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

EGU General Assembly
Vienna, Austria
April 9, 2018

Kaylin Bugbee[1], Valerie Dixon[2], Rahul Ramachandran[2], Dana Shum[3], Jeanne le Roux[1], Adam Sisco[1], Patrick Staton[1], and Betzy Hernandez[1]
(1) University of Alabama in Huntsville, (2) NASA,
(3) Raytheon Company Riverdale

NASA

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Raytheon

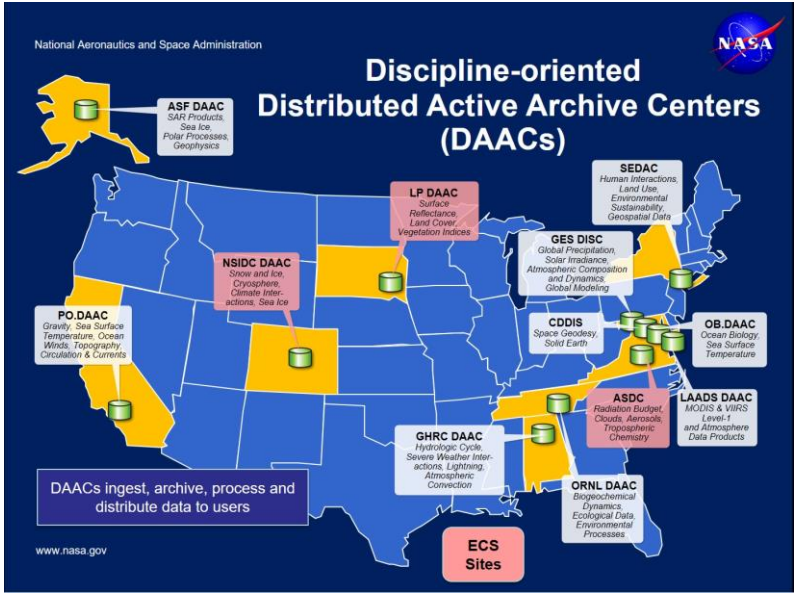
DSIG

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)

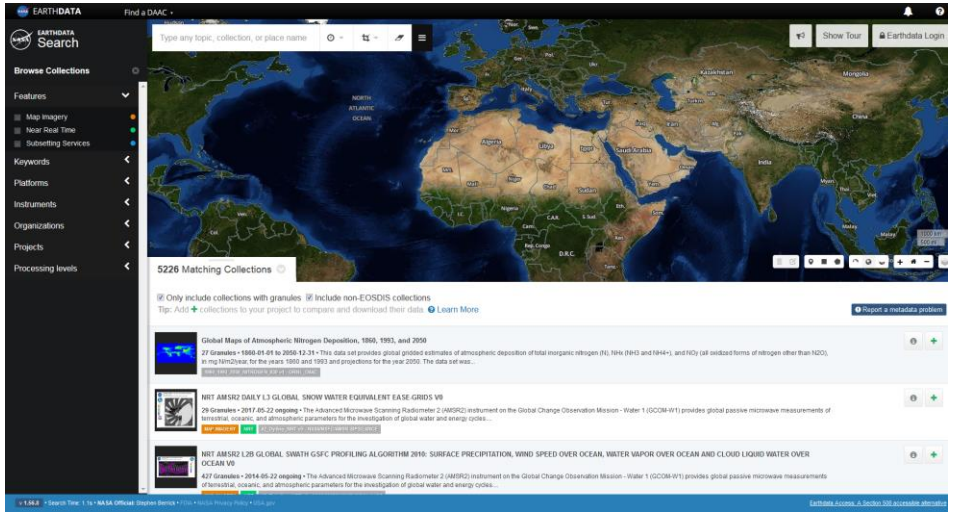


Earthdata Search

One stop shop for NASA Earth Science Data

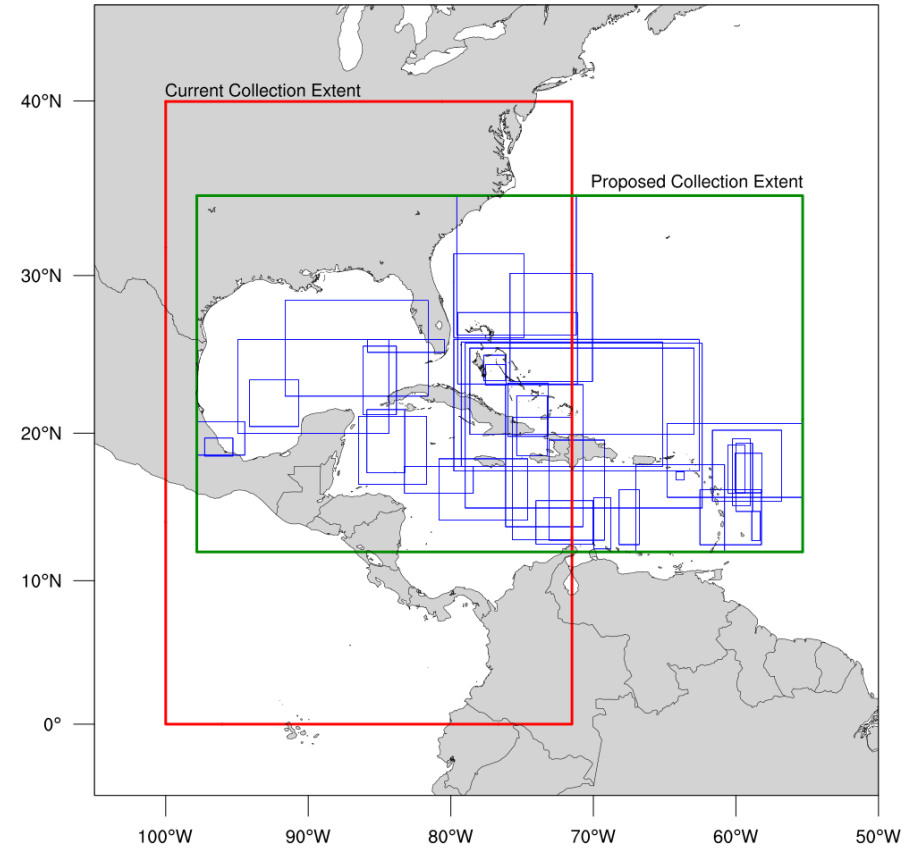
Uses metadata in the CMR to help users find the information they are looking for

Functions best when metadata is complete, consistent, accurate

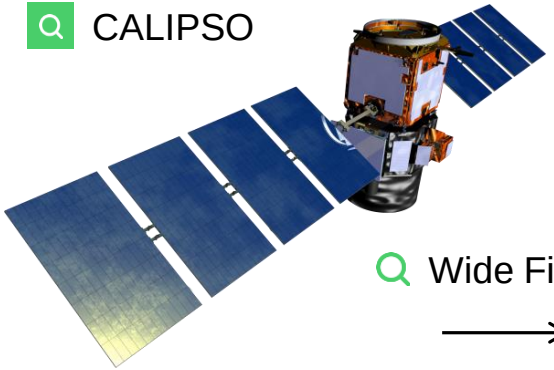


Search and Discovery

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.”

Digital Curation Center, Edinburgh, Scotland

Metadata curation

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

Analysis and Review of CMR (ARC) Team

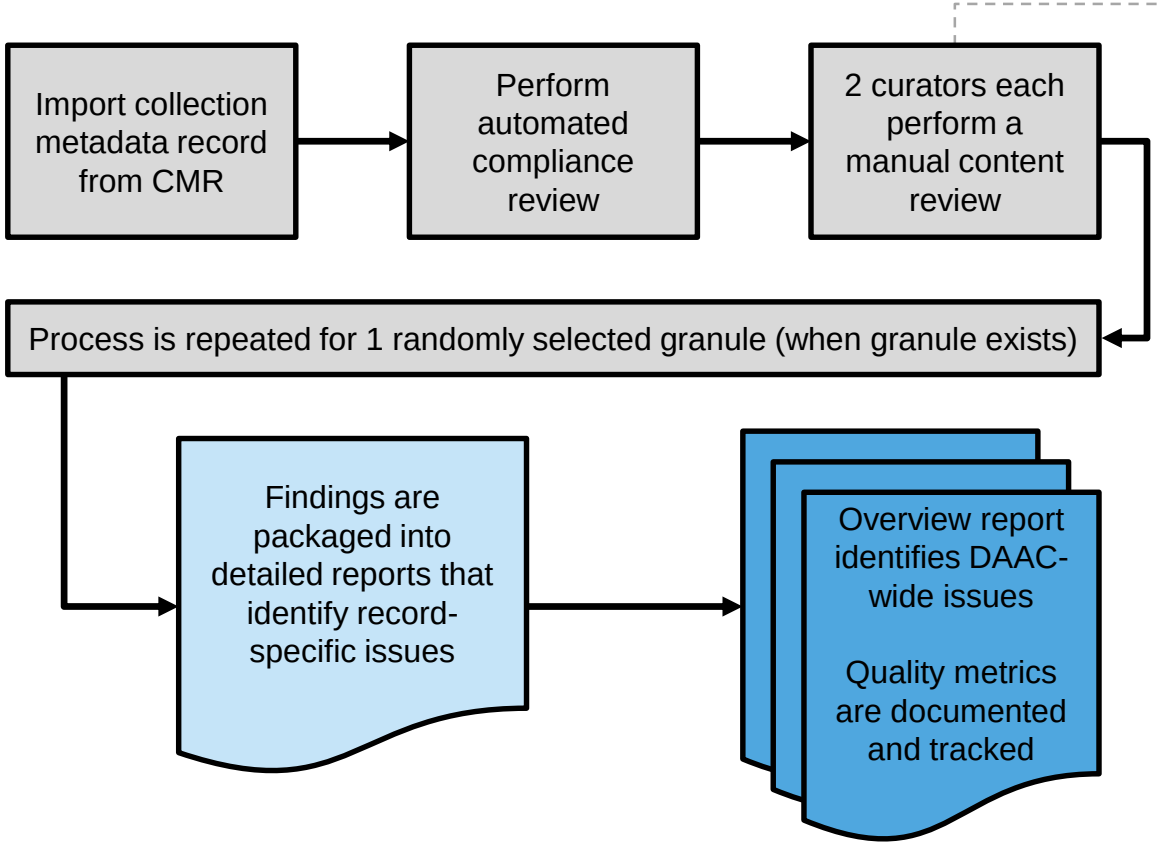
Team is comprised of Earth Science data and metadata specialists

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

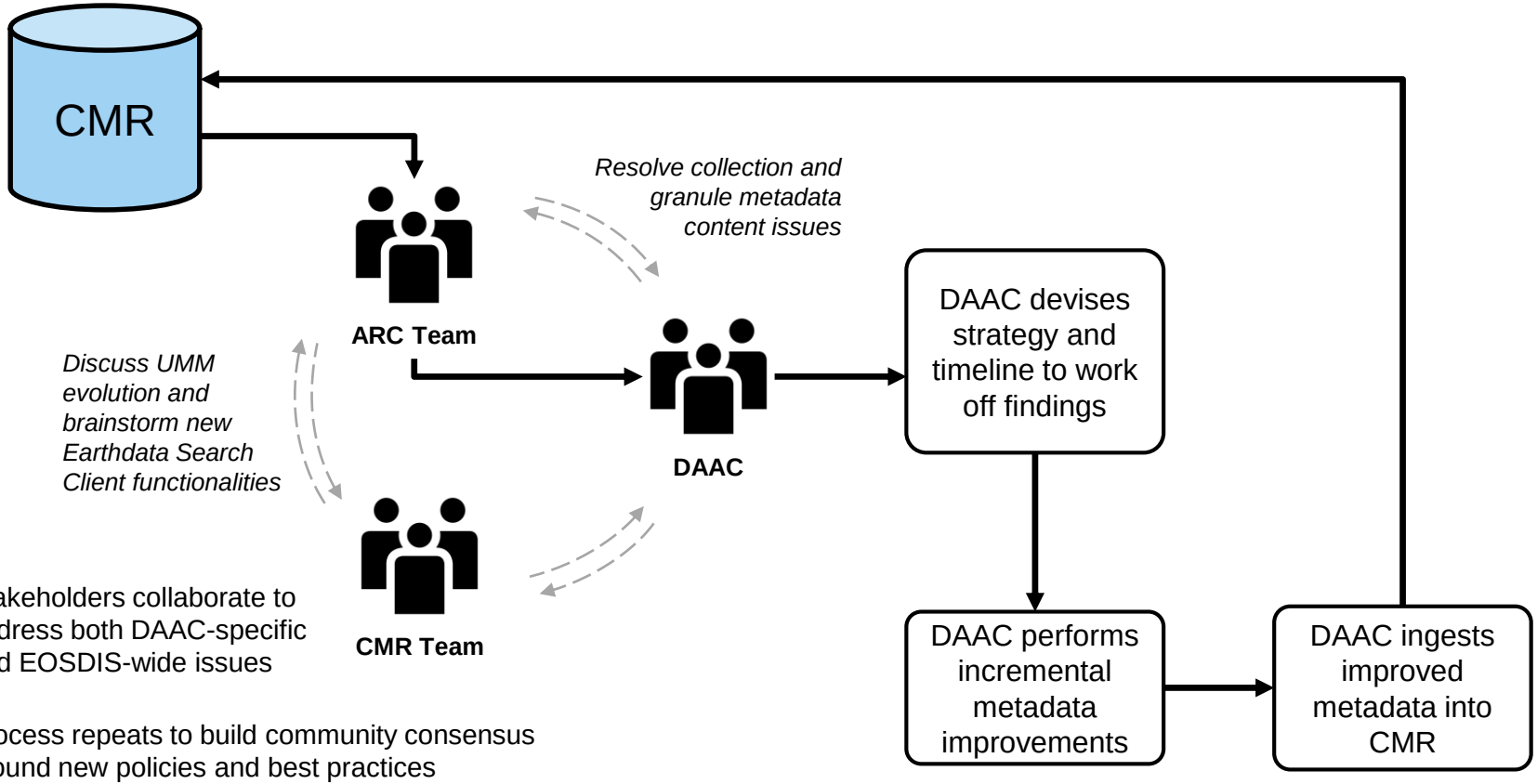


High	• Inaccurate, incomplete, or missing content • Broken URLs and invalid collection-granule relationships
Med	• Revisions of existing content • Addition of new information
Low	• Minor consistency issues

Priority classification scheme

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

ARC Curation Process



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

Process repeats to build community consensus around new policies and best practices

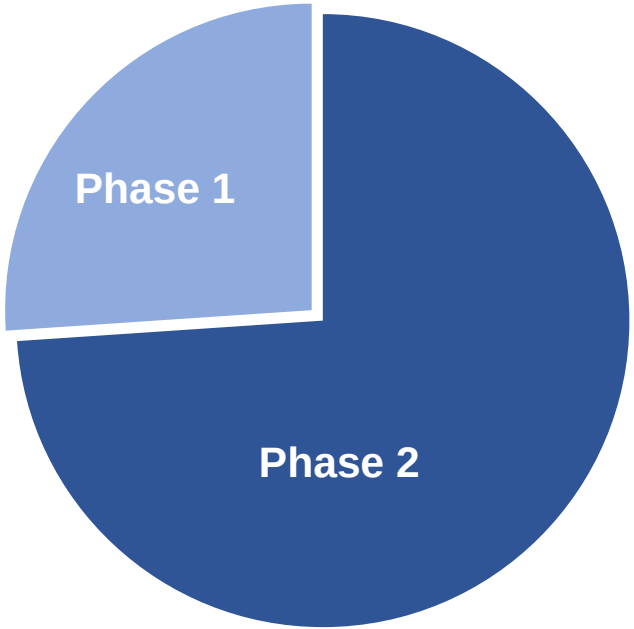
Phase I Metrics

Reviewed over 25% of collection level records in CMR

- Records from all 12 data centers reviewed

Reviewed metadata in 4 dialects

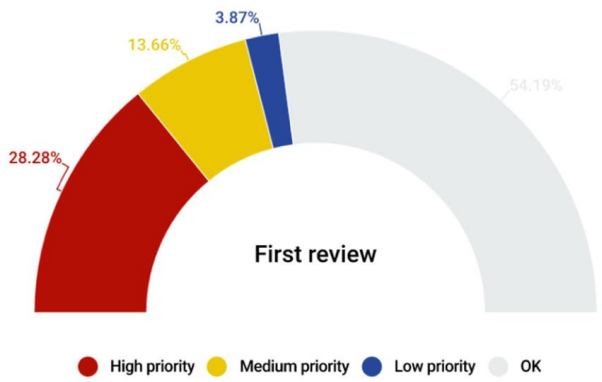
Supported two data centers in the generation of brand new collection and granule metadata



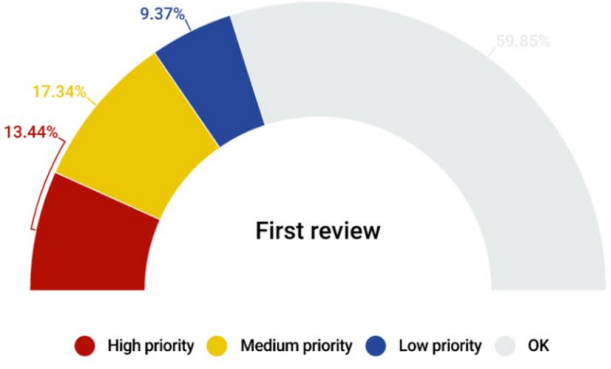
1,959 collections reviewed

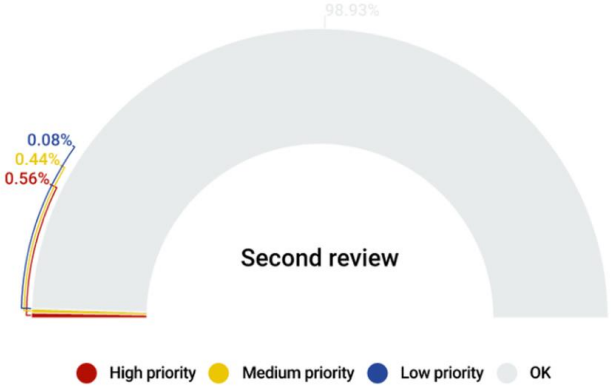
Key Outcomes from Phase I

ORNL



SEDAC

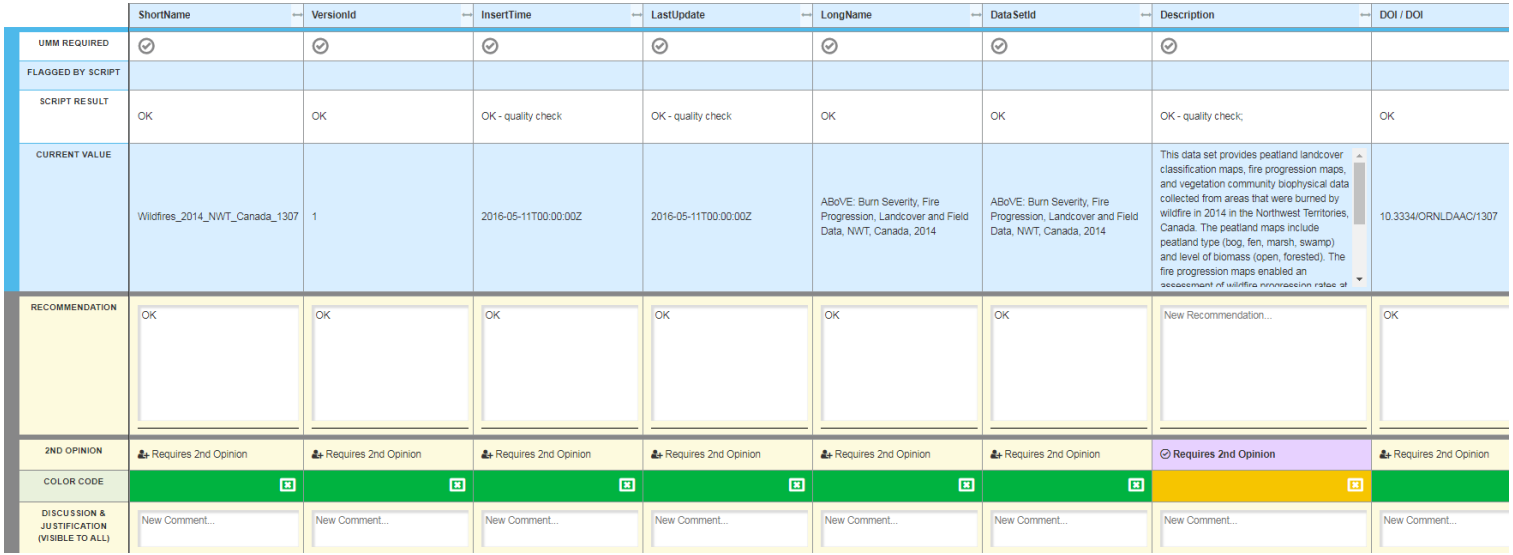




Metadata Curation Dashboard

Goals of the dashboard include:

- Providing an environment for ARC reviewers and metadata curators to collaborate on metadata improvements
- Streamlining communication between ARC and the data centers
- Creating easy to read reports of ARC's metadata findings
- Generating metadata quality metrics



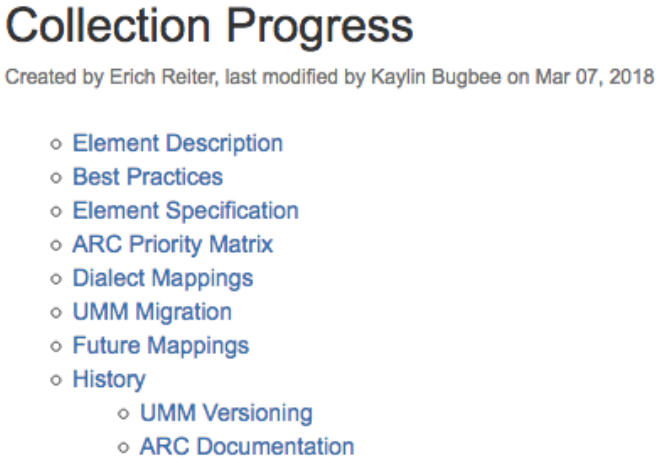
	ShortName	VersionId	InsertTime	LastUpdate	LongName	DataSetId	Description	DOI / DOI
UMM REQUIRED	✓	✓	✓	✓	✓	✓	✓	
FLAGGED BY SCRIPT								
SCRIPT RESULT	OK	OK	OK - quality check	OK - quality check	OK	OK	OK - quality check;	OK
CURRENT VALUE	Wildfires_2014_NWT_Canada_1307	1	2016-05-11T00:00:00Z	2016-05-11T00:00:00Z	ABoVE: Burn Severity, Fire Progression, Landcover and Field Data, NWT, Canada, 2014	ABoVE: Burn Severity, Fire Progression, Landcover and Field Data, NWT, Canada, 2014	This data set provides peatland landcover classification maps, fire progression maps, and vegetation community biophysical data collected from areas that were burned by wildfire in 2014 in the Northwest Territories, Canada. The peatland maps include peatland type (bog, fen, marsh, swamp) and level of biomass (open, forested). The fire progression maps enabled an	10.3334/ORNLDAAC/1307
RECOMMENDATION	OK	OK	OK	OK	OK	OK	New Recommendation...	OK
2ND OPINION	Requires 2nd Opinion	Requires 2nd Opinion	Requires 2nd Opinion	Requires 2nd Opinion	Requires 2nd Opinion	Requires 2nd Opinion	Requires 2nd Opinion	Requires 2nd Opinion
COLOR CODE								
DISCUSSION & JUSTIFICATION (VISIBLE TO ALL)	New Comment...	New Comment...	New Comment...	New Comment...	New Comment...	New Comment...	New Comment...	New Comment...

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



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

Thank You!

Questions?
Contact me at:
Kaylin Bugbee
Kaylin.m.bugbee@nasa.gov

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

NASA

Raytheon