



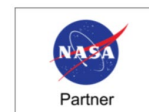
ADS Backoffice Update

The ADS Team

ADS Users Group Meeting, 20-21 Nov. 2025



CENTER FOR
ASTROPHYSICS
HARVARD & SMITHSONIAN



This work is openly licensed via [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/).



Search

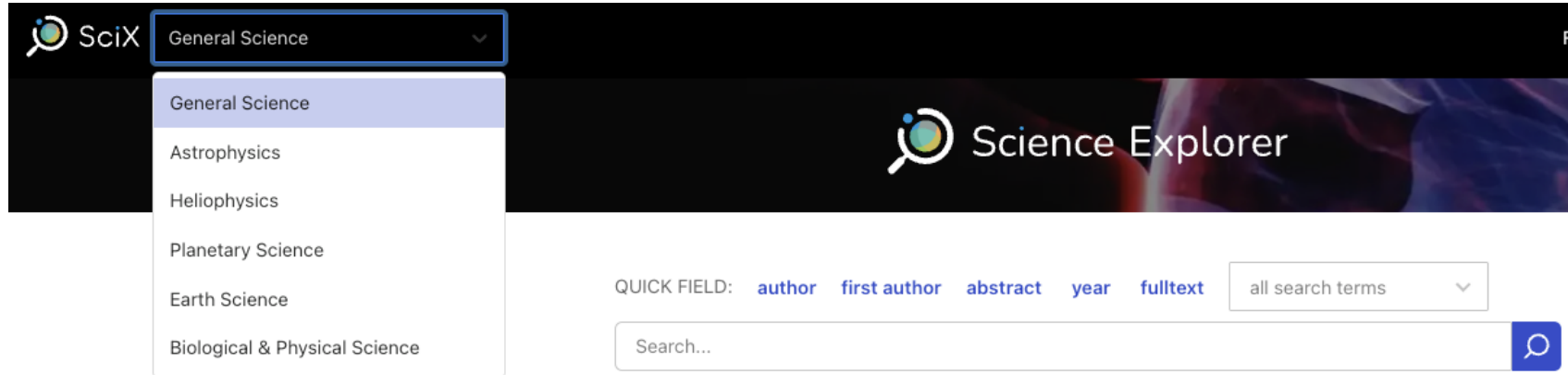
- **Partial results enabled to reduce query timeouts**
 - Allows solr to return partial results if query would otherwise fail due to timeout
 - Shows warning to users if partial results returned
- **Solr 9**
 - Upgraded from Solr 7, which was no longer receiving security updates
 - Enables future work with dense vector searches (i.e. semantic search)
- **OSC Consultation and Recommendations**
 - Solr Cloud
 - Established potential upgrade path from standalone to solr cloud
 - Identified potential complications in upgrade
 - Helped establish proof of concept implementation
 - Delivered implementation recommendations for citation cache replacement
 - Development of search evaluation framework tool





Classifier Pipeline

- The Classifier Pipeline is automating the placement of records into appropriate collections
 - **Astrophysics, Planetary Science, Heliophysics**, Earth Science, Physics, Other - Allowing multi-disciplinary classification
 - Collections allow relevancy boosts in domain specific searches
 - Utilizes Language Model classification, and will later include journal heuristics and citation graph information
 - LM-based classification is running in the Development Environment
 - Classified 3M records in the classic defined Astronomy Collection
 - Feedback is useful regarding how we display the collections, do we keep top-level Astronomy category





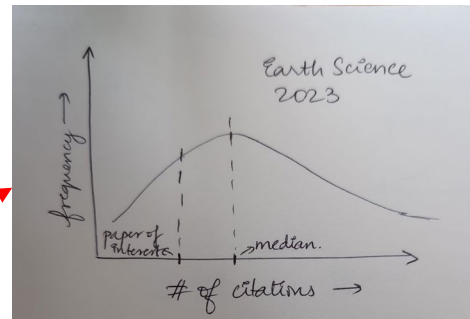
Discipline and Doctype Boosting

1/2

- **New Relevance Scores in SciX – Default ranking order for SciX, customized for each discipline selected in the drop-down menu in the UI**

- Implemented a brand new pipeline to compute boost factors
- Types of boost:
 - **[prod] Discipline-specific boost:** different collection rankings for each discipline; ensures that “seismology” search for astronomers and earth scientists will be different from each other, but more relevant to that audience
 - **[prod] Refereed boost:** peer-reviewed articles have higher weight
 - **[prod] Doctype boost:** articles have highest priority compared to other doctypes like conference abstract, etc.
 - **[WIP] Recency boost:** Exponentially decaying weight as a function of age to boost recent articles (cut off at 24 months)
 - **[WIP] Citation boost:** Computed as the percentile of a distribution of citations for the respective collection and publication year. Records in multiple collections will have some combination of different cite_boost factors
- Works remarkably well to 1st order

Replacing
“score”
computation
from ADS
classic





Discipline and Doctype Boosting

2/2

Earth Science

General Science

Astrophysics

Heliophysics

Planetary Science

Earth Science

Biological & Physical Science

abstract year fulltext

all search terms

Relevance

☐ Select All

Bulk Actions

Explore

What is a geomagnetic storm?

1

Gonzalez, W. D.; Joselyn, J. A.; Kamide, Y.; Kroehl, H. W.; Rostoker, G.; Tsurutani, B. T.; Vasyliunas, V. M.; [show details](#)

1994/04 · Journal of Geophysical Research · cited: 1840

No Highlights

Modeling production and climate-related impacts on ¹⁰Be concentration in ice cores

2

Field, Christy V.; Schmidt, Gavin A.; Koch, Dorothy; Salyk, Colette; [show details](#)

2006/08 · Journal of Geophysical Research (Atmospheres) · cited: 129

No Highlights

Bulk magnetization of the heavy rare earth titanate pyrochlores - a series of model frustrated magnets

3

Bramwell, S. T.; Field, M. N.; Harris, M. J.; Parkin, I. P.; [show details](#)

2000/01 · Journal of Physics Condensed Matter · cited: 162

Bulk **magnetization** of the heavy rare earth titanate pyrochlores - a series of model frustrated **magnets**
In these materials **magnetism** arises from the rare earth ions, while Ti is **non-magnetic**.

Author

☐ Wang, Y 5.8k >

☐ Wang, J 4.9k >

☐ Zhang, Y 4.8k >

☐ Liu, Y 4.7k >

☐ Wang, X 4.3k >

☐ Li, Y 4.1k >

☐ Zhang, X 3.8k >

☐ Zhang, J 3.6k >

☐ Wang, Z 3.5k >

☐ Li, J 3.5k >

1865

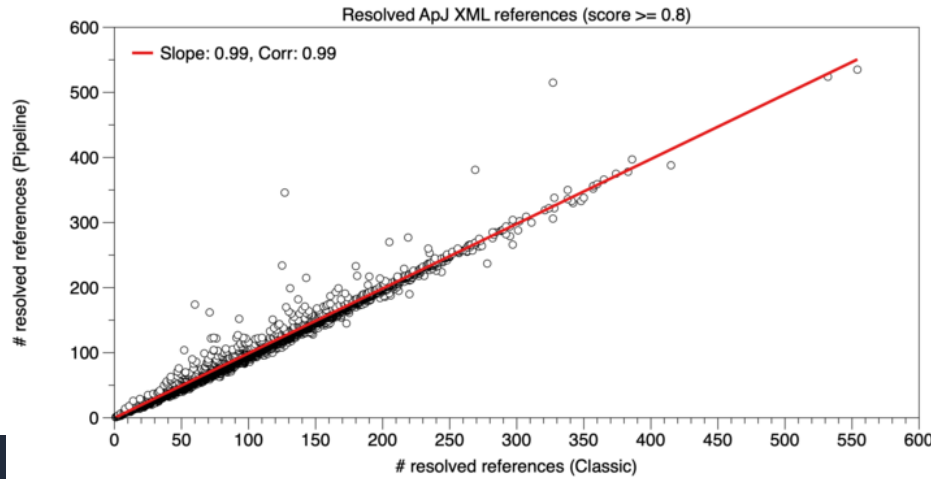
2026

5



Transition to the Reference Pipeline:

- Resolves references for citation graph and exposure through the User Interface
- Developed a transition plan to switch from the classic reference resolving process
- Remove technical debt with a more robust pipeline built on current technology
- Currently testing, documenting, and identifying and fixing issues in the new Pipeline
- Processed current arXiv, ApJ, AJ, ApJS and MNRAS holdings
- Meets and sometimes exceeds the performance of the classic process



Back to Results

Full Text Sources
ADS

Data Products

Related Materials

Abstract

Citations 195

References 33

Creates

Mentions

Co-Reads

Similar Papers

Volume Content

Graphics

Metrics

Download Citation

Papers referenced by
Spots on T Tauri stars.

View as search results

- 1 Spectroscopy and infrared photometry of southern T Tauri stars.
Appenzeller, I.; Jankovics, I.; Krautter, J.; [show details](#)
1993/08 - Astronomy and Astrophysics Supplement Series - cited: 109
- 2 Stellar Activity in Synchronized Binaries. II. A Correlation Analysis with Single Stars
Basri, Glibor; [show details](#)
1987/05 - The Astrophysical Journal - cited: 85
- 3 Long-period variations in R CrA, S CrA, T CrA and R Mon.
Bellingham, J. G.; Rossano, G. S.; [show details](#)
1980/05 - The Astronomical Journal - cited: 22
- 4 Accretion Disks around T Tauri Stars
Bertout, Claude; Basri, Glibor; Bouvier, Jerome; [show details](#)
1988/07 - The Astrophysical Journal - cited: 588
- 5 DN Tauri : a spotted T Tauri star.
Bouvier, J.; Bertout, C.; Bouchet, P.; [show details](#)
1986/04 - Astronomy and Astrophysics - cited: 45



Sitemaps & Search Engines

- **Better Search Engine Optimization (SEO) and discoverability**
 - The purpose of the sitemaps is to help search engines discover and index our records more efficiently, by proactively pointing them to the URLs for ~30M papers
 - **Search drives approximately a third of our traffic**
 - Are part of our efforts to **retire classic infrastructure**
- **For the first time, sitemaps include SciX as well as ADS**
 - Deployed to production in November
 - Fully automated sitemaps generation
 - All papers in SciX and ADS should soon become discoverable via Google searches

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8"?>
<sitemapindex xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <script id="epplocemhnm1bhjplcgkorfciiegomcon"/>
  </script>
  <sitemap>
    <loc>https://scixplorer.org/sitemap/sitemap_static.xml</loc>
    <lastmod>2025-11-12</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://scixplorer.org/sitemap/sitemap_bib_44.xml</loc>
    <lastmod>2025-11-10</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://scixplorer.org/sitemap/sitemap_bib_455.xml</loc>
    <lastmod>2025-11-12</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://scixplorer.org/sitemap/sitemap_bib_83.xml</loc>
    <lastmod>2025-11-10</lastmod>
  </sitemap>
  <sitemap>
    <loc>https://scixplorer.org/sitemap/sitemap_bib_304.xml</loc>
    <lastmod>2025-11-12</lastmod>
  </sitemap>
</sitemapindex>
```




- **Events Based Successor to Classic Backoffice**
 - Complete ground-up rewrite of backoffice scripts and pipelines
 - Removes technical debt and ensures a maintainable system for the future
- **New Harvesters**
 - 3 new harvester (APS, FTPs, SFTP)
- **SciX Identifiers (SciX IDs)**
 - Successor to bibcodes
 - Bibcodes will still be minted when applicable
 - Allows us to mint identifiers for records where the traditional bibliographic model doesn't apply
 - Will be of the form scix:XXXX-XXXX-XXXX with X being an alphanumeric character
 - Based on [Cool DOI](#) suffix spec
 - No semantic meaning
 - SciX ID generation currently being tested in existing infrastructure.
 - Full explainer blog in prep
- **Identifiers Table Prototype**
 - Matches external identifiers with internal UUIDs and assigned SciX ID
 - Will allow for easier cross-linking and record matching
- **Updates to Fulltext Data Model**
 - Fielded footnotes, captions, tables, instruments, grants, etc
 - Fielded multilingual elements (e.g. abstracts), language tagging for papers not in English
 - Parser for our most common delivery format in beta, others to follow.





API Gateway

- **The API-Gateway is the successor to our Python 2.7 API Gateway, adsws**
 - Lives in AWS and manages all connections handled by ADS/SciX website/API
 - Built from the ground up by WinterWay in Python 3.10
 - Built with modern libraries with clear upgrade paths
 - Implements stronger password requirements
 - More granular rate limiting
 - Higher performance
 - Allows for better metrics collection
- **Deployed to Production in April 2025 (With an aborted deployment in December 2024)**
 - December deployment was reverted due to a database field missing an index resulting in high DB computational load
 - This resulted in service interruptions that occurred between Christmas and New Year's Day.
 - The issue was fixed prior to deployment



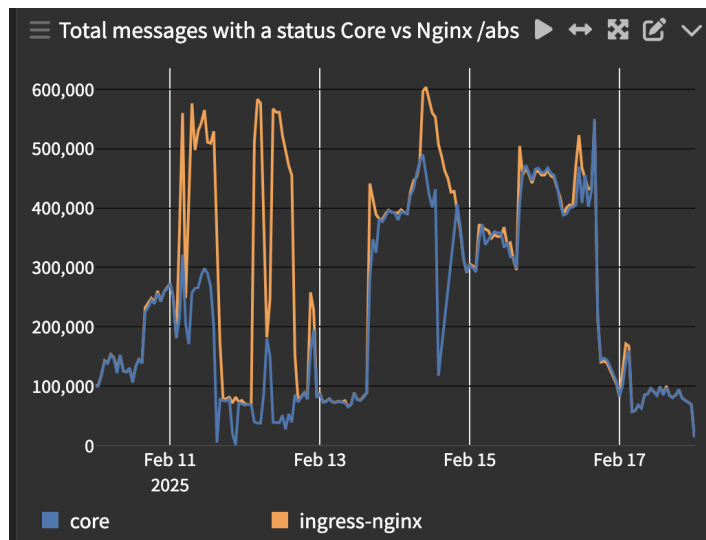
February DDoS Incident

- Beginning the week of February 10th we began to see an increase in traffic to the ADS /abs endpoint.

- Traffic came in sudden bursts: sometimes hundreds of thousands of anonymous clients per minute (our usual rate of anonymous clients is ~5000 per minute).
- These requests overwhelmed our system and caused outages for users
- Discovered and blocked 3 subnets that were responsible for the majority of the new traffic
- We believe this traffic to be related to issues other websites experienced during the period as a result of uncontrolled [AI bot scraping](#).

- Lessons Learned**

- We have completely overhauled our pod scaling and node scaling
- externally hosted status page (See the Frontend Presentation for details)
- Configured new AWS WAF rules that better monitor for bot traffic
- Permanently blocked the offending domains
- Set alerts for database connections, core resource usage and pod count so we are alerted to these situations before they become a major incident.
- Examining additional long-term protections





ADS/SciX Hardware and Infrastructure

- **Remote Backups at CDS**
 - Recommendation for off-site backups from Sidrat DevOps audit
 - CDS has provided storage and server resources
- **GPU Servers**
 - 2 nodes (8 x L40S)
 - Purchased, racked, OS installed
 - Server room UPS was upgraded to handle load (Sep 2025)
 - Need new PDUs to upgrade power to 220V
- **SciX High-memory node**
 - Additional node added to SciX Neuron Cluster
 - Purchased, racked, OS installed
 - Incorporated into SciX Neuron Kubernetes cluster
- **Automated Deployment**
 - Implementing ArgoCD to handle pipeline and microservice deployments
 - In testing on Neuron Cluster
 - Argo Workflows managed by ArgoCD for GPU workflows
 - Will be deployed once GPU PDUs are in place
 - Will free up DevOps team time from having to manually deploy images





Migration to ITS

- All ADS systems are being migrated to CfA's Information Technology Services (ITS) management from CXC's syshelp
 - Currently under ITS management:
 - ADS netapp appliance (main on-premise data storage)
 - SciX Neuron Machines
 - GPU nodes
 - ITS has begun the process of fully migrating the netapp
 - Mounted on SciX Neuron machines
 - Need to transfer from syshelp to ITS NIS tables
 - Still need to create an ITS DMZ
 - Required for serving sites from on premises servers
 - SIMBAD mirror
 - Search index replication
 - Existing development and production environment remain under syshelp management
 - 8 servers still in need of migration (3 in syshelp DMZ, 5 for current production and dev environments)



Classifier Pipeline - Metrics

- Applied the LM-based classifier to the classically defined Astronomy collection

- 3M records in the classic defined Astronomy Collection
- Curators hand validated ~4000 classified records
- Similar Precision as the Classic classification (0.61)

Astrophysics Classification Metrics

True Positive: 1029

True Negative: 1989

False Positive: 629

False Negative: 341

Precision: 0.62

Recall: 0.75

F1: 0.68

F2: 0.72

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

