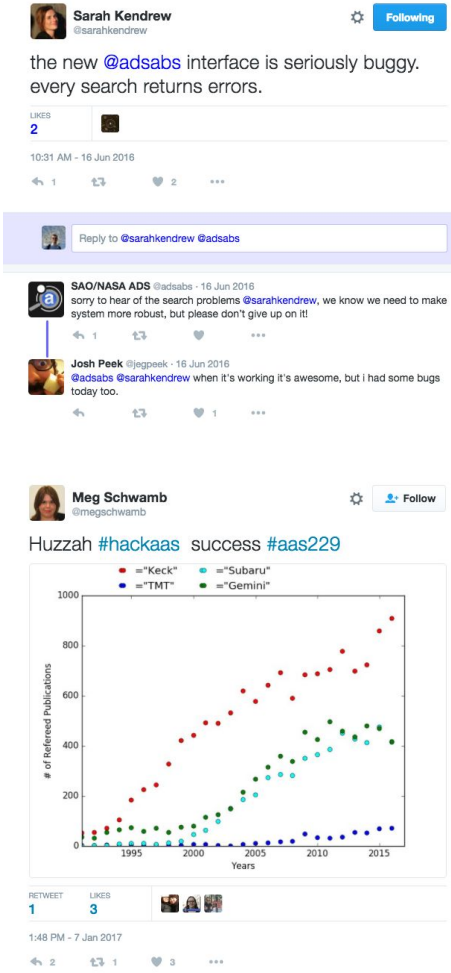
Challenges, Strengths, Threats & Opportunities

Alberto Accomazzi

ADS Users Group Meeting
January 17-18, 2017

Challenges - Technical

- Keeping ADS updated and Classic running during transition -- this limits our ability to make sweeping changes to our system, free up staff time
- Making cloud computing architecture and services robust -- this means investing time and effort in understanding AWS & microservices in-depth
- Scaling services up by one order of magnitude each year for the next two years -- we need to be fully prepared for the day the majority of users transition
- Tackling new initiatives and projects while infrastructure is in transition -- we cannot afford to be idle while the world of scholarly publishing changes at such a rapid pace



Challenges - Staff hiring and Retention

Skills and nature of jobs very marketable

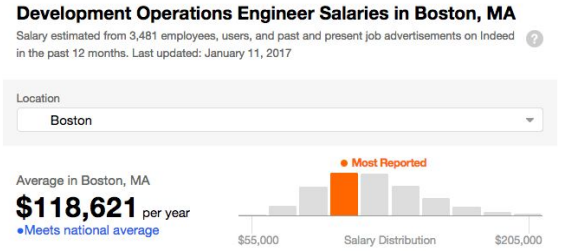
- Competing with IT industry which offers better pay
- Junior people require greater oversight, more likely to go
- Astronomy-oriented applicants need background, training
- Sweet-spot: astro-coder, data science person?

Hiring process slow and cumbersome

- SAO hires based on academic timescales, government regulations and requirements
- Very little flexibility on salary offering, qualifications

Effect

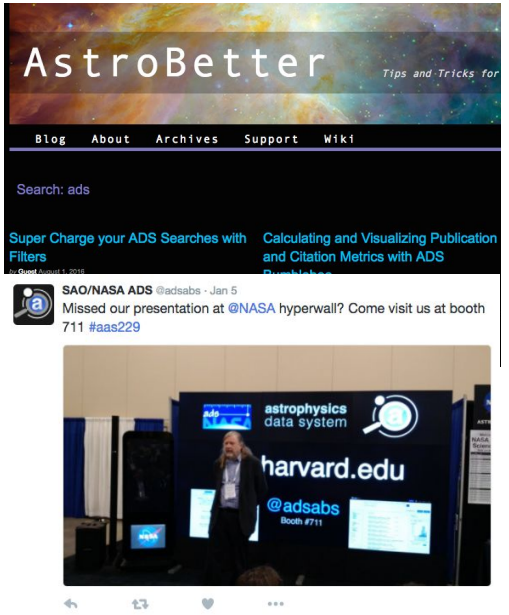
- Staff resignations costly -- long period of time required between hires
- Retaining talent difficult -- career path not always available for new and existing members
- Vacancy fatigue -- every year at least one resignation for last 4 years, hard to make progress



	ResDate	Vacancy
Web dev	8/13	6m
Full stack	7/14	6m
Dev Ops	8/15	7m
Back end	7/16	5m*
Front end	12/16	4m*

Challenges - Communication

- Spreading the news -- most ADS users still don't know there is a new interface
- Meeting expectations -- while we want people to use Bumblebee, we don't want to oversell it or push it too hard before it's complete / solid
- Managing user suggestions -- not always easy to prioritize feedback based on tweets, but we always listen & try to report back
- Finding the right venue -- social media popular with younger crowd, but this is only a fraction of our user base



Strengths - Community Support

I have just discovered your "Explore" button in your adsbeta interface, and I was blown away by the usefulness of the plots! I must congratulate with you for conceiving and providing those features.

I particularly like the Explore/ResultGraph feature, to quickly identify trendy papers. The presentation of your plots is really clear. I appreciated the ability to quickly change between the Read-Time, Citation-Time and Read-Citation views, the possibility of changing between linear and log axis, and the combination of both colors and disk sizes, to visualize four quantities at the same time.

Keep up with this fantastic work, which further expands the invaluable contribution that ADS has been providing to our research for many years!

My only suggestion is to better advertise these feature: I keep using ADS multiple times every day, but I discovered the features almost by chance...

-- M.C., UK, 6 Jan 2017

Strengths - Collaborations

Archives

- All NASA archives use ADS for data discovery, enhance ADS by sharing bibliographies, data links
- ADS seen as gateway to having content citable

CDS

- ADS hosts mirror sites for SIMBAD, Vizier
- Indexing of SIMBAD objects, Vizier metadata in ADS search engine
- Possible use of shared technology for text extraction from manuscripts

Publishers, Societies

- Excellent relationship with most publishers
- Strong partnership with arXiv, community projects (e.g. ASCL)
- Leading projects with AAS (Software citation, Astrolabe, Astronomy Rewind, UAT)

Librarians

- Working with CfA library staff on developing tools in support of bibliography curation
- Incorporating historical content (e.g. Harvard Plates collection), PhD Theses (via Zenodo)

Threats

Alternative scholarly search engines, social research platforms

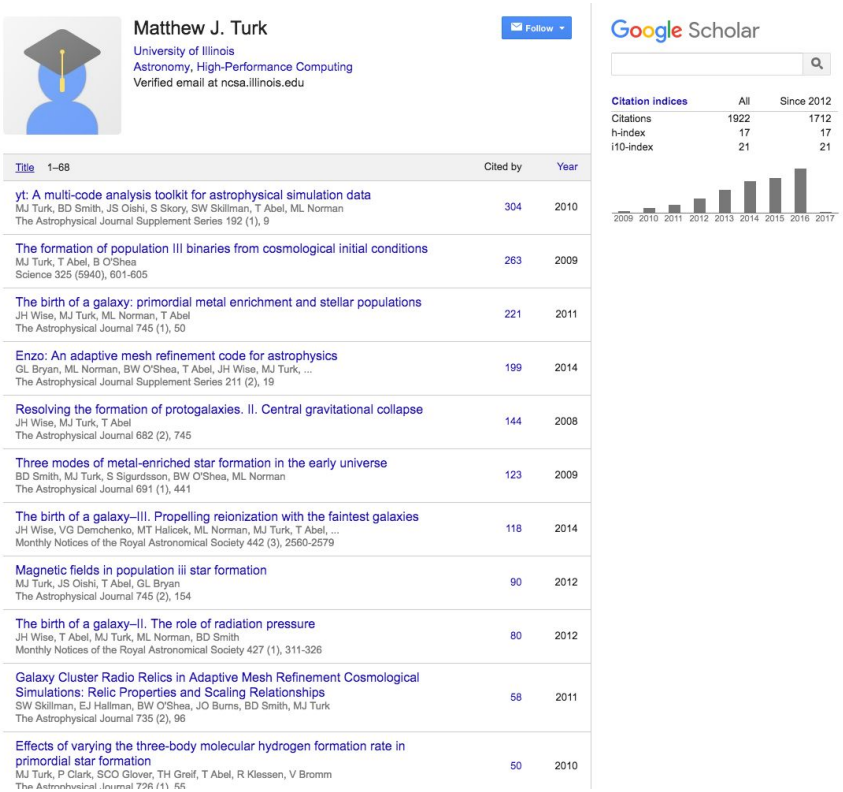
- Google Scholar not as good, but shows greater citation count…
- Mendeley, Research Gate, and other reference management systems feature user profiles, sharing
- New systems such as Semantic Scholar offer novel ranking features

Rapid changes in publishing environment may require changes in ADS

- Traditional paper model is dying, with new “articles” being updated over time (e.g. arXiv, soon AAS)
- More varied content now being cited in papers: software, data, blog entries, etc; what to index?
- Annotations are being proposed for peer-review, commenting, crowdsourced curation of content
- What is the paper of the future and how should it be indexed in ADS?

Major changes from funding agencies

- OA initiatives and mandates from US & EU agencies are creating new infrastructure, disruption
- Funding climate stable now, but single source of funding could prove to be a weakness long-term



Matthew J. Turk

Follow

University of Illinois
Astronomy, High-Performance Computing
Verified email at ncsa.illinois.edu

Title 1–68	Cited by	Year
yt: A multi-code analysis toolkit for astrophysical simulation data MJ Turk, BD Smith, JS Oishi, S Skory, SW Skillman, T Abel, ML Norman The Astrophysical Journal Supplement Series 192 (1), 9	304	2010
The formation of population III binaries from cosmological initial conditions MJ Turk, T Abel, B O'Shea Science 325 (5940), 601-605	263	2009
The birth of a galaxy: primordial metal enrichment and stellar populations JH Wise, MJ Turk, ML Norman, T Abel The Astrophysical Journal 745 (1), 50	221	2011
Enzo: An adaptive mesh refinement code for astrophysics GL Bryan, ML Norman, BW O'Shea, T Abel, JH Wise, MJ Turk, ... The Astrophysical Journal Supplement Series 211 (2), 19	199	2014
Resolving the formation of protogalaxies. II. Central gravitational collapse JH Wise, MJ Turk, T Abel The Astrophysical Journal 682 (2), 745	144	2008
Three modes of metal-enriched star formation in the early universe BD Smith, MJ Turk, S Sigurdsson, BW O'Shea, ML Norman The Astrophysical Journal 691 (1), 441	123	2009
The birth of a galaxy–III. Propelling reionization with the faintest galaxies JH Wise, VG Demchenko, MT Halicek, ML Norman, MJ Turk, T Abel, ... Monthly Notices of the Royal Astronomical Society 442 (3), 2560-2579	118	2014
Magnetic fields in population iii star formation MJ Turk, JS Oishi, T Abel, GL Bryan The Astrophysical Journal 745 (2), 154	90	2012
The birth of a galaxy–II. The role of radiation pressure JH Wise, T Abel, MJ Turk, ML Norman, BD Smith Monthly Notices of the Royal Astronomical Society 427 (1), 311-326	80	2012
Galaxy Cluster Radio Relics in Adaptive Mesh Refinement Cosmological Simulations: Relic Properties and Scaling Relationships SW Skillman, EJ Hallman, BW O'Shea, JO Burns, BD Smith, MJ Turk The Astrophysical Journal 735 (2), 96	58	2011
Effects of varying the three-body molecular hydrogen formation rate in primordial star formation MJ Turk, P Clark, SCO Glover, TH Greif, T Abel, R Klessen, V Bromm The Astrophysical Journal 726 (1), 55	50	2010

Google Scholar

Citation indices	All	Since 2012
Citations	1922	1712
h-index	17	17
i10-index	21	21

2009 2010 2011 2012 2013 2014 2015 2016 2017

Semantic Scholar

Search

SIGN IN

Matthew J. Turk

Share

72
Citations

Citations Per Year

Authors who most influenced Matthew J. Turk:

Judith Segal 2
Marian Petre 2
Tom Abel 2
Karl Fogel 2
Aleksandra Pawlik 2

Matthew J. Turk

Authors most influenced by Matthew J. Turk:

4 Tom Abel
3 Matthew G. Knepley
2 Wolfgang Bangerth
2 Barry Smith
2 Jed Brown

Filter Results:

Publication Year
2009 2013

Publication Type

Co-author

Key Phrase

Publication Venue

Sort by: Highly Influential Citations | Citations | Acceleration | Velocity | Recency

Learn More

Yt: a Multi-code Analysis Toolkit for Astrophysical Simulation Data

Matthew J Turk, Britton D Smith, Jeffrey S Oishi, Stephen Skory, Samuel W Skillman, Tom Abel +1 other · 2010

The analysis of complex multiphysics astrophysical simulations presents a unique and rapidly growing set of challenges: reproducibility, parallelization, and vast increases in data size and complexity chief among them. In order to meet these challenges, and in order to open up new avenues for collaboration between users of multiple simulation platforms, we… (More)

View PDF | Related Publications | More...

How to Scale a Code in the Human Dimension

Matthew J. Turk · ArXiv · 2013

As scientists' needs for computational techniques and tools grow, they cease to be supportable by software developed in isolation. In many cases, these needs are being met by communities of practice, where software is developed by domain scientists to reach pragmatic goals and satisfy distinct and enumerable scientific goals. We present techniques that have… (More)

View On ArXiv | Related Publications | More...

Opportunities - Leveraging Identity

Personalizations

- ADS already has user accounts, libraries, identity claiming, readership info
- Connecting your ID with the content you have written lets us know what you have worked on
- Connecting your ID with the content you are reading lets us know what you're interested in
- Connecting with the citation network: we know what you write, who you cite, who cites you
- Connecting with readership network: we know what you read, who reads your papers

Social

- ADS also knows who you have published with, so we can suggest connections
- ADS accounts will support sharing of libraries (private collections) with lists of collaborators
- Connecting co-authors helps disambiguate people

Recommendations: you may be interested in this content because it's something:

- Similar to what you've written
- Similar to what you've read
- Similar to what you have saved
- Your collaborators are reading, citing
- Your readers are reading, citing
- People similar to you are reading, citing

Opportunities - Leveraging Content

Topic Analysis

- Use Latent Dirichlet Allocation (LDA) to determine what topics are discussed in a paper
- Use LDA space to classify and cluster documents, provide recommendations
- Process is unsupervised and uses a generative statistical model

Text Mining

- Use text extraction techniques to find specific concepts mentioned in a paper
- Concepts could be terms from a controlled list (such as the Unified Astronomy Thesaurus)
- They can also be things such as instrument or telescope names, grant IDs, etc…

Semantic Search

- Broaden / Narrow results using related terms, subtopics
- Allow (some) removal of ambiguity: is SST the “Spitzer Science Telescope,” the “Spin-State Transition” or “Sea Surface Temperature”?

Opportunities - Expanding Coverage

Disciplinary Expansion

- Planetary Sciences
- Heliophysics
- Earth Sciences
- Geophysics / all of Physics

Data / Software / Gray Literature

- What data products belong in ADS? How should they be searchable?
- What software should be represented and what strategy should be followed to index it?
- Should ADS become a repository for any of this content? Should it find partners to work with?
- Currently indexing some collections in Zenodo (CERN), more content is out there (e.g. Figshare)
- Should we put more effort in indexing observing proposals, talks, presentations, etc?

OA Repositories and Initiatives

- Currently indexing all of arXiv, but use of system is uneven in wider community
- Some OA content in agency-sponsored repositories such as CHORUS, PMC

Opportunities - Annotations

Authoring and Reviewing

- Comments and questions shared among authors of a paper
- Referee's comments during peer review process
- New platforms available (e.g. Authorea, eLife, AGU)

Bookmarking, Tagging and Commenting

- Adding records to a collection (e.g. an ADS library)
- Associating a label a paper ("to read")
- Journal clubs (e.g.VoxCharta), blog posts, social media
- PubPeer, eLife, arXiv experimenting

Curating

- Librarians creating bibliographic databases (this article is written by a CfA researcher)
- Archivists linking to data (article mentions CXC ObsID #123)
- Curators recording claims in database (article claims redshift of Abell 1689 is 0.1828)

Strengths

- Supportive User Community
- Publisher Relationships
- Stable Funding
- Small, Agile Team
- Key Collaborations
- Uniquely Complete Content

Weaknesses

- Small User Community
- Single Funding Source
- Staff Retention / Redundancy
- Multiple Systems to Maintain

Opportunities

- Serve Related Disciplines
- Greater Personalization
- Increased Data Integration
- Micropublishing / Annotations

Threats

- Alternative Search Engines
- Alternative Social Platforms
- Changes in Publishing Paradigm
- Agency Mandates

Discussion