The NASA Astrophysics Data System

Alberto Accomazzi

Special Libraries Association
14 June 2015
Overview

- The Role of ADS in research and policy
- ADS Data Holdings and Curation activities
- ADS Usage and Access
- Development History and System Evolution
- Usability Studies and System Evaluation
- Highlights of new System Functionality
- Prioritized Archive Objectives (PAOs)
- Budget Discussion
- Responses to Panel Questions
ADS’s mission

- 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
- Provide services for curators and librarians involved in maintaining bibliographies, linking literature and data products, measuring impact
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)
- Questions which ADS can help answer:
  - Which papers do not have OA version available one year after publication?
  - Which papers have NASA data associated with them?
  - Am I required to deposit a version of my paper in PMC?
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
  ○ 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
ADS Data Holdings

- **Bibliographic Data**
  - 11.2M records (2.1 Astronomy) -- up 27% since 2011
  - Ingested historical records from SPIE, AIP, Elsevier

- **Links**
  - 77M citation links -- up 71% since 2011
  - 596K links to data products, SIMBAD & NED objects

- **Full-text archive**
  - 640K articles (4.8M pages) digitized and hosted by ADS
  - 4.5M full-text documents indexed in ADS -- up 275% since 2011
    - All of ADS’s scanned content
    - 1M articles downloaded from arXiv
    - 3M 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
Data Links to 2012 articles

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

  - 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
# A Day in the life of ADS: Events

<table>
<thead>
<tr>
<th>Events:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page views</td>
</tr>
<tr>
<td>Curation activities</td>
</tr>
<tr>
<td>Search</td>
</tr>
</tbody>
</table>

### Daily use of the ADS, by type

<table>
<thead>
<tr>
<th>Event Description</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot queries</td>
<td>One per second</td>
</tr>
<tr>
<td>User queries</td>
<td>One per minute</td>
</tr>
<tr>
<td>Abstracts viewed on screen</td>
<td>One per hour</td>
</tr>
<tr>
<td>Scanned pages viewed on screen</td>
<td>One per day</td>
</tr>
<tr>
<td>Full Text downloads</td>
<td>One per week</td>
</tr>
<tr>
<td>Citation links added</td>
<td></td>
</tr>
<tr>
<td>Lists (citations, references, ...) viewed</td>
<td>One per minute</td>
</tr>
<tr>
<td>Page views from California</td>
<td></td>
</tr>
<tr>
<td>Metadata changes</td>
<td></td>
</tr>
<tr>
<td>New articles added</td>
<td></td>
</tr>
<tr>
<td>Date (direct, Simbad, ...) links followed</td>
<td></td>
</tr>
<tr>
<td>Page views from Austin, TX</td>
<td></td>
</tr>
<tr>
<td>Page views from Bulgaria</td>
<td></td>
</tr>
<tr>
<td>Page views from San Antonio, TX</td>
<td></td>
</tr>
<tr>
<td>New links to ADS from Wikipedia</td>
<td></td>
</tr>
<tr>
<td>Help desk queries</td>
<td></td>
</tr>
<tr>
<td>Page views from Fort Davis, TX</td>
<td></td>
</tr>
<tr>
<td>Page views from Senegal</td>
<td></td>
</tr>
<tr>
<td>Page views from Tajikistan</td>
<td></td>
</tr>
<tr>
<td>Page views from Burkina Fasci</td>
<td></td>
</tr>
<tr>
<td>Page views from Wacaxachie, TX</td>
<td></td>
</tr>
<tr>
<td>Page views from Burundi</td>
<td></td>
</tr>
<tr>
<td>Page views from Cuero, TX</td>
<td></td>
</tr>
</tbody>
</table>
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, 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, visualizations. Now available in beta at: [http://ui.adsabs.harvard.edu](http://ui.adsabs.harvard.edu)
1992 - ADS Classic
2011 - ADS Labs Streamlined Search
2013 - ADS Labs 2.0
2015 - ADS Bumblebee
1992 - ADS Classic

SAO/NASA Astrophysics Data System (ADS)

Query Results from the ADS Database

Retrieved 200 abstracts, starting with number 1. Total number selected: 313. Total citations: 23327

<table>
<thead>
<tr>
<th>#</th>
<th>Bibcode</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2001ApJ...553..67F</td>
<td>Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant</td>
</tr>
<tr>
<td>2</td>
<td>1998ApJ...528..101</td>
<td>A view of the universe</td>
</tr>
<tr>
<td>3</td>
<td>1993ApJS...85..255H</td>
<td>A survey of galaxy redshifts. IV - The data</td>
</tr>
<tr>
<td>5</td>
<td>1994ApJ...427..63S</td>
<td>The Hubble Space Telescope Extragalactic Distance Scale Key Project: 1: The discovery of Cepheids and a new distance to M31</td>
</tr>
</tbody>
</table>

Authors:
- Freedman, Wendy L.
- Madore, Brian F.
- Gibson, Break K.
- Fomnese, Laura
- Kelson, Daniel D.
- Saha, A.
- McNeil, Jeremy R.
- Kozasa, Robert C.
- Ford, Holland C.
- Graham, John A.

Access Control Help
2011 - ADS Labs Streamlined Search
Usability Studies (2011/12)

Usability studies by Library & Information Science Simmons students on Streamlined Search in 2011, 2012
User Experience Study (2014)

User Feedback -- ADS 2.0 vs. Classic

- Comprehensive & broad search
- Metros
- Visualization
- Better features
- Hard to transition
- Faster
- Graphics
- Web design or interface better
- Get extra info
- Easier for broader search
- Refreshed look
- One-box search
- Filter results in various ways
- Filtering confusing
- Price/fee/search
- Search tactics
- Semantic searching
- No advantages

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)

<table>
<thead>
<tr>
<th>Question</th>
<th>ADS 2.0</th>
<th>Bumblebee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which interface do you find easier to use?</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Which of these retrieval mechanisms better facilitates your overall searching?</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Which interface had the most useful search results layout?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Survey | Question             | P1 | P2 | P3 | P4 | P5 |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-session</td>
<td>ADS 2.0</td>
<td>Ease of use</td>
<td>6</td>
<td>n/a</td>
<td>n/a</td>
<td>5</td>
</tr>
<tr>
<td>Post-session</td>
<td>ADS 2.0</td>
<td>Ease of use</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Post-session</td>
<td>Bumblebee</td>
<td>Ease of use</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

Usability testing of Bumblebee vs. ADS Labs 2.0 (2014)
Example Queries

First Author search

Abstract search

Search full-text

Acknowledgments

Affiliation search

Unfielded search (AND)

Positional searches

Citation search

Remove self-citations

For more information, please see: http://adsabs.github.io/help/search/search-syntax/
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) 
  \[\text{citations(title:DAOPHOT and author:stetson)} \text{ OR full:DAOPHOT}\]
- 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: \text{ack:ADS}
ORCiD Integration Work
ADS Serving the Community

- **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
Resources

ADS “Bumblebee” UI: https://ui.adsabs.harvard.edu
OSTP Open Access Memo: https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Presentations by ADS PM to the IVOA Data Curation and Preservation Interest Group:
Q&A