Platform Transition: Next Steps

Alberto Accomazzi & the ADS Team aaccomazzi@cfa.harvard.edu @aaccomazzi

ADS Users Group Meeting - 11/3/207







Completing the Transition

- Getting rid of the ADS Classic search engine is the first, big step
 - Migrate applications (API)
 - Allow crawlers to index new pages
- Other major components that require reimplementation:
 - The ingest and curation system (code and workflow)
 - The reference resolver (building the citations database)
 - The notification service (myADS personalized alerts)
 - The article archive (serving content digitized by ADS)
- Until all these components are replaced, dependencies on the ADS Classic still exist (even if just under the hood)

Ingest and Curation

- Scope of work:
 - Revamp data ingest system developed in the early days of the project, which relies on an obsolete technology stack
 - Manage current data holdings, consisting of just a bunch of files on disk
 - Recode / replace 200K lines of (mostly) PERL code supporting data harvesting, normalization, validation, and ingest in ADS classic
- Goals:
 - Use modern technology stack for harvesting, curation workflows, pipelines; ideally do this by re-using an existing DL framework
 - Update metadata model to accommodate expanded content types (data and software)
 - Make ingest more, not less, efficient than current ADS classic

Reference Resolver

- Machinery which is used to extract, parse, and identify references, build the citation database
 - Existing code consists of 50K lines of mostly python code, fully reliant on ADS classic backoffice data and original metadata model (bibcode)
 - Thanks to the work being done on the new platform, we now have a better metadata store to use for reference validation
 - Still, a lot of the "secret sauce" is found in data extraction and parsing heuristics, which needs a major rewrite
- Goals
 - Collaborate with other projects, use existing reference extraction frameworks to get a jump start on extraction and parsing rewrite
 - Replace heuristics with ML approaches, train on ADS core journals

Notifications and Recommendations

- ADS provides notifications via the • myADS service
- Subscribers receive daily or weekly \bullet notifications based on their profile
- Stay up to date on topics, people, citations, trending articles
- Currently used by 16k accounts total, \bullet 6k astro, 4.1k physics, 5.3k weekly arXiv, 5.2k daily arXiv
- Will need to be updated to use new \bullet search engine, recommender system

mvADS Personal Notification Service Alberto Accomazzi Tue Oct 31 08:25:11 2017 daily arXiv e-prints database



* ADS - Recent Pape



2012PASP, 124, 212S; Sada,+; Extrasolar Planet Transits

vorite Authors - Recent Pape

Astrometric and Photometric Characteristics of Commercial Scanners in their Application for the Scientific Purpose 2013PASP. 125. 306F: Foreman-Mackey,+: emcee: The MCMC Hammer

Observed at Kitt Peak National Observatory

Article Archive

- ADS provides open access archival copies of all journal and conference proceedings articles in astronomy
- Current holdings consist of 663k articles, 4.9M scanned pages, 2TB
- While our digitization program is essentially over, we occasionally add some unique material (e.g. ICRC)
- Access to this content is also based on ADS Classic
- Will need to find new framework to move this content to

70L. I.	CAMBRIDGE, NOVEMBER 2, 1849.	NO. 1.
	No. 1.	
AST	RONOMICAL JOUR	NAL.
	THE	
	HAVERFORD COLLEGE HAVERFORD, PA	
	LIBRARY OF	

The enthusiasm of astronomers and the liberality of friends of science in America have enabled me to commence the Arranowawara. Joursan, with the full conviction that it will be permanently supported. Of its importance, — its necessity, indeed, — for the properiod development of astronomy in our contrary, there can be but one opinion. Astronomy has already reached a stage of development in America, which entities it to claim a higher position than has yet been accorded it, and which requires a larger roope for its future growth. The influence which a purely scientific journal, devoted exclusively to astronomy and its kindred desartments of inonity, may exert upon the future process of the science is very reart: and it is there-

fore, with difficience, but without hesitation, that I begin the work. Such a work ought to support the dignity of a pure science, striving for the extension of the realm of human intellect; it should formish the means of publication and prompt dissemination of discoveries and researches; and should promote harmony among astronomers, laboring for a common end, — while it furfishes an opportunity for the magly expression of differences of opinion.

It will be distinctly understood that the publication of statements or opinions implies no indersement of them by the Editor. No communication will be published without the name of the author; and I desire to be held answerable for the accuracy of such articles or researches only, as may be published with my name or initial.

In the earnest hope that the establishment of the Arranovourca. Joursat. may be hereafter referred to, as an era for astronomy in America, I commend it to the sympathy and cooperation of the lovers and votaties of science. EEX. APTIONP GOULD 74.

DEVELOPMENT OF THE PERTURBATIVE FUNCTION OF PLANETARY MOTION.*

BY BENJAMIN PEIRCE, LL. D., PERKINS PROPERSOR OF ANTRONOM AND NATHEMATICS IN HARVARD UNIVERSIT

in.

B

.6

Tux following development of the perturbative function has powers of the elements. It will be observed that, in the higher been made with the grantest care, and has been subject to powers the development differs maturally from others which such tests and checks as would seen to insure in perfect ashave been published.

This important paper was communicated in June last to Lieutenant Charles Henry Davis, Superintendent of the Manical Almanac, who had
commenced the publication of it in an independent form, for the use of his office; but who has, at the request of the Editor, had the goodness
communicate it to him, as forming as fitting commencement for the Atronomical Journal.

© American Astronomical Society • Provided by the NASA Astrophysics Data System

Build out Plan

- Keep track of user searches, enable saving searches and push notifications based on them
- Leverage on personalizations, ADS libraries to provide recommendations
- Leverage on ORCID to provide claim notifications
- Enable sharing of collections (libraries) to support collaborations
- Integrate with third-party manuscript editing systems

adsbeta 🗿	🗩 Feedback 🛛 🔞 ORCID - 😤 Learn - 🐁 Acco
My Home Page	ADS Libraries Customize Settings
ADD A LIBRARY	ADS edit description edit
VIEW ALL BRARIES LIBRARY HELP	ADS-related papers Date Created: Date Last Modified: Owner: 30 May 31 2016, 5:11pm Oct 27 2017, 11:31am aaccomazzi Image: View editing options (You have owner privileges) Output Output <t< td=""></t<>
	 Wew Library A Manage Access Chatton Helper
	Sort: Date desc \$
	1. 2017arXV170505840K 2017/05 cited: 1 Knowledge discovery through text-based similarity searches for astronomy literature Kerzendorf, W. E.
	2. 2014AAS22325525A 2014/01 I I I I I I I I I I I I I I I I I I I
	3. 2014AAS22325503A 2014/01 📄 📰 🗮 🗶 🗙 Introducing ADS 2.0 Accomazzi, Alberto; Kurtz, M. J.; Henneken, E. A. <i>and 5 more</i>
	 2013ASPC.4757M 2013/10