

National Aeronautics and  
Space Administration



# Open Source Science and Open Science Policy at NASA

Steve Crawford  
Science Data Officer  
November 2, 2022

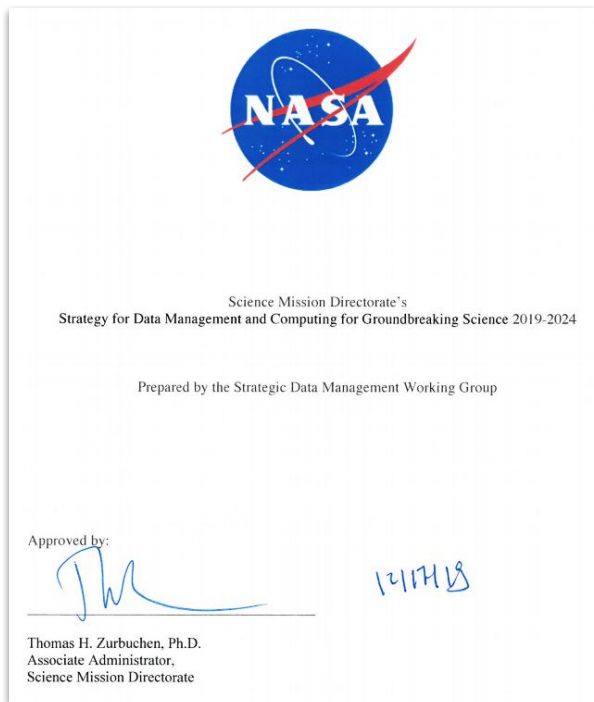


# Overview

---

- Open Source Science
- Support for Open Source
- Transform to Open Science

# What is the SMD Strategy for Data and Computing?



**An SMD-approved strategy to enable transformational open science** through continuous evolution of SMD's science data and computing systems.

**Goal 1:** Develop and Implement Capabilities to Enable Open Science

**Goal 2:** Continuous Evolution of Data and Computing Systems

**Goal 3:** Harness the Community and Strategic Partnerships for Innovation



# SMD Strategy for Data Management and Computing for Groundbreaking Science

SMD identifies strategic data management and science computing as a priority (February 2018)



Archives Processing and Data Exploitation Meeting (GRC) (August 2018)



SMD's Strategy for Data Management and Computing (December 2019)

SMDWG KICKOFF

RFI ACTIVITY

WORKSHOP #1

WORKSHOP #2

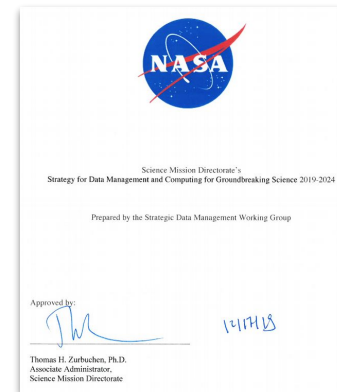
STRATEGY DEVELOPMENT



67 RFI Responses with five common themes (July-September 2018)



Maximizing the Scientific Return of NASA Data (DC) (October 2018)



# Open Source Software Policy Options



National Academies report on recommendations and best practices for open source software for Earth and Space Science. Some of the recommendations included:

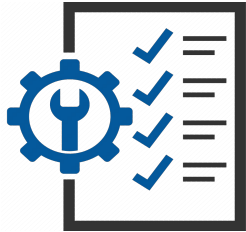
- Recognize the scientific value of open source software and incentivize its development and support, with the goal that open source science software becomes routine scientific practice
- train researchers in open software practices
- support the infrastructure, governance, and maintenance of a healthy open source community
- foster career credit for scientific software
- Updating internal policies on software release

<http://nap.edu/25217>

# NASA Open Source Science Initiative (OSSI)

Making the process of science open

<https://science.nasa.gov/open-science-overview>



## Policy development, education, compliance tool development

Updating NASA policies on scientific information to better enable the activation of open science (SPD-41A)



## Core Services for Science Discovery

Developing core data & computing services, tools, and training to enable open science



## ROSES Elements

Supporting open-source software, tools, frameworks, libraries, platforms, and training with over \$5 million dollars in grants



## Community Building & Partnerships Transform to Open Science (TOPS)

Fostering and accelerating adoption of open science practices across the scientific community

# SMD Scientific Information Policy (SPD-41)

SPD-41 was released in Aug. 2021

SPD-41 brought together existing NASA and Federal guidance on open data, software, and publications

Applies to all new SMD-funded activities starting Sept. 2021 and later

SPD-41 does not apply to restricted information

SPD-41: Scientific Information Policy

<https://go.usa.gov/xtNTJ>



Scientific Information Policy Website & FAQ

<https://science.nasa.gov/researchers/science-data/science-information-policy>



[https://science.nasa.gov/researchers/science-data/science-information-policy\\_faq](https://science.nasa.gov/researchers/science-data/science-information-policy_faq)

# SMD Scientific Information Policy (SPD-41A)

SPD-41A draft released in Nov. 2021.

Community RFI soliciting feedback closed Mar. 2022; main concerns around software, support, and further guidance

Final version expected to be released in December 2022.

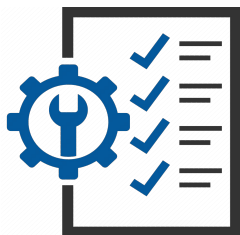
## SPD-41: Scientific Information Policy

<https://go.usa.gov/xtNTJ>

## Scientific Information Policy Website & FAQ

<https://science.nasa.gov/researchers/science-data/science-information-policy>

[https://science.nasa.gov/researchers/science-data/science-information-policy\\_faq](https://science.nasa.gov/researchers/science-data/science-information-policy_faq)



### Policy development, education, compliance tool development

Updating NASA policies on scientific information to better enable the activation of open science (SPD-41A)

- ★ SPD-41A is forward looking and will be released soon
- ★ Compliance for new activities no earlier than Jan. 2025
- ★ TOPS OpenCore modules will provide relevant training
- ★ Training events will be held during 2023 Year of Open Science
- ★ Further guidance on implementation will be provided
- ★ SMD Core Services aim to provide support for implementation
- ★ SMD Divisions will release companion policies in 2023



# What are the **new** proposed changes in the SPD-41a draft?

## Data

**Scientific data** should be FAIR and shall be made publicly available with a clear, open, and accessible data license no later than the publication of the research, and be citable.

**Mission data** shall be openly available with no period of exclusive access.

## Software

**Research software** shall be publicly available no later than the publication of the research, assigned a permissive software license, and be citable.

**Mission software** shall additionally be developed openly in a publicly accessible, version-controlled platform that allows for contributions and engagement from the community.

## Publications

**Manuscripts** versions of as-accepted manuscripts shall be deposited in a NASA repository and made publicly available within 12-months. Publishing as open access is supported and posting preprints is encouraged.

**Mission publications** shall additionally be made publicly available at the time of their publication.

**Science workshops and meetings** shall be open to broad participation and documented in public repositories.

# Open-Source Science Policy for Earth Science Missions

- A. All mission data, metadata, software, databases, publications, and documentation shall be available on a full, free, open, and unrestricted basis starting in Phase B with no period of exclusive access.
- B. Science workshops and meetings shall be open to broad participation and documented in public repositories.

**1** Software shall be developed openly in a publicly accessible, version-controlled platform using a permissive software license allowing for community use and contributions.

**2** Manuscripts shall be published with open access licenses; versions of as-accepted manuscripts shall be made available as open preprints and deposited in a NASA or [Partner] repository upon publication.

**3** All mission data, calibration information, and simulated products supporting development and validation of algorithms shall be made available without any conditions to use.

**4** Scientific data, metadata, software, publications and documentation shall be archived and made available by NASA and/or [Partner] starting in Phase B.

**5** NASA and [Partner] software, documentation and data shall be properly marked, cited, and/or attributed. Metrics to measure and acknowledge open-source science contributions will be developed.

**6** NASA and [Partner] will mutually develop an Open-Source Science Plan that specifies details of collaboration.

Collaborative, accessible, inclusive, transparent, and reproducible from the beginning.

# Overview of the implementation SPD-41a

---

Future implementation plans include:

- Improving Software release process for NASA scientist
- Guidance for awards, contracts, solicitations, and Announcement of Opportunities; PIs should include these costs in proposals
- Incentives for the community to make the transition - e.g ROSES22 F8. Supplement for Open Source Software

SPD-41a is ***forward looking*** - it is meant to apply to work going forward. Existing missions and investigations should adopt parts of this policy consistent with available resources.

# Open Source Software Best Practices

Here are some lessons learned on best practices on Open Source Software:

- *Release with a widely-used, permissive, open license*: Increase accessibility with clear guidelines on use (Example: Apache, BSD 3 Clause, MIT)
- *Release Early*: Allow for community to see and contribute to the software. Researchers can plan for their own science based on what is available (Example: JWST)
- *Use and contribute to open source libraries*: Do not reinvent the wheel and provide solutions for everyone (example: Astropy, Pangeo)
- *Welcome Community*: Include a Code of Conduct and a Contributors Guide so people know how to contribute to your project.

# Open Science Guidance

---

Developing processes to incorporate community best practices and community-led development of Open Science guidelines

- Adopt practices already in place in our community especially those developed through standards groups (IVOA, IPDA, SPASE, RDA, FAIR)
- Establish community-led guidelines for how to practice open science for our researchers, missions, and data repositories

**Example:** Open Science Guidelines working group (Lead: Mark Parsons) developed a request for comment DOI policy for SMD Data Repositories.

**Now out for comment. Please review and comment!**

**<https://github.com/nasa/smd-open-science-guidelines>**

# Supporting Open Source Science

Supporting open source software, open data, and open science practices in the SMD community.

- ROSES20 E.7 Open Source Tools, Frameworks, and Libraries selected 16 proposals supporting 22 different projects.
- ROSES22 F.8 Supplement for Open Source Software
- ROSES F.14 High Priority Open Science



Full description of supported projects is available on [NSPIRES](#)

# What is Transform to Open Science (TOPS)?

---

TOPS is a 5-year NASA SMD initiative to foster adoption of Open Science practices across the scientific community.



## Strategic Objectives:

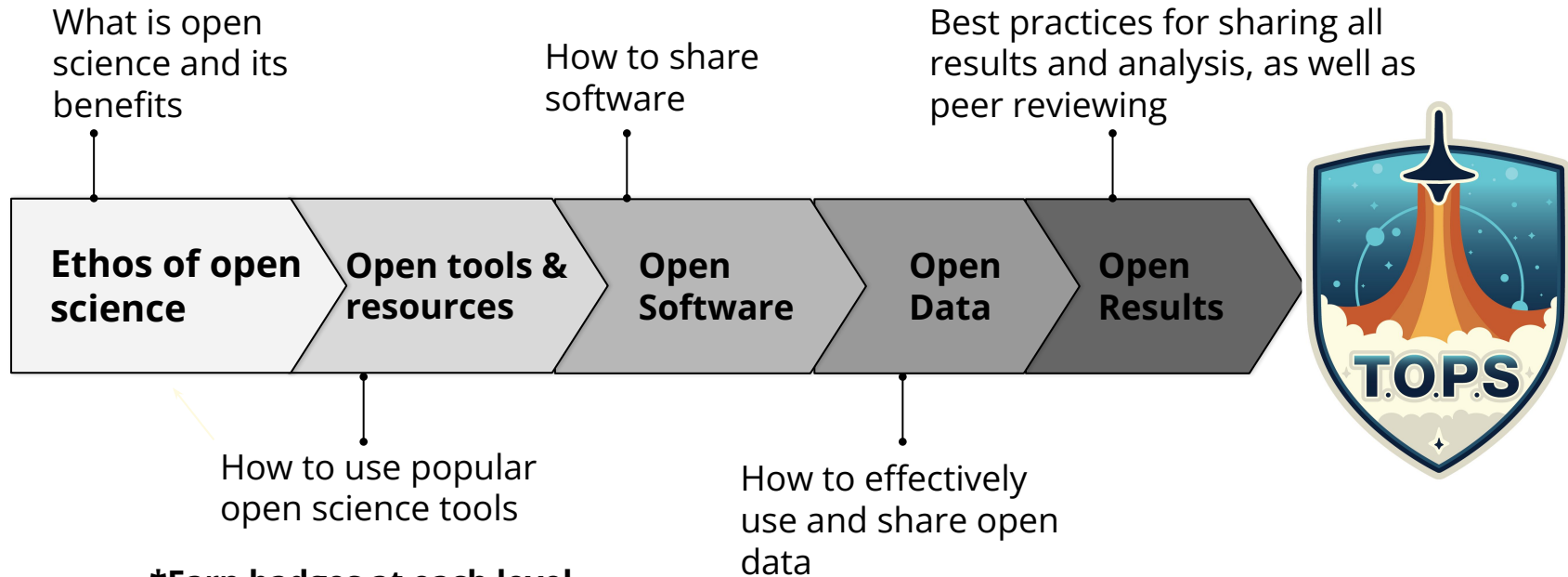
- Increase understanding & adoption of open science
- Accelerate major scientific discoveries
- Broaden participation by historically underrepresented communities

<https://nasa.github.io/Transform-to-Open-Science/>

# OpenCore: Introducing Open Science



**OpenCore** is a core curriculum (5 modules) that introduces open science to a general community

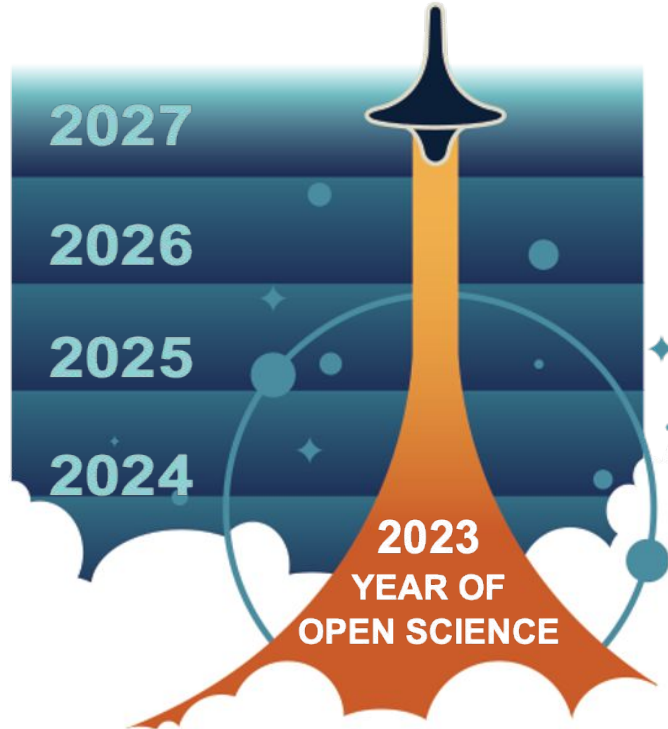


**\*Earn badges at each level**

**\*Completed in person, independently, and in virtual cohorts**



# Year of Open Science 2023



To kick off Transform to Open Science, SMD has declared 2023 as a Year of Open Science.

## Key Activities:

- *High visibility*: Publications, resources, and awards on Open Science
- *Conferences*: TOPS at major conferences and events supporting open science
- *Training*: Summer schools, virtual workshops, and targeted events to provide training on Open Science

## Get Involved:

<https://nasa.github.io/Transform-to-Open-Science/>

Back up slides

# SMD Strategy for Data Management and Computing for Groundbreaking Science 2019-2024

## Goal 1: Develop and Implement Capabilities to Enable Open Science

**1.1** Develop and implement a **consistent open data and software policy** tailored for SMD

**1.2** Upgrade capabilities at existing archives to **support machine readable data access using open formats and data services**

**1.3** Develop and implement a SMD data catalog to support discovery and access to complex scientific data across divisions

**1.4** Increase transparency into how science data are being used through a free and open unified journal server

## Goal 2: Continuous Evolution of Data and Computing Systems

**2.1** Establish **standardized approaches for all new missions** and sponsored research that encourage the adoption of advanced techniques

**2.2** Integrate investment decisions in High-End Computing with the strategic needs of the research communities

**2.3** Invest in capabilities to use commercial cloud environments for open science

**2.4** Invest in the tools and training necessary to enable breakthrough science through application of AI/ML

## Goal 3: Harness the Community and Strategic Partnerships for Innovation

**3.1** Develop **community of practice and standards group**

**3.2** Partner with **academic, commercial, governmental and international organizations**

**3.3** Promote opportunities for continuous learning as the field evolves through collaboration

# Context for Data and Computing Architecture Study

---

The CSDO is conducting two activities to develop cyberinfrastructure to support the Strategy for Data Management and Computing and SPD41:

## 1. Defining Core Data and Computing Services Requirements

Common SMD IT policies, software and computing capabilities to support:

- **Moving to hybrid cloud environments:** computing, storage, cybersecurity, networking, and business processes
- **Open-Source Science/SPD-41 requirements:** Research Data and Software Archive, User Registration, Data Set Search, Journal Search, AI/ML models, and more

## 2. Data and Computing Architecture Study

- Study to evaluate architecture options for scientific data and computing elements of Core Services infrastructure.
- Produce recommendations for a **Hybrid Cloud Infrastructure** for SMD (mixed computing, storage, and services environment made up of on-premises infrastructure, private cloud services, high-end computing, and a public cloud)

**Core Services funding initiates in FY24 and ramps up fully in FY25.**

# Timeline for Core Services

Divisions continue to support their existing data and computing activities

CSDO supports pilot cloud environments for divisions

Study: Scientific Data and Computing Architecture

Develop Core Services requirements and cost models

Refine and approve

Initiate HQ and Center Offices

Develop and deploy initial capabilities

Transition plan execution

Core Services Operational

FY22

FY23

FY24

FY25

FY26

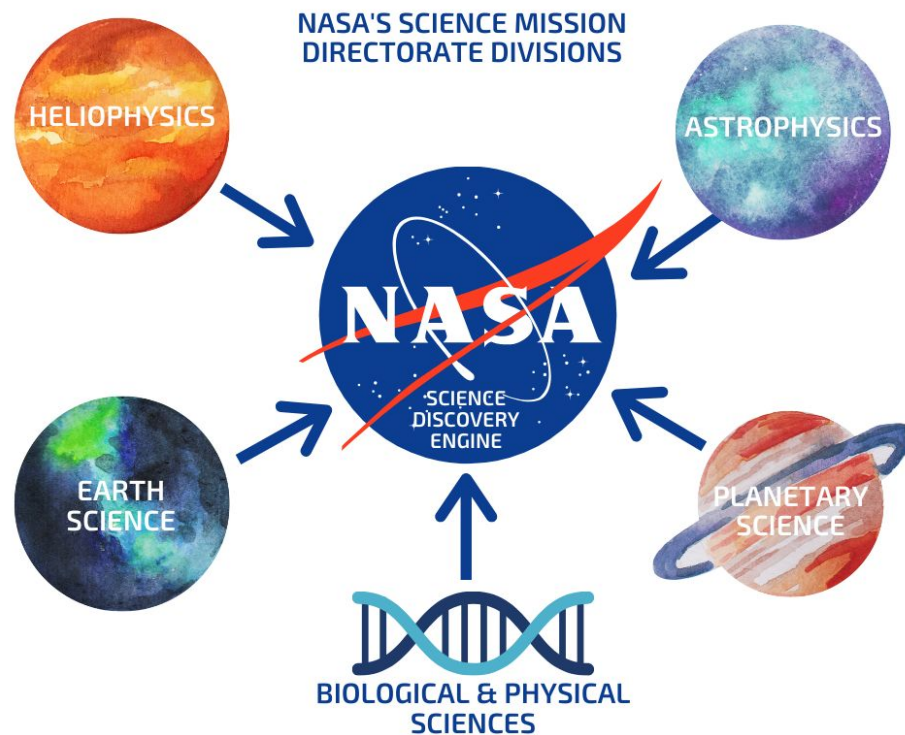
today

Open-Source Science Initiative Council and SMAc will guide and approve Core Services.

# SMD Science Discovery Engine

Create an SMD discovery capability to enable open source science. Scope includes:

- Astrophysics: **NAVO registry**
- BPS: GeneLab, Life Sciences Data Archive
- Earth Science: Common Metadata Repository
- Heliophysics: SPASE registry, Events Knowledgebase
- Planetary Science: PDS API
- + Models, software, tools and other contextual information from all 5 divisions
- Over **1 million documents & metadata** included at this time.
- Incorporated 3 SMD relevant facets into the interface
  - Platforms
  - Instruments
  - Missions



# SMD Science Discovery Engine



Welcome Bugbee, Kaylin M. (MSFC-ST11)



Search for...



kbugbee

Settings

Cross Portal Navigation



Saved Queries

Folders

Alerts

Active Filters

<empty search>

Additional Terms

Refine your search

Filter boxes will not display when not applicable to search results.

Science Knowledge Sources

|  |         |
|--|---------|
| <a href="#">Astrophysics</a>                     | 50,672  |
| <a href="#">Biological and Physical Sciences</a> | 4,195   |
| <a href="#">Earth Science</a>                    | 55,978  |
| <a href="#">Heliophysics</a>                     | 89,016  |
| <a href="#">NASA Science Documents</a>           | 299,939 |
| <a href="#">Planetary</a>                        | 925,902 |

All (1,425,734) [Data](#) (760,086) [Models](#) (5,679) [Documentation](#) (567,171) [Software and Tools](#) (45,891) [Missions and Instruments](#) (6,855)

1,425,734 results

[Threshold-Crossing-Event and Kepler-Object-of-Interest table definitions](#)

[Astrophysics > Software and Tools > NExSci > API Search](#)

<https://exoplanetarchive.ipac.caltech.edu/docs/TableColumnDescriptors.html>

Modified on Sep 26, 2022, 10:43:27 AM

Document Details

[Settings and Parameters](#)

[Astrophysics > Software and Tools > NExSci > API Search](#)

[https://exoplanetarchive.ipac.caltech.edu/docs/pgram/pgram\\_parameters.html](https://exoplanetarchive.ipac.caltech.edu/docs/pgram/pgram_parameters.html)

Modified on Sep 26, 2022, 10:43:26 AM

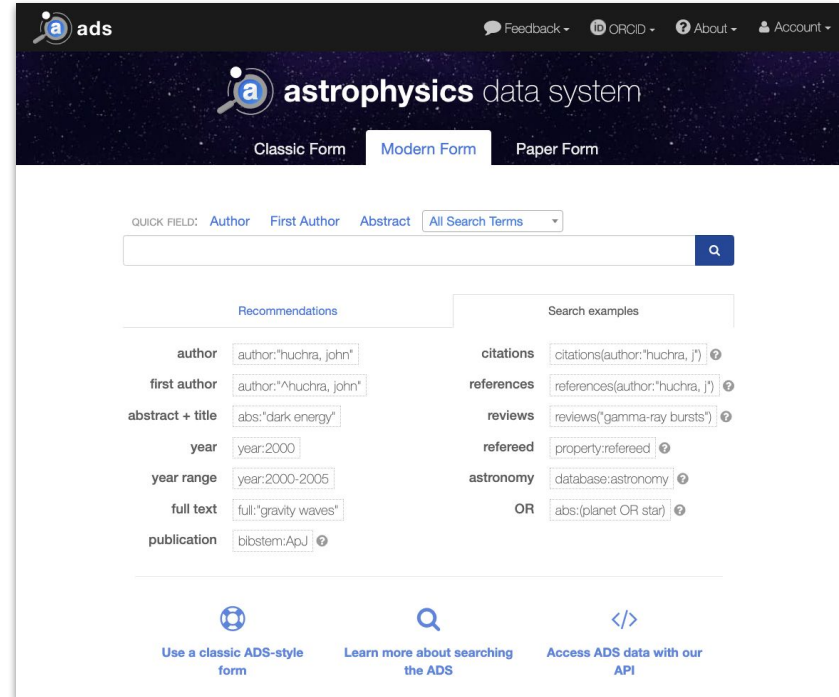
Document Details

[Periodogram Algorithms](#)

# The NASA Astrophysics Data System (ADS)

NASA-funded is funding an expansion of ADS beyond Astronomy & Physics to included:

- Heliophysics
- Earth Science
- Planetary Science
- NASA studies in Biological & Physical Sciences



The screenshot shows the ADS search interface. At the top, there's a navigation bar with the ADS logo, a search icon, and links for Feedback, ORCID, About, and Account. Below this is the main header with the text "astrophysics data system" and three tabs: Classic Form, Modern Form (selected), and Paper Form. A search bar is present with a "QUICK FIELD:" dropdown menu showing options like Author, First Author, Abstract, and All Search Terms. Below the search bar are two columns of search examples. The "Recommendations" column includes fields for author, first author, abstract + title, year, year range, full text, and publication. The "Search examples" column includes fields for citations, references, reviews, refereed, astronomy, and OR. At the bottom, there are three icons with corresponding text: a classic ADS-style form, learning more about searching the ADS, and accessing ADS data with the API.

<https://ui.adsabs.harvard.edu>