astrophysics data system

Development: 2026 Roadmap

Kelly Lockhart and the ADS Team

ADS Users Group Meeting, 20-21 November 2025

SciX

CENTER FOR ASTROPHYSICS HARVARD & SMITHSONIAN

SMITHSONIAN INSTITUTION

NASA Partner

This work is openly licensed via CC BY 4.0.

2026 Development Strategic Planning 1/2

Within the development team:

- Brainstormed about current and planned tasks on their plates
- Grouped together similar tasks into projects
- Filled in gaps, including work needed from other team members

Cross-team:

- Brainstormed with the curation team and project scientists about their existing projects that require development work (in progress and planned)

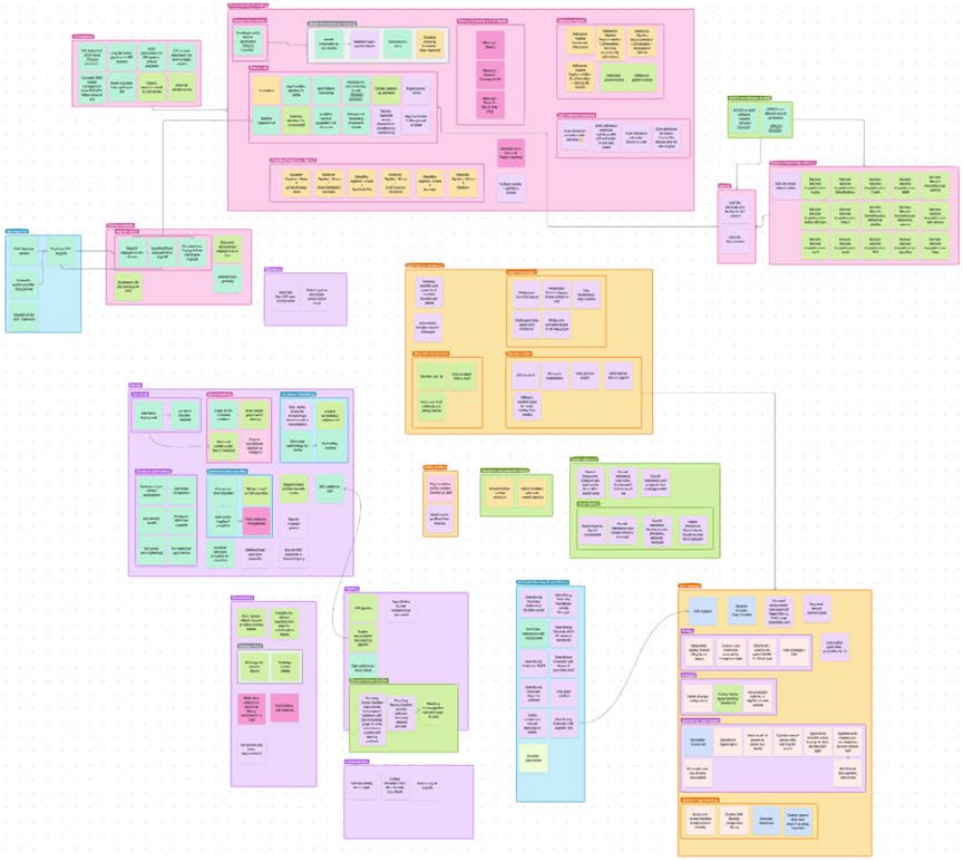
Prioritization:

- Place each project on a prioritization matrix (impact vs effort) to narrow scope

Planning

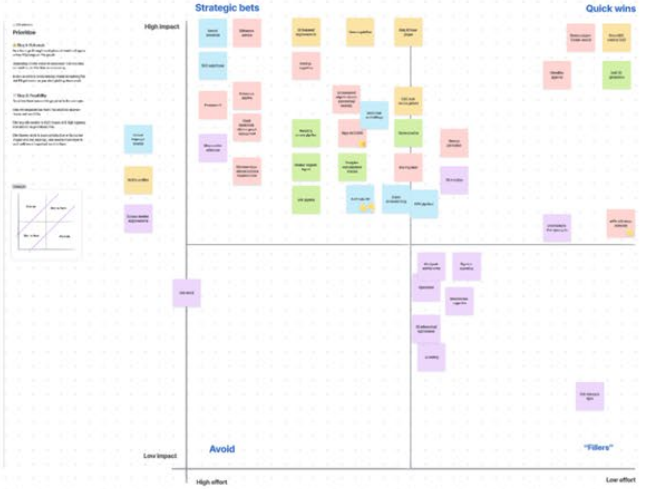
- Place each selected project on a planning matrix (impact vs urgency) to determine timing
- Grouped by theme, color-coded

2026 Development Strategic Planning



2026 Themes:

1. retiring and modernizing infrastructure
2. enhancing the user experience
3. expanding and enriching content
4. advancing AI & data driven capabilities
5. strengthening core operations



Scenario 1 (ADS+SciX): 2026 Priorities

- Retiring and modernizing infrastructure
- Enhancing the user experience
- Expanding and enriching content
- Advancing AI & data-driving capabilities
- Strengthening core operations and reliability

Retiring and modernizing infrastructure 1/3

The majority of the projects in this theme revolve around replacing 30-year-old code, much in Perl, and outdated infrastructure (no databases!). The goal is to finish porting this functionality into more modern languages and architectures before the current Classic maintainers retire.

- **Scan Explorer in production (Nov 2025):** This will allow us to fully decommission the old Classic search, years after the Classic interface was retired
- **Continue work on Honeycomb, our new ingest pipeline architecture (multi-year project), including new harvesters and augment pipelines**
- **Reference resolver pipeline into production**
- **arXiv reference extraction in production (Q1 2026)**
- **Classifier pipeline phase 1 in production (Dec 2025), begin work on phase 2**
    - Phase 1 will enable heliophysics and planetary science collections

Retiring and modernizing infrastructure 2/3

Other projects in this theme resolve other sorts of technical debt.

- **Finish internal server migration from CXC's syshelp to ITS (Q3 2026):** This is mandated by the CfA director's office
- **Automated deployments for pipelines and microservices (Q2 2026):** A custom-built deployment system is no longer functional and deployments are currently done manually, which takes valuable developer time

These projects resolve technical debt related to our expanding corpus.

- **Take Solr's citation graph out of memory (multi-month project for single search dev)**: As the size of our corpus grows, so too do the number of edges in the citation graph. The current implementation is reaching its technical limits
- **Generate SciX IDs for all records (Jan 2026)**: This will allow us to ingest records that we can't generate a bibcode for
- **Remove bibcode dependencies from microservices (ongoing)**: This is needed to support records that have only SciX IDs, no bibcodes

Enhancing the user experience

Now that SciX is out of beta, the projects within this theme focus on increasing user growth and adding features that are especially useful to brand-new users.

- **SciX post launch growth & user acquisition:** Improve SciX SEO by adding product pages (e.g. a UAT landing page) and improve journey for new users
- **Move ADS users to SciX**
  - Soft launch: add “see this page in SciX” links in ADS (Dec 2025), create ADS → SciX help pages
  - Hard launch with redirects (Scenario 2: Q1 2026, scenario 1: EOY 2026)
- **New features for the SciX user interface:** Add exploratory interfaces to help new users understand our corpus, add more autocompletes to make searching easier
- **Author profiles R&D**: Begin work on creating author profiles to help with disambiguation, leveraging ORCID and other data
- **SciX help pages**: Focused around making these more navigable for new users, we need to streamline and overhaul, and possibly implement a chat assistant

Expanding and enriching content

This theme is focused on projects most relevant to non-astronomy content and users.

- **Ingest new SciX ID records that are incompatible with bibcodes**: Our existing modern ingest pipeline will need to be retrofitted to allow ingest of these records.
- **Maintain discipline completeness metrics:** This is a joint project with the curation team, to hand-curate bibliographies for a subset of authors to check completeness of their publications in SciX vs competitor systems.
- **Improve relevance sort:** A larger, more interdisciplinary corpus requires more fine-tuned information retrieval. Current projects include tweaking the new disciplinary boosts and implementing a citation boost and a recency boost
- ***Improve search relevance tooling***: Fine-tuning relevance calculations is currently done qualitatively; OSC helped us develop more quantitative methods but more work is needed. On hold as most of the work was done by Stephanie, our former planetary science project scientist.
- ***Planetary names pipeline in production***: Currently on pause, given the departure of Golnaz, the main developer for this project, and Stephanie

Advancing AI & data-driven capabilities 1/2

These projects focus on AI/ML R&D, though some infrastructure work is also included to support both R&D and production capabilities.

R&D

- **Document embeddings**: Finalize the calculation method & framework to evaluate whether to replace the current slow text-based similar() with embeddings equivalent
- **NER metadata enrichment:** Our desired list of NER categories for metadata enrichment is long; this year we'll likely begin with software, data, and/or telescopes/facilities
- **KAILAS**: Finalize the UAT keyword assignment model

Advancing AI & data-driven capabilities 2/2

These projects focus on AI/ML R&D, though some infrastructure work is also included to support both R&D and production capabilities.

Engineering

- **Utilize new GPUs for pipelines and R&D**: The new GPUs need a framework to efficiently run models (astroBERT-sized 100M to LLM-sized 14B parameters) and code
- **SciX Labs**: Incorporate the student-created SciX Labs UI into our infrastructure and spin up product pages for pilot projects
- **User intent understanding framework:** Implement better logging and a more modern metrics storage/calculation service via our new API Gateway, to enable near real-time data analysis and feedback

Strengthening core operations

Generally these tasks have less impact on users, so will be worked on as time allows, or if external deadlines increase their urgency

- **Solr maintenance**: Upgrade our ANTLR grammar to the latest version, test index compaction to allow less-frequent reindexing, work on startup time optimizations
- **Port to Solr Cloud:** Follow-on work from the OSC contract work. Our current Solr infrastructure will lose support in the coming years, so this will need to be done eventually
- **Synonym management improvements:** Our current synonym system is one-size-fits all; to enable discipline-specific synonyms some infrastructure work needs to be done
- **Microservice improvements and redesigns**: Things on our “nice to have” list: libraries feature improvements, redesign older (vis, metrics) services for speed/efficiency
- **Pipeline improvements and redesigns:** Turn the UAT keyword assignment script into a pipeline, port text mining from scripts to pipeline, for better maintainability
- **UI maintenance:** Testing improvements, refactoring, analytics
- **Management operations:** Implement shared team-wide performance monitoring, automate API spec maintenance

Scenario 2 (ADS only): 2026 Priorities

In the event of SciX's funding being cut, we assume that ~50% of our activities would need to be cut as well. Our reduced priorities would focus on:

- **Retiring and modernizing infrastructure**
- **Enhancing the user experience**
- ~~**Expanding and enriching content**~~
- ~~**Advancing AI & data-driving capabilities**~~
- **Strengthening core operations and reliability *(as time allows)***

To reduce overhead, we'd also need to shutter the ADS interface in Q1 2026, fully moving users to the SciX interface.

SciX funding cuts: long term consequences

Reducing activities and personnel by ~50% have severe long-term consequences:

- **Development of Honeycomb will slow to a crawl, preventing us from resolving this technical debt and reliance on 30-year-old ingest software**
  - The maintainers of the Classic ingest software are mostly quite senior staff; retirements start to become a worry in the next 5-10 years
- **Lack of knowledge and skill redundancy across the team**
  - We were very close to hiring a second search engineer; funding cuts caused us to withdraw this offer
- **Lack of diversity of career stage across the team, which will only worsen if we're unable to hire**
  - Lose long-term succession plan