# The NASA Astrophysics Data System: Overview

Alberto Accomazzi, Michael J. Kurtz

ADS Users Group Meeting January 17, 2017







### **Overview**

- ADS's mission
- Assessment and recommendations from Review Panels
- ADS Data Holdings and Curation activities
- ADS Usage and Access
- Development History and System Evolution
- System Requirements and User Evaluation
- Highlights of new System Functionality
- Priorities and Implementation Plan
- Q&A

### Who is ADS?

- Alberto Accomazzi, PI & Program Manager
- Michael J. Kurtz, Project Scientist
- Carolyn S. Grant, Data Ingest and Curation
- Edwin A. Henneken, System Development and Operations
- Donna M. Thompson, Data Curation Librarian
- Roman Chyla, System Architecture and Development
- Steve McDonald, System Development and Operations
- Vacant, User Interface and Front-end Development
- Vacant, Back-End Software Development (pipelines)
- New, Back-End Software Development
- New, System Operations and Cloud Computing

## **ADS's Mission**

## ADS's mission (1/2)

- Maintain a comprehensive, timely and complete database of the scholarly literature in Astronomy & Astrophysics
- Provide discovery services to support research in Astrophysics and related fields
- Promote the use of NASA Astrophysics data by integrating bibliographies and links to data products generated by NASA missions and hosted by NASA archives

## ADS's mission (2/2)

- Provide services for curators and librarians involved in maintaining bibliographies, linking literature and data products, measuring impact
- Interface with publishers and the community to facilitate the implementation of agency policy and government mandates related to Open Access publishing
- Make its efforts in software development freely available under an open-source software license

## **Unique Community Focus**

- Editorial policies reflect community views
  - Making decisions daily on journal inclusion, refereed status
  - Inclusion of gray literature, e.g. conference proceedings
  - Indexing of non-traditional content (catalogs, observing proposals, software) is result of evolution in astronomy scholarly publishing
- Features, Services based on community needs
  - Increasing volume and specialization in the field requires better discovery and analytical services
  - Full-text search essential for maintenance of bibliographies, analytics
  - Additional functionality often requested:
    - ORCID integration, affiliation normalization
    - Citation analysis, visualizations, notifications

## **Unique ADS functionality**

- Comprehensiveness, timeliness, accuracy, focus
  - The only literature system where all of Astrophysics is represented
  - Properly manages eprint and published content, metrics
  - Includes areas of Physics at the boundary with Astrophysics
- NASA Astrophysics data, scientific output exposed
  - o Includes observing proposals for most missions, archives
  - Links to data products, integration of bibliographies
  - Allows search of NED & SIMBAD objects, high-level data catalogs
- Supports wider NASA programs and goals
  - Science: Earth, Planetary Sciences, Heliophysics covered by ADS
  - R&D: mission planning, instrument building, program evaluation

# Assessment and Recommendations from Senior Review Panels

## 2008 Senior Review (1/2)

Rating: Ranked #1

Overall assessment and recommendations:

The Panel applauds the ADS team for providing an outstanding service to the astronomical community and for extending the ADS service to many international mirror sites. The ADS team made the compelling case that the system was "aging." In addition to new hardware, a new (open source) database system should be implemented to ensure reliability, maintainability, and long term stability of the ADS. The Panel recommends that NASA continue to fund the ADS at the full level including the "over-guide" budget.

## 2008 Senior Review (2/2)

#### Additional Comments:

- ADS is so extensively used by the entire professional astronomy community that it is hard to imagine existing without it. By one calculation referenced in the proposal, the efficiency increase to astronomy research in 2002 through the use of ADS is estimated to be approximately 736 full time researchers (compared with otherwise obtaining the information in libraries). The ADS is an enabling tool for research in all of NASA astronomical research programs.
- ADS needs to find a way to educate (busy) astronomers on its secondary capabilities and search features to enable them to separate the 'wheat from the chaff'. This may be helped by providing alternate interfaces (new looks?) to the user.

## **2011 Senior Review (1/2)**

Rating: Ranked #1

Overall assessment and recommendations:

The new ADS interface and functionality is a significant improvement to the old system, and should be able to meet the standards and expectations of the most web-savvy end-user for at least the next five years. The obviously strong relationship that the ADS team has established with the user community is commendable; the panel encourages the ADS to continue to allow user feedback to help shape and direct the ADS design in the future, particularly with regards to the new tools being made available in this Summer 2011 release. The panel recommends that NASA continue to fund the ADS at the full in-guide budget.

https://smd-prod.s3.amazonaws.com/science-green/s3fs-public/atoms/files/ApArchSR\_2011report\_final.pdf

## **2011 Senior Review (2/2)**

#### **Additional Comments:**

- [T]he NASA relevancy of the ADS archive comes from its inherent ability to enable science that is closely aligned to NASA SMD's Science and Strategic plans. The ADS provides important information that contributes to all stages of scientific inquiry, beginning with the preparation of proposals that lead to data acquisition and/or analysis in support of NASA-related science investigations and ending with the final publication and dissemination of results.
- The team has recognized the need for long-term strategic planning, and is encouraged to initiate the development of a 10-year plan that addresses, among other factors, sustainability of in-house expertise on the staff and maintaining a competitive edge in a rapidly-evolving world of electronic information and new methods of media distribution.

## 2015 Senior Review (1/2)

Rating: Excellent

Overall assessment and recommendations:

The panel in general agrees with the ADS prioritized list of tasks and provides the following additional guidance: (1) Maintain continuous current services, (2) Complete transition to new system including transition to the new Ingest Pipeline and incorporate functionalities from the ADS classic to the new system, (3) Improve ADS services incorporating the new database, search/indexing engine, etc., release the new user-interface, incorporate additional functionalities such as the visualization interface, and links to social interfaces such as ORCID needs to be explored.

## 2015 Senior Review (2/2)

#### **Additional Comments:**

- The panel recommends that the ADS take the lead and coordinate the following activities amongst all the data archives: (1) provide tools and infrastructure (with MAST) for creating and registering digital object identifiers (DOI), and (2) work with the journals to provide direct linkages to data sets from manuscripts.
- The panel recommends that ADS sets up a user group, comprised of a representative user community including a member of the NASA archive community that provides guidance to ADS on (1) annual operations/development plans, (2) prioritization of new tools and infrastructure improvements, (3) applicability to science and (4) access to data.

# Data Holdings and Curation activities

## What ADS Aggregates

- We harvest and merge bibliographic data from multiple sources (arXiv, CrossRef, publishers, Astronomy archives, ASCL)
- We enrich metadata via text-mining of the fulltext sources (extract references, acknowledgments, keywords, plots and images)
- We generate and maintain citation and usage networks
- We cross-correlate content (arXiv & published paper, translations, re-publications, Vizier catalogs, observing proposals)
- We collect and maintain external links to publishers, archives (SIMBAD, Vizier, NED, MAST, ESO, etc.)
- We incorporate bibliographies from institutes and archives

## **ADS Data Holdings**

### Bibliographic Data

- 11.8M records (2.2 Astronomy, 8.2 Physics) -- up 25% in last 5 years
- Includes all records from arXiv and publications in relevant journals

#### Links

- 90M citation links -- up 80% in last 5 years
- 667K links to data products, SIMBAD & NED objects

#### Full-text available "from" ADS

- 660K articles (4.9M pages) digitized and hosted by ADS (OA)
- 1.2M articles hosted by arXiv and linked from ADS (OA)
- 8.6M articles hosted by publishers and linked from ADS (mixed)

#### Full-text indexed in ADS

- All of ADS's scanned content & 1.2M articles from arXiv (OA)
- 3.5M digital full-text documents from all major publishers

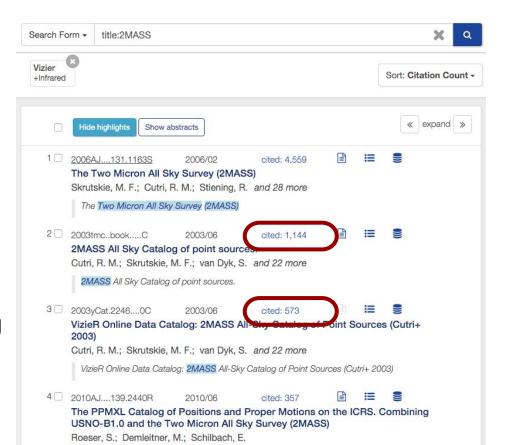
## Ingest of articles in 2012

ADS's ingestion policies are designed to maximize efficiency and coverage of relevant content

- Astronomy -- broadest possible coverage:
  - 40,835 articles (10 pubs with n > 1,000; 51 with n > 100) not refereed
  - 27,540 articles ( 5 pubs with n > 1,000; 53 with n > 100) refereed
- Physics -- core refereed literature:
  - 97,732 articles (8,753 → astro) not refereed
  - 235,257 articles (9,844 → astro) refereed
- General -- multidisciplinary publications:
  - 36,976 articles ( 229 → astro) not refereed
  - 57,988 articles (1,166 → astro) refereed

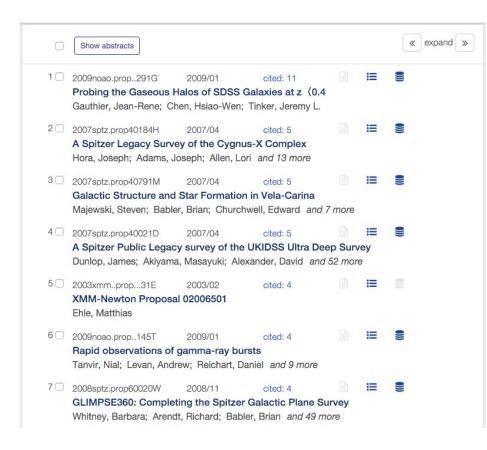
## High-level Data Products Indexed in ADS

- Important datasets are often described in "data" papers
- But can also be available as electronic catalogs
- Greatest majority are from Vizier (close to 10,000 records)
- Once in ADS, they become easily discoverable, citable
- This is how our community has dealt with "data citation" all along



## Observing Proposals Indexed in ADS

- Proposals contain early descriptions of current and ongoing science activities
- They provide a direct link to existing or planned observations
- HST, IUE, CXC, NOAO, XMM,
   KOA, Spitzer, ATNF, Subaru, ...
- 36,000 records, 38,000 data links, 300 citations
- Ongoing ingest rate is 1,000 records/year



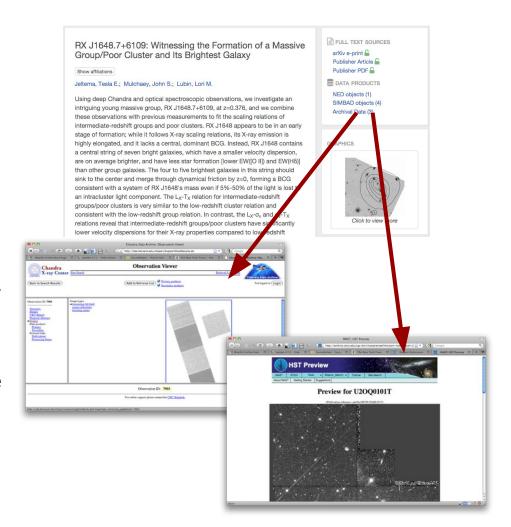
## **Bibliographies**

- Institutional bibliographies, highlighting scientific output from research center or project
- "Telescope" bibliographies, identifying papers related to their data products
- About 30 bibliographic groups so far, over 330K records
- Help with scientific evaluation of projects and institutions, but also useful in disambiguation

ALMA	ISO	ROSAT
ARI	IUE	SDO
CfA	JCMT	SMA
CFHT	Keck	Spitzer
Chandra	Leiden	Subaru
ESO	LPI	Swift
Gemini	Magellan	UKIRT
Herschel	NOAO	USNO
HST	NRAO	XMM

## **Data Links**

- Have existed between Data Centers and ADS since 1994
- Maintained by librarians, data archivists, harvested by ADS
- Bibcode-URL pairs, linking to either individual observations or aggregates
- Often part of data center's bibliographies, used to compute metrics



## Data Links to 2012 articles

Thanks to our ingest of links to archives and objects, the Astronomy collection is particularly "data rich"

- Core journals (AJ, ApJ, ApJL, ApJS, A&A, MNRAS):
  - 8,942 total articles; 6,499 articles with any data links (157,525 citations)
    - 3,145 articles with links to data products (62,833 citations)
    - 6,291 articles with links to SIMBAD objects
    - 2,008 articles with links to NED objects
- All other Astronomy articles:

54,692 total articles; 5,049 articles with any data links

- 2,270 articles with links to data products
- 2,987 articles with links to SIMBAD objects
- 65 articles with links to NED objects

# Usage and Access

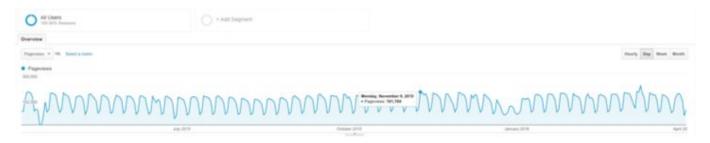
# **Usage: Population**

- 10M Total ADS users (cookies), of which...
- 3.9M Returning users (within a year), of which...

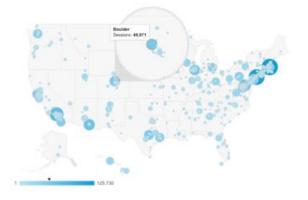
- 55K "Regulars" (multiple visits/month), of which...
- 40K Have created a login account, of which...
- 17K Have signed up for notifications (myADS)
- 3.5K Have active ADS libraries

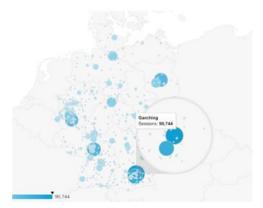
# **Usage: Sessions (1/2)**

ADS use is stable, with ~200K pageviews on heavy days



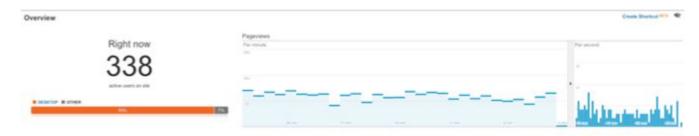
ADS use is concentrated in astronomy centers



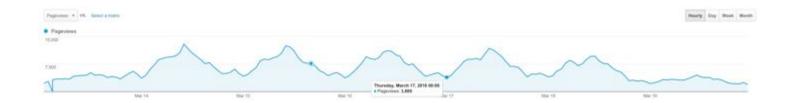


# **Usage: Sessions (2/2)**

ADS routinely has 300-400 simultaneous users



During the work week ADS never averages below one pageview per second



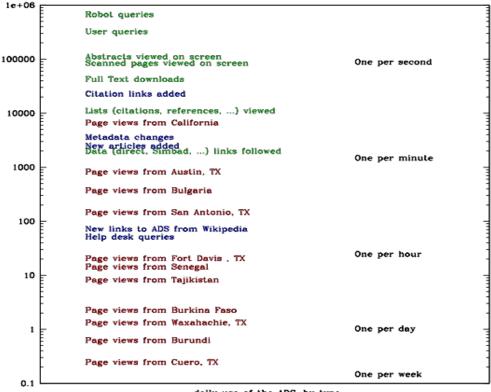
## A Day in the life of ADS: User Sessions

Worldwide Use of ADS Classic on April 8, 2015

Midnight EST



## A Day in the life of ADS: Events



#### **Events:**

Page views
Curation activities
Search

daily use of the ADS, by type

# Development History and System Evolution

# **System Evolution**

1992 ADS Classic: Custom-built search, limited to metadata fields (title, authors, abstract)
 2011 ADS Labs Streamlined Search: a new "skin"

over ADS Classic, introduces facets (filters) of top N results for query refinement and selection

2013 ADS Labs 2.0: Invenio-based metadata store.

ADS Labs 2.0: Invenio-based metadata store, new search engine, full-text search functionality, scalable facets over collections, API

2015 ADS Bumblebee: mongoDB data store, microservices API, client-side dynamic page loading, responsive design, cloud platform

SR Recommendations

"New interface" "wheat from chaff" (2008)

"10-year plan" "keep competitive edge" (2011)

"complete transition"
"improve services"
"release new interface"
(2015)

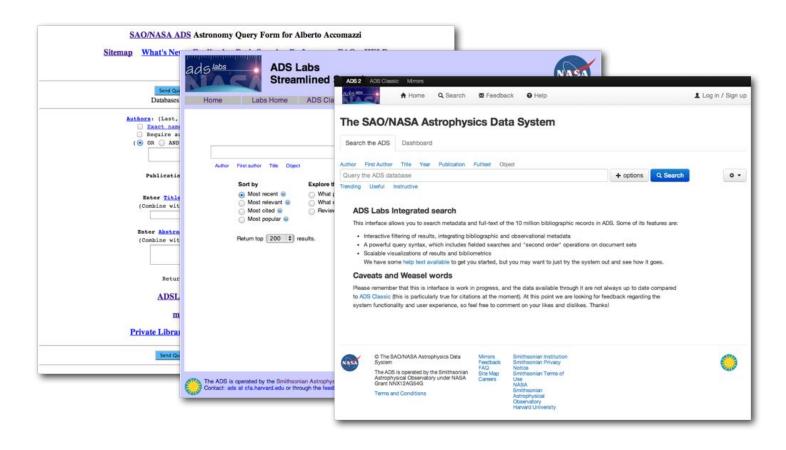
## 1994 - ADS Classic

Sitemap What's New Feedback Basic Search Preferences FAQ HELP  Check out ADS Labs!
Send Query Return Query Form Store Default Form Clear  Databases to query:
Authors: (Last, First M, one per line)
Enter Abstract Mords/Keywords Require text for selection (Combine with: • OR AND simple logic boolean logic)
Return 200 items starting with number 1
ADSLabs Full Text Search: Search within articles
myADS: Personalized notification service
Private Library and Recently read articles for Alberto Accomazzi

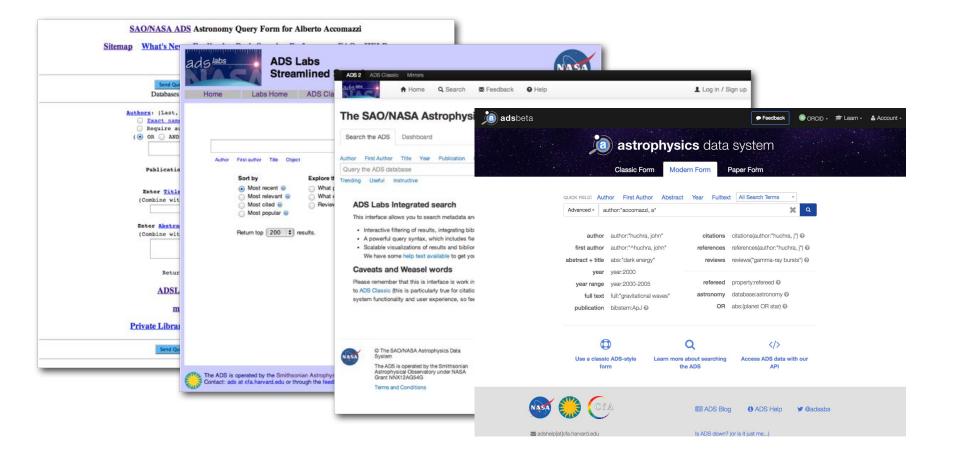
## 2011 - ADS Labs Streamlined Search



### 2013 - ADS Labs 2.0



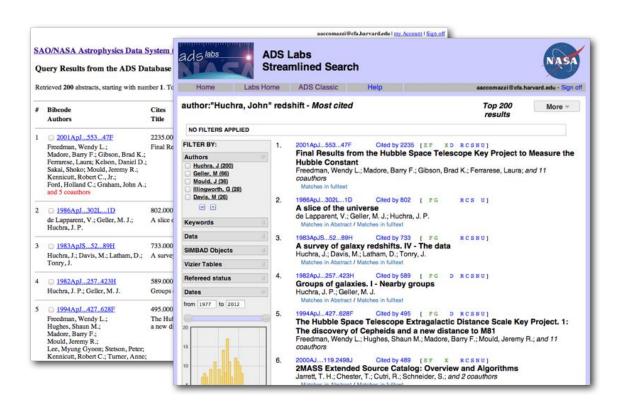
## 2015 - ADS Bumblebee



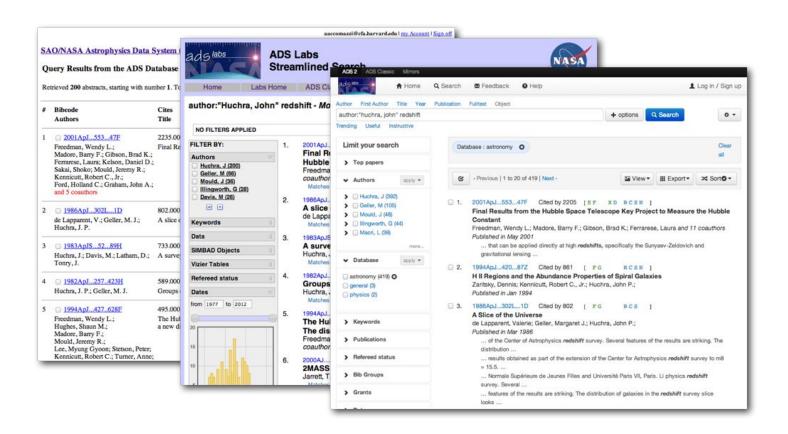
## 1994 - ADS Classic

								accoma	zzi@cfa.h	urvardædu i my Account i Sign of
SA	O/NASA Astrophysics Data S	Title Access Control Help    Access Control Help								
Q	uery Results from the ADS Da	tabase								Go to bottom of page
Re	strieved 200 abstracts, starting with nun	nber 1. Tota	l number sele	ected: 31	3. Total cita	tions: 23	3327			Sort options 1
#	Bibcode	Cites	Date	List of	Links					
	Authors	Title		Access	Control H	elp				
1	□ 2001ApJ55347F	2235.000	05/2001	AZI	EF LX	D	R	C S	N	UН
	Freedman, Wendy L.; Madore, Barry F.; Gibson, Brad K.; Ferrarese, Laura; Kelson, Daniel D.; Sakai, Shoko; Mould, Jeremy R.; Kennicutt, Robert C., Jr.; Ford, Holland C.; Graham, John A.; and 5 coauthors	Final Resu	lts from the l	Hubble S	Space Teleso	ope Ke	y Project t	o Measu	ire the H	lubble Constant
2	□ 1986ApJ302L1D	802.000	03/1986	AZ	FG		RC	S	<u>и</u> н	ı
	de Lapparent, V.; Geller, M. J.; Huchra, J. P.	A slice of	he universe							
3	□ 1983ApJS5289H	733.000	06/1983	ΔZ	FG		RC	SN	QΨ	
	Huchra, J.; Davis, M.; Latham, D.; Tonry, J.	A survey o	f galaxy red	shifts. IV	- The data					
4	□ 1982ApJ257423H	589.000	06/1982	ΔZ	FG	D	RC	SN	O U	
	Huchra, J. P.; Geller, M. J.	Groups of	galaxies. I -	Nearby s	groups					
5	□ 1994ApJ427628F	495.000	06/1994	ΔZ	FG	D	RC	SN	QU	
	Freedman, Wendy L.; Hughes, Shaun M.; Madore, Barry F.; Mould, Jeremy R.; Lee, Myung Gyoon; Stetson, Peter; Kennicutt, Robert C.; Turner, Anne;	a new distance to M81								

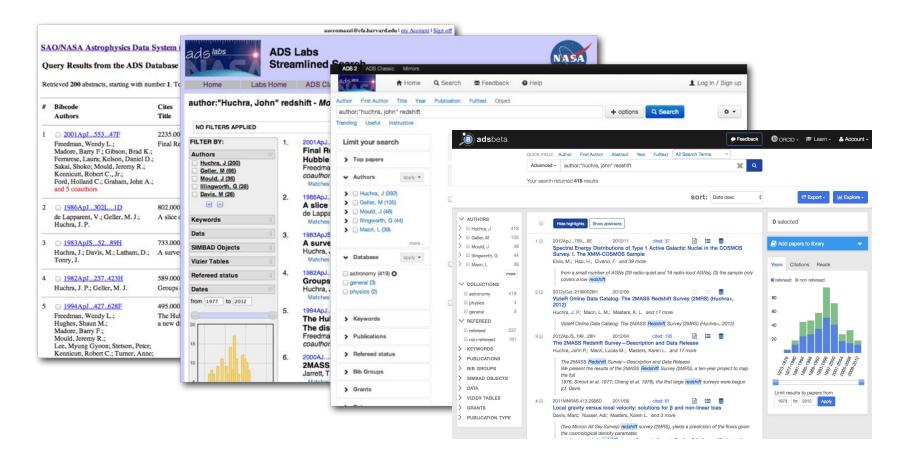
## 2011 - ADS Labs Streamlined Search



## 2013 - ADS Labs 2.0



## 2015 - ADS Bumblebee



### 1994 - ADS Classic

#### SAO/NASA ADS Astronomy Abstract Service · Find Similar Abstracts (with default settings below) Toggle Highlighting · Custom Format · Electronic Refereed Journal Article (HTML) Full Refereed Journal Article (PDF/Postscript) FIND IT O HARVARD arXiv e-print (arXiv:astro-ph/0012376) On-line Data References in the article · Citations to the Article (2235) (Citation History) Refereed Citations to the Article SIMBAD Objects (133) NED Objects (144) · Also-Read Articles (Reads History) HEP/Spires Information Translate This Page Title: Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant Authors: Freedman, Wendy L.; Madore, Barry F.; Gibson, Brad K.; Ferrarese, Laura; Kelson, Daniel D.; Sakai, Shoko; Mould, Jeremy R.; Kennicutt, Robert C., Jr.; Ford, Holland C.; Graham, John A.; Huchra, John P.; Hughes, Shaun M. G.; Illingworth, Garth D.; Macri, Lucas M.; Stetson, Peter B. AA(The Observatoires, Carnegie Institution of Washington, Pasadena, CA 91101.), AB(The Observatoires, Carnegie Affiliation: Institution of Washington, Pasadena, CA 91101; NASA/IPAC Extragalactic Database, California Institute of Technology, Pasadena, CA 91125.), AC(Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia.), AD(Rutgers University, New Brunswick, NJ 08854.), AE(Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington, DC 20015.), AF(National Optical Astronomy Observatoires, PO Box 26732, Tucson, AZ 85726.), AG(Research School of Astronomy and Astrophysics, Australian National University, Weston Creek Post Office, Weston, ACT, Australia 2611.), AH(Steward Observatory, University of Arizona, Tucson, AZ 85721.), AI(Department of Physics and Astronomy, Bloomberg 501, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218.), AJ(Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington, DC 20015.), AK(Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138.), AL(Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK.), AM(Lick Observatory, University of California, Santa Cruz, CA 95064.), AN(Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138.), AO(Dominion Astrophysical Observatory, Herzberg Institute of Astrophysics, National Research Council, 5071 West Saanich Road, Victoria, BC V8X 4M6, Canada; Guest User, Canadian Astronomy Data Centre, which is operated by the Herzberg Institute of Astrophysics, National Research Council Publication: The Astrophysical Journal, Volume 553, Issue 1, pp. 47-72. (ApJ Homepage) Publication Date:

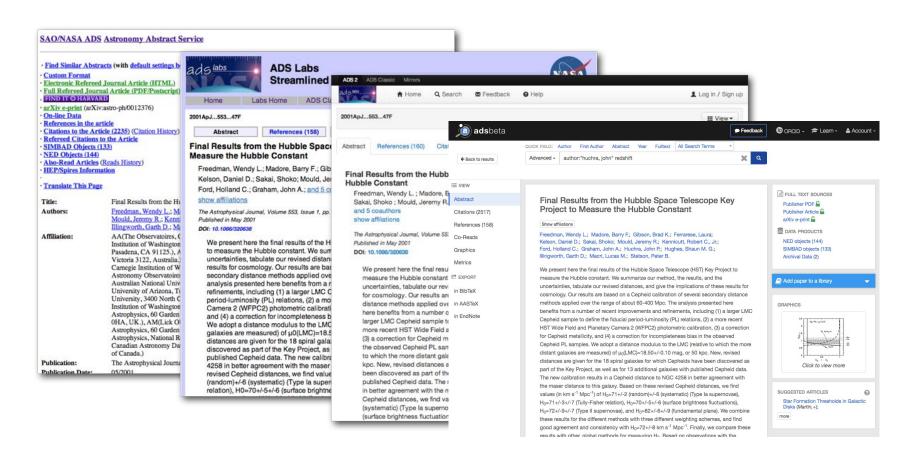
## 2011 - ADS Labs Streamlined Search



## 2013 - ADS Labs 2.0



## 2015 - ADS Bumblebee



	Classic	Streamlined	2.0	Bumblebee
Metadata	Custom	Custom	MARC	Custom
Content	metadata	metadata	full-text	full-text
Search Technology	C + Perl	C + Perl	Invenio/SOLR	Custom SOLR
Search Results	List	List + limited facets	List + scalable facets	List + facets + metrics
Serialization	HTML	XML	XML	JSON
Templating	C + HTML	XSLT, webpy	flask, jinja2, bootstrap	flask, backbone, bootstrap, d3
Storage	filesystem	filesystem	mySQL	MongoDB, postgres
UI Markup	static HTML	CSS, HTML4	CSS2, XHTML	fluid design, CSS3, semantic HTML5

# System Requirements and User Evaluation

## **Search Functionality**

First Author search author: "^Accomazzi, Alberto"

Abstract search abs:(gravitational lensing)

Search full-text full:(HST or JWST)

Acknowledgments ack:ADS

Affiliation search aff:(Harvard or HCO or SAO or Smithsonian)

Unfielded search (AND) Accomazzi ADS bibliography

Positional searches pos(aff:SAO, 2)

Citation search citations(author: "Kurtz, M")

Remove self-citations citations(author: "Kurtz, M") -author: "Kurtz, M"

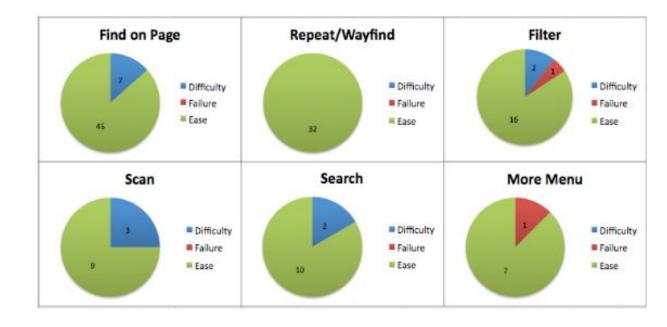
For more information, please see: <a href="http://adsabs.github.io/help/search/search-syntax/">http://adsabs.github.io/help/search/search-syntax/</a>

## **Usability Studies (2011/12)**

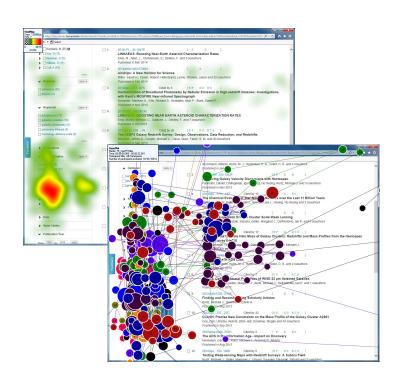


Jennifer Prentice and Jeremy Guillette

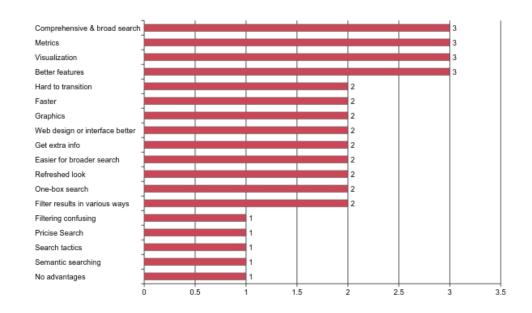
November 27, 2012



# **User Experience Study (2014)**



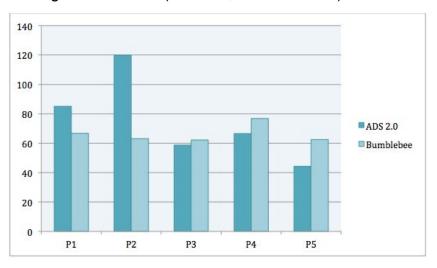
#### User Feedback -- ADS 2.0 vs. Classic



ADS 2.0 User Experience and Eye Tracking Study (2014). Prof. Rong Tang, Simmons College

# A/B Testing (ADS 2.0 vs. Bumblebee)

#### Average time on task (seconds; lower is better)



#### Interface Preference (post-session survey)

Question	ADS 2.0	Bumblebee		
Which interface do you find easier to use?	0	5		
Which of these retrieval mechanisms better facilitates your overall searching?	0	5		
Which interface had the most useful search results layout?	1	4		

Survey	Question	P1	P2	P3	P4	P5
Pre- session	ADS 2.0 Ease of use	6	n/a	n/a	5	5
Post- session	ADS 2.0 Ease of use	5	4	5	5	5
Post- session	Bumblebee Ease of use	6	5	6	5	6

# Highlights of New System Functionality

https://ui.adsabs.harvard.edu

# Implementation Plan

**Priorities and** 

## First Priority: Maintain Current Services

- Community dependence on ADS require 24/7 operations
  - Critical for research, paper, proposal writing
  - Only way to keep up with current literature given rate of publication
  - ADS used by people and applications around the clock
- Current system vulnerable and not sustainable
  - ADS Classic is not maintainable technology -- 20yr old code lacks scalability and is hard to maintain
  - Project requires stability -- redundancy of skills best insurance against resignations, illnesses
- Rate and variety of data indexed expected to increase
  - Publications, citation rates indicate sustained growth in content
  - Non-traditional content now being cited will require ADS indexing

## **Second Priority: Complete Transition**

- Reproduce ADS Classic front-end functionality
  - Integration of NED object search
  - Support all export options, bibliographic report for grant submission
  - o Implement personalizations -- fully featured user libraries, notifications
  - Increase reliability and service capacity
- Implement back-end functionality to support new system
  - Metadata, full-text processing workflow
  - Reference resolution framework
  - Integration of links and observational metadata
  - Processing and serving of full-text scans
- Adapt to changing industry standards, new requirements
  - Use of DOIs instead of journal, volume, page causes nomenclature issues
  - More complex metadata needed to support ORCiD, FundRef, etc.

## **Third Priority: Improve Services**

- Support Data, Software, non-traditional Scholarly Content
  - Identification of software and data products required for reproducibility
  - Formal citation and indexing essential for providing credit
  - o Blogs, posters, presentations are now being cited in scholarly literature
  - Community must vet content, ADS will provide discovery ("if it's not in ADS it doesn't exist")
- Expanded use of ADS libraries
  - Librarian use for creating bibliographies, tracking data products
  - Library sharing will support collaborative curation, publishing of collections
- Improving search via disambiguation and classification
  - Author disambiguation, claiming via ORCiD (supported by AAS, others)
  - Institution disambiguation via Ringgold standard
  - Concept extraction using Unified Astronomy Thesaurus (AAS/ADS)

# **Programmatic Highlights for FY17-18 (1/2)**

#### Data Curation and Indexing

- a. Update metadata model to allow for non-bibliographic content indexing (software, data)
- b. Complete affiliation normalization and mapping
- c. Expand metadata aggregation to include additional OA repositories, article versions
- d. Implement flexible metadata enrichment pipeline
- e. Import ORCID mappings from arXiv and other collaborators

#### Service Hosting and Deployment

- a. Create robust mirroring and failover strategy for services in AWS
- b. Optimize response for user queries vs. API requests
- c. Scale up service capacity and responsiveness by two orders of magnitude

#### System Architecture

- a. Make metrics and visualization services scalable on large collections
- b. Enable third-party authentication via industry standards
- c. Provide interoperability with emerging online authoring environments

# **Programmatic Highlights for FY17-18 (2/2)**

#### Search functionality

- a. Implement NED searches, facets
- b. Improve relevancy, add sort options (by author name, # of authors, norm. citations, etc.)
- c. Implement highly used export formats natively, (custom format, bibtex, XML)
- d. Implement reporting tools (e.g. author/affiliation page for proposal writing)
- e. Implement linkout service for fulltext, other internal/external resources

#### Personalizations

- a. Keep track of recent searches, recently read articles, saved searches
- b. Upgrade the myADS notification system to use new search engine and user accounts
- c. Integrate ORCiD claiming into user profiles, notifications

#### Transition ADS classic Search Engine infrastructure

- a. Update curation and management of ancillary knowledge bases (synonyms, schema)
- b. Make website search-engine and third party application friendly
- c. Optimize UI for mobile apps, crawlers, web applications and widget embedding
- d. Retire ADS Classic user interface and search engine

# **Notional Program for FY19-22**

#### 1. Operations and Development

- a. Continue system maintenance focusing on reliability, currency, completeness
- b. Re-engineer data ingest pipeline and decouple it from ADS Classic legacy code
- c. Implement best practices in curation, including record-level and field-level provenance
- d. Support metadata enrichment at database scale through supervised text mining
- e. Re-implement reference parsing and resolution based on new API
- f. Update the ADS citation management system using modern relational databases technology

#### 2. New Efforts and Initiatives

- a. Provide real-time harvesting, aggregation, and indexing of resources
- b. Use machine-learning techniques to improve document classification, recommendations
- c. Support collaborative research environments and next-generation e-publishing systems
- d. Leverage annotations in support of distributed curation and indexing
- e. Implement context-sensitive auto-complete suggestions
- f. Support researcher, institution, funder focused pages based on ORCID, Fundref standards

### Resources

- ADS "Bumblebee" Interface: <a href="https://ui.adsabs.harvard.edu">https://ui.adsabs.harvard.edu</a>
- "Recent Developments and Initiatives in Scholarly Publishing" (May 2016):
   <a href="http://wiki.ivoa.net/internal/IVOA/InterOpMay2016-DCP/IVOA2016PublishingUpdate.pdf">http://wiki.ivoa.net/internal/IVOA/InterOpMay2016-DCP/IVOA2016PublishingUpdate.pdf</a>
- "Automatic Construction of Evaluation Sets and Evaluation of Document Similarity Models in Large Scholarly Retrieval Systems" (Jan 2016): <a href="https://ui.adsabs.harvard.edu/#abs/2016arXiv160101611K/abstract">https://ui.adsabs.harvard.edu/#abs/2016arXiv160101611K/abstract</a>
- "ADS Services in support of the Discovery, Management and Evaluation of Science Data" (Dec. 2015): <a href="https://ui.adsabs.harvard.edu/#abs/2015scop.confE...3A/abstract">https://ui.adsabs.harvard.edu/#abs/2015scop.confE...3A/abstract</a>
- "Measuring Metrics A forty year longitudinal cross-validation of citations, downloads, and peer review in Astrophysics" (Oct. 2015): <a href="https://ui.adsabs.harvard.edu/#abs/2015arXiv151009099K/abstract">https://ui.adsabs.harvard.edu/#abs/2015arXiv151009099K/abstract</a>
- "Introduction to DOIs and some thoughts on data citation" (Oct. 2015):
   <a href="http://wiki.ivoa.net/internal/IVOA/InteropOct2015DCP/Accomazzi-DOI-Overview.pdf">http://wiki.ivoa.net/internal/IVOA/InteropOct2015DCP/Accomazzi-DOI-Overview.pdf</a>
- "The White House Open Access Mandate: Implications for Astronomy" (May 2013): <a href="http://wiki.ivoa.net/internal/IVOA/InterOpMay2013DCP/OA\_DCP.pdf">http://wiki.ivoa.net/internal/IVOA/InterOpMay2013DCP/OA\_DCP.pdf</a>
- "Annotations, Paper Claiming and ORCID" (May 2013):
   <a href="http://wiki.ivoa.net/internal/IVOA/InterOpMay2013NewTech/NewTechAccomazzi.pdf">http://wiki.ivoa.net/internal/IVOA/InterOpMay2013NewTech/NewTechAccomazzi.pdf</a>
- "The Unified Astronomy Thesaurus" (Oct 2012):
   <a href="http://wiki.ivoa.net/internal/IVOA/InterOpOct2012DCP/UAT-Update.pdf">http://wiki.ivoa.net/internal/IVOA/InterOpOct2012DCP/UAT-Update.pdf</a>
- "Data Citation in Astronomy" (Oct 2011): <a href="http://wiki.ivoa.net/internal/IVOA/InterOpOct2011DCP/DataCitations.pdf">http://wiki.ivoa.net/internal/IVOA/InterOpOct2011DCP/DataCitations.pdf</a>
- NASA Open Access Plan in response to OSTP Memo (Dec 2014):
   <a href="http://science.nasa.gov/media/medialibrary/2014/12/05/NASA\_Plan\_for\_increasing\_access\_to\_results\_of\_feder\_ally\_funded\_research.pdf">http://science.nasa.gov/media/medialibrary/2014/12/05/NASA\_Plan\_for\_increasing\_access\_to\_results\_of\_feder\_ally\_funded\_research.pdf</a>

# Q&A

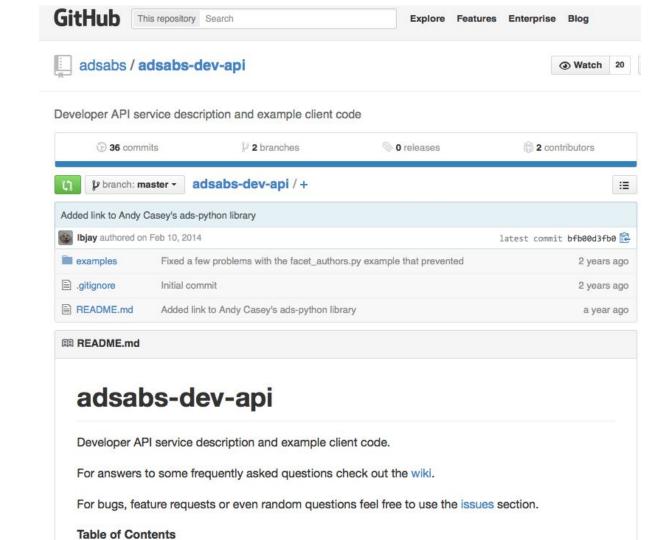
**Backup Material** 

## **Panel Questions**

- What strategies do the Archives use for providing citations to the data they provide?
   ADS has a well-established system in place to track citation to all of its records, including papers, observing proposals, catalogs, software.
- What user/community outreach activities do the archives do to advertise their data products?
  - ADS attempts to have a presence at the major conferences in Astronomy (AAS, DPS, etc), and Library Science (LISA, SLA). Our overguide proposal includes an increase in instructional material and on-line presence to promote and support the new system and its features.
- What is the data backup and hardware/software refresh policies do the archives perform?
   On average, upgrade of storage on a 3-year cycle and major computing components on 2-year cycle. With move to the cloud comes much greater flexibility in provisioning hardware. Goal for software is continuous integration (currently applies to new system components).
- Show examples of cross-archive linkages

  ADS has data links to: Chandra (10,840), Spitzer (2,133), HEASARC (14,079), NED (65,483),

  MAST (20,775), NExScI (1,393), plus more; cross-archives links can be discovered via ADS.
- What ongoing activities are the Archives doing to conform to community standards?
   PM was chair of IVOA Data Curation & Preservation IG, promoting use of DOIs, indexing of data & software products; ADS early partner of arXiv, ORCID, FundRef, OA initiatives



# A Day in the life of ADS: Sample Queries

Topic Queries From Around the World on ADS Classic: April 8, 2015 1:00 EST NGC6819 Regularized Macdivilarisble Symplectic Rs Oph photo conductivity studies of polymen "reflections on relativity" nuclear reactions magnetars

# Open Access: ADS, the OSTP Mandate, and NASA Policy

- ADS already performs most of the goals outlined in the 2013 Office of Science & Technology Policy for NASA Astrophysics
- Our community is compliant because of delayed open access, existence of arXiv and ADS as repository linking to Open Access and non-OA full-text
- Decision by NASA to have full-text deposited in PubMed Central does not affect the need for disciplinary repositories such as ADS, INSPIRE, PubMed (all providing services for specialized search and discovery)

# **ADS and Scholarly Publishing**

## New industry-wide initiatives and standards

- ORCiD (standard author ID): adopted by all major journals, ORCiD search and claiming now integrated in ADS
- FundRef (standard funding reporting): adopted all major agencies in US and abroad, will be used for impact analysis in ADS
- DataCite: DOIs assigned to data products, software, gray literature

## Google Scholar

- Comprehensive index across a wide range of disciplines, but no or negligible curation effort
- Uses ADS as an indexing source, among others
- Unknown indexing policies, low accuracy of metrics
- Business model, long-term future unknown

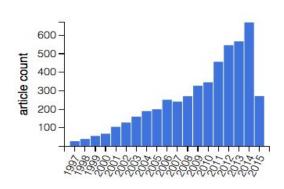
## **Support for Discovery and Curation**

Many data products, software, instruments, are not formally cited in literature. Discovery and evaluation can be done by selecting papers citing the core articles, as well as ones mentioning the product(s). Examples:

- DAOPHOT: 6929 papers (3921 citations / 3008 mentions not in citations)
   citations(title:DAOPHOT and author:stetson) OR full:DAOPHOT
- RVSAO: 662 papers (402/260)
- Advanced Camera for Surveys: 9620 papers (925/8695)
- HIRES: 3145 papers (888/2257)

Acknowledgement searches are also possible now that we have full-text (with limitations):

Acknowledgments to ADS: <u>ack:ADS</u>



Guideline and Augmented FTE requests	FY16		FY17		FY18		FY19		FY20	
	Guide	Aug								
1. Bibliography Support	2.73	0.80	2.73	0.80	2.73	0.80	2.57	0.96	2.41	1.12
a. Bibliographic ingest	1.73	0.40	1.73	0.40	1.73	0.40	1.57	0.56	1.41	0.72
b. Indexing /archiving/databases	1.00	0.40	1.00	0.40	1.00	0.40	1.00	0.40	1.00	0.40
2. Development	3.25	1.00	3.00	1.25	2.50	1.75	2.50	1.75	2.50	1.75
a. search functionality	0.80	-	0.80	-	0.80	-	0.80	-	0.80	-
b. user tools	0.68	-	0.55	0.13	0.30	0.38	0.30	0.38	0.30	0.38
c. software maintenance	1.78	1.00	1.65	1.13	1.40	1.38	1.40	1.38	1.40	1.38
3. User Support	0.73	0.20	0.73	0.20	0.73	0.20	0.69	0.24	0.65	0.28
4. Hardware and Licenses	-	-	-	-	-	-	-	-	-	-
5. Management	1.98	-	1.98	-	1.98	-	1.98	-	1.98	-
Total	8.68	2.00	8.43	2.25	7.93	2.75	7.73	2.95	7.53	3.15
Grand Total		10.68		10.68		10.68		10.68		10.68