

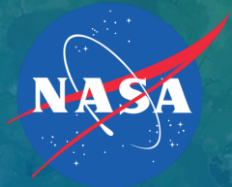
# Improving Access to NASA's Earth Science Data through Collaborative Metadata Curation

---

AGU Fall Meeting  
New Orleans, LA  
December 11, 2017

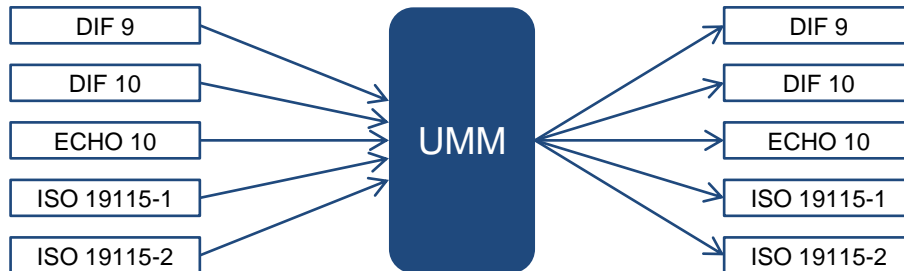
Adam W. Sisco<sup>1</sup>, Kaylin Bugbee<sup>1</sup>, Dana Shum<sup>2</sup>, Katie Baynes<sup>3</sup>, Valerie Dixon<sup>3</sup>, and Rahul Ramachandran<sup>4</sup>

(1) University of Alabama in Huntsville, (2) Raytheon Company Riverdale,  
(3) NASA Goddard Space Flight Center, ESDIS, (4) NASA Marshall Space Flight Center



# EOSDIS and CMR

- Earth Observing System Data and Information System (EOSDIS) manages NASA's Earth science data
- Ever growing collection of 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)
- Data is described using a number of different metadata standards, and core elements of each standard are mapped to and from a common model – the Unified Metadata Model (UMM)



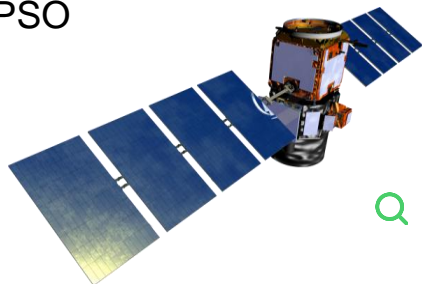
# Earthdata Search

- The Earthdata Search Client uses metadata in the CMR to **present users with the information they are looking for** and **hand users off to more specific applications**
  - Are users finding the information they are looking for? If not, why?
  - Are users being handing off to more specific applications? If not, why?
- Poor quality metadata is often the answer
- The CMR functions best when the metadata it houses is complete, consistent, and accurate
- Let's examine real examples of “less than ideal” metadata and the consider the consequences

# Search and Discovery

*Collection metadata must accurately describe all, not some, of the child granules.*

## Q CALIPSO



Q Wide Field Camera (WFC) → 165K granules

Q Imaging Infrared Radiometer (IIR) → 436K granules

Q Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) → 1 granule

**LIDAR**  
**1.8M granules**

## Q Terra/MISR

1999-12-18

Collection metadata range

0 hits

2007-06-01

40,000 hits

Actual granule range

2014-12-18

15,000 hits

2017-12-11


More than ¼ of the granules are not described by the parent collection.

1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017

# Accessibility

- Can I access the data via direct download?
  - Served correct data?
  - Served all data requested?
- 

# Usability

- Does the metadata enable users to be handed off to online documentation?
  - User's guides, README files, ATBDs, FAQ pages, product quality assessments, etc.
- 

# What is metadata curation?

Traditional curation



Information Age web content curation



---

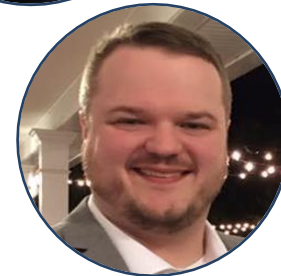
## Digital curation

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

“...curation enhances the long-term value of existing data by making it available for further high quality research.”

# Analysis and Review of CMR (ARC) Team

- All have been or currently are users of NASA Earth Science data for research applications
- 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's approach to digital curation

## Compliance

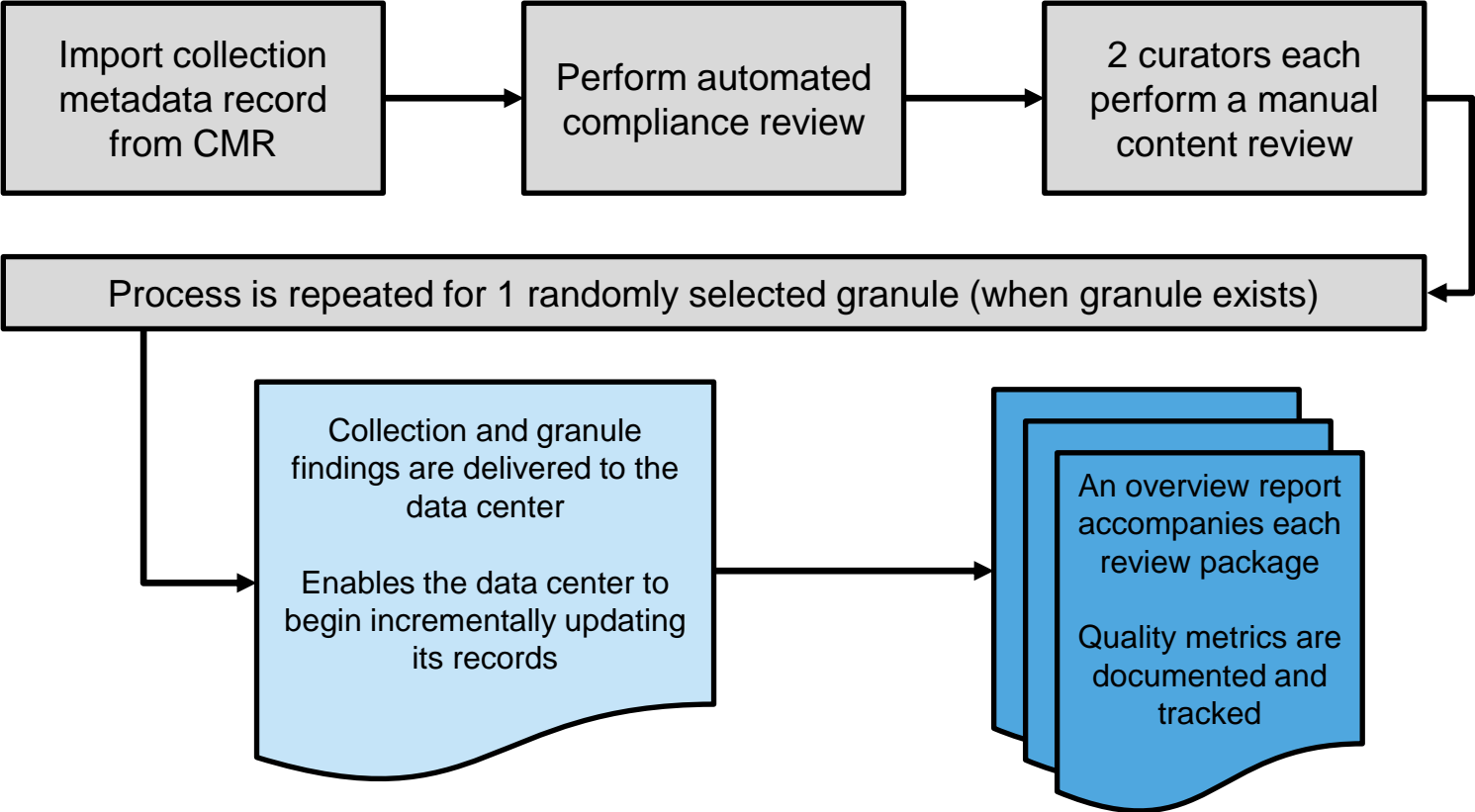
- Required elements
- Controlled vocabulary
- Broken URLs
- UMM usage
- DOIs

## Compliance + Content

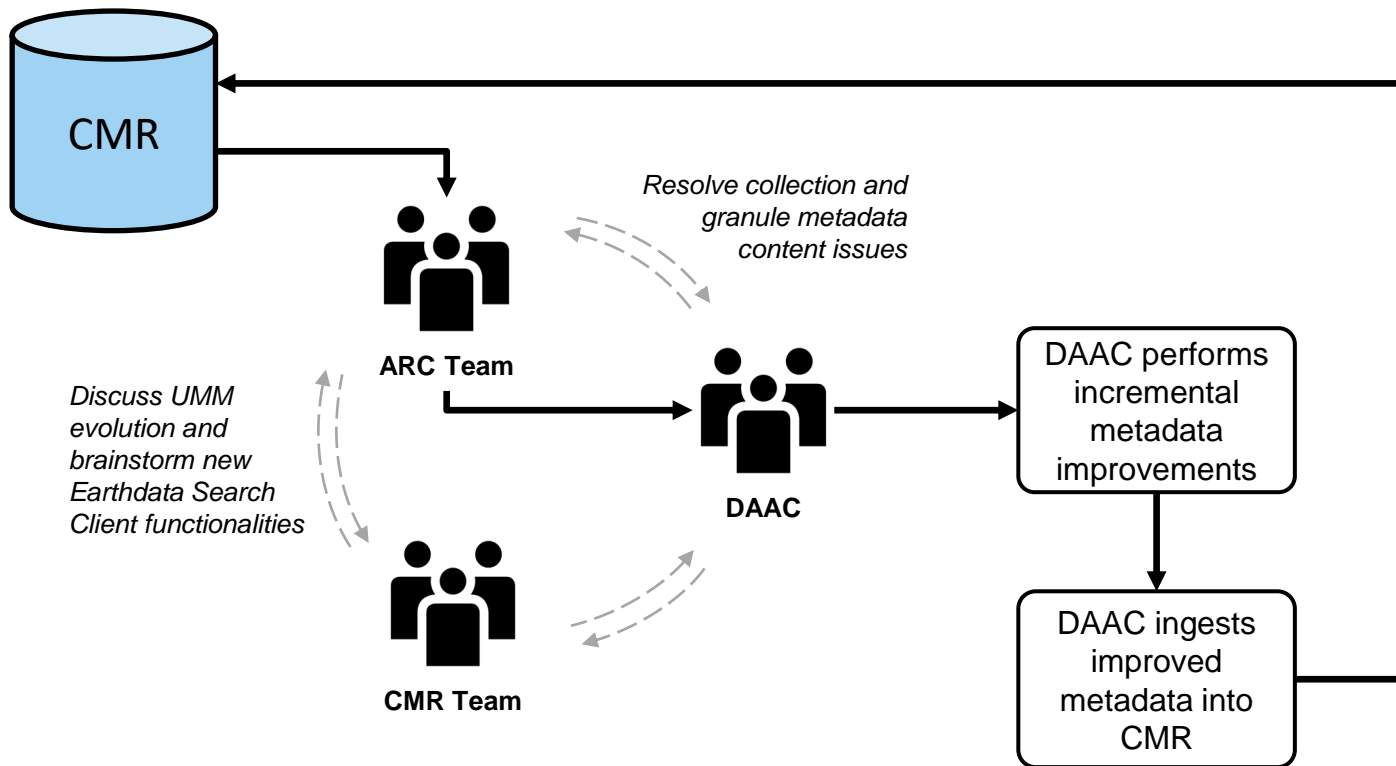
- Accuracy
- Consistency across collections
- Addition of new information
- Comprehensibility
- Keyword relevancy



# ARC Curation Process



# ARC Curation Process



*Stakeholders collaborate to address both DAAC-specific and EOSDIS-wide issues*

# ARC Curation Process

- Priority classification scheme
  - Assist DAAC in formulating a strategic plan to address findings
  - Track resolution of issues

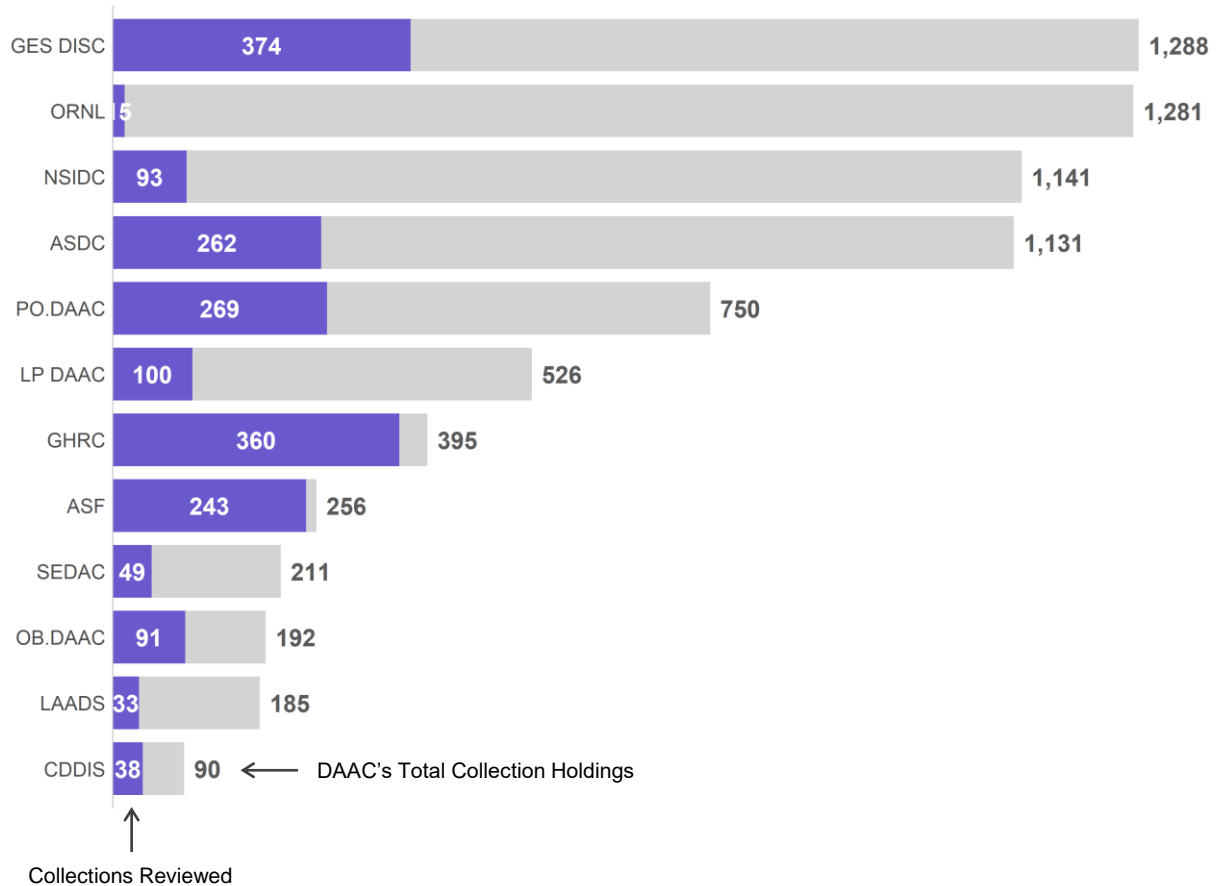
High	<ul style="list-style-type: none"><li>• Inaccurate, incomplete, or missing content</li><li>• Broken URLs and invalid collection-granule relationships</li></ul>
Medium	<ul style="list-style-type: none"><li>• Revisions of existing content</li><li>• Addition of new information</li></ul>
Low	<ul style="list-style-type: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

- ARC submits finding to DAACs
  - Overview report (Identifies DAAC-wide issues)
  - Detailed reports (Identify record-specific issues)
- DAAC submits a report to ESDIS on a strategy and timeline devised to work off findings
- DAAC works off findings with the ARC and CMR teams available for support
- DAAC alters internal processes as needed to ensure adherence to EOSDIS policies and best practices moving forward

# Phase I

- Mid 2016 to late 2017
- Records from all 12 DAACs reviewed
- 1,959 collections reviewed
- GHRC, ASF, and CDDIS fully reviewed
- Supported CDDIS and SEDAC in the generation of brand new collection and granule metadata

## ARC Collection Reviews Ending December 2017



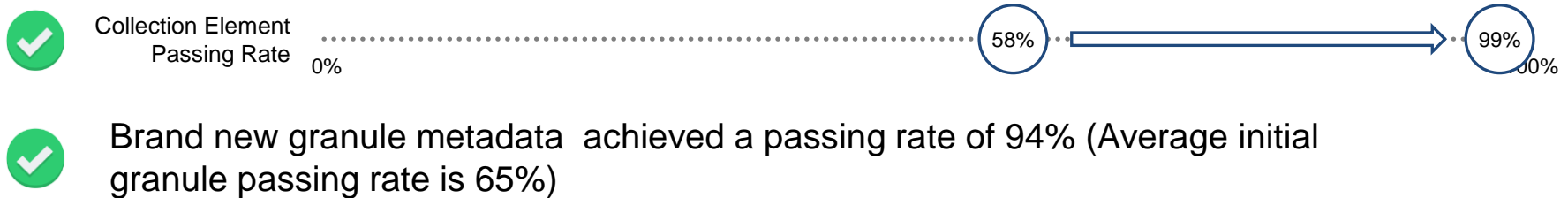
# Key Outcomes from Phase 1

- Evaluation of updated metadata for ORNL and SEDAC

## ORNL

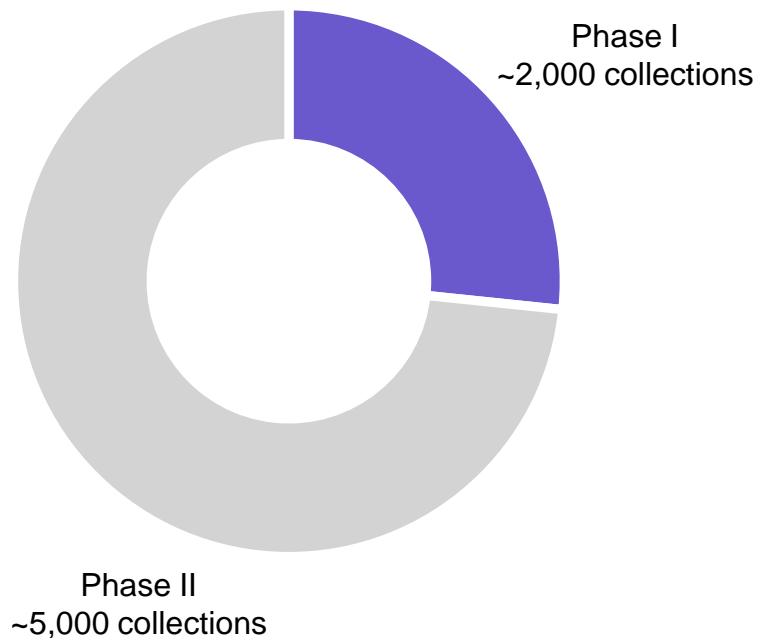


## SEDAC



# Phase II

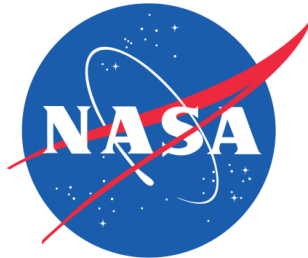
- ARC reviews will transition to an online dashboard environment
  - Improve ARC/DAAC communication
  - Enable automated metric tracking
- Implement a more strategic approach to ARC delivery of findings
- Track DAAC improvements from Phase I
- Improve UMM documentation and provide new reference resources for metadata authors
- Document and disseminate best practices that have emerged from the curation effort



# Questions

Adam Sisco

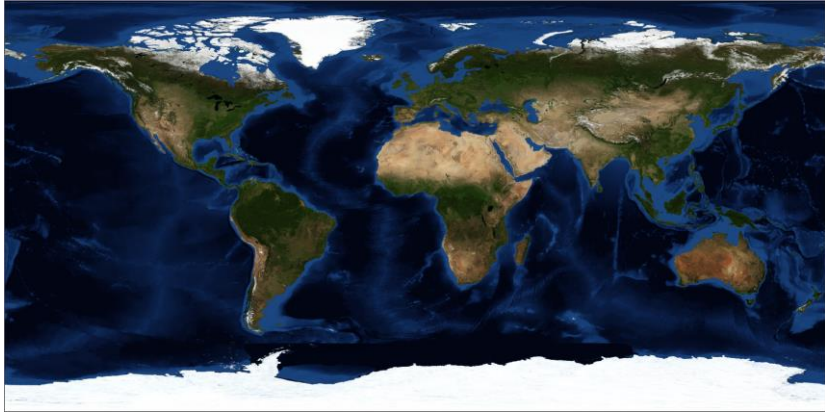
[adam.sisco@nsstc.uah.edu](mailto:adam.sisco@nsstc.uah.edu)



**Raytheon**

# Scratch Slides





**ARC Team**



**DAAC**



**CMR Team**

