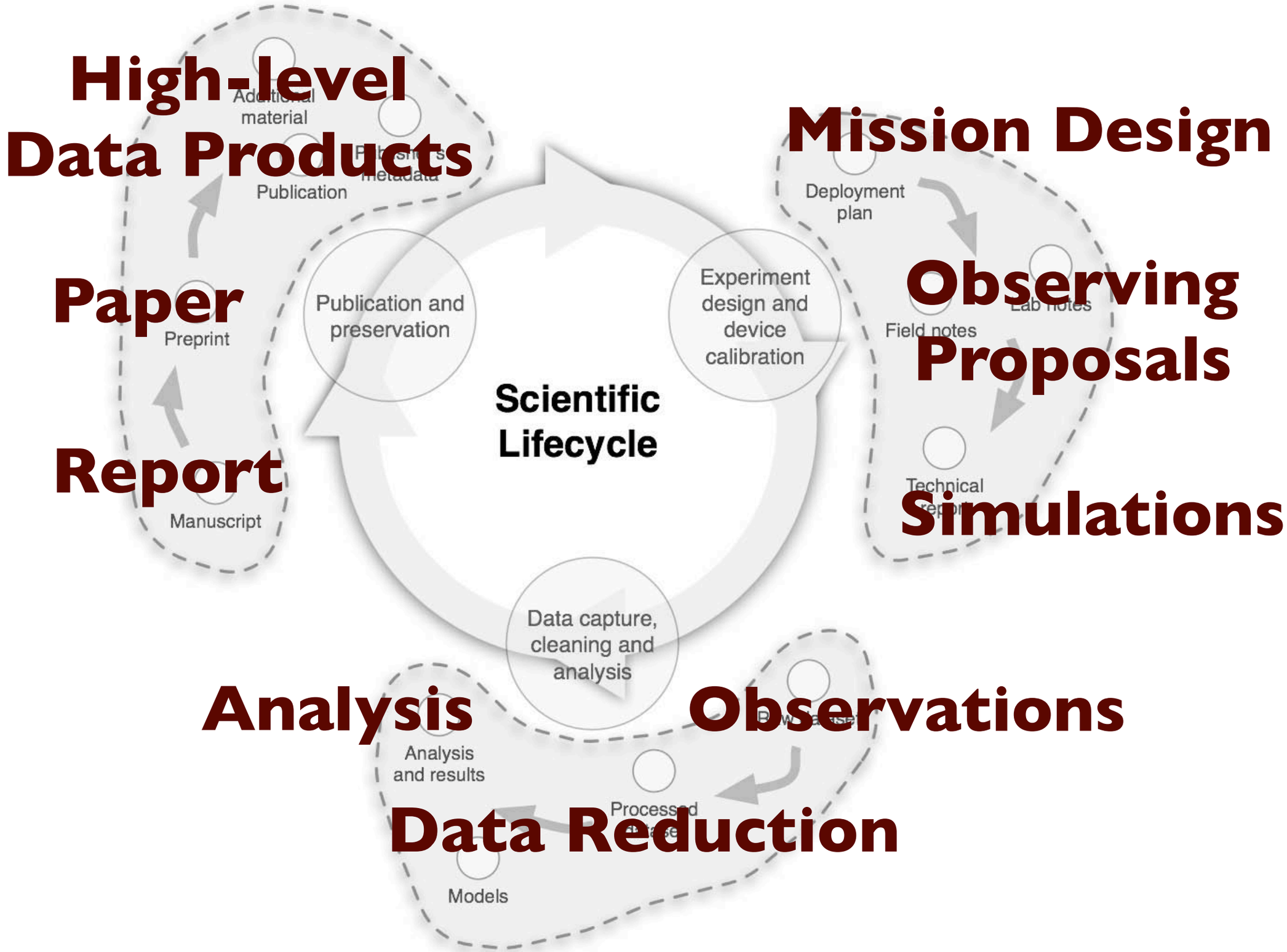
Astroinformatics 2010, Pasadena -- June 18 2010

Overview

- Preserving the Astronomy research lifecycle
- The changing electronic publishing landscape
- The roles of Users, Journals and Archives in Data Curation & Preservation
- Future opportunities for the VO

The Big Picture

- Scientific research requires *repeatability*
- The lifecycle of a research project should be documented by *capturing all artifacts*
- Data, Processes, Results need to be properly described, accessible, and *linked* together
- *Provenance* information should be attached to metadata throughout the process



Pepe et al, JASIST (2009)

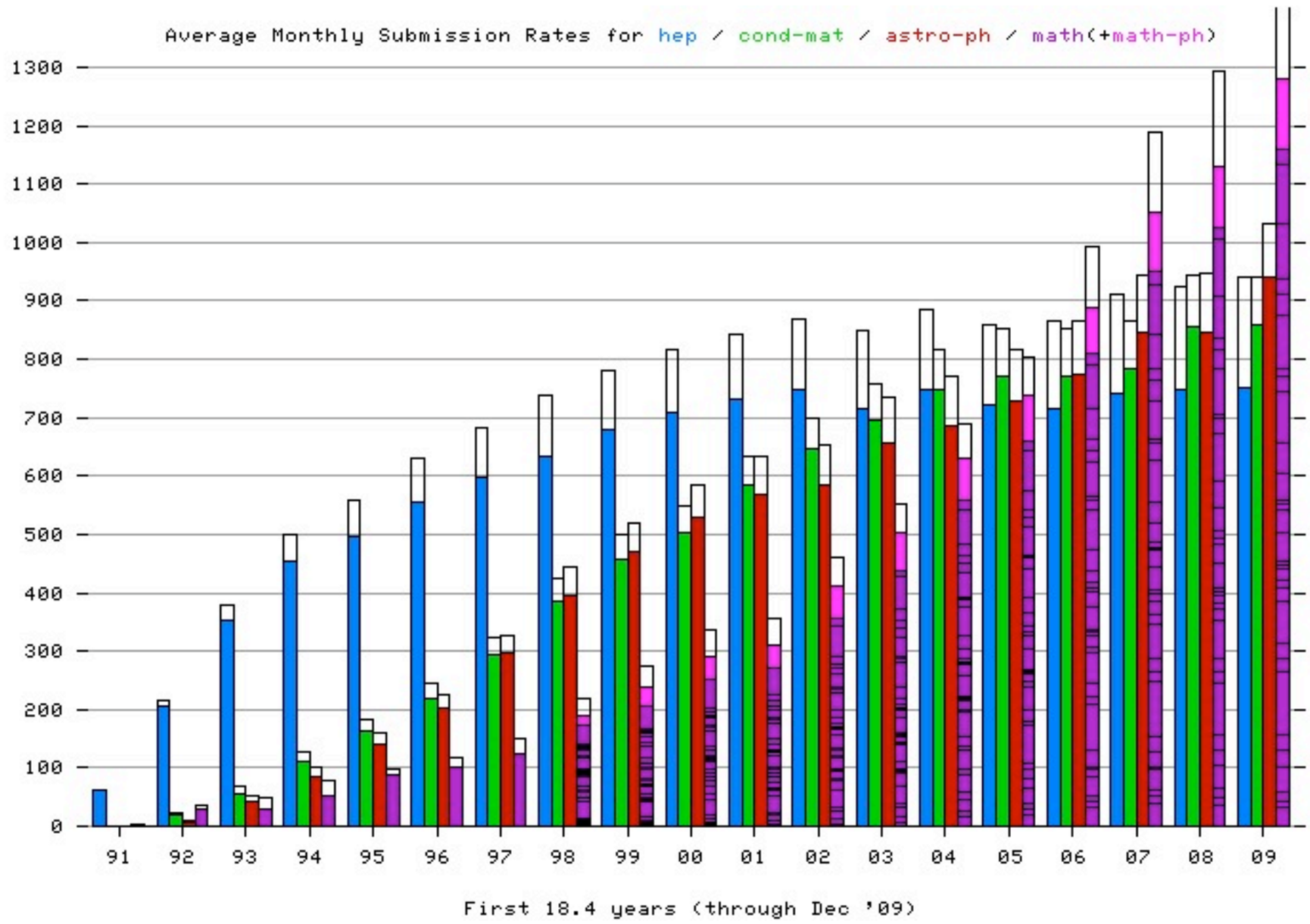
Astronomical Resources

- ADS, curator of bibliographic metadata, provider of fulltext for historical publications
- Journals from Publishers & learned societies, arXiv
- SIMBAD, NED & Vizier, curators of observational data and metadata
- NASA, NSF & European data Archives (MAST, Chandra, HEASARC, ESO)
- The VO provides the glue to tie observational resources together, provide tools for analysis

The e-print revolution

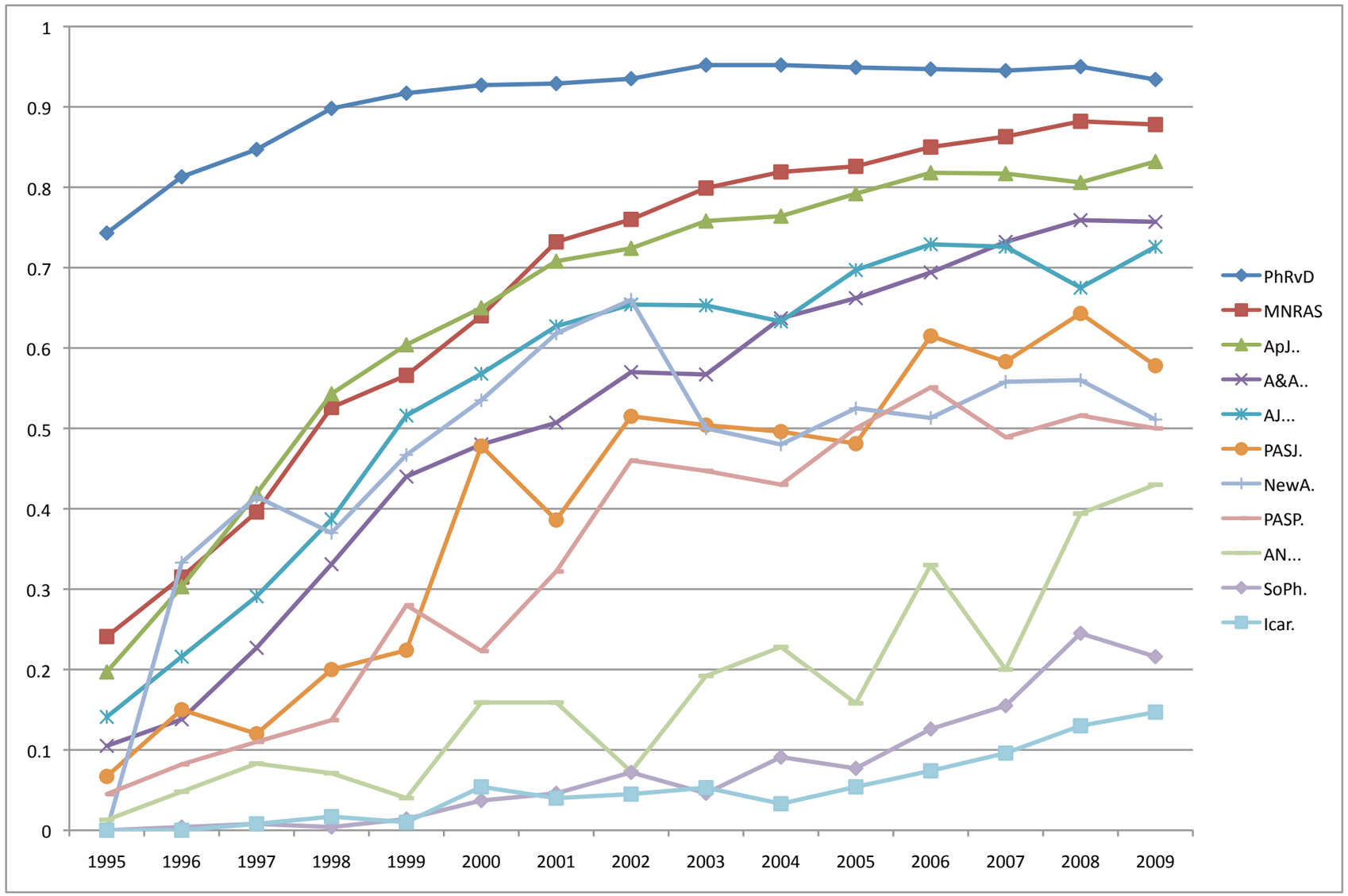
- arXiv has revolutionized the way we publish and read the literature
- Now primary destination for current research papers in astronomy
- Promoted and enabled Open Access in the physical sciences without a big fight
- A “Killer app” for physicists and astronomers

Community Adoption



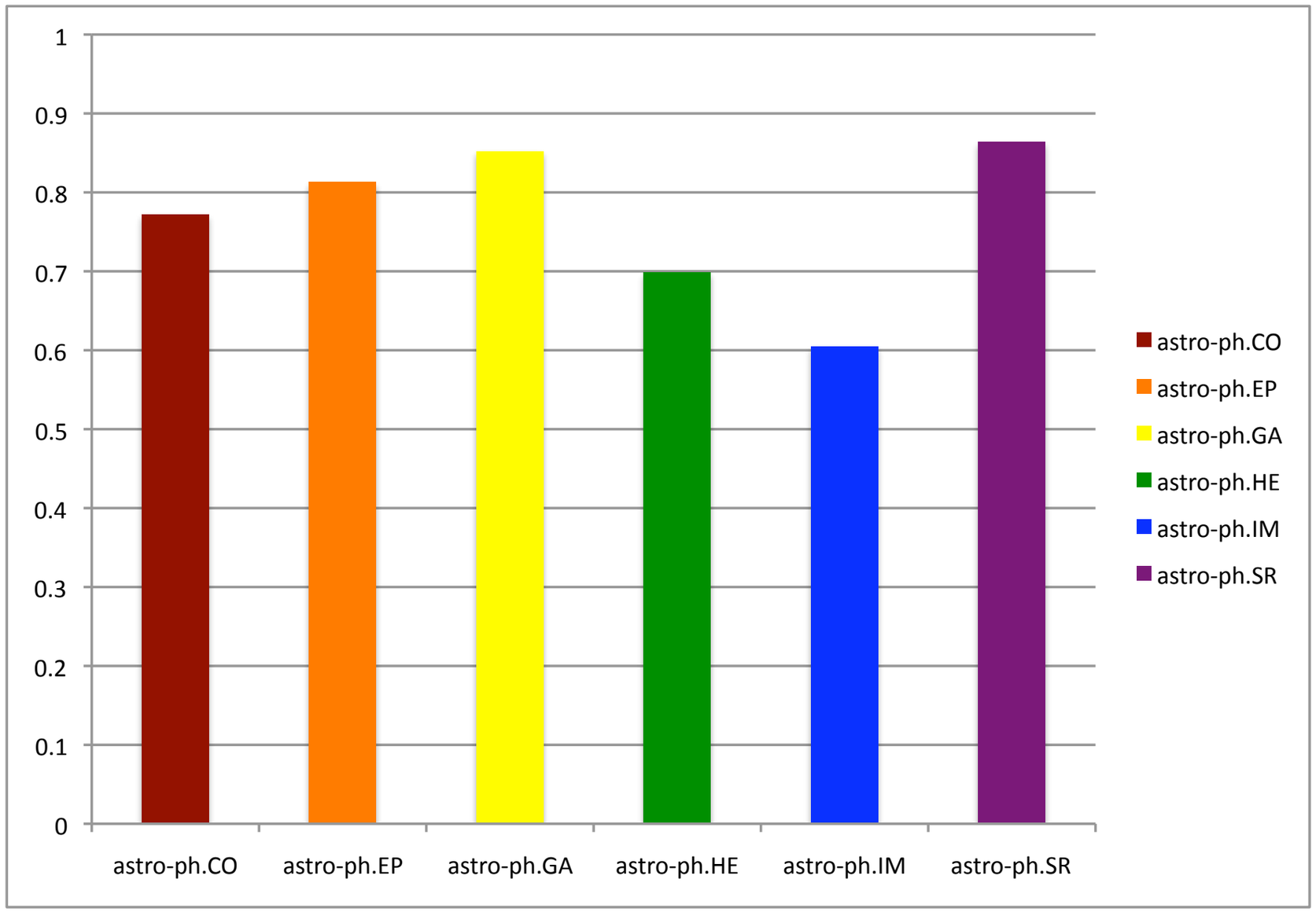
astro-ph submissions still growing. Credit: arxiv.org

Fraction of e-printed papers



fraction of e-prints matched to ADS articles. Credit: E. Henneken

More things are "published"



Fraction of arXiv eprints corresponding to published papers in ADS. Credit: E. Henneken

What about the Journals?

- Adopted Delayed Open Access, may be further reduced by Government mandate
- Under pressure to keep costs low, publication times low, quality high
- What added value do they provide?
  - Peer Review, editorial improvements
  - Article enhancements

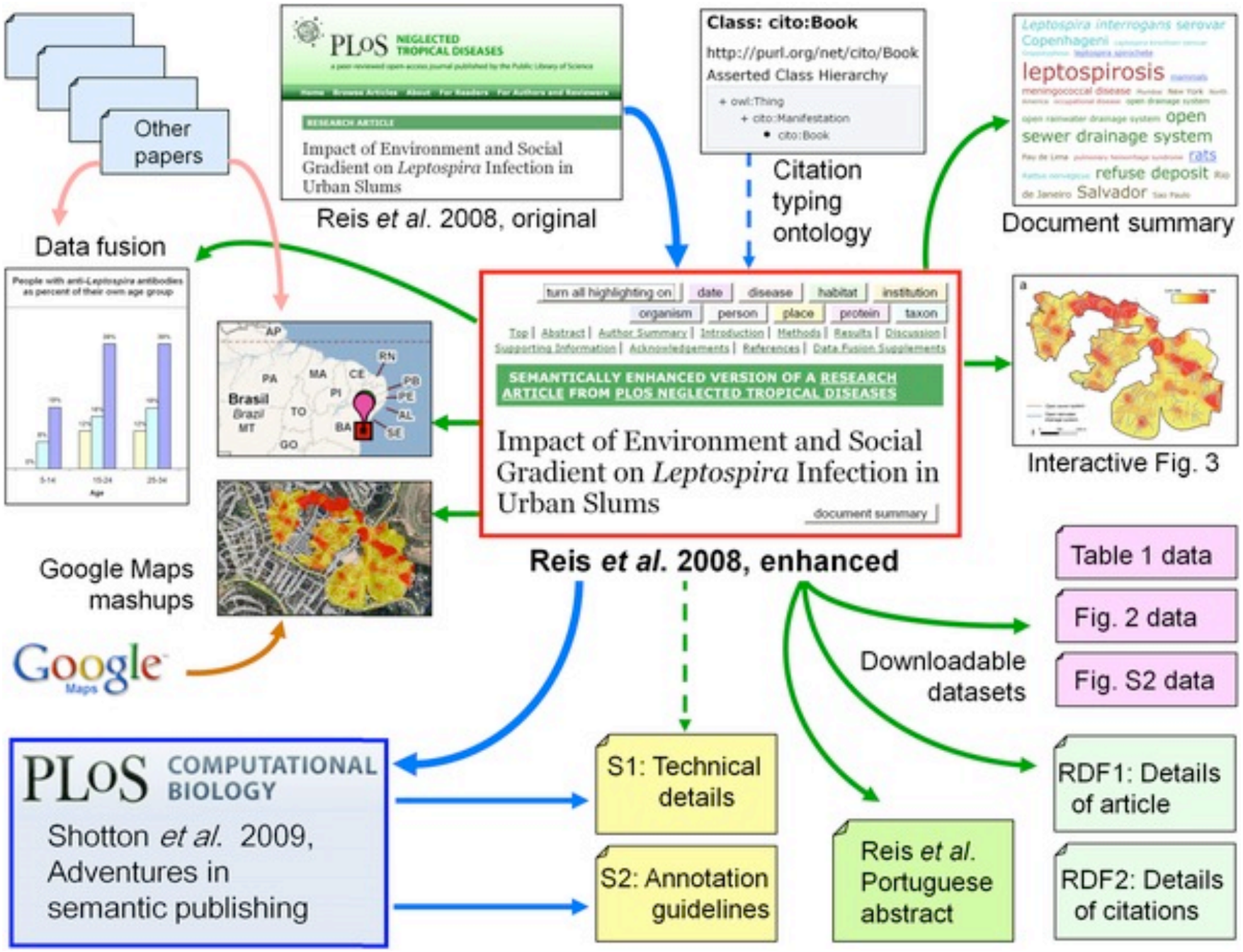
Peer Review, Editorial Process

- Peer Review increasingly being cited as THE reason why we still need the journals (PR as a service?)
- Supplementary material to on-line article reason to prefer publisher article over e-print
- HTML article provides enhancements, but HTML version not as popular as PDF!

The “article of the future”

- Semantic markup of relevant entities (RDF/A)
- Links to repositories, support for annotations
- Interactive plots, linked images, access to data behind them
- Mashups, embedded multimedia content
- Experiments by Elsevier, PLoS (Shotton et al.)

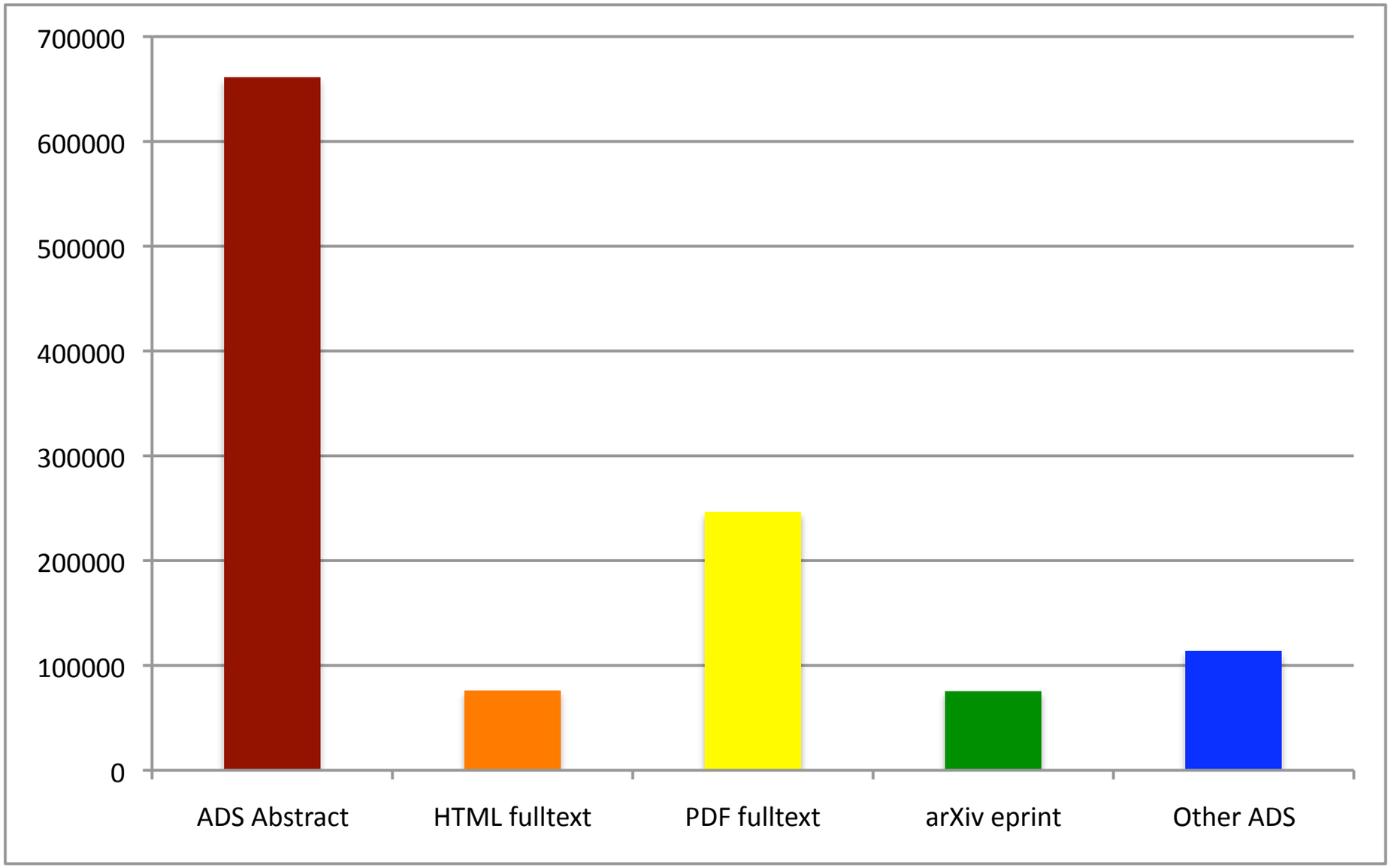
Shotton D, Portwin K, Klyne G, Miles A (2009)
"Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article"
PLoS Comput Biol 5(4): e1000361. doi:10.1371/journal.pcbi.1000361



If we build it will they come?

- What amount of effort is required for the "article of the future"?
- Who should do the work? What are the rewards? What are the gains?
- Is the publisher's website the proper place for this rich content?
- How do we bring the future to the past? Or: what about the rest of the literature?

What users do



ADS user clicks for 2007-2010 ApJ papers from Sep 2009 - Jan 2010. Courtesy: M. Kurtz

In other words...

- Users show a preference for *reading* (PDF) rather than *interacting* with article on publisher's websites by a factor of 3:1
- When searching, browsing, sorting, selecting, they use a few *well-established, trusted archives* rather than disparate journal websites
- When given a choice, users prefer the *copy of record* rather than an arXiv e-print

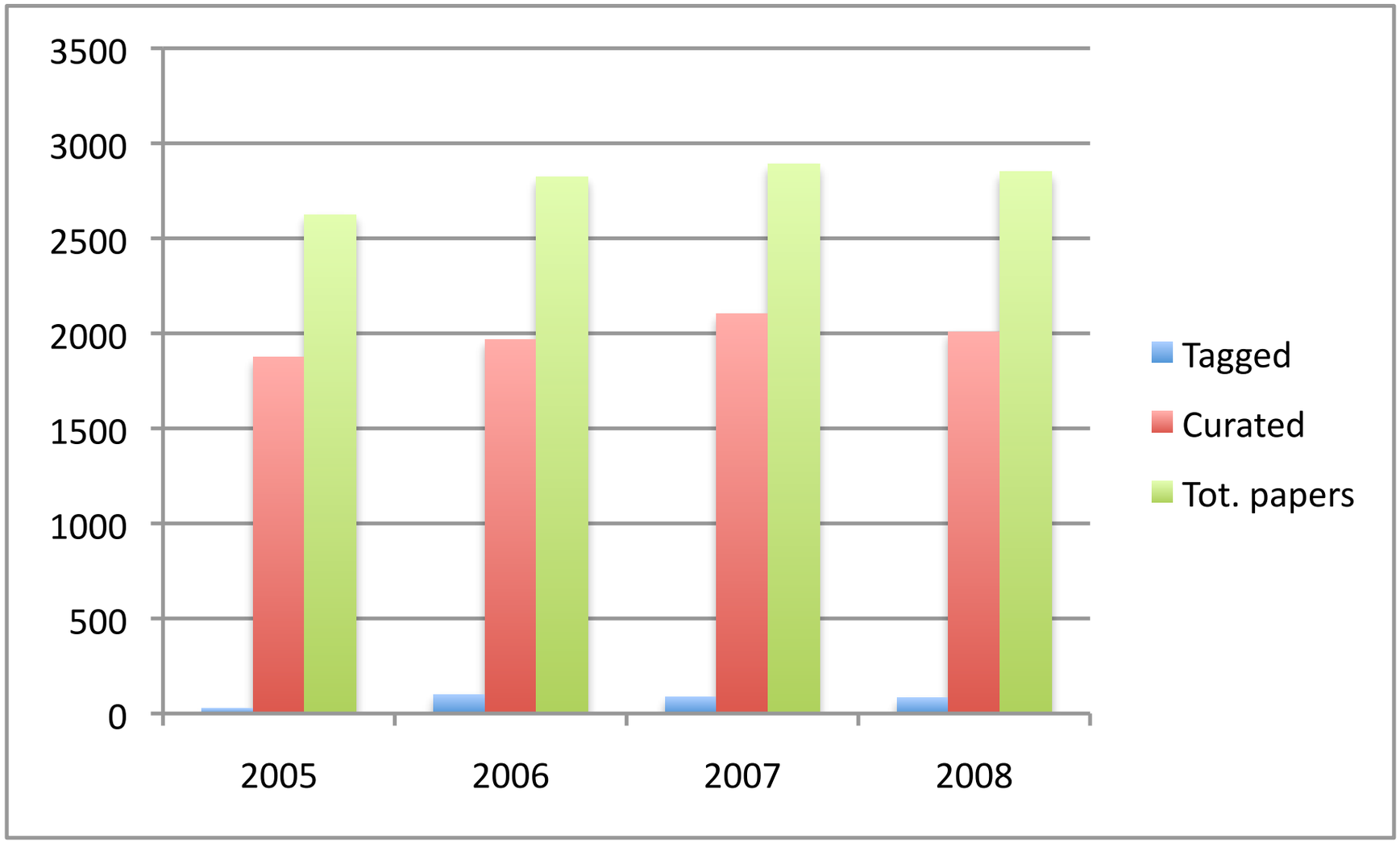
AAS Journals Enhancements

- Collaboration between ADS, NED, CDS, NASA ADEC, AAS journals
- Since 2004, AASTeX 5.2 supports additional markups in papers by authors:

  \object{}
  \dataset{}
  \facility{}

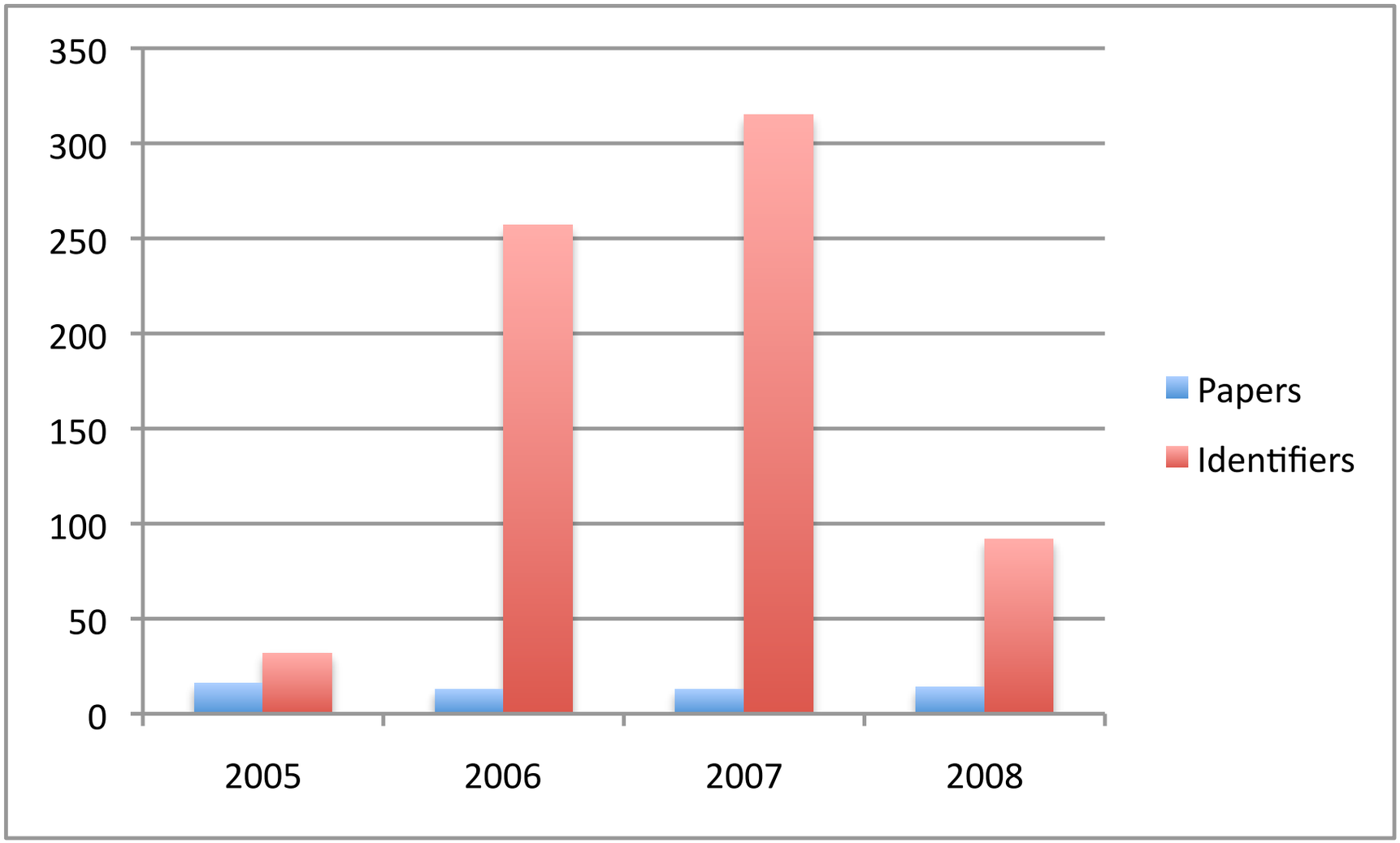
- What has been the rate of adoption by users and editors?

Links to objects



Number of 2005-2009 ApJ papers with links to objects

Links to datasets



Number of 2005-2009 ApJ papers with links to dataset identifiers

Lessons learned

- Limited adoption of user markup despite support from data centers, societies
- Lack of awareness from users, enforcement by editors, no reward system for scientists
- Multiple parties involved, multiple points of failure
- User submission *no substitute for curation!*

A Way Forward

- Recast linking in wider scope, the *semantic interlinking of digital assets in astronomy*: articles, objects, datasets
- Enable *upload of high-level data products* (plots, tables) to trusted, community-curated digital repository
- Provide infrastructure to enable minting of identifiers, *exposing metadata*, persistent link resolution
- Capture identifiers and their inter-relationships *during and after* the peer review process, *publish all metadata in machine-readable format*

What metadata integration can do

- Create a *semantic knowledge base*, allowing integration of heterogenous metadata, tracking of provenance, attribution
- Enable UIs integrating *views of bibliographies, objects, datasets*, other resources interacting together
- Provide a *recommender / alert service* for data products, objects based on bibliographic use/citations
- Support the *retrofitting of semantics* over the existing body of curated knowledge, papers

Astronomy article of the future?

Paper Objects Data Figures Tables
Provenance Annotations

TAOS

020.00141
121.00043
051.00014

SIMBAD

V* gam Dor
HD 64813
HR 418
M 37
V* GM Leo
V* del Sct
V* HN Dra
V* bet Cep
V* SX Phe

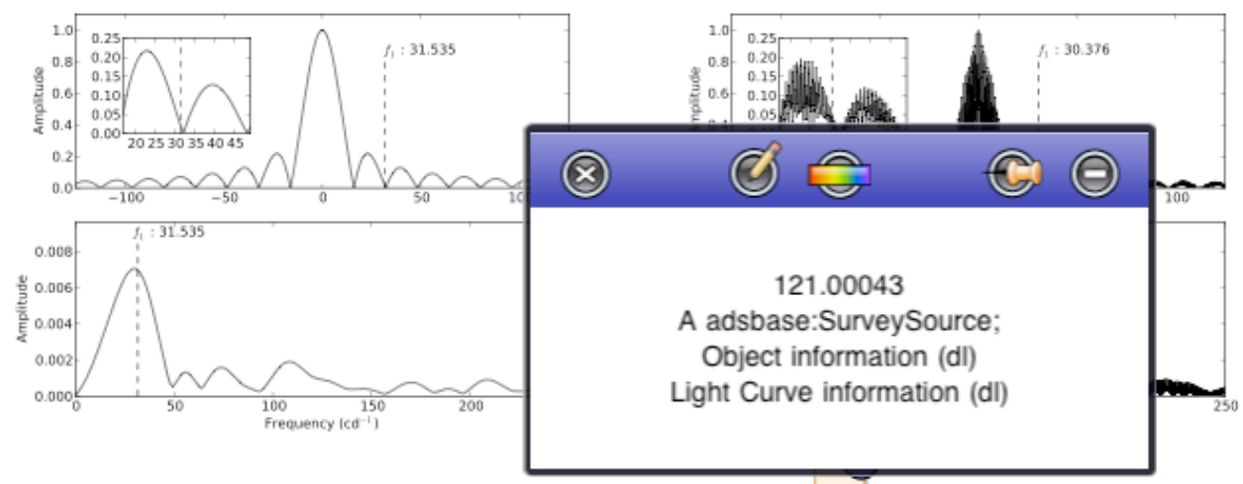
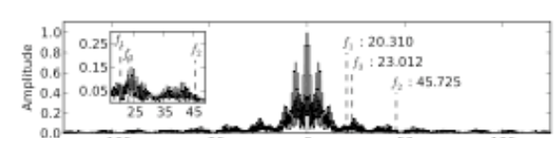


Fig. 4.— The spectral window and the power spectrum of the star ID 020.00141. The top panel shows the spectral window and the bottom panel shows the power spectrum of the star. The dashed line shows the detected frequency. In the top panel, we magnified the spectral window to clearly show the detected frequency.

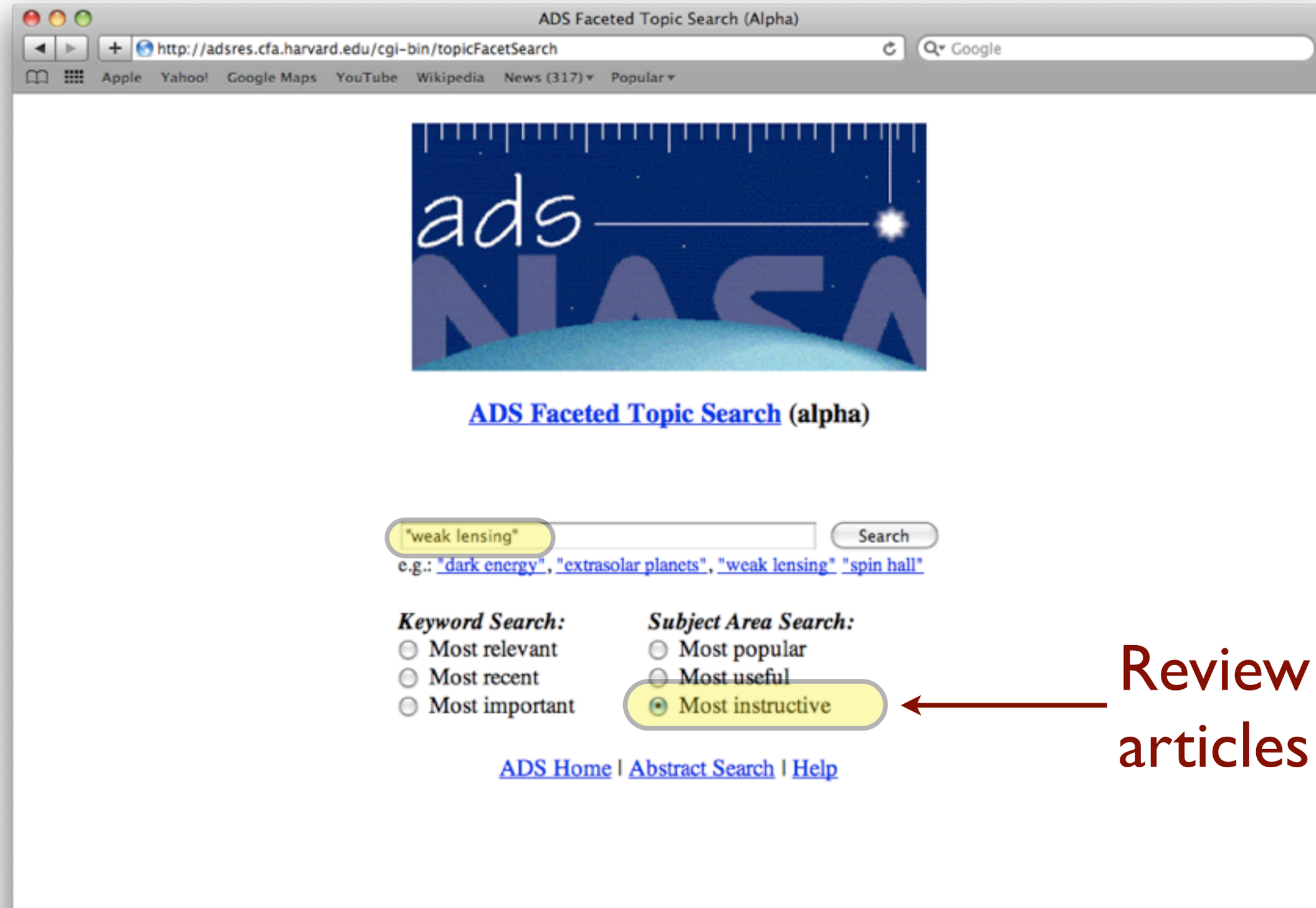
Fig. 5.— The spectral window and the power spectrum of the star ID 121.00043. The top panel shows the spectral window and the bottom panel shows the power spectrum of the star. The dashed line shows the detected frequency. In the top panel, we magnified the spectral window to clearly show the detected frequency.

& Garrison 1987). B type stars do not show such strong Ca II K line. The spectrum also shows weak metallic lines (e.g. Ca I and Mg II line) which are usually presented in A type stars. Based on the strength of Ca II K line, hydrogen lines and metallic lines, the star is likely



A mockup of a semantically annotated ApJ article. Credit: R. Dave, MSR

Example: object views on papers



Review articles

ADS Query Results

http://adsres.cfa.harvard.edu/cgi-bin/topicFacetSearch?q=%22weak%20lensing%22;qtype=INSTRU

Apple Yahoo! Google Maps YouTube Wikipedia News (284) Popular

SAO/NASA Astrophysics Data System (ADS)

objects in papers

Query Results from the ADS Database

Go to bottom of page

Related Objects

- NAME HDF (12)
- NAME Chandra Deep Field-South (8)
- ACO 1689 (8)
- ACO 901 (8)
- ZwCl 0024+1652 (7)
- ACO 902 (7)
- ACO 2390 (6)
- ZwCl 1455+2232 (5)
- NAME SGP (4)
- ClG 2137.3-2353 (4)
- ACO 2261 (4)
- ACO 2219 (4)
- ACO 2218 (4)
- ACO 1835 (4)
- ACO 1763 (4)
- ACO 370 (4)
- ACO 267 (4)
- ACO 209 (4)
- ACO 68 (4)
- 1E 0302.5+1717 (4)
- ZwCl 1358+6245 (3)
- NAME HDF-S (3)
- NAME COSMOS FIELD (3)
- NAME CBI (3)
- NAME ACO 901A (3)
- NAME ACO 901-902 SUPERCLUSTER (3)
- ClG 1224.7+2007 (3)
- ClG 1054.4-0321 (3)
- ACO 1942 (3)
- ACO 1364 (3)

Selected and retrieved **200** abstracts. Sort options

#	Bibcode / Authors	Score / Title	Date	List of Links / Access Control Help							
1	2006glsw.book..269S Schneider, P.	81.000 Weak Gravitational Lensing	n/a 2006	A	E		X	R	C	c	U
2	2008PhR...462...67M Munshi, Dipak; Valageas, Patrick; van Waerbeke, Ludovic; Heavens, Alan	68.000 Cosmology with weak lensing surveys	Jun 2008	A	E		X	R	C	c	U
3	2003ARA&A..41..645R Refregier, Alexandre	61.000 Weak Gravitational Lensing by Large-Scale Structure	n/a 2003	A	E	F	X	R	C	c	U H
4	2008ARNPS..58...99H Hoekstra, Henk; Jain, Bhuvnesh	51.000 Weak Gravitational Lensing and Its Cosmological Applications	Nov 2008	A			X	R	C	c	U
5	2003astro.ph..6465S Schneider, Peter	44.000 Gravitational lensing as a probe of structure	Jun 2003	A			X	R	C	c	U H
6	2006MNRAS.368.1323H Heymans, Catherine; Van Waerbeke, Ludovic; Bacon, David; Bernstein, Jack; Bernstein	41.000 The Shear Testing Programme - I. Weak lensing analysis of simulated ground-based observations	May 2006	A	E	G	X	R	C	c	U

Microsoft WorldWide Telescope Web Client

http://www.worldwidetelescope.org/webclient/default.aspx?wtml=http%3a%2f%2fwww.worldwidete

Apple Yahoo! Google Maps YouTube Wikipedia News (284) Popular

Explore Guided Tours Search View Settings

Collections > Open Collections > Link Collection >

1 of 1

NAME HDF

Finder Scope

Classification:
Cluster Of Galaxies
in Ursa Maior

Hubble Deep Field Image at Full Resolution

RA:	12h36m48s	Magnitude:	n/a
Dec:	62 : 13 : 01	Distance:	n/a
Alt:	16 : 17 : 01	Rise:	Circumpolar
Az:	11 : 08 : 59	Transit:	Circumpolar
		Set:	Circumpolar

Image Credits:
Credit: R. Williams (STScI), the Hubble Deep Field Team and NASA

Name:Hubble Deep Field Image at

Information
Imagery
Virtual Observatory Searches

Set as Foreground Imagery
Set as Background Imagery

Properties
Copy Shortcut

Look up on SIMBAD
Look up on SEDS
Look up on Wikipedia
Look up publications on ADS
Look up on NED
Look up on SDSS

Look At: Sky
Imagery: Digitized Sky Survey (Optical)

Ursa Major
Hubble Deep Field

1 of 1

N
Ursa Major 00:03:58
RA : 12h36m48s
Dec : 62:12:15

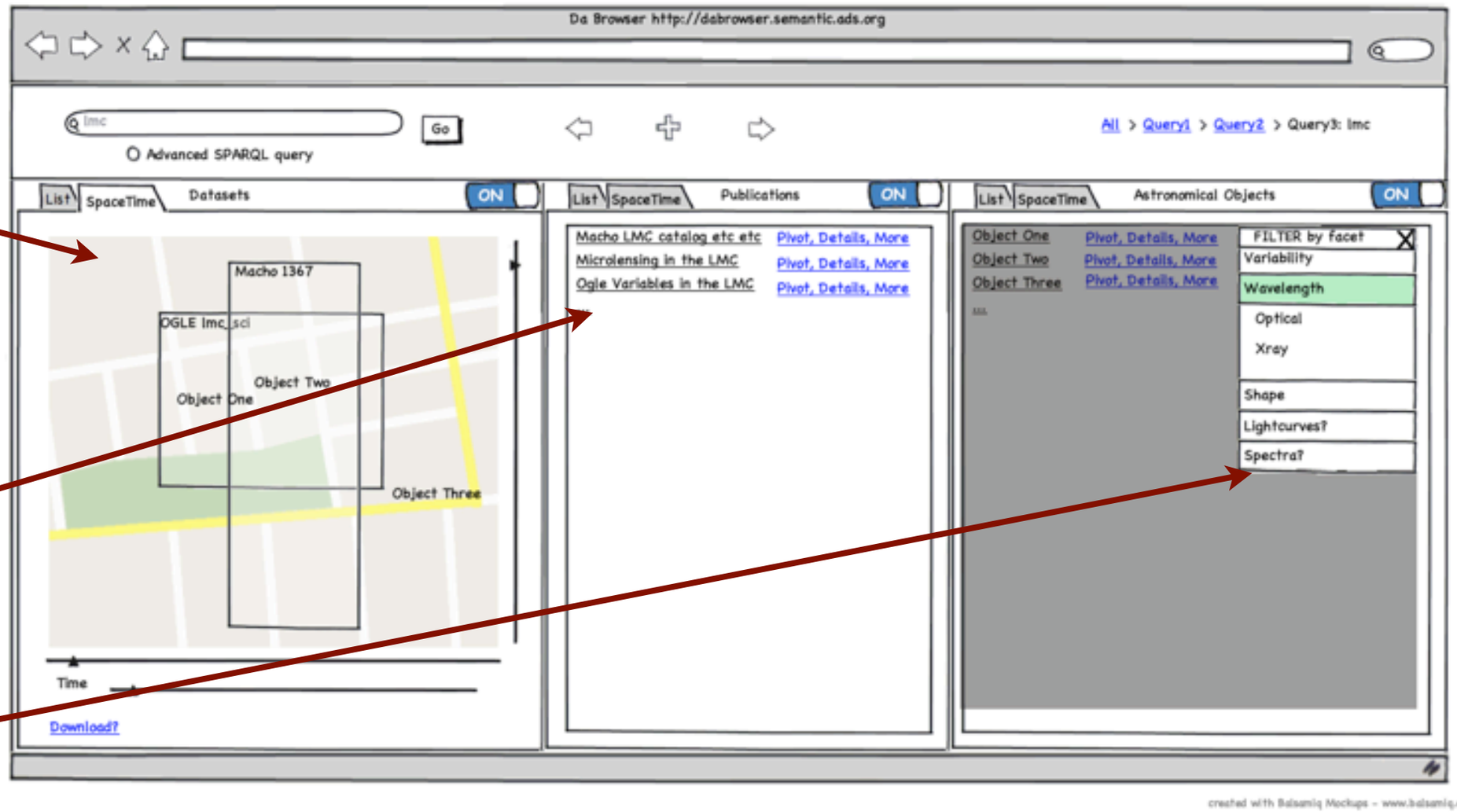
WWT links back to ADS

Example: APOD.App

Spacetime view

List view

History



Datasets Publications Objects

Why Semantics rocks

- Integration creates useful view of *VO resources*
- Tackles *outstanding issues* with dataset IDs, relates to ongoing VAO and IVOA DC&P activities
- Provides semantic layer over existing resources, *crossing boundaries* between archives
- Enhances existing services, enables novel *discovery* interfaces, seamless search / browse

Conclusions

- Electronic publishing of scholarly literature *still evolving*, arXiv usage and features still growing
- Great opportunities exist in *mining full-text* of current and historical publications
- Better tools, policies no substitute for *curation and preservation* by archives
- Semantic interlinking of astronomy resources and services will enable *seamless astronomy* applications (A. Goodman)