

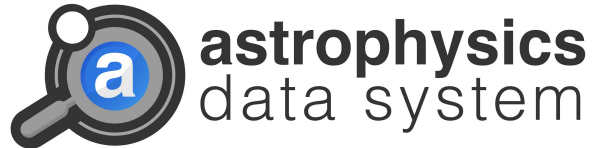
Platform Transition: Removing “beta” from Bumblebee

Alberto Accomazzi & the ADS Team

aaccomazzi@cfa.harvard.edu

@aaccomazzi

ADS Users Group Meeting - 11/2/2017



Panel Recommendations

The ADS Users Group recommends that ADS proceed with an expedited move to the ADS Bumblebee from the classic platform, and discontinue support for the classic platform. Such a transition should be fairly smooth since Bumblebee has available tabs for both the “Classic Form” and the “Modern Form.” [...]

We recommend that the ADS group prioritize the launch of Bumblebee with its classic functionality and engage the public to use it. Additional features (e.g., user’s profiles) can be deferred.

We recommend that a schedule be created for this very much needed transition, especially given the short-staffing situation on ADS (see below). We acknowledge that a robust restructuring of infrastructure may be needed to accommodate dramatically larger numbers of users on Bumblebee. We note that the discontinuation of ADS Classic will significantly free up resources to enable further development

Rollout Plan

- Build out Bumblebee to achieve **feature parity** with ADS Classic
(add features to search interface and corresponding microservices)
- Achieve **content parity** with ADS Classic
(improve data migration pipeline to new data store)
- Provide a **more reliable user experience**
(minimize usability problems and search errors)
- Increase **system capacity** to match expected load from ADS Classic users
(2 orders of magnitude increase)
- **Manage transition** plan and expectations
(be prepared for questions / complaints / unforeseen problems)

Feature Parity - What

- ADS Classic and Bumblebee use very different search engines, which expose different APIs and features, so 100% parity is neither possible nor desirable
- Therefore, parity should be evaluated based on required functionality and priority given to the most frequently used features
- Usage data and direct user feedback have guided us in prioritizing the implementation of functionality

aacomazzi@efa.harvard.edu | [my Account](#) | [Sign off](#)

[SAO/NASA ADS](#) Astronomy Query Form for Alberto Accomazzi

[Sitemap](#) [What's New](#) [Feedback](#) [Basic Search](#) [Preferences](#) [FAQ](#) [HELP](#)

Need a more powerful search? Try [ADS Bumblebee!](#)

Databases to query: [Astronomy](#) [Physics](#) [arXiv e-prints](#)

Authors: (Last, First M, one per line) [SIMBAD](#) [NED](#) [ADS Objects](#)

[Exact name matching](#) [Object name/position search](#)

Require author for selection Require object for selection

(OR AND [simple logic](#)) (Combine with: OR AND)

Publication Date between and
(MM) (YYYY) (MM) (YYYY)

Enter [Title Words](#) Require title for selection
(Combine with: OR AND [simple logic](#) [boolean logic](#))

Enter [Abstract Words/Keywords](#) Require text for selection
(Combine with: OR AND [simple logic](#) [boolean logic](#))

Return items starting with number

Search within articles using [ADS Bumblebee](#)

[myADS](#): Personalized notification service

[Private Library](#) and [Recently read articles](#) for Alberto Accomazzi

Feature Parity - Why

- We don't want users to go back to their old habits because some feature is missing from Bumblebee
- We want to have a clean "cutover" so that we stop relying on Classic even as a back-end service (e.g. for exporting records)
- Some rarely used features in the Classic search engine are needed for layered services (e.g. myADS notifications)

FILTERS

Select References From:

- All bibliographic sources
- All refereed articles
- All non-refereed publications

Select only [articles](#)

Select/deselect publications: (',' separated list)

Select References With:

- A bibliographic entry
- At least one of the following (OR):
- All of the following (AND):
- None of the following (NOT):

<input type="checkbox"/> Abstracts	<input type="checkbox"/> Data Links	<input type="checkbox"/> Electronic Articles
<input type="checkbox"/> Full Text Articles	<input type="checkbox"/> Scanned Articles	<input type="checkbox"/> Mail Order Links
<input type="checkbox"/> arXiv e-print	<input type="checkbox"/> Table of Contents	<input type="checkbox"/> Other related articles
<input type="checkbox"/> References	<input type="checkbox"/> Citations	<input type="checkbox"/> PDS Information
<input type="checkbox"/> SIMBAD Objects	<input type="checkbox"/> NED Objects	<input type="checkbox"/> Also-read
<input type="checkbox"/> Author Comments	<input type="checkbox"/> Library Links	
<input type="checkbox"/> Multimedia	<input type="checkbox"/> HEP/SPIRES Links	

Select References In:

- All Groups
- At least one of the following groups (OR):
- All of the following groups (AND):

<input type="checkbox"/> ARI	<input type="checkbox"/> CfA	<input type="checkbox"/> CFHT	<input type="checkbox"/> Chandra
<input type="checkbox"/> ESO/Lib	<input type="checkbox"/> ESO/Telescopes	<input type="checkbox"/> Gemini	<input type="checkbox"/> Herschel
<input type="checkbox"/> HST	<input type="checkbox"/> ISO	<input type="checkbox"/> IUE	<input type="checkbox"/> JCMT
<input type="checkbox"/> Keck	<input type="checkbox"/> Leiden	<input type="checkbox"/> LPI	<input type="checkbox"/> Magellan
<input type="checkbox"/> NOAO	<input type="checkbox"/> NRAO	<input type="checkbox"/> NRAO/Telescopes	<input type="checkbox"/> ROSAT
<input type="checkbox"/> SDO	<input type="checkbox"/> SMA	<input type="checkbox"/> Spitzer	<input type="checkbox"/> Subaru
<input type="checkbox"/> Swift	<input type="checkbox"/> UKIRT	<input type="checkbox"/> USNO	<input type="checkbox"/> VSGC
<input type="checkbox"/> XMM			

Entry Date:

Since: Before:
Day(DD) Month(MM) Year(YYYY) Day(DD) Month(MM) Year(YYYY)

Min Score:

Example: Abstract Search Feature Parity

Search fields

- Authors
- Objects (SIMBAD, NED, ADS)
- Title words
- Abstract words

Filters / modifiers

- Database
- Publication Date
- Number of items
- Refereed
- Article only
- Specific journals
- "References with" (property selection)
- "References in" (bibliographic group selection)
- Entry date
- Minimum score

Formatting

- All formats natively supported by BBB (Note: these are only available from the export service rather than search engine)

Sorting

- Hybrid Score (technically implemented but not yet exposed)
- Normalized score (not available as native sort)
- Citation count
- Normalized citation count ($\text{citation_count} / \text{author_count}$)
- First author name
- Number of authors
- Publication Date
- Entry date
- Page (~ bibcode)

Settings

- Require field for selection: combination of "and" across fields or implemented via "+" syntax per field
- Synonym replacement: "done" for most fields
- Relative weights: not being implemented
- Use for Weighting: same for this one
- Weighted scoring: also let go

Key: done, in progress, todo, not being implemented

Content Parity - What

- Bumblebee currently depends 100% on ADS Classic for data ingest
- A Data migration pipeline takes content from ADS classic and pushes it to new system
- Occasional ingest problems do occur due to the new system's stricter requirements in metadata encoding and schema
- Currently working on improving error detection and recovery so we can react to them quickly

The screenshot displays the ADSbeta search interface. At the top, there's a search bar with 'Advanced' and '**' filters, and a search button. Below the search bar, it indicates 'Your search returned 12,628,006 results with 99622185 total citations'. The main content area shows a list of search results, including titles like 'Generalized Gradient Approximation Made Simple' and 'Self-Consistent Equations Including Exchange and Correlation Effects'. On the right side, there's a sidebar with a bar chart showing 'Citations' over time from 1974 to 2018. The chart compares 'referenced' (blue) and 'non-referenced' (green) citations, showing a significant increase in total citations over the years.

Content Parity - Why

- We definitely don't want to give the impression that Bumblebee is lacking content wrt Classic
- For weekly updates, a few missed records can be fixed, reingested over the next couple of days
- For daily arXiv updates, slow ingest becomes a blocker to achieving parity
- Just now we are starting to work on simple metadata ingest for daily updates which is Classic-independent

The screenshot shows the myADS Notification interface for user Alberto Accomazzi. The page is titled "myADS Notification (Astronomy database)" and "ADSmyADS". It displays the user's profile, a list of notification categories, and a search form. The notification categories include:

- myADS Notifications:** View in browser, Update myADS Settings, myADS Help
- ADS Query Forms:** ADS Bumblebee, Astronomy, Physics, arXiv e-prints, FAQ
- Current Tables of Contents:** Astronomical Journal, Astronomy & Astrophysics, Astrophysical Journal, Astrophysical Journal Letters, Astrophysical Journal Supplements, Monthly Notices of the Royal Astronomical Society, Recent astro-ph
- Search Recent Papers:** Author, Subject, Send Query
- ACCOMAZZI, ALBERTO - Citations: 357 (total 418):** 2017IAUS...325...361Y: Vavilova,+, UkrVO Astronomical Software and Web-services, 2017IAUS...325...83G: Gangler: The European perspective for LSST, 2017A&C...19...1T: Tapiador,+, Enabling data science in the Gaia mission archive. The present-day mass function and age distribution, 2017AJ...153...37S: Steer,+, Redshift-independent Distances in the NASA/IPAC Extragalactic Database: Methodology, Content, and Use of NED-D, 2017AdAst...2017E...5C: Cano,+, The Observer's Guide to the Gamma-Ray Burst Supernova Connection
- ADS - Recent Papers:** 2017A&A...606A.105R: Rauch,+, Stellar laboratories. IX. New Se v, Sr iv-iii, Te vi, and I vi oscillator strengths and the Se, Sr, Te, and I abundances in the hot white dwarfs G191-B2B and RE 0503-289
- ADS - Most Cited:** 2009Info...3...1H: Henneken,+, Use of astronomical literature - A report on usage patterns, 2015ASPC...492...189A: Accomazzi,+, ADS: The Next Generation Search Platform, 2006EPL...9...22: Henneken,+, Effect of E-printing on Citation Rates in Astronomy and Physics, 2007arXiv:0709.0896K: Kurtz,+, Open Access does not increase citations for research articles from The Astrophysical Journal, 2000A&AS...143...41K: Kurtz,+, The NASA Astrophysics Data System: Overview
- ADS - Most Cited:** 2016MNRAS...459...789T: Turner,+, Ground-based near-UV observations of 15 transiting exoplanets: constraints on their atmospheres and no evidence for asymmetrical transits, 2014Natur...505...69K: Kreidberg,+, Clusters in the
- Favorite Authors - Recent Papers:** No new articles found
- "MACHINE LEARNING", etc - Recent Papers:** 2017GeoPL...44.9276R: Rouet-Leduc,+, Machine Learning Predicts Laboratory Earthquakes, 2017AcASn...57...26C: Cai,+, The Satellite Clock Bias Prediction Method Based on Takagi-Sugeno Fuzzy Neural Network, 2017AcASn...57...39L: Li,+, Preliminary Study on the Classification of Cassini ISS Images' Availability for Astrometry
- "MACHINE LEARNING", etc - Most Popular:** 2010IJMPD...19.1049B: Ball,+, Data Mining and Machine Learning in Astronomy, 2017ApJS...230...20A: Anyan,+, Classifying Radio Galaxies with the Convolutional Neural Network, 2017MNRAS...464.4463K: Kim,+, Star-galaxy classification using deep convolutional neural networks, 2015arXiv:151203385H: He,+, Deep Residual Learning for Image Recognition, 2014arXiv:1406.2661G: Goodfellow,+, Generative Adversarial Networks
- "MACHINE LEARNING", etc - Most Cited:** 2000ApJ...536...571B: Benitez: Bayesian Photometric Redshift Estimation, 1986Natur...323...533R: Rummelhart,+, Learning representations by back-propagating errors, 2016MNRAS...455.1171M: Marshall,+, SPACE WARPS - I. Crowdsourcing the discovery of gravitational lenses, 2015ApJ...811...20C: Collett: The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys, 2015MNRAS...450.1441D: Dieleman,+, Rotation-invariant convolutional neural networks for galaxy morphology prediction

User Experience

- Make the Bumblebee application more robust
- Provide on-line documentation which includes FAQ, quickstart, transition questions
- Minimize errors from API services and the SOLR search engine
- When things go wrong, provide reasonable feedback to the user
- Perform local user testing before launch to help iron out problems

The screenshot shows the adsbeta search interface. At the top, there is a navigation bar with the 'adsbeta' logo, a 'Feedback' button, and links for 'ORCID', 'Learn', and 'Account'. Below this, a red error message states 'Library could not be created'. The search bar contains the query 'author:acomazzi' and shows 'Your search returned 191 results'. The results are sorted by 'Date desc'. A left sidebar contains a navigation menu with categories like 'AUTHORS', 'COLLECTIONS', 'REFEREED', 'KEYWORDS', 'PUBLICATIONS', 'BIB GROUPS', 'SIMBAD OBJECTS', 'DATA', 'VIZIER TABLES', 'GRANTS', and 'PUBLICATION TYPE'. The main content area displays a list of search results, including titles like 'New ADS Functionality for the Curator', 'NASA's Long-Term Astrophysics Data Archives', and 'Stephen S. Murray (1944-2015)'. A right sidebar shows '0 selected' items, an 'Add papers to library' button, and a bar chart titled 'Years' showing the distribution of 'referenced' (blue) and 'not referenced' (green) papers over time.

Execution

- Critical requirements must be satisfied before “Beta” removed from Bumblebee
 - All readiness criteria met: content & feature parity, error-free experience
 - User testing indicates no major problems
 - Scalability issues fully solved and system ready for on-demand growth
- Messaging
 - Communicate to users the need for transition to ADS Classic its advantages
 - Prepare mitigation plan and support material
- Plan timeline
 - Achieve Bumblebee readiness at time T, freeze code
 - Advertise locally and conduct user testing on sample population at T + 2 weeks
 - Work on bug-fixes, evaluate user feedback and fix major issues
 - Deprecate use of ADS Classic to the general public at T + 3 months
 - Discontinue use of the ADS Classic search at T + 15 months

Where we are

- Working as fast as we can on multiple fronts
 - Implementation of missing system functionality in new microservices (e.g. link redirection, native export, author/affiliation forms, NED object search)
 - UI changes to include additional functionality (e.g. custom export, linking)
 - API architecture being upgraded to avoid congestion problems
 - Ingest pipeline upgraded and being tested as we speak
- Bumblebee Readiness not yet reached
 - System components still being developed and tested
 - Known issues with search still to be tackled
 - Some UI changes still in progress
- Currently working on reaching readiness on Jan 1st, 2018
 - Note: this is a tight schedule with little room for contingencies
 - ADS Classic deprecation expected in early Q2 2018