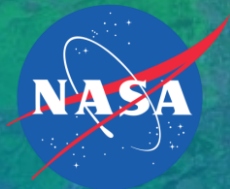


NASA's Collaborative Metadata Curation Activity to Improve Earth Science Data Discovery

EGU General Assembly
Vienna, Austria
April 9, 2018

Kaylin Bugbee¹, Valerie Dixon², Rahul Ramachandran², Dana Shum³, Jeanne le Roux¹, Adam Sisco¹, Patrick Staton¹, and Betzy Hernandez¹

(1) University of Alabama in Huntsville, (2) NASA,
(3) Raytheon Company Riverdale

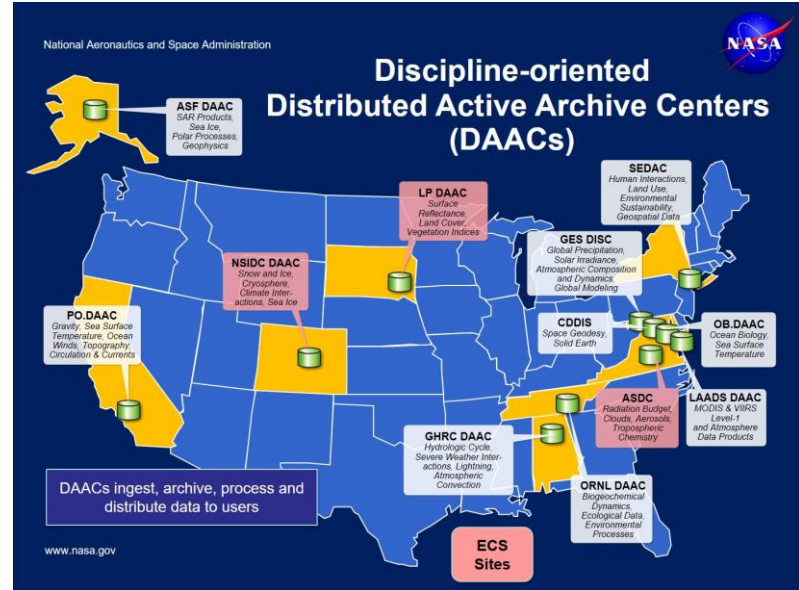


NASA Earth Science Data

NASA's Earth Observing System Data and Information System (EOSDIS)

Data is archived and distributed by 12 Distributed Active Archive Centers (DAACs)

Nearly 7,000 collections and 370 million granules are described by metadata housed in the Common Metadata Repository (CMR)

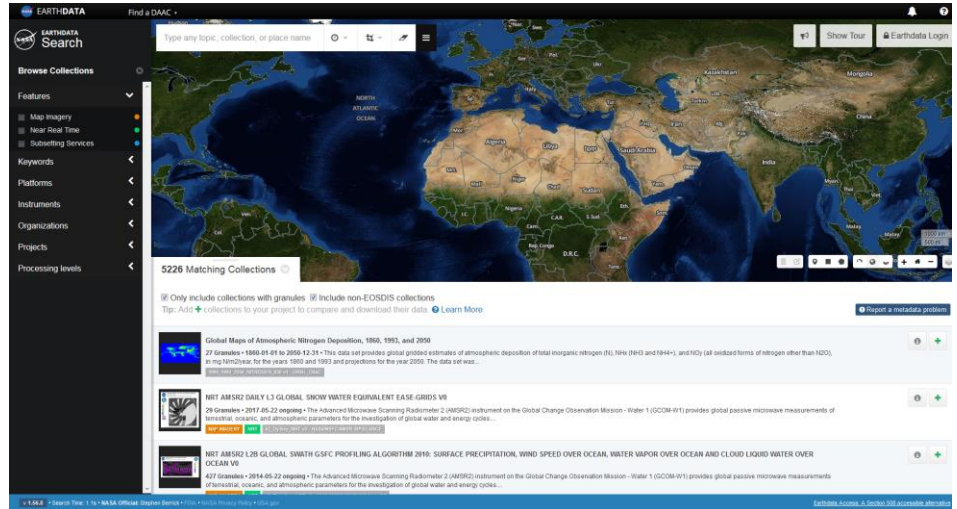


Earthdata Search

One stop shop for NASA
Earth Science Data

Uses metadata in the
CMR to help users find
the information they are
looking for

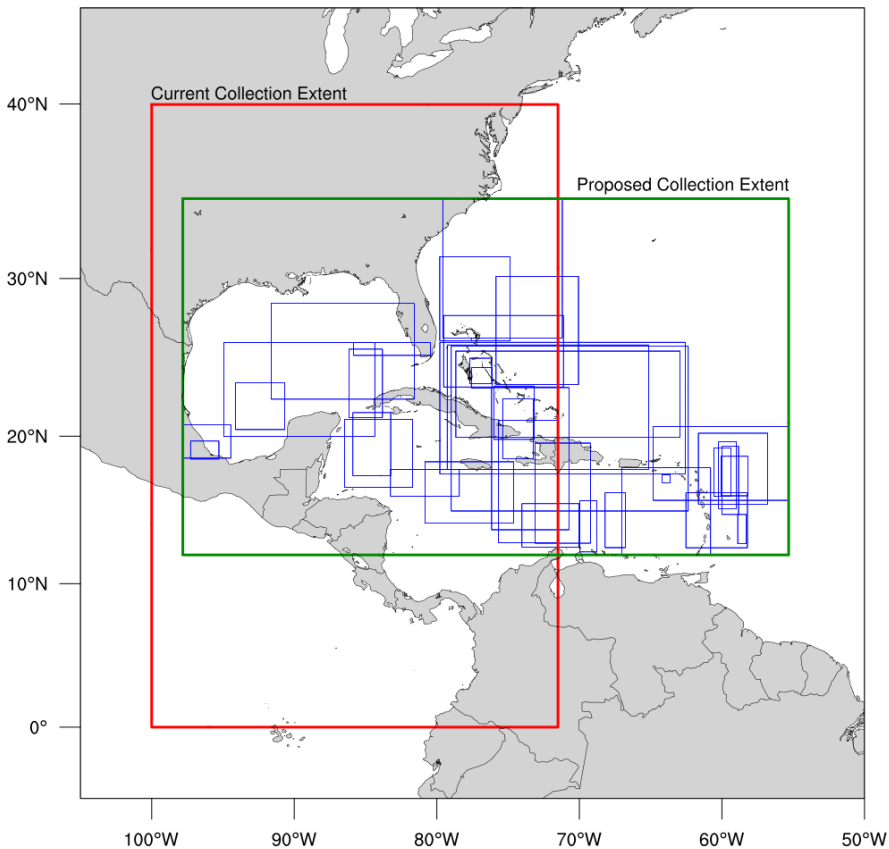
Functions best when
metadata is complete,
consistent, accurate



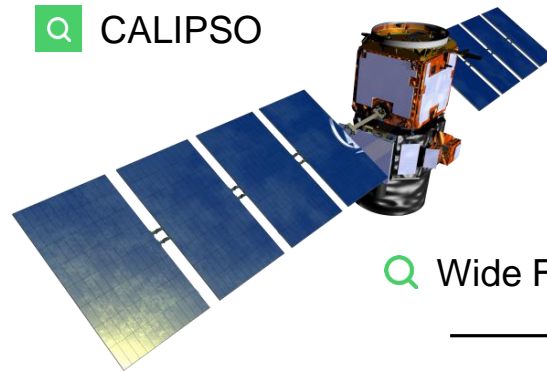
The screenshot displays the Earthdata Search web application. At the top, there is a search bar with the text "Find a DAMC". Below the search bar is a map of the Atlantic Ocean and surrounding regions. To the left of the map is a sidebar with navigation options: "Browse Collections", "Features", "Map Imagery", "Near Real Time", "Subsetting Services", "Keywords", "Platforms", "Instruments", "Organizations", "Projects", and "Processing levels". Below the map, it shows "5226 Matching Collections". There are several filters and options, including "Only include collections with granules" and "Include non-EOSDIS collections". Below the filters, there is a list of search results. The first result is "Global Maps of Atmospheric Nitrogen Deposition, 1980, 1991, and 2000" with 27 granules. The second result is "M1T AMSR2 DAILY 1.3 GLOBAL SNOW WATER EQUIVALENT (A SE-GRIDS V0" with 26 granules. The third result is "M1T AMSR2 L2B GLOBAL SWATH GSFC PROFILING ALGORITHM 2016: SURFACE PRECIPITATION, WIND SPEED OVER OCEAN, WATER VAPOR OVER OCEAN AND CLOUD LIQUID WATER OVER OCEAN V0" with 427 granules. At the bottom of the page, there is a footer with "1.1842 Search Time 1.16s NASA Official Stephen Berman" and "Earthdata Search 1.1842".

Search and Discovery

🔍 Spatial Coverage



🔍 CALIPSO



🔍 Wide Field Camera (WFC)

—————> 170K granules

🔍 Imaging Infrared Radiometer (IIR)

—————> 449K granules

🔍 Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP)

—————> 1 granule

LIDAR 2M granules

What is metadata curation?

Traditional curation



Digital curation

“Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle.”

Metadata curation

Supports the research data lifecycle by ensuring the correctness, completeness and consistency of metadata

Analysis and Review of CMR (ARC) Team

Team is comprised of Earth Science data and metadata specialists

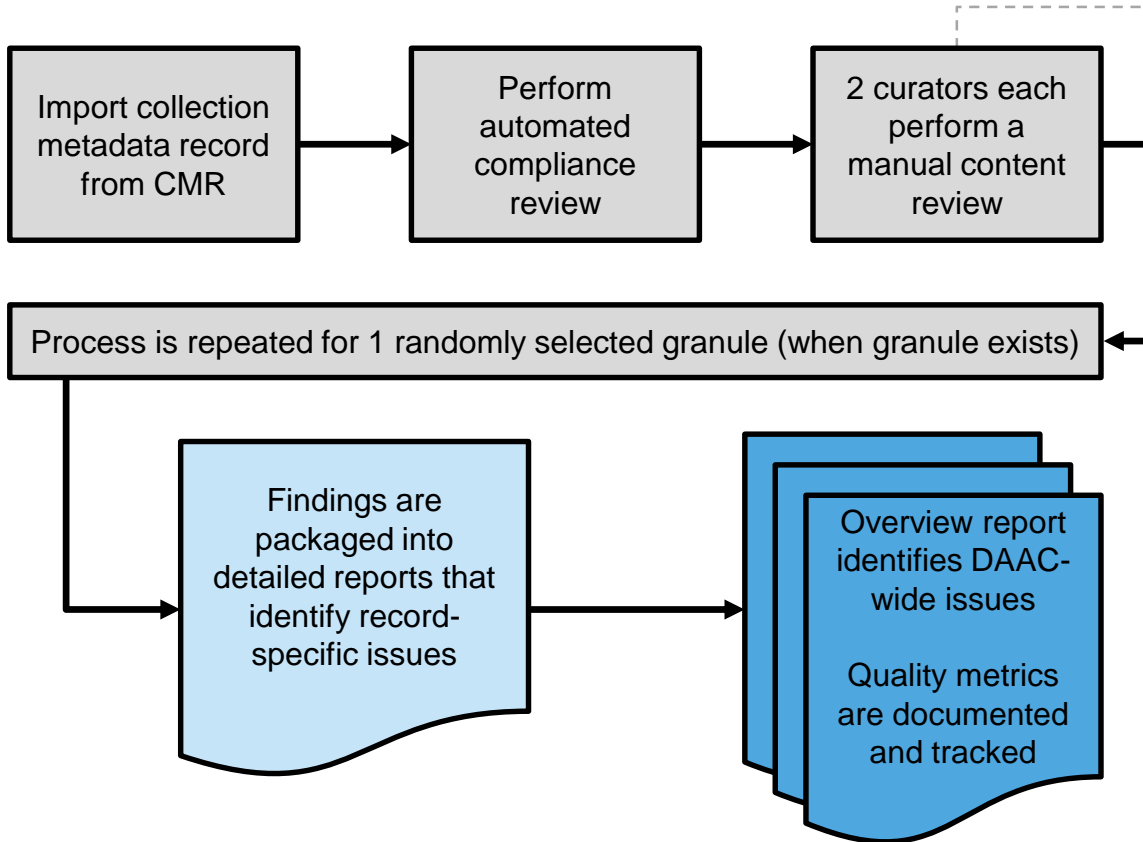
Backgrounds in Earth science, atmospheric science, space science, and remote sensing

Previous experience from the Climate Data Initiative (CDI)

- Review of 850 metadata records for quality and accessibility



ARC Curation Process

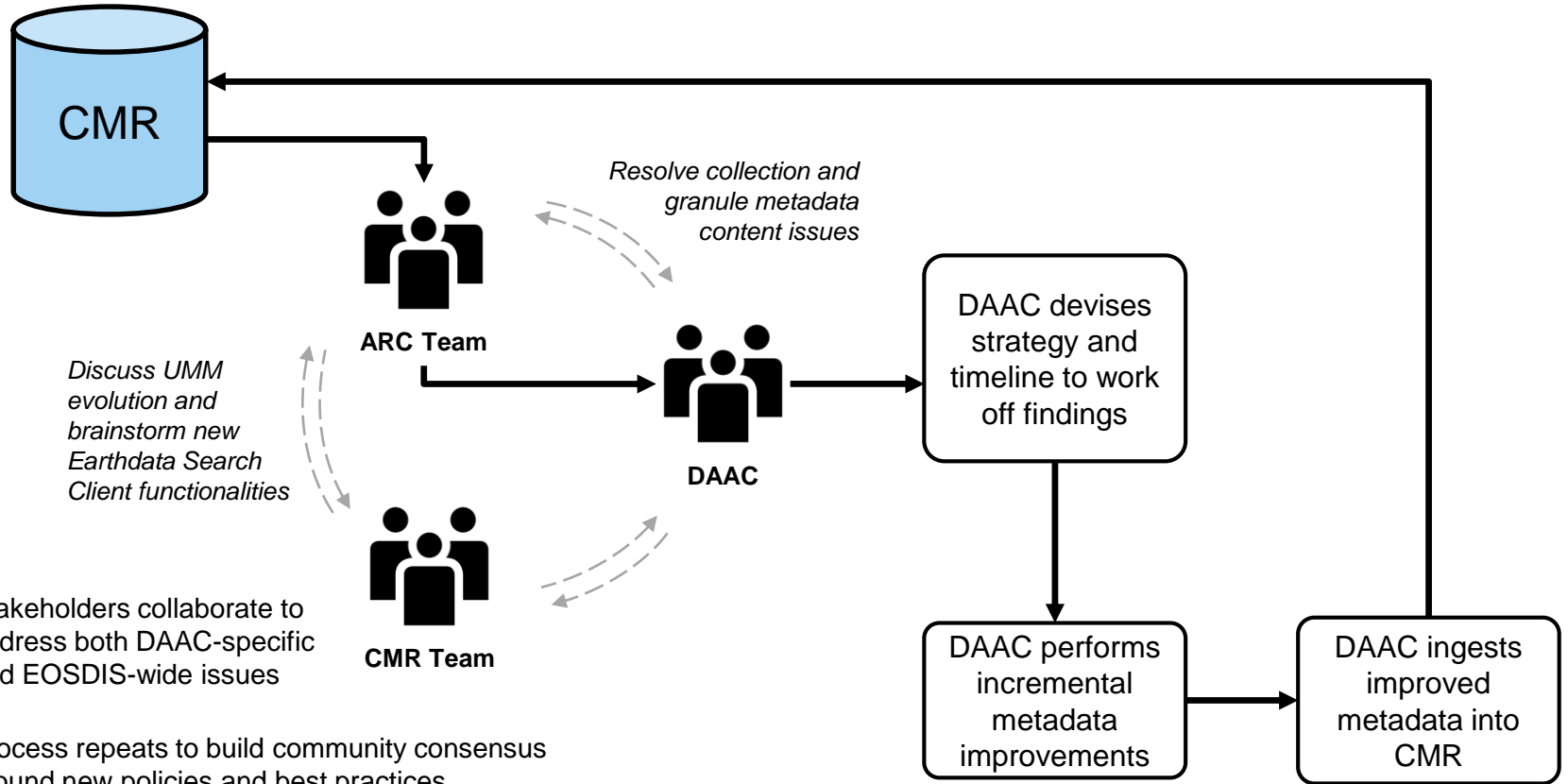


High	<ul style="list-style-type: none">• Inaccurate, incomplete, or missing content• Broken URLs and invalid collection-granule relationships
Med	<ul style="list-style-type: none">• Revisions of existing content• Addition of new information
Low	<ul style="list-style-type: none">• Minor consistency issues

Priority classification scheme

1. Assists DAAC in formulating a strategic plan to address findings
2. Used to track resolution of issues

ARC Curation Process



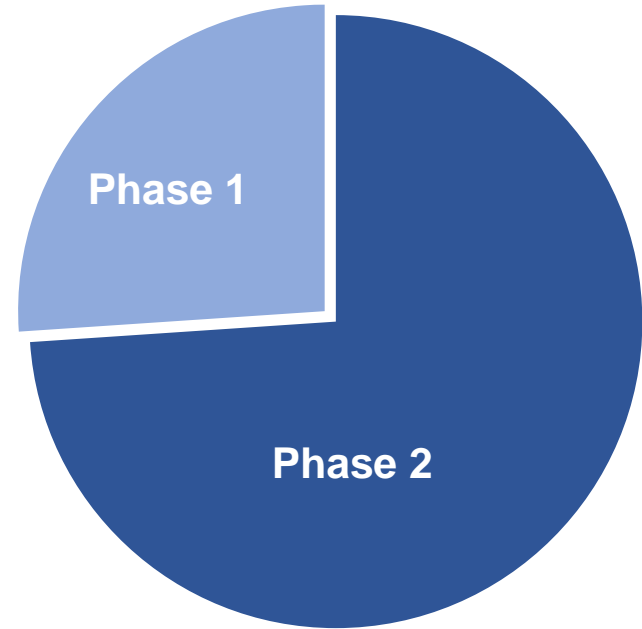
Phase I Metrics

Reviewed over 25% of collection level records in CMR

- Records from all 12 data centers reviewed

Reviewed metadata in 4 dialects

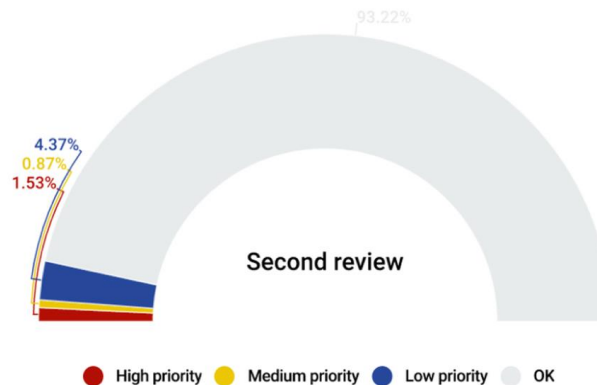
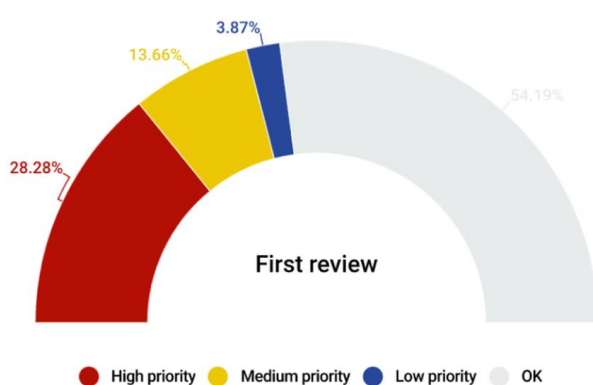
Supported two data centers in the generation of brand new collection and granule metadata



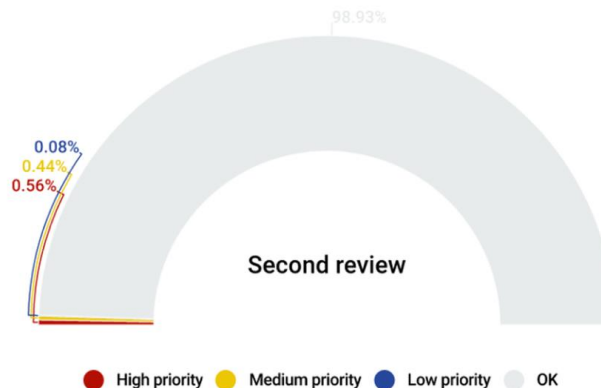
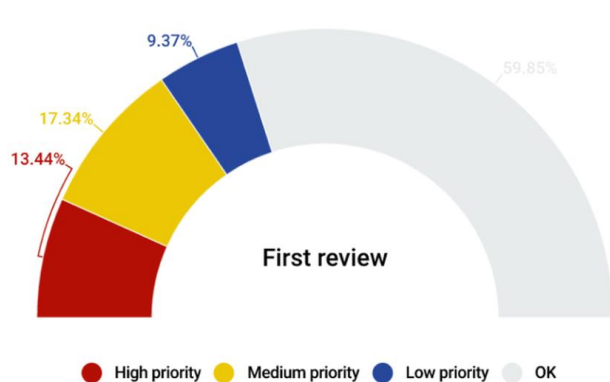
1,959 collections reviewed

Key Outcomes from Phase I

ORNL



SEDAC



Metadata Curation Documentation

Developing easy to understand guidance or best practices on metadata

Metadata curation wiki space will include detailed information on each element

Will support metadata curation dashboard and other tools

Collection Progress

Created by Erich Reiter, last modified by Kaylin Bugbee on Mar 07, 2018

- Element Description
- Best Practices
- Element Specification
- ARC Priority Matrix
- Dialect Mappings
- UMM Migration
- Future Mappings
- History
 - UMM Versioning
 - ARC Documentation

<https://wiki.earthdata.nasa.gov/display/CMR/Collection+Progress>

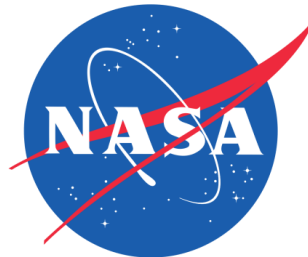
Thank You!

Questions?

Contact me at:

Kaylin Bugbee

Kaylin.m.bugbee@nasa.gov



Raytheon