

# Scientific Content Curation In An Open Science Era

---

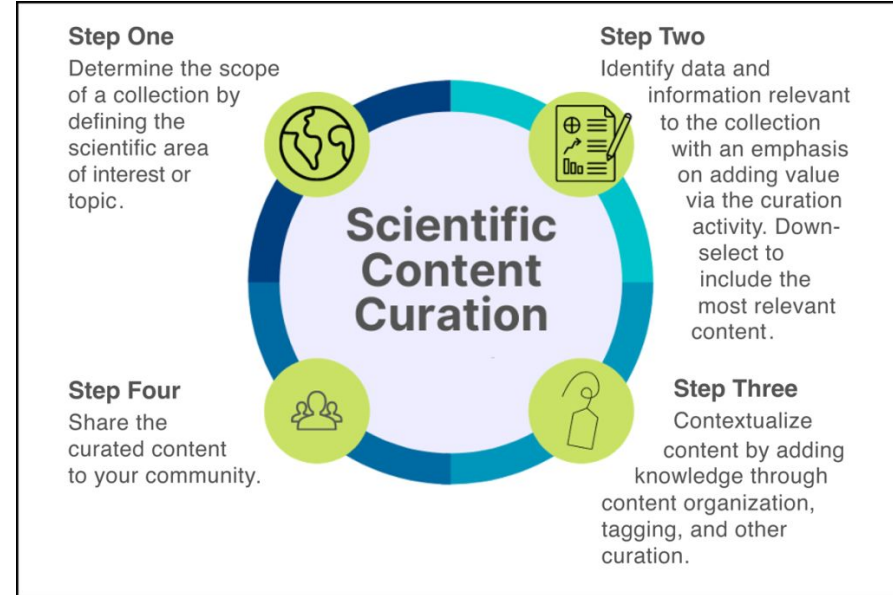
Kaylin Bugbee<sup>1</sup>, Deborah Smith<sup>2</sup>,  
Stephanie Wingo<sup>2</sup> and Emily Foshee<sup>2</sup>

1: NASA Marshall Space Flight Center  
2: University of Alabama in Huntsville



# What is Scientific Content Curation?

- Act of involving authoritative experts who identify, gather, validate, synthesize, organize, and present contextual details necessary to discover, understand, and use scientific data and knowledge effectively
- A value-adding effort that includes any activity or process that accelerates progress toward actionable science or makes it easier for diverse audiences to digest scientific information



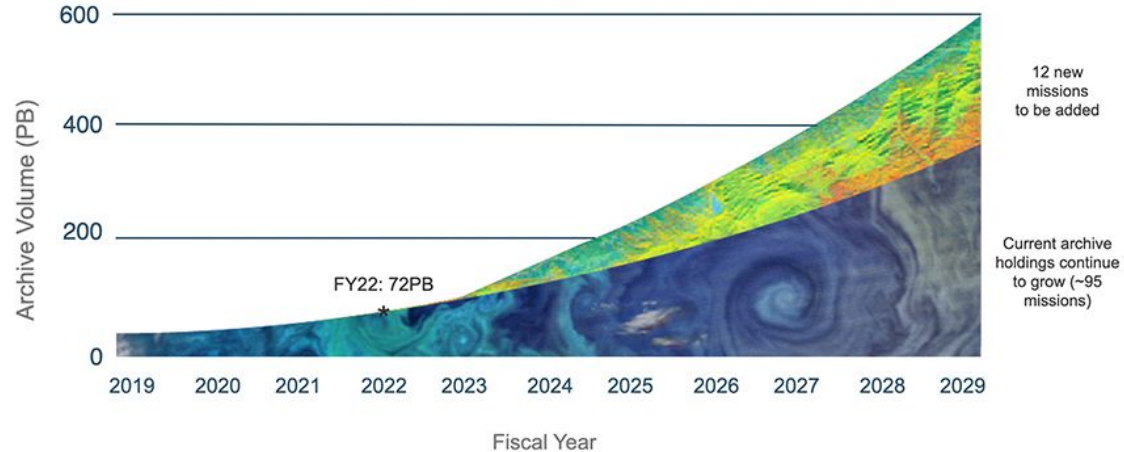
Citation: Bugbee, K., D. Smith, S. Wingo, and E. Foshee (2023), The art of scientific curation, Eos, 104, <https://doi.org/10.1029/2023EO230201>. Published on 19 May 2023.



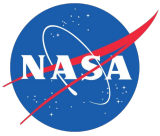
# Why Do We Need Scientific Content Curation?

- Explosive data volume growth
- Increased information available outside of traditional channels
  - Reports
  - Policy literature
  - White papers
  - Web sites
  - Blogs
- Proliferation of mis- and disinformation
  - Predatory publishers
  - Sham societies
  - Bogus conferences

## Earth Science Data Archive Growth Projection

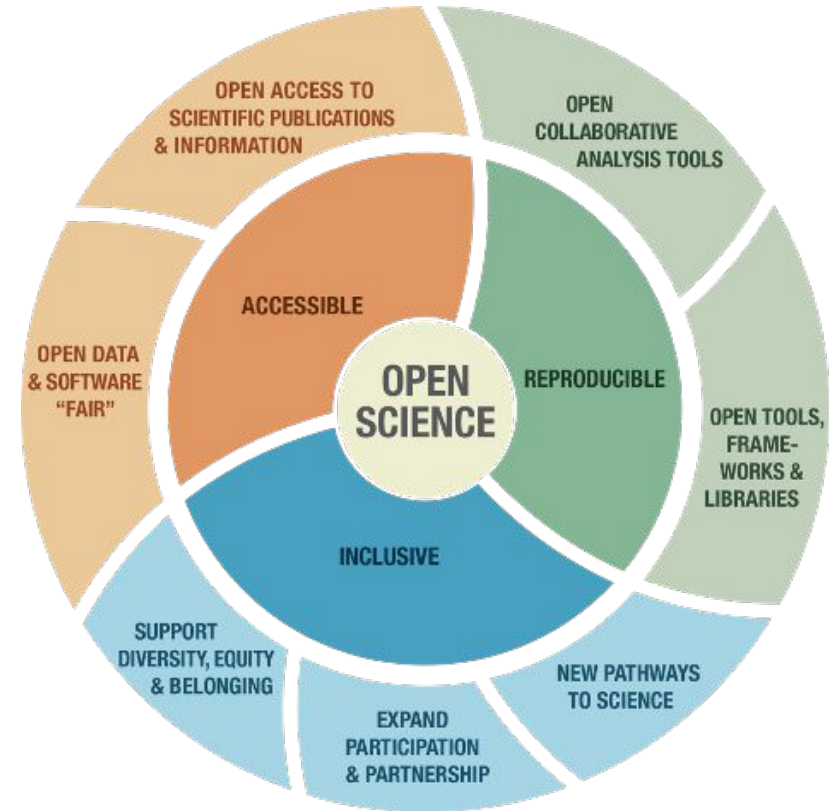


Credit: NASA EMS



# Why Do We Need Scientific Content Curation?

- Open Data, Software, Access
  - More data, code, documentation and publications will be shared or made openly available as more people embrace open science principles
  - We need a way to help everyone find the information they are looking for
- As science becomes more inclusive, we need to provide more equitable access to data and information
  - Curated content for different audiences and scientific communities will be needed to help promote a sense of belonging and to expand participation





# Airborne Data Management Group (ADMG)

The **Airborne Data Management Group** (ADMG) was established by the NASA Earth Science Data System management in Sep 2018 as part of IMPACT located at NASA MSFC in Huntsville, AL.

## ADMG Primary Tasks:

- Take a **full assessment** of NASA Airborne and Field Earth Science data
- Construct a public, centralized, **metadata-rich inventory** of airborne and field investigations, platforms, instruments, and data product access
- Develop **systematic approaches** and best practices that bring consistency and expediency to airborne and field data stewardship
- Maintain a **knowledge center** containing important information and document access, and simplified access to airborne tools and use cases
- **Improve communication** among DAACs, airborne campaign investigators, ADMG, and other stakeholders



Image credits: NASA



# CASEI: The Catalog of Archived Suborbital Earth Science Investigations



- CASEI contains **detailed information** of airborne and field campaigns and accesses data products using DOIs
- Additional contextual details provide users with **flexible discovery** and a greater understanding of campaigns, platforms, and instruments
- Users can explore and discover the inventory via the **highly-linked web-based user interface** to find information across all campaigns, data centers, people, topics, etc.
- Filters provide users with **many different ways to search** through the linked information

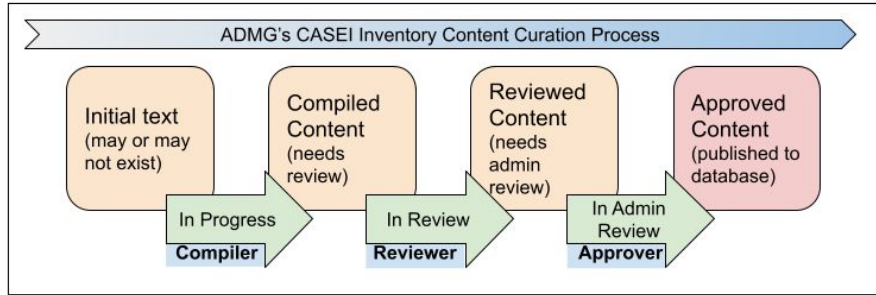


CASEI was developed by the ADMG with help from Development Seed and other IMPACT team members

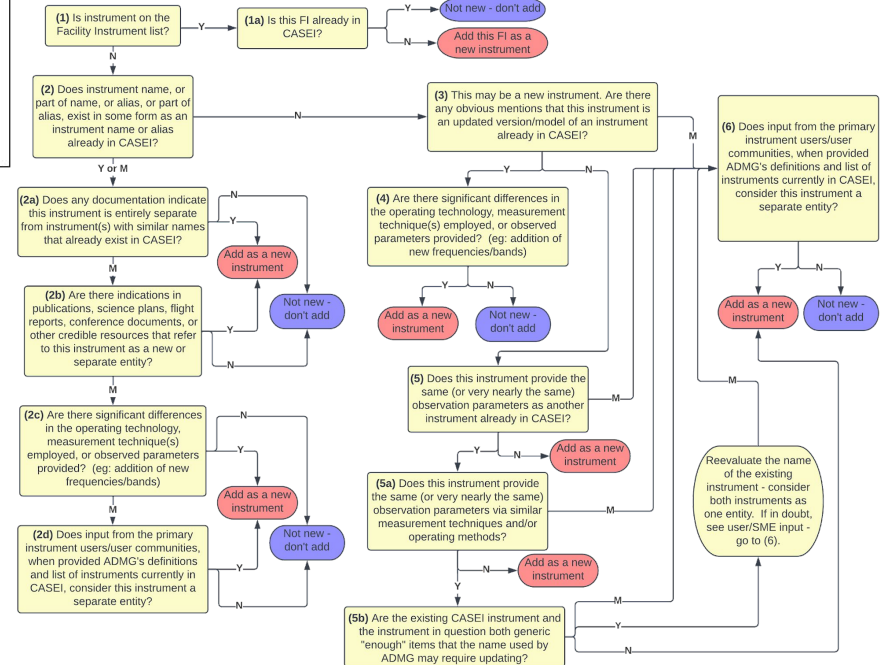
<https://impact.earthdata.nasa.gov/casei/>



# CASEI: A Simplified Content Curation Process



## Sample Decision Tree Workflow

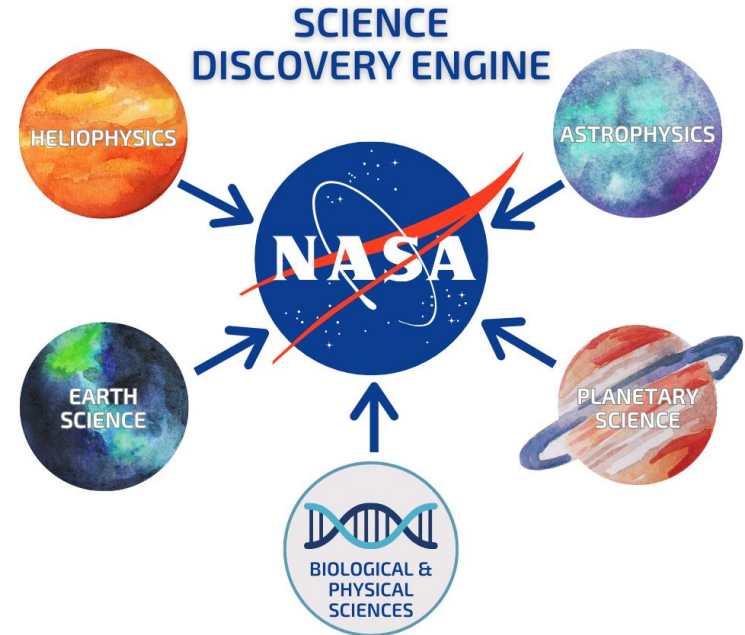


- 3 person vetting on all added metadata
- Decision trees used to ensure consistent process by all curators
- Reliance on authoritative sources of information
- Content organized for consistent representation



# Science Discovery Engine (SDE)

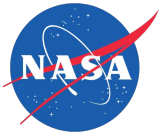
- The SDE is a source of trusted and curated open science data and information
- The SDE includes:
  - Metadata about science data
  - Code
  - Documentation
  - Images
  - Tutorials
  - Mission and instrument information
- Includes over 84,000 metadata records and thousands of documents from hundreds of sources across NASA science



Access at:

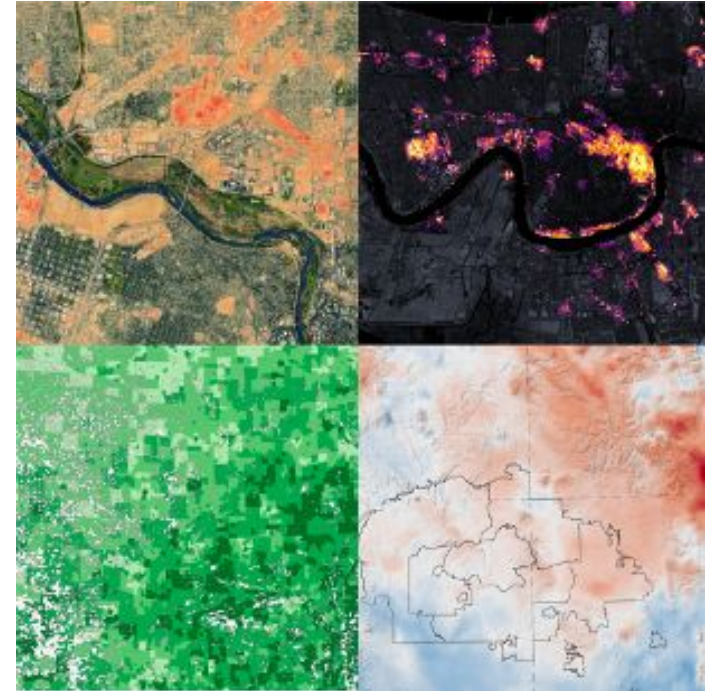
<https://sciencediscoveryengine.nasa.gov/>

Image credit: SDE team



# SDE: Enabling Curated Search Applications

- Scientific communities have a need to curate and share data and information around their specific use case
- Prototype NASA Environmental Justice Data Search portal supports EJ communities by expanding awareness, accessibility, and use of Earth science data by providing a curated, trusted, easy-to-use search interface
- Curated data catalog serves as the basis for the initial collection
- Data and information organized by:
  - Environmental Justice indicator (such as disasters, urban flooding, extreme heat, and food availability)
  - Possible intended use cases for each dataset such as 'explore data' or 'analyze data'



Examples of how NASA Earth science data are being applied to environmental justice. Clockwise from upper left: Identifying vulnerable populations using remotely-sensed urban heat island data in Sacramento, CA; detecting power outages using nighttime lights imagery in New Orleans, LA; assessing Navajo Nation water resources using NASA's Drought Severity Evaluation Tool (DSET) in the U.S. Southwest; helping policy officials identify communities in need using U.S. Social Vulnerability Index data from NASA's Socioeconomic Data and Applications Center (SEDAC). Credit: NASA ESDS.



SDE

?

BETA

# ENVIRONMENTAL JUSTICE DATA SEARCH INTERFACE

🔍 Type to search...



Search

In order to search within a focus area, type your query and click the focus area below.  
Or, click a focus area to view all datasets within.

CLIMATE CHANGE

DISASTERS

EXTREME HEAT

FOOD AVAILABILITY

HUMAN DIMENSIONS

HEALTH & AIR QUALITY

URBAN FLOODING

WATER AVAILABILITY



Image credit: SDE team



# Future of Scientific Content Curation



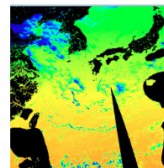
- Aggregating existing information relevant to a use case
- Thoughtfully distilling aggregated content to guide users to enhanced understanding (as in Earthdata Pathfinders)
- Bringing together new or previously unconnected content (e.g., curation around a topic like “landslides across the solar system”) to create new perspectives
- Developing new content to enhance existing curation efforts (e.g., adding campaign timelines to CASEI)
- Applying AI and Large Language Model (LLM) techniques to improve curation efficiency

## Find the Data



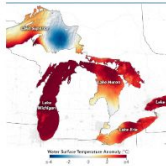
### Ocean Color

Ocean color assessments are used for answering fundamental questions about phytoplankton blooms, the aquatic food web, fisheries, and the storage of carbon in the ocean.



### Photosynthetically Available Radiation (PAR)

PAR is important for understanding biogeochemical cycles and addressing global climate change issues



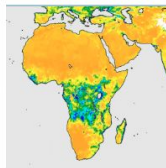
### Water Surface Temperature

Water surface temperature is important for the health of microscopic organisms and controls algal blooms.



### Precipitation

The amount and intensity of precipitation is a key aspect of soil moisture, runoff, turbidity, and water availability.



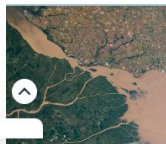
### Soil Moisture

The amount of moisture in the top levels of soil determines vegetation health and productivity.



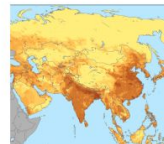
### Runoff

Runoff consists of the chemicals, pollutants, minerals, and sediments picked up and transported by water.



### Turbidity

The rate at which water flows and the nutrients carried in this flow impact water quality and color.



### Human Impacts/Socioeconomic

NASA socioeconomic data can help identify populations that may be more vulnerable to or most impacted by compromised water quality.

Earthdata Water Quality Pathfinder:

<https://www.earthdata.nasa.gov/learn/pathfinders/water-quality-data-pathfinder>



# Discussion



- Ensuring that human expertise is available is a challenge
- We need to recognize and reward contributors' efforts by providing credit
- The data management community needs to focus on providing enhanced services to user communities
- Institutions should expect funding requests in scientific proposals to support curation activities
- Through such efforts, we can improve the long-term sustainability of well-organized, curated scientific repositories that
  - Help make sense of the vast information landscape
  - Open more equitable access to information
  - Foster the interdisciplinary work needed to address many challenges facing the world today



Exploring Climate Data in a New Way. Credit: NASA  
<https://www.nasa.gov/image-detail/52996510071-c5abce524b-o/>

The background features a light blue architectural blueprint with various lines and structures. Overlaid on the right side is a large, semi-transparent DOI number: 10.1029/2023EO230201. The text 'Thank you!' is centered on the left side in a dark blue, sans-serif font.

# Thank you!

This presentation originates from our EOS article:

Citation: Bugbee, K., D. Smith, S. Wingo, and E. Foshee (2023), The art of scientific curation, *Eos*, 104, <https://doi.org/10.1029/2023EO230201>. Published on 19 May 2023.