

Annotations, Paper Claiming and ORCID

*Alberto Accomazzi (ADS)
Harvard-Smithsonian Center for Astrophysics*

IVOA
New Technologies Session
13 May 2013

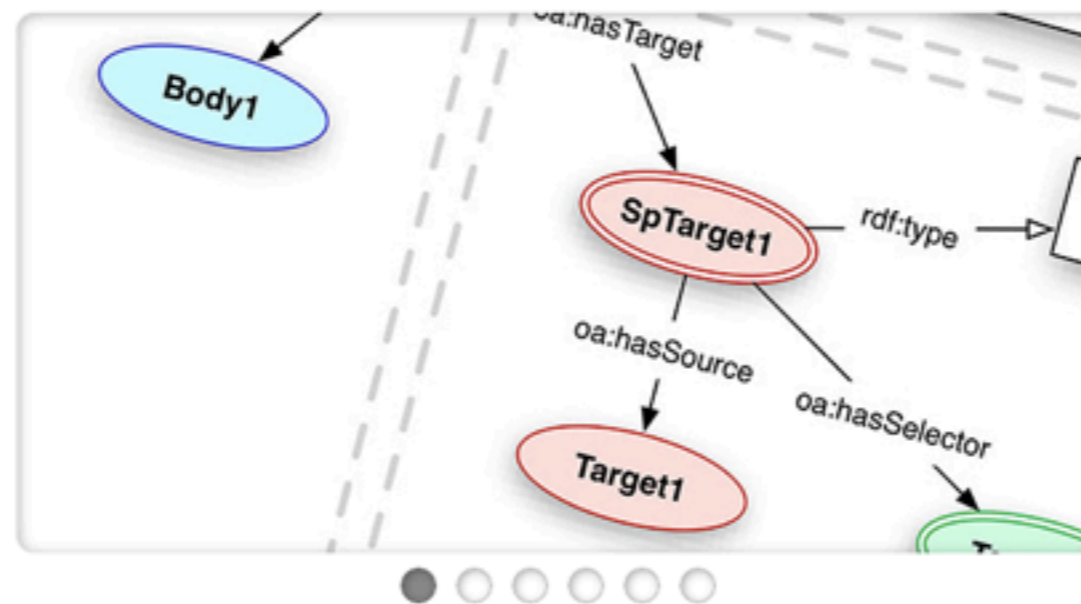


Annotations are Back!

I ANNOTATE

ANNOTO ERGO SUM

APRIL 10-12, 2013, SAN FRANCISCO, CA



BUILDING A COMMUNITY TO ENABLE THE
ANNOTATION OF THE WORLD'S KNOWLEDGE

<http://iannotate.org>

A Brief History

1993: Annotation capability added *and removed* from Mosaic

2001: W3C starts work on Annotea

2009: Google introduces Sidewiki (discontinued in 2011)

2009: Open Annotation Collaboration (OAC) created

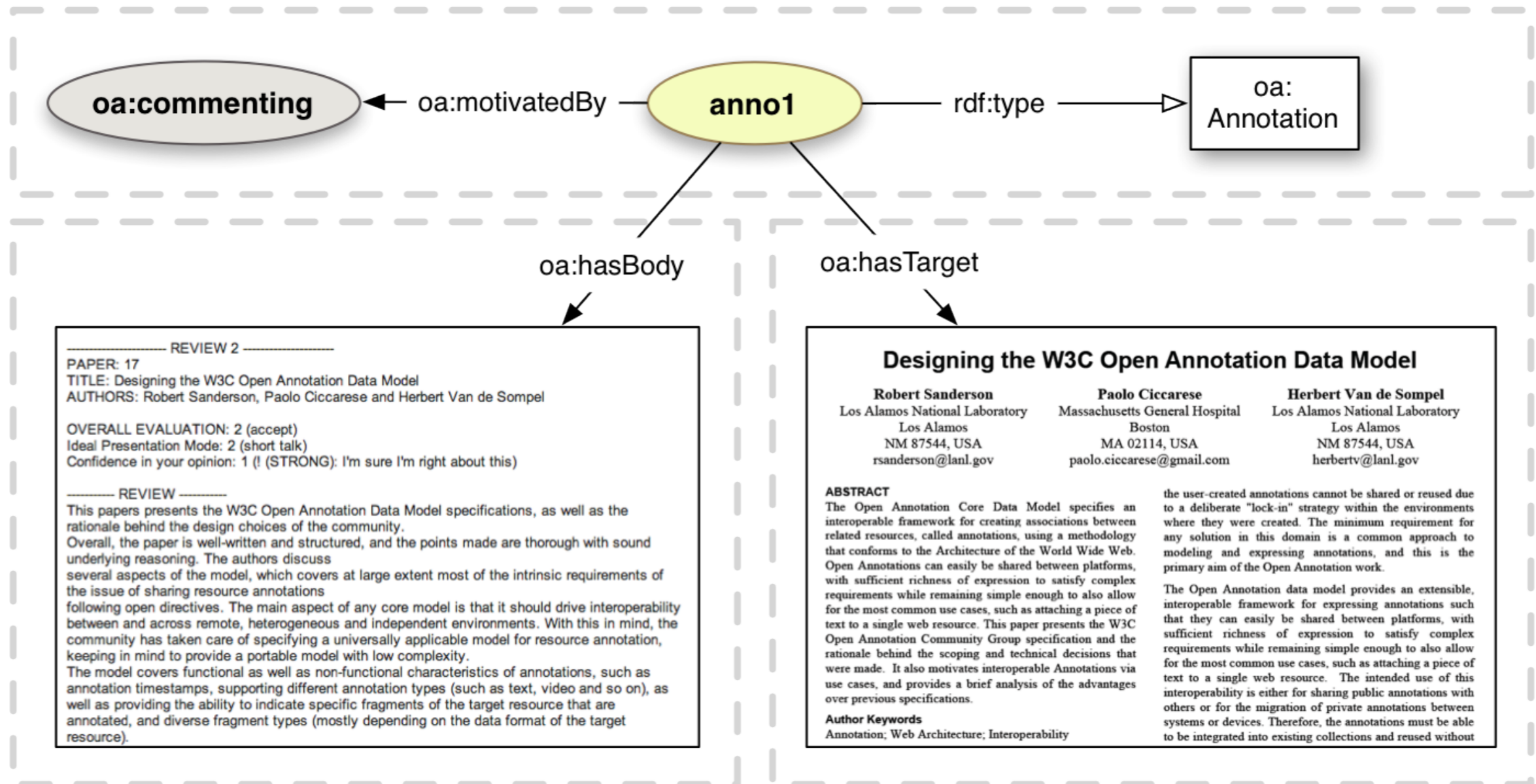
2009: Annotation Ontology (AO) published

2011: W3C Open Annotation Community Group established

2012: First draft of Open Annotation Data Model

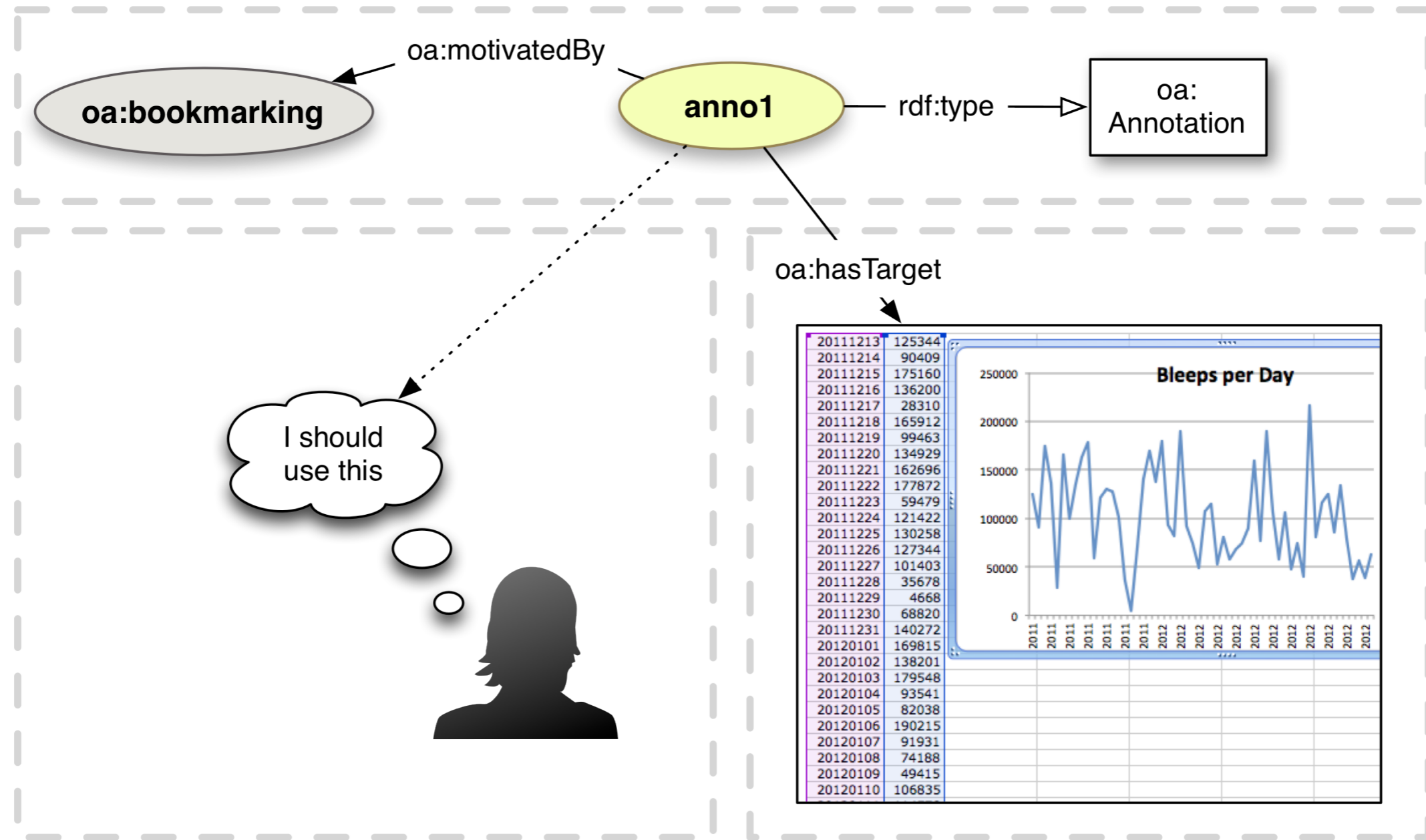
2013: Second Draft, community days in Boston, SF, Manchester

Use Case: Commenting/Refereeing



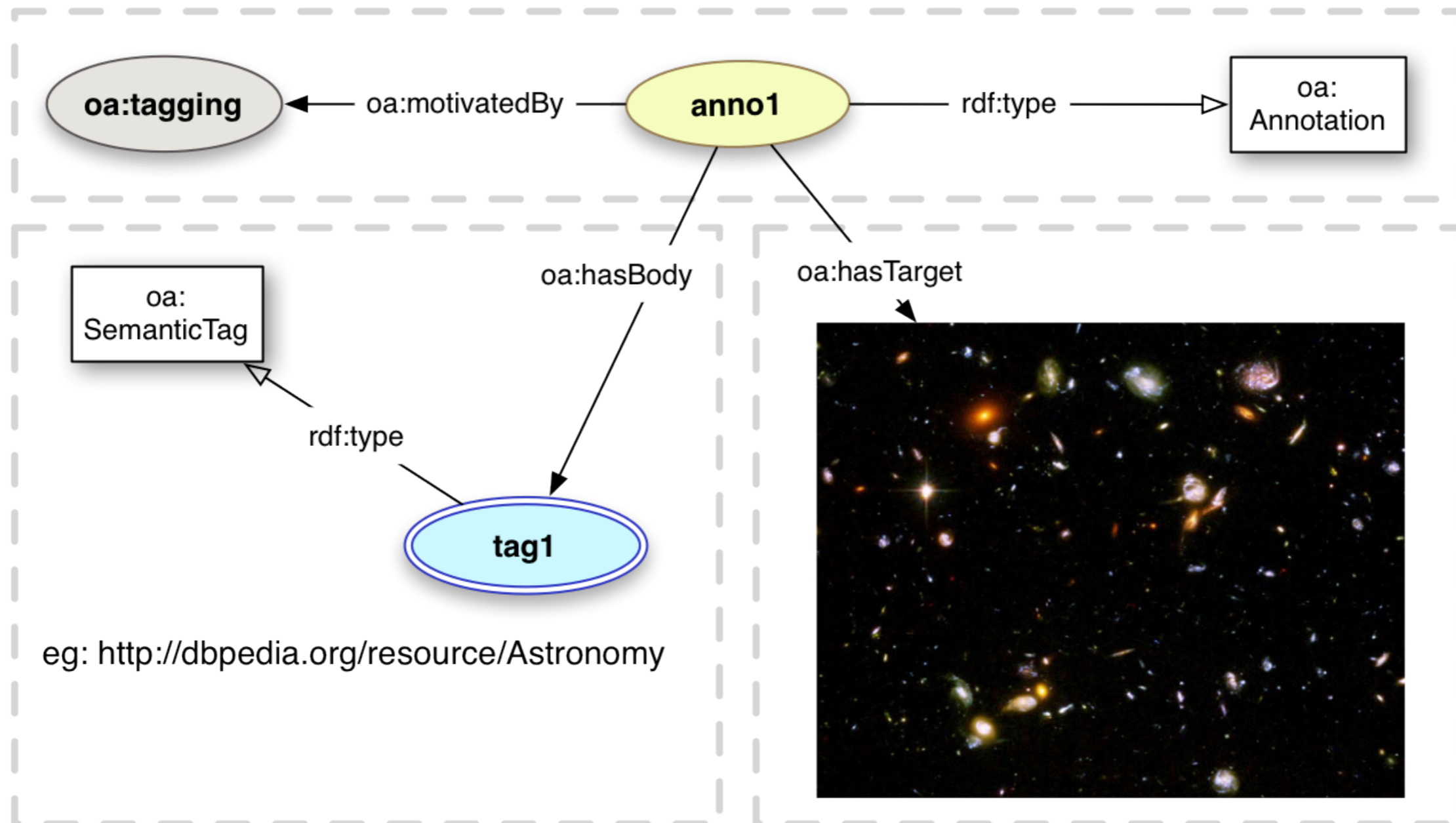
Rob Sanderson, "Open Annotation Community Group" iAnnotate workshop, 10-12 April 2013

Use Case: Bookmarking



Rob Sanderson, "Open Annotation Community Group" iAnnotate workshop, 10-12 April 2013

Use Case: Tagging



Rob Sanderson, "Open Annotation Community Group" iAnnotate workshop, 10-12 April 2013

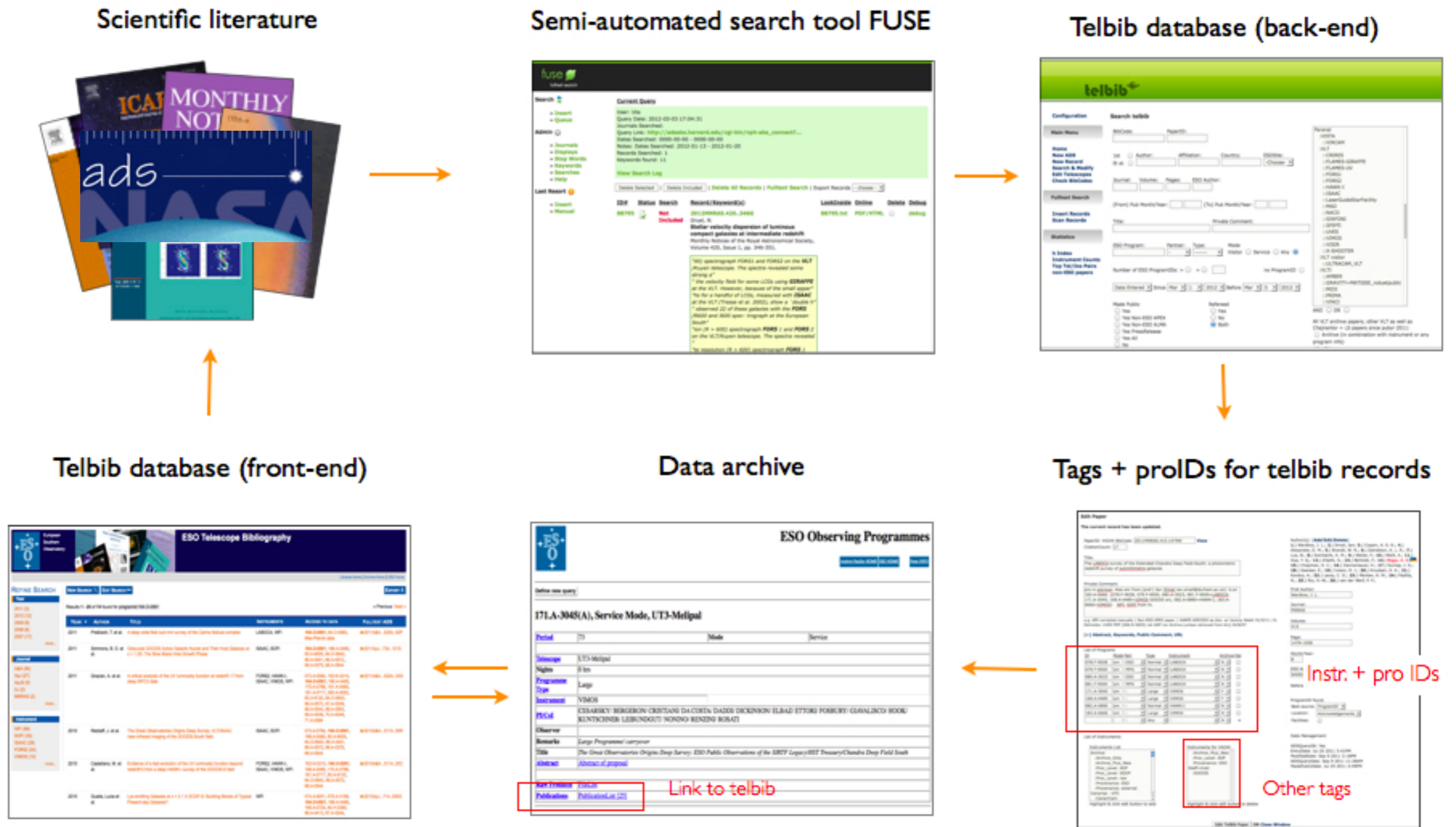
Astronomy Curation and Annotation

- Literature is printed by publishers, curated and enriched by ADS; bibliographies are created by users, institutions, projects, data archives using ADS as a discovery platform
- Astronomical databases and archives curate and maintain datasets, catalogs, measurements for astronomical objects
- Connections between resources are established and shared across all of the above both **pre-** and **post-**publication by publishers, archivists, users, librarians, e.g.
 - ▶ this article claims the redshift of Abell 1689 is 0.1828
 - ▶ this article analyzes data from Chandra Obsid #234
 - ▶ this article reviews multi-wavelength studies of interstellar medium

Problems

- Many of curation workflows, technologies used have been developed over a decade ago; librarians lack the skills to refresh the stack
- Typically, curation happens in closed environments, knowledge is stored in silos and little of it is exposed
- As a result, there is no holistic view of paper, data, entities, claims (no real LOD)
- ADS mediates the problem by aggregating a partial view of the curated bibliography, but most of the information gathered during the process is locked away or lost

Sample bibliography workflow



Meakins & Grothkopf, 2011: Linking Publications and Observations: The ESO Telescope Bibliography. ASP Conf. Proc. 461, 767 <http://adsabs.harvard.edu/abs/2012ASPC..461..767M>

Semantic Tagging of arXiv

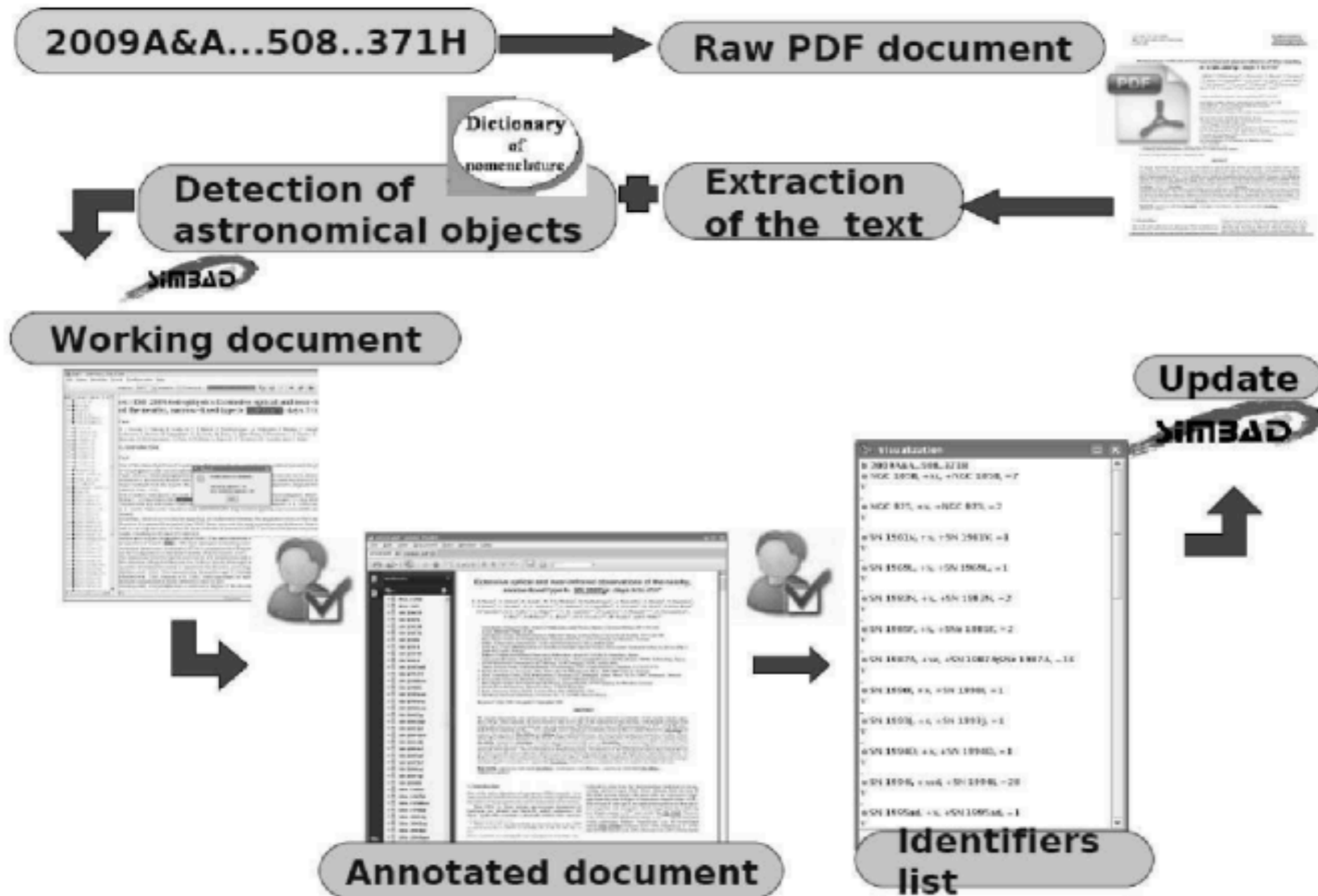
The screenshot shows a web browser window with the URL `sciencewise.info/articles/new`. The page title is "New arXiv.org articles submissions — ScienceWISE". The navigation bar includes "ScienceWISE", "Ontology", "Bookmarks", "New articles", "News", "Login", and "Register".

Filters for "last week", "Yesterday", and "Today" are visible. A row of category dropdowns includes "astro-ph" (7), "cond-mat" (0), "cs" (0), "hep" (0), "math" (0), "nucl" (0), "phys" (0), "q-bio" (0), and "q-fin" (0). A "Specify research interests" section contains a text box with the instruction: "Order latest arXiv.org articles according to your research interests. Use the link above to specify your interests." and a "Replacements" button.

The main content area displays "83 new articles from arXiv.org" with a "Recommend" button. Three articles are shown:

- 1304.3058[PDF]**
Alberto Petriella, Sergio Paron, Elsa Giacani
An X-ray study of the supernova remnant G20.0-0.2 and its surroundings
Tag: `astro-ph.GA`
Abstract:
Aims: We study the supernova remnant G20.0-0.2 and its surroundings in order to look for the high energy counterpart of the radio nebula and to find evidence of interaction between the shock front and the interstellar
Tags: `Supernova remnant`, `Molecular cloud`, `Young stellar object`, ...
Buttons: `+ Add`, `Replacements`
- 1304.3055[PDF]**
Alejandro H. Córscico, Leandro G. Althaus, Alejandra D. Romero
The pulsating low-mass He-core white dwarfs
Tag: `astro-ph.SR`
Abstract:
Recent years have witnessed the discovery of many low-mass ($\lesssim 0.45 M_{\odot}$) white dwarf (WD) stars --- expected to harbor He cores--- in the field of the Milky Way and in several galactic globular and open clusters.
Tags: `White dwarf`, `Star`, `Sloan Digital Sky Survey`, ...
- 1304.3038[PDF]**
F. Taddia, M. D. ...
Carnegie Supernova Project: Observations of Type IIc Supernovae
Tags: `astro-ph.CO`, `astro-ph.SR`

Astronomical Object Extraction



Lesteven et al, 2010: DJIN: Detection in Journals of Identifiers and Names. ASP Conf. Proc. 433, 317 <http://adsabs.harvard.edu/abs/2010ASPC..433..317L>

We could do better...

- Provide the technology and platform to facilitate the curation efforts already taking place, using the open annotation framework
- Web-based portal supporting the creation of user profiles and curatorial roles (scientist, librarian)
- Advanced discovery tools supporting curator-supplied terms to allow targeted search, fine-grained annotations, review and validation of results
- API to support private and public sharing of annotations, integration in third party platforms (publishers websites, other collaborators)

Don't forget the Data!

- John Kunze (CDL): research data are ripe for crowd-sourced annotations
- On the one hand, do we really have crowds eager to annotate astronomy datasets?
- However, it's true that (heterogeneous datasets) are hard to describe and find, hence annotations provide a good vehicle for exposing metadata
- What can we learn from SIMBAD and AstroDAbis annotation experiences with objects and catalogs?

Researcher ID & Paper Claiming

- With an increasing body of literature being published on a regular basis, author name ambiguity is becoming a serious problem
- Over 24,000 papers by “Wang, Y” in ADS alone. How many individuals does that represent?
- Paper claiming is the process that allows a researcher to assert authorship of an article
- arXiv has been doing this since 2009; Thomson-Reuters’s ResearcherID 2008; all publishers want in
- This can fit nicely in an Annotation-based system

Enter ORCID

Connecting Research
and Researchers

- ORCID = Open Researcher and Contributor ID
- Non-profit organization (similar to Crossref)
- Goal: create unique identity system and framework for its dissemination and use
- Broad support from publishers, institutions, funding agencies
- Go get yours at <http://orcid.org>; I am 0000-0002-4110-3511
- Use it when you write papers, apply for grants, publish data, share software

How will it work?

- “Soon” publishers will start asking for ORCID IDs upon paper submission
- Right after that, funding agencies will ask to enter ORCID IDs when submitting proposals
- Some systems (ADS included) will be ORCID aware so that you can search for `orcid:XXXX`
- A service for claiming papers is already available in ORCID (still clunky), claims will be exported to ADS, etc.
- Result: an unambiguous list of {papers | grants | datasets | code | metrics} for unambiguous people
- Stay tuned for updates; full deployment probably ~ 1 yr away

For Data Providers

- Consider implementing ORCID authentication via Auth2
- Look at ORCID's API if you plan to manage researcher-based profiles
- Participate in discussions to shape ORCID's policies and technical services

<http://orcid.org/about/community>