

Content & Curation Team

Edwin Henneken and the ADS Team

ADS Users Group Meeting, 05-06 Dec. 2024





ADS/SciX Content & Curation Team

Carolyn
Grant

Data Ingestion
Lead

Daniel
Chivvis

Digital
Technologies
Development
Librarian

Edwin
Henneken

Content,
Curation &
Collaborations
Lead

Jenny
Koch

Digital
Technologies
Development
Librarian

Matt
Templeton

Data
Integration
Manager

???

Librarian

Tom Allen

Liaison with
Tech Team

Harry Blom

Senior
Publishing
Liaison



TOPICS

- **Some highlights from previous year**
- **Priorities from the Roadmap for 2025:**
 - Enhancing the SciX collections
 - Removing dependency on legacy components
 - Improved infrastructure and tools
- **User Support and Feedback**

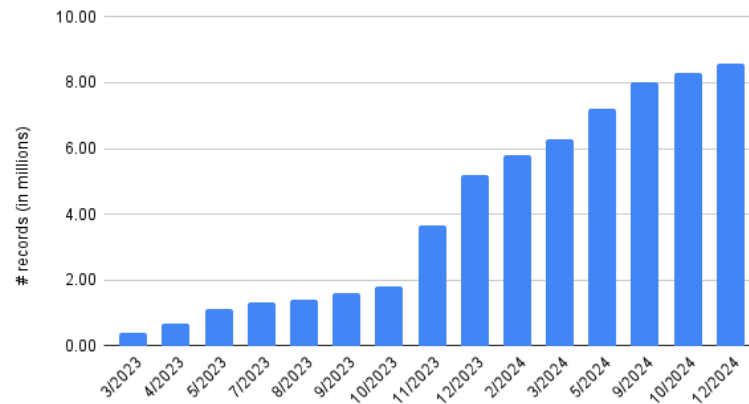


Some highlights from previous year

# citations added	20,198,612	~300k/wk!
# records added	6,303,647	~100k/wk!
# records added (Earth Science)	4,864,477	
# ES records added with full text	2,032,761	
# preprint matches added	313,640	
# data links added	34,205	

- The Earth Science collection was formalized as one of the main SciX collections
- Indexing and matching of new preprint sources: EarthArXiv, ESSOAr
- Reprocessing of ISI reference data resulted in ~500k additional citations
- Agreements with publishers. Examples: IEEE and ACS
- Records for IPAC datasets
- Deposit of scanned AAS content in Portico (preservation)
- Three part blog on data linking and data citations in SciX
- Implementation of BibVO (Bibliographic Interfaces in the Virtual Observatory - protocol for registering/harvesting data links in the VO Registry)

Earth Science Collection



Notable changes in the Curation Team

- Donna retired on September 6 after 25 years on the team
- Daniel got permanent position
- Matt assumed the role of Data Integration Manager
- Tom Allen will focus his efforts on curation/ingestion operations
- We will hire one new team member
 - Librarian (supervisor: Carolyn)



Enhancing the SciX Collections

- **Earth Science (focus)**
 - Define "Core" set of Earth Science journals and be complete for these
 - Lay down content-based requirements for removing "beta" label of Earth Science collection
 - "Low-hanging" grey literature
 - Bibliographies: NASA Earth Science DAACs, USGS
 - Citation analysis of unmatched references from Earth Science journals
 - Support of AGU copy-paste export format (together with Tech Team)
- **Heliophysics & Planetary Science**
 - Link to data products
 - Improve coverage of Heliophysics grey literature (books, conferences)
 - Better coverage of cited literature
- **Astrophysics**
 - Business as usual
 - Data links from Data Availability Statements
 - Contribute to finalizing of the UAT model and pipeline
 - Add missing IAUC/CBET
 - Ingest highly cited papers on individual basis
 - Better coverage of astrobiology, astrochemistry and laboratory astrophysics
 - More eBooks



Priorities and Roadmap for 2025

2/2

Removing dependency on legacy components

Cross-team Effort with Tech Team

- Preprint matching ("docmatching") using new architecture
- Reference matching with the new reference pipeline
- Replace Classic Scan Viewing interface with Scan Explorer (**Astrophysics**)
- Reference extraction from arXiv preprints using new pipeline
- Replace Classic classifier with new (enhanced) classifier across all SciX disciplines



Priorities and Roadmap for 2025

Improved infrastructure and tools

- Implement workflows to remove broken links
- Begin work on curation dashboard
- Add publisher data to journals database / public display of journals database content
- Mine license information from full text documents



User Support and Feedback

User Support

Totals by topic/tag

AWS

1.2%

Duplicate

1.7%

Question

1.9%

Curation/correction

3.4%

System messages

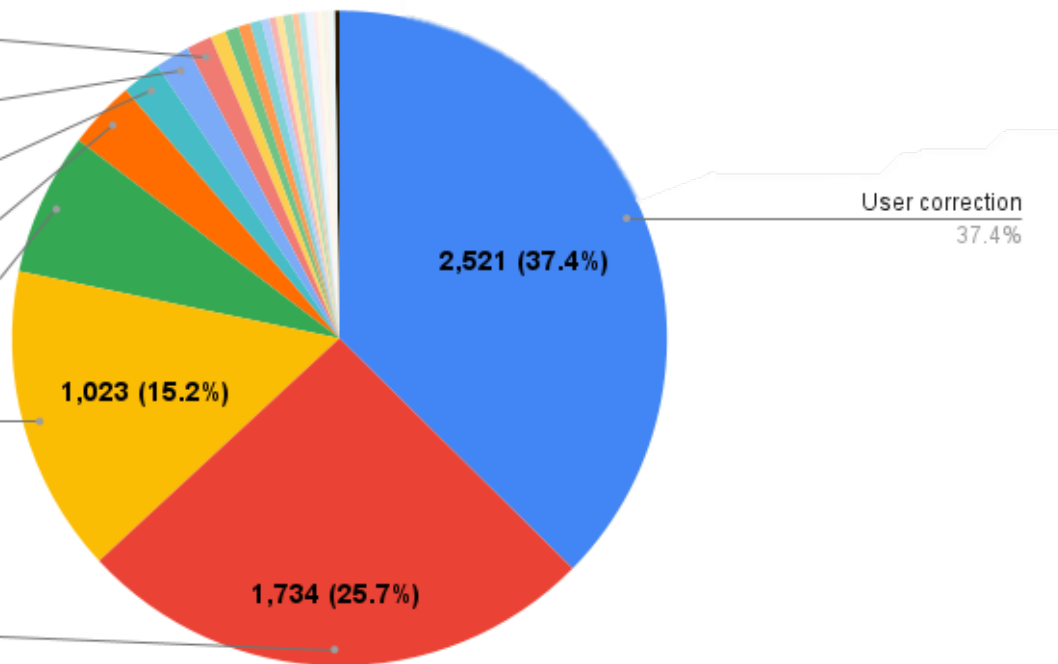
7.0%

User association

15.2%

User submission

25.7%



There were 43,800 references submitted in the last year, compared to 12,148 the year before