

ADS Overview

Alberto Accomazzi & the ADS Team

aaccomazzi@cfa.harvard.edu

@aaccomazzi

ADS Users Group Meeting - 11/28/2018



Overview

- ADS's mission, functionality and focus
- Recommendations from last Users Group meeting
- Status and highlights for FY18
- Program for FY19-FY20
- Staffing and Management

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 Functionality and Focus

- Editorial policies reflect community views
 - Making decisions daily on content inclusion, refereed status
 - Indexing of non-traditional content (catalogs, proposals, software)
- Features, Services based on community needs
 - Full-text search essential for maintenance of bibliographies, analytics
 - SIMBAD and NED objects, ORCID integration, metrics, visualizations
- Comprehensiveness, timeliness, accuracy, focus
 - The only literature system where *all* of Astrophysics is represented
 - 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
 - Supports wider NASA programs and goals: Planetary Sciences, Heliophysics, mission planning, instrument building, program evaluation

ADSUG Recommendations (1/4)

http://ads.harvard.edu/adsug/2017b/ADSUG_Report_Jan2018.pdf

Platform Transition:

ADSUG approves of this aggressive timeline and expects to see significant progress toward Classic deprecation before the next ADSUG meeting.

It is reasonable to expect complaints by the community upon transition, but the reality is that ADS does not have resources to continue to support two interfaces, nor should they do so.

ADSUG recommends that ADS move forward without entertaining calls from the community to extend maintenance of ADS Classic for any longer period.

(covered in next session)

ADSUG Recommendations (2/4)

Staffing:

With a growing team, the management load is significant, and we would recommend the team look at structures that distribute this more widely. The group felt attention should be paid to the medium to long term transition of project leadership, which has been stable for many years.

To attract a broad field of excellent candidates, the project scientist position should be advertised with a substantial component of guaranteed research time. To accomplish this, the team should begin exploring funding options with NASA and, if necessary, SAO and CfA

(covered in this presentation)

ADSUG Recommendations (3/4)

Planetary:

ADSUG continues to believe that the ADS services would be of tremendous value to the entire PS community. [...] It is already the case that a large number of Planetary Scientists use ADS regularly even though the current holdings are not complete. [...]

We encourage the ADS project to engage the agency and the PDS nodes in discussions of funding opportunities and coordination between the community and the Divisions.

As a first step, ADSUG recommends that ADS begin with a focused effort to improve coverage of exoplanets and related topics, ensuring complete coverage of exoplanets within the core content and services.

(covered in next presentation)

ADSUG Recommendations (4/4)

Architecture:

ADSUG recognizes that technical documentation, including code, APIs, and high-level architectural documentation, is a valuable project outcome. ADSUG endorses committing resources for such documentation. These products will be valuable to peers in the open access scholarly publishing community, and ADS may find it useful for onboarding of technical staff in the future

(covered in Transition/Development/Architecture sessions)

External Funding and Collaborations:

ADSUG encourages ADS to continue to pursue opportunistic collaborations with peers, and to further develop external funding for those projects as appropriate.

(covered in next talk, Content/Curation and ADS2025 sessions) 9

Project Status and Highlights for FY19/20

- Status of transition to new platform (FY2018)
 - Development status
 - System Updates
 - Operations
- Highlights for FY19/20
 - Platform transition
 - Backoffice processing and curation
- Staffing and Management
 - Updated staffing roles
 - New Hires
 - Awards

Move to new System - Current status

- Enhanced Bumblebee to achieve **feature parity** with ADS Classic
(added most features to search interface and corresponding microservices)
- Achieved **content parity** with ADS Classic
(daily arXiv updates and weekly astronomy content available in new ADS)
- Worked hard to provide a **more reliable user experience**
(displaying search progress, feedback to user when errors occur)
- Increasing **system capacity** to match expected load from ADS Classic users
(major effort in back-end architecture to scale up by factor of 2 when needed)
- **Manage transition** plan and expectations
(preparing for questions / complaints / unforeseen problems)

System Updates

- Feature Parity
 - SIMBAD and NED object indexing, search
 - Native exports and link-out services (removes dependency on classic)
 - Coauthor-affiliation reports
 - User clicks now merged with ADS Classic usage
 - Added personalization options (custom formats, soon search history)
 - ADS Classic Query translator released in May 2018 as a “bridge” to new interface
- Other
 - Created FAQ and Quickstart guides
 - Scheduled blog posts to help with UI transition
 - Updated Word Cloud visualization
 - Implementing library set operations
 - Updated functionality to work with ORCID API v2.0
 - Updating paper claiming to work efficiently with large profiles, allowing bulk add / delete

Operations

- Content Parity: daily updates from arXiv and weekly updates from publishers now available in new platform within hours of ingest, soon will appear in new system ahead of Classic
- New API used regularly for text mining (e.g. CDS, NRAO, STScI, ESO)
- New architecture in Amazon Web Services based on Kubernetes (vendor independent open source cloud computing platform)
- Upgrades to microservice infrastructure leading to increased reliability
- Main ADS Classic node has 99% uptime, 100% including 11 mirrors
- ADS Classic continues to be in “maintenance mode,” no additional resources being put in its infrastructure going forward

Highlights for FY19/20 (1/2)

Overall Theme: Complete system transition, update back-end components

- Move to new system
 - Deprecate ADS Classic in Jan 2019, EOL in Summer 2019 ([ongoing](#))
 - Document features and operations in literature, tech blogs ([ongoing](#))
 - Provide transition path to new API for ADS classic crawlers, applications ([ongoing](#))
 - Make libraries, analytics and visualization services scalable for power users
- Data Curation and Indexing
 - Improve coverage of exoplanet literature, data, within astrophysics journals/archives ([ongoing](#))
 - Improve indexing and replication speed for daily/weekly ingests ([ongoing](#))
 - Enrich records with normalized affiliations, keywords, ORCIDs ([ongoing](#))
- User Interface and Personalizations
 - Update the myADS notification system to use new search engine and user accounts ([ongoing](#))
 - Enable collaborative curation of ADS Libraries
 - Update UI for mobile apps, crawlers, web applications and widget embedding

Highlights for FY19/20 (2/2)

- Ingest and Curation System
 - 25yo data ingest system developed relies on an obsolete technology stack
 - Manages ADS Classic data holdings, consisting of just a bunch of flat files on disk
 - Scope of work: recode / replace 200K lines of (mostly) PERL code
- Citation Resolver
 - 20yo machinery which is used to extract, parse, and identify citations ([ongoing](#))
 - Implements “secret sauce” and heuristics built over the years to optimize data extraction
 - Scope of work: replace 50K lines of python code relying on ADS Classic ([ongoing](#))
- ADS Article Archive
 - 25yo system used to process and visualize digitized scans of historical literature
 - Current holdings consist of 670k articles, 5M scanned pages, 2TB
 - Scope of work: replace 50K lines of C/PERL, decouple from ADS Classic, build into UI

Highlights for FY19/20 (2/2)

- Ingest and Curation System
 - 25yo data ingest system developed relies on an obsolete technology stack
 - Manages ADS Classic data holdings, consisting of just a bunch of flat files on disk
 - Scope of work: recode / replace 200K lines of (mostly) PERL code
- Citation Resolver
 - 20yo machinery which is used to extract, parse, and identify citations (ongoing)
 - Implements “secret sauce” and heuristics built over the years to optimize data extraction
 - Scope of work: replace 50K lines of python code relying on ADS Classic (ongoing)
- ADS Article Archive
 - 25yo system used to process and visualize digitized scans of historical literature
 - Current holdings consist of 670k articles, 5M scanned pages, 2TB
 - Scope of work: replace 50K lines of C/PERL, decouple from ADS Classic, build into UI

Until we replace these components, we are still dependent on ADS Classic

Staff Changes and Planning

- Suffered one resignation (DevOps engineer) in May
- Used NASA budget underrun and secured additional funding guarantees to allow for an expanded team
- Identified lack of redundancy and skills in team, leading to two additional positions (Search & Text Mining, UI/UX)
- Successfully filled two positions (DevOps, Search & Text Mining), still looking for UI/UX developer
- Promoted two team members to support PI in management of Software Development effort, as well as Content, Curation and Collaborations
- Recruiting for Project Scientist position still in early stages, plan is to engage NASA and CfA in coming year; still waiting to hear/see what ADS will look like in the future wrt expansion

ADS Staff and Roles

- Alberto Accomazzi, PI & Program Manager
- Michael J. Kurtz, Project Scientist
- Edwin A. Henneken, Content, Curation and Collaborations Lead
- Sergi Blanco-Cuaresma, Software Development and Cloud Operations Lead
- Carolyn S. Grant, Senior Curator
- Roman Chyla, System Architect and Senior Developer
- Donna M. Thompson, Data Curation Librarian
- Steve McDonald, Pipeline Development and Operations
- Tim Hostetler, User Interface and Front-end Developer
- Golnaz Shapurian, Senior Developer, Applications & Services
- Matthew Templeton, Data Ingest and Curation
- Kelly Lockhart, Back-End Development and Community Outreach
- Nathan Rapport, System Operations, Cloud Computing and Search Support
- Kris Bukovi, Text Mining and Search Support
- [TBD], User Experience and User Interface Development

JASIST Paper of the Year Award

The Journal of the Association for Information Science & Technology (JASIST), the leading journal in its field, has selected the paper “Measuring metrics - a 40-year longitudinal cross-validation of citations, downloads, and peer review in astrophysics” by ADS team members Mike Kurtz and Edwin Henneken for its best paper of 2018 award.



“The paper was selected for its detailed examination of academic success, its longitudinal analysis based on the records of 922 scholars in astrophysics, and its well-reasoned discussion and analysis. The paper is highly readable and adds to the literature on sociometrics and altmetrics, as well as to the understanding of academic success measures.”

<https://www.asist.org/news/kurtz-henneken-win-best-jasist-paper-award/>

2018 PAM Division Award

The Physics-Astronomy-Mathematics (PAM) Division of the Special Library Association gave its 2018 Division Award to the ADS project. The award is given for an “outstanding contribution to the literature of physics, astronomy, or mathematics or to honor work that improves the exchange of information in one or more of these areas.”



“ADS is not only an outstanding provider of information but an integral and involved part of the physics and astronomy communities and those of us in these communities are richer – and spoiled – due their tireless efforts. ADS has earned, through hard work and attention to their user base, the PAM Division Award.”

<http://pam.sla1.org/report-of-the-awards-presentation/>