

Expansion and Enhancement of FAIR Content in the ADS

Alberto Accomazzi¹, Michael Kurtz¹, Edwin Henneken¹, and Sergi Blanco-Cuaresma¹

¹Harvard-Smithsonian Center for Astrophysics

November 24, 2022

Abstract

The NASA Astrophysics Data System (ADS) is the primary Digital Library portal for researchers in Astronomy and Astrophysics. It is also used extensively by the broader community of Space Science Researchers. In addition to the scientific literature, the ADS has for a long time included in its database non-traditional scholarly resources such as research proposals, software packages, and high-level data products, making them discoverable and easily citable. Over the next three years, in response to NASA's efforts supporting interdisciplinary research across Science Mission Directorate Divisions, the ADS will expand its coverage of Planetary Science and Heliophysics content. During this time, an in-depth analysis of this disciplinary content will provide us with opportunities to improve our coverage of the literature as well as linked research objects such as datasets and software, improving their discoverability and citability. In this presentation I provide an overview of the ADS system, its distinguishing features, and then focus on our efforts to support and promote the goals of Open Science.

Expansion and Enhancement of FAIR Content in the ADS

Alberto Accomazzi, Michael J. Kurtz, Edwin A. Henneken, Sergi Blanco-Cuaresma & the ADS Team
Center for Astrophysics | Harvard & Smithsonian



ABSTRACT

The NASA Astrophysics Data System (ADS) is the primary Digital Library portal for researchers in Astronomy and Astrophysics. It is also used extensively by the broader community of Space Science Researchers. In addition to the scientific literature, the ADS has for a long time included in its database non-traditional scholarly resources such as research proposals, software packages, and high-level data products, making them discoverable and easily citable. Over the next three years, in response to NASA's efforts supporting interdisciplinary research across Science Mission Directorate Divisions, the ADS will expand its coverage of Planetary Science and Heliophysics content. During this time, an in-depth analysis of this disciplinary content will provide us with opportunities to improve our coverage of the literature as well as linked research objects such as datasets and software, improving their discoverability and citability. In this presentation we provide an overview of the ADS system, its distinguishing features, and then focus on our efforts to support and promote the goals of Open Science.

PRESENTED AT:

MOTIVATION

The NASA Astrophysics Data System (ADS) is the primary Digital Library portal for researchers in Astronomy and Astrophysics. It is also used extensively by the broader community of Space Science Researchers. In addition to the scientific literature, the ADS has for a long time included in its database non-traditional scholarly resources such as research proposals, software packages, and high-level data products, making them discoverable and easily citable. Over the next three years, in response to NASA's efforts supporting interdisciplinary research across Science Mission Directorate Divisions, the ADS will expand its coverage of Planetary Science and Heliophysics content. During this time, an in-depth analysis of this disciplinary content will provide us with opportunities to improve our coverage of the literature as well as linked research objects such as datasets and software, improving their discoverability and citability. In this presentation we provide an overview of the ADS system, its distinguishing features, and then focus on our efforts to support and promote the goals of Open Science.

One of the distinguishing features of the ADS is its collection and curation of bibliographies and links to software and data products. For example, a majority (60%) of recent papers published in the core astronomy journals have links to data products (catalogs and observations), while 12% of them explicitly cite software packages. The advantages of these links are well known: by connecting the literature with data and software products, the ADS increases discoverability of both and promotes their use. With the inclusion of Planetary Science and Heliophysics in our core collection, we will work with journals and archives to improve coverage of data links for these disciplines.

Access to the ADS is provided via its portal website: <https://ui.adsabs.harvard.edu>

SPACE SCIENCE LITERATURE

The ADS is well equipped to expand its coverage of Planetary Science and Heliophysics in the next two years, starting with the current effort to identify what missing content we should focus on in order to fully support the two communities and where we can obtain it from. The success of the harvesting and indexing steps required to ingest this content will depend on our ability to successfully create partnerships with the relevant stakeholders, which may take time to develop and not be completed until the end of 2022. The priority for the initial content expansion will be to cover the current literature, but historical content will be added on an ongoing basis once the relevant sources make it available to us.

Discoverability is significantly improved by curation and the use of discipline-specific semantics, which help identify the different ways in which concepts are expressed and linked to each other in the research literature. As an example, the dwarf planet Eris's original designation by the MPC was "2003 UB313." It is expected that a proper retrieval system be capable of finding all papers mentioning this object irrespective of the nomenclature used in them and that these papers be connected to articles sharing similar concepts drawn from disciplinary knowledge bases such as the Unified Astronomy Thesaurus.

Achieving complete coverage of the Planetary and Heliophysics content will involve a sustained effort for at least three years, during which we expect to have identified and ingested current and past refereed journal literature as well as gray literature such as PhD Theses, conference proceedings, presentations, and reports. Similarly, on the development side we start with the development of harvesting and ingesting workflows and then enhance the system through semantic search and metadata enrichment. We expect to deliver this content and capabilities to the community by 2024.

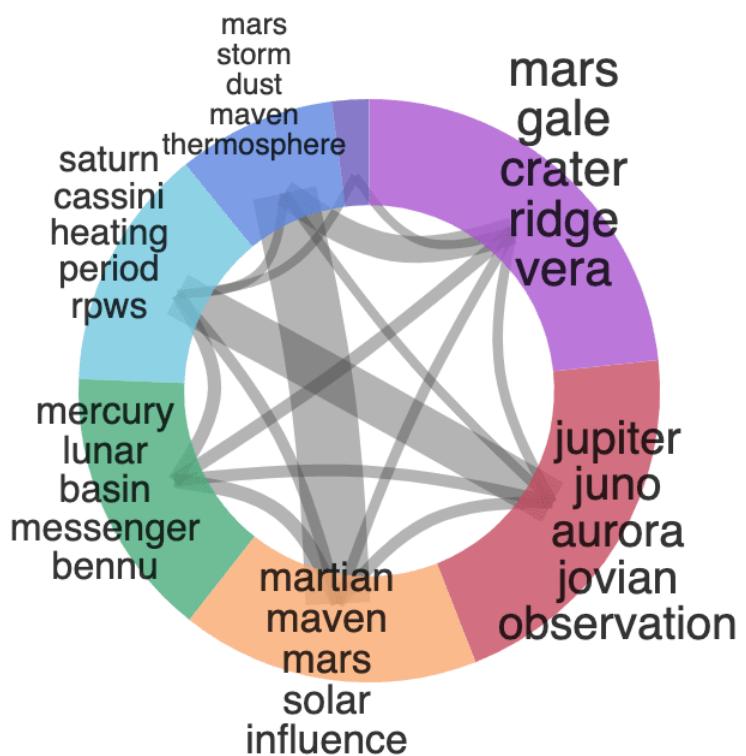
SUPPORTING FAIR GOALS

With the publication of its "Strategy for Data Management and Computing for Groundbreaking Science 2019-2024," the NASA Science Mission Directorate (SMD) recognized the unique value of ADS as a discovery platform and its role in furthering several of NASA's goals. The ADS has for a long time included in its database non-traditional scholarly resources such as astronomical catalogs, observing proposals, and research software packages. Two important reasons for including these scholarly artifacts are the need to make them easily discoverable and citable, two essential steps in making them compliant with the FAIR data principles (Findable, Accessible, Interoperable, Reusable). In addition, the ADS has long provided access to and connections between astronomy publications, datasets, and software hosted by external archives. Achieving a similar level of integration is one of the goals of the expansion in Planetary Science and Heliophysics.

Initial efforts in promoting FAIR data access are already underway: by indexing high-level data products from the PDS small body node and mining links from open research statements in AGU journals, we have started linking articles in the ADS with the major Planetary data archives. The figure to the right shows a paper network representing the connections between papers which have links to PDS nodes (descriptive labels have been assigned to clusters of articles identified in the citation network). As more data sources are added, an increasingly large number of records in ADS will have direct links to the software and data products that were mentioned in them. By indexing these data links in the system, the ADS will allow users to search by topic while restricting results to a particular set of data archives. For example, the search:

jupiter data:PDS

returns records about Jupiter and that use data products hosted at a PDS node.



ROADMAP

While our initial efforts are already underway, a lot more work will be required over the next three years to accomplish the goals of the expansion. Here are the three main areas we will be working on.

Improving Content

As is the case for other disciplines connected to mainstream Astronomy and Astrophysics, ADS does already cover the main journals of interest to researchers in Planetary Science and Heliophysics, although with some unevenness due to its greater focus on disciplines traditionally more connected to Astronomy and effort required to obtain additional scholarly material. For instance, currently Solar Physics is better represented than Space Weather, as is Planetary Exploration compared to Astrobiology. Coverage of the gray literature is much more sparse, and we will do our best to rapidly expand our coverage of it. As an example, we will make sure that ADS properly indexes all scholarly material produced by the AGU. Thanks to our

decades-long collaboration with the society, we already have all of their refereed publications as well as their meeting abstracts. However, we do not yet have content currently indexed in ESSOAr, which is on our roadmap for ingestion in 2022.

Improving Connections

Improving the usefulness of ADS beyond Astrophysics to better serve Planetary Science and Heliophysics requires collaborative work. Following the FAIR data policies from publishers, authors of scientific papers should identify software and data products discussed in their studies, and cite them appropriately. This will allow ADS to automatically detect and link relevant articles with data and software.

To extend the linkages between literature and data, we encourage the PDS and other data archives to work with ADS to develop curatorial programs which maintain authoritative lists of research articles and the relevant data products. Sharing these links with the ADS would then allow our system to provide the connections between data and literature that have been a defining feature of the Astrophysics Data Archives for the past 25 years.

Improving Semantics

In order to improve the usefulness of ADS for scientists involved in Planetary Sciences and Heliophysics research, we plan to enhance our system so that it becomes aware of the specific terminology used by practitioners in the disciplines. In order to do this, we will rely on existing knowledge bases such as:

- Existing and developing concept schemes such as the Unified Astronomy Thesaurus and the AGU index terms
- Metadata schema such as the PDS4 data dictionary and ARMS (Astrobiology Resource Metadata Standard)
- Taxonomies of solar system objects as aggregated by the Quaero API hosted at IMCCE (Paris Observatory)
- Data models and registries maintained by the Space Physics Data Facility (SPDF) and the Virtual Solar Observatory (VSO)

If you are aware of additional resources that are used to help discovery of this content please let us know.