Discovering and Accessing Planetary Sciences Literature with the New Astrophysics Data System (ADS). Ed-win A. Henneken and the ADS Team, Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, MA 02138, USA,
ehenneken@cfa.harvard.edu.

Introduction: The NASA Astrophysics Data System (ADS) has been around for over 2 decades, helping professional astronomers and planetary scientists navigate, without charge, through the increasingly complex environment of scholarly publications. As boundaries between disciplines dissolve and expand, the ADS provides powerful tools to help researchers discover useful information efficiently. In its new form, code-named ADS Bumblebee (https://ui.adsabs.harvard.edu), it may very well answer questions you didn't know you had! While the classic ADS (http://ads.harvard.edu) focuses mostly on searching basic metadata (author, title and abstract), today's ADS is best described as a an "aggregator" of scholarly resources relevant to the needs of researchers in astronomy and planetary sciences (PS), and providing a discovery environment on top of this. In addition to indexing content from a variety of publishers, data and software archives, the ADS enriches its records by text-mining and indexing the full-text articles (about 4.7 million in total, with 130,000 from planetary science journals), enriching its metadata through the extraction of citations and acknowledgments. Recent technology developments include a new Application Programming Interface (API), a new user interface featuring a variety of visualizations and bibliometric analysis, and integration with ORCID services to support paper claiming. The new ADS provides powerful tools to help you find review papers on a given subject, prolific authors working on a subject and who they are collaborating with (within and outside their group) and papers most read by by people who read recent papers on the topic of your interest. These are just a couple of examples of the capabilities of the new ADS. We currently index most journals covering the planetary sciences and we are striving to include those journals most frequently cited by planetary science publications [1].

The new ADS: In contrast to the monolithic architecture of ADS Classic, the new ADS, codenamed Bumblebee, is built on a framework of individual microservices communicating via application programming interfaces, or APIs. Each component of Bumblebee, such as libraries or metrics, lives in its own independent microservice. This modularity allows functionality to be added to individual components or for each component to scaled up as needed to support user traffic without affecting the rest. Kubernetes is used to manage the microservice loads and to scale their availability in real time to match current usage.

In fact, the connection between the back-end search engine and the front-end search interface is also mediated by an API. The web-based interface and the search engine exist as separate components, unlike in Classic. The same search engine API that the web-based interface uses can be accessed by regular users via command line scripts. More information can be found on our help pages and our Github-based instruction manual (https://github.com/adsabs/adsabs-dev-api).

The new ADS uses a modern, open-source, enterprise-level search engine, Apache Solr [2], which is built on the Apache Lucene library and is used by both academic institutions and commercial companies. Out of the box, Solr enables a variety of advanced search options, included fielded and unfielded queries, approximate searches, and wildcard searches. Its high-performance, scalable architecture allows us to index not just the basic article information (metadata) but also the full-text contents of each article, a highly requested feature from the library community. Our implementation of Solr has been heavily modified for our use, including the addition of second-order operators such as `trending()` and familiar ADS Classic operators like the first-author search caret (^). It also support a powerful query language, allowing for queries ranging from a very general nature to fine-grained ones that find matches based on many types of metadata we hold in our index. The most visible part of this query language is the availability of a large set of search modifiers. An exhautive list of modifiers can be found in the online Help (https://adsabs.github.io/help/search/comprehensive-solr-term-list). A significant difference between queries in the old ADS versus those in the new ADS is the switch from OR to AND logic, as is common with all major search engines.

The new ADS is also now open-source: the microservices, back-end data pipelines, and front-end codebase are available on Github. The service is no longer hosted on local servers or on mirror servers, but is instead cloud-based and hosted by Amazon Web Services. This should improve the speed and reliability of the new ADS over Classic.

Overall, these improvements to the structure and hosting of ADS bring searching the astronomy literature into the 21st century. The new ADS has reached feature parity with Classic and work is ongoing to continue to improve it and add new functionality. If you have any

feedback or suggestions, let us know using the Feedback form available in Bumblebee.

Other improvements have been made, such as integration with ORCID, allowing users to claim their own papers or to search using a known ORCID ID. This allows for easier assembly and sharing of personal bibliographies. In addition, graphics and figures are available for articles when provided by the publisher. This provides a quick preview of the data discussed in the paper and its main results. The new ADS also offers new ways to visualize data, including author networks and related papers.

Planetary Science Content in the ADS: The ADS indexes many scholarly journals crucial for research in the planetary sciences, both core PS, those with a significant PS content and the journals mosted cited/citing articles in these journals. The role planetary science litature currently holds in the ADS content and curation efforts is illustrated in figure 1: the ADS content and curation efforts can be represented as rings representing different levels of these efforts.
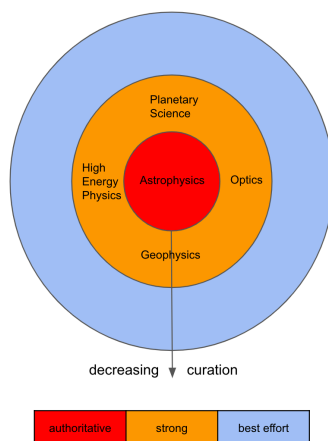

Figure 1 ADS content and curation

In the core collection, representing astrophysics, we strive to maintain our collection at the highest level of completeness (both publications and associated citations) and invest commensurate curation efforts in order to achieve this. Here we endeavour to be the authoritative source. We try to be fully complete in collecting and indexing the literature, not just the refereed journals, but also books, conferences, reports, PhD thesis, the so called gray literature. Here we put substantial effort into collaborating with outside groups (CDS, NED, MAST, HEASARC, ESO, etc.); we work with data centers and archives to link papers in our database to the raw and reduced data behind them; we maintain a database of name changes. Our aim is to achieve complete coverage with essentially 100% precision and recall for content belonging to this collection. Here we take a

leadership role. Next out, the inner ring, are documents which are likely to be used/cited by authors of documents in the bullseye. Here we have nearly every refereed article in physics, optics, geophysics and Planetary Science. We also have many of the larger conference series from the major publishers (e.g. AIP). We do not attempt to curate this content at the same level of the bullseye. We do not seek the kind of close collaborations which we have in the bullseye core.

Below is an example query that finds refereed papers, published in the period 2000 through 2018, with "planet", "evidence" and "water" or "ice" in either their abstracts, titles or keywords

```
abs:(planet evidence (water OR ice))
year:2000-2018 property:refereed
```

Currently, this query finds about 1,000 matches. Figure 2 shows the result generated by the Paper Network in the new ADS. The paper network detects groups of papers based on shared references between those papers.
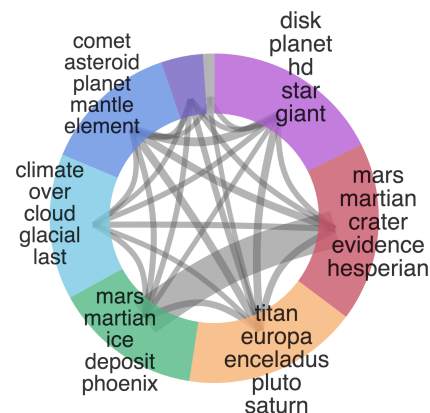

Figure 2 A subject matter clustering for refereed papers, published in 2000-2018, discussing the evidence of planetary water or ice

The articles found for this query were published in journals like *JGRE*, *Icarus*, *Geophysical Research Letters*, *Earth & Planetary Science Letters*, *Planetary and Space Science*, Space Science Reviews, *Astrobiology* and core astronomy journals (like the ApJ, AJ, MNRAS and A&A).

References:

[1] Kurtz et al. (2018), *Merging the Astrophysics and Planetary Science Information Systems*. arXiv e-prints arXiv:1803.03598. [2] Chyla et al. (2015), *Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV), 401*.