

National Aeronautics and
Space Administration

EARTHDATA

State of ESDIS Search and Discovery Tools

Presentation Date 1/21/2025

Organizer: Valerie Dixon ¹

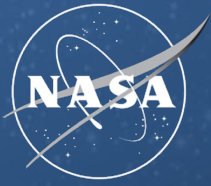
Authors: Ryan Abbott ², Erich Reiter ³,
Christopher D. Gokey ⁴

1-GSFC; 2-Element84, 3-Raytheon, 4-KBR Wyle

©2025 Raytheon Company. All rights reserved

GOVERNMENT RIGHTS NOTICE

This work was authored by employees of Raytheon, KBR Wyle and Element84 under Contract No. 800 SFC21CA001 with the National Aeronautics and Space Administration. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, worldwide license to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, or allow others to do so, for United States Government purposes. All other rights are reserved by the copyright owner.



National Aeronautics and
Space Administration

EARTHDATA

Common Metadata Repository & STAC

Presentation Date 1/21/2025

Ryan Abbott

Search & Discovery Architect



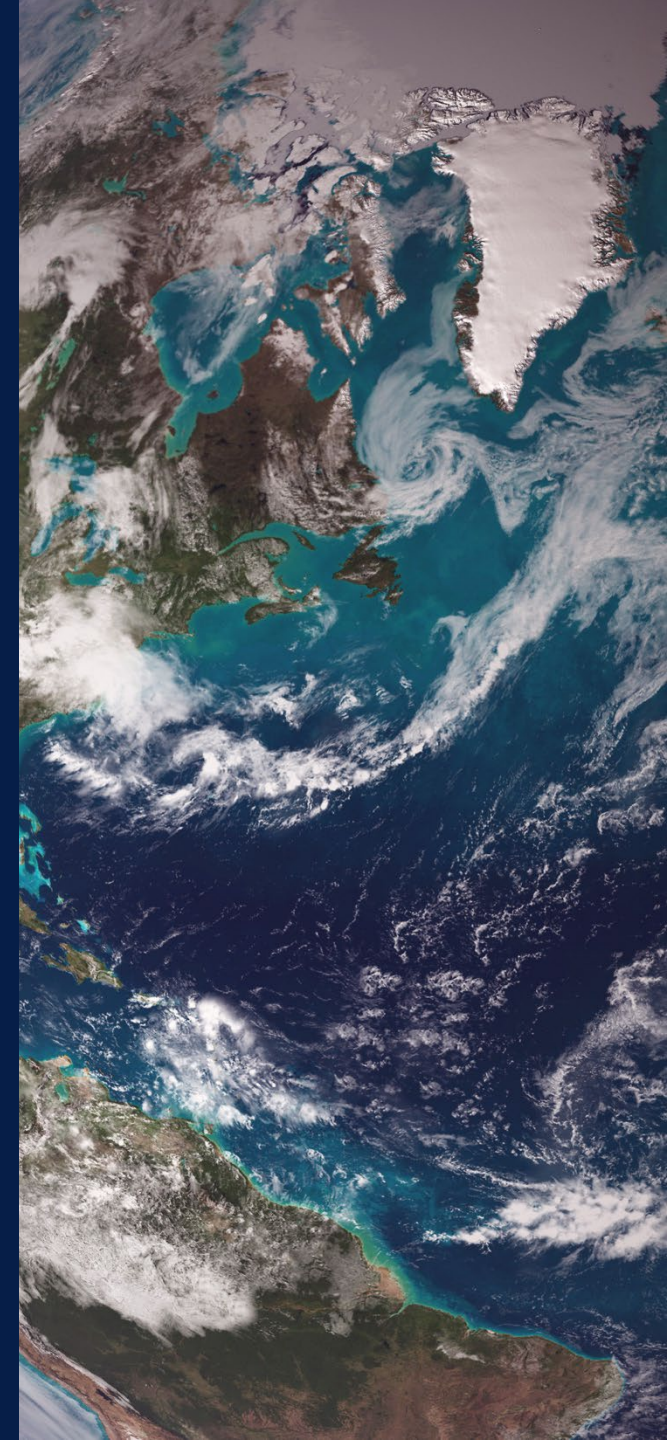
Table of Contents

History

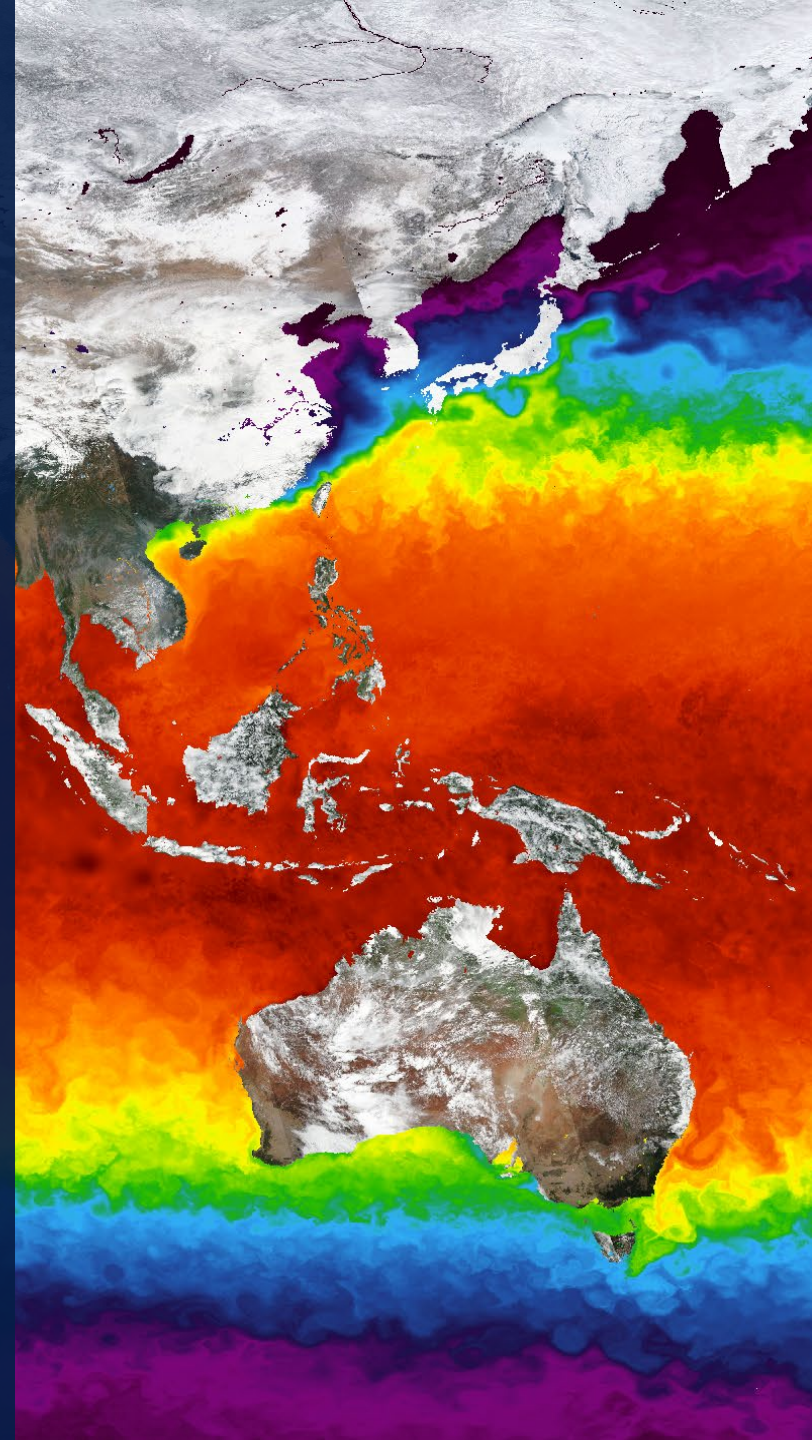
The Fumble

Course Correction

Future



History of CMR STAC

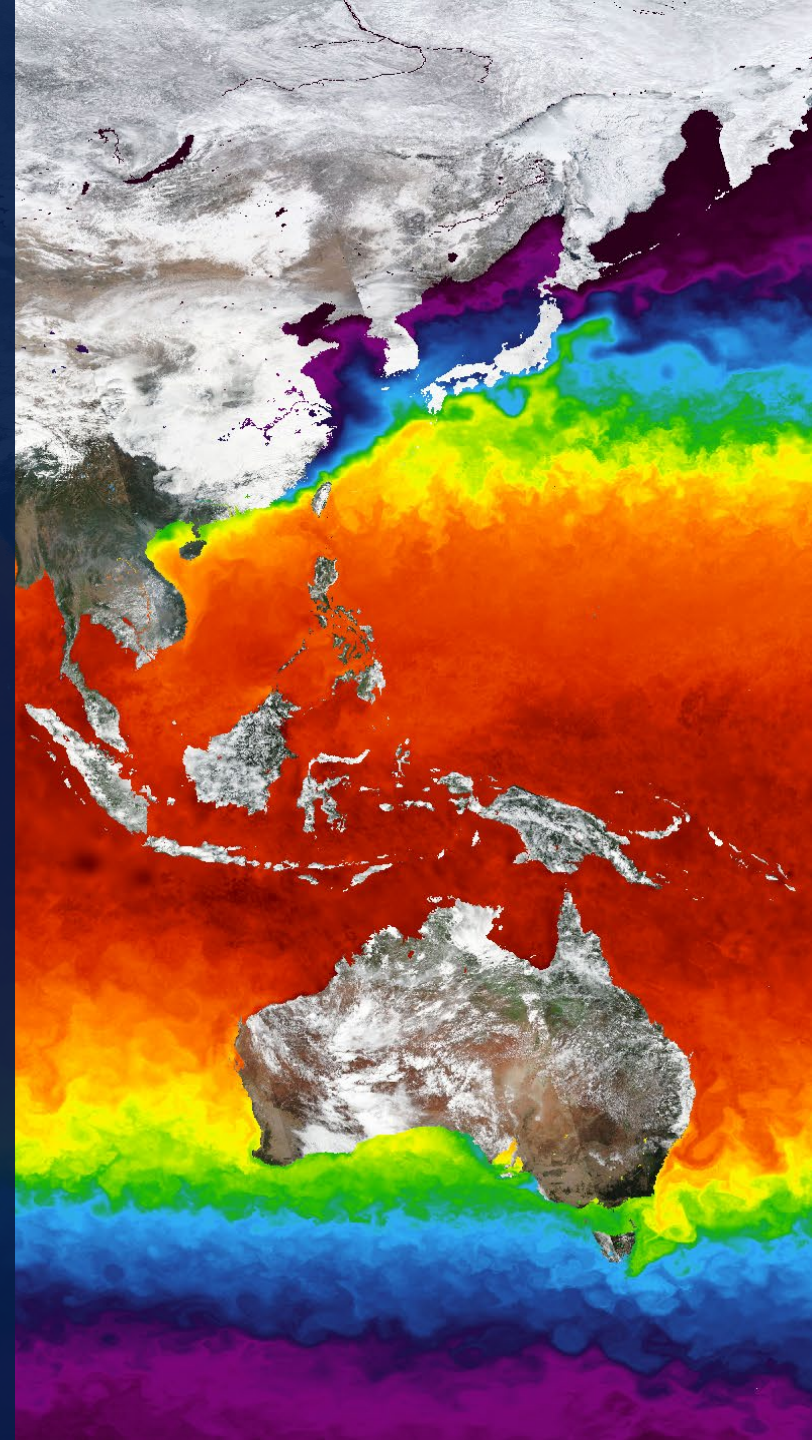


History

- Contracted work during active inception of STAC
 - Combined with minimal knowledge of CMR
- Hindered by functionality of CMR
 - Performance / feature implications
- Ambiguity on best way to offer the API to clients
- Reached an official release



The Fumble

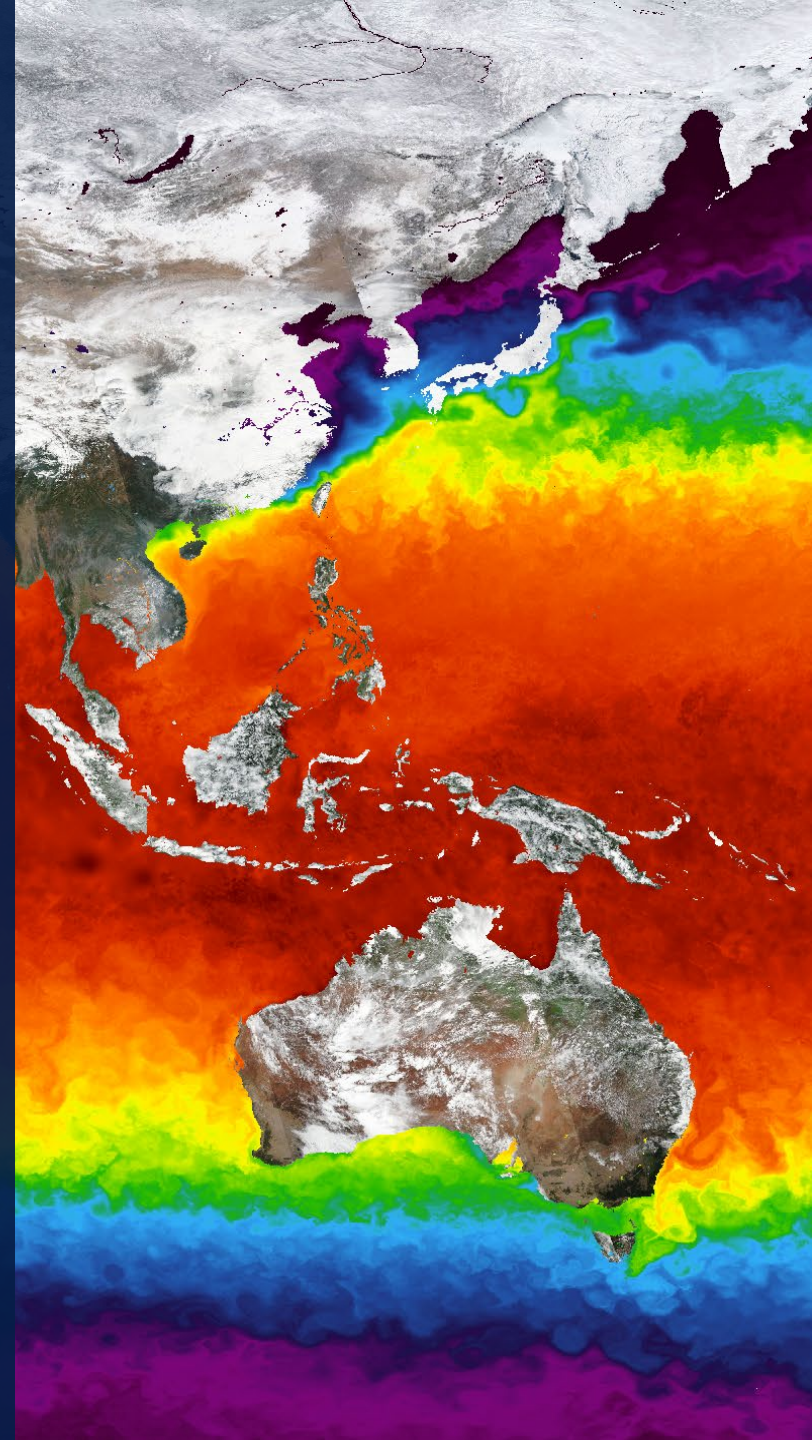


What Happened?

- Experienced some performance challenges
- Development resources became unavailable
- Fell behind use cases and community



Recalibrating

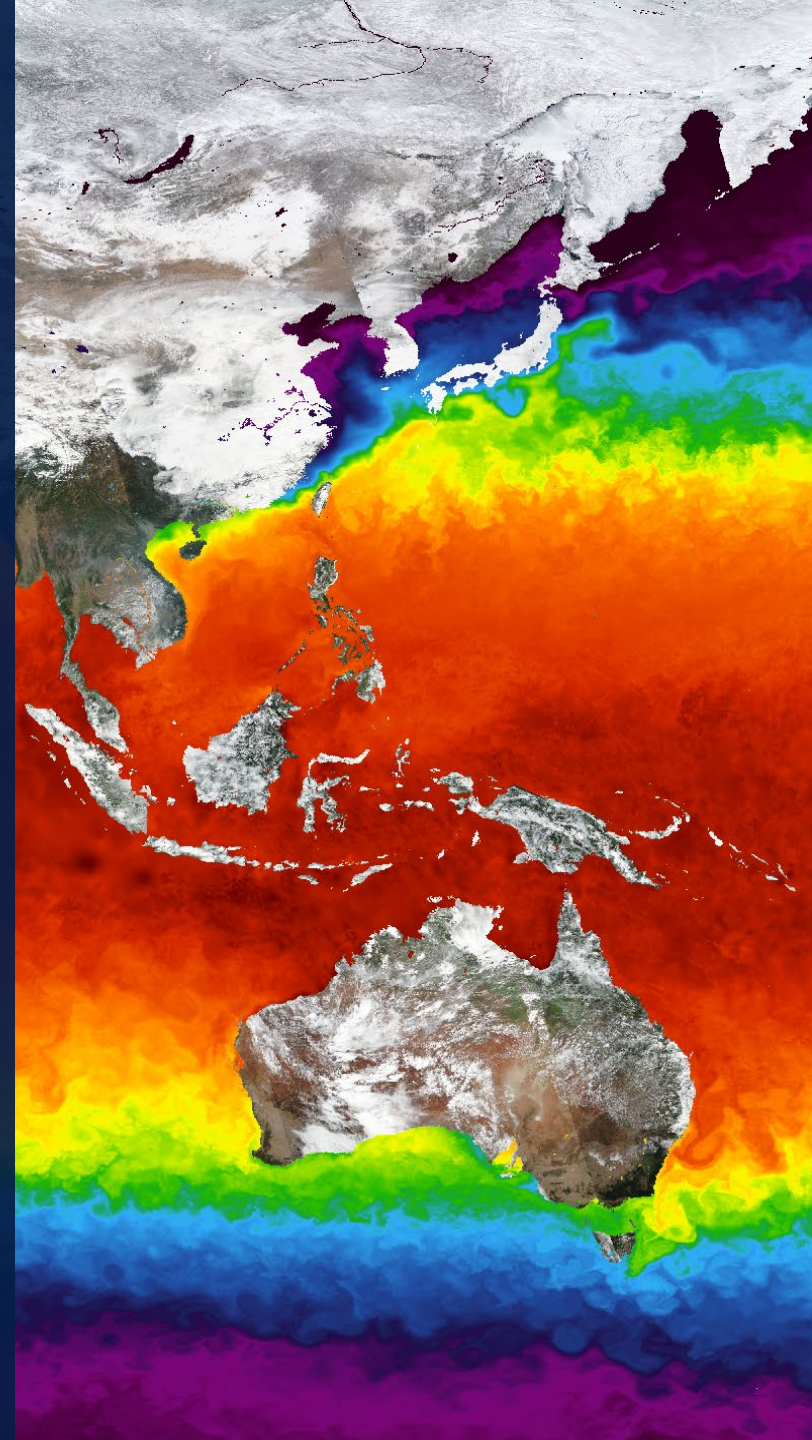


Course Correction

- Made strides in catching up
 - Added searching capabilities
 - Cleaned up many broken use-cases
- Using STAC Browser as a test
- Frequently communicating with community via GitHub
 - <https://github.com/nasa/cmr-stac>



What's Next?



BiG STAC

- Started Prototyping the next step of CMR STAC
- Goals
 - Performantly search 8+ billion records
 - Reduce costs as much as possible
 - Do it *with* the community
 - <https://github.com/nasa/bigstac>

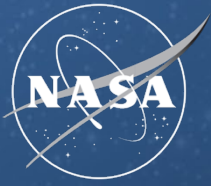


EARTHDATA

earthdata.nasa.gov

Thank you!

This work was supported by NASA/GSFC under Raytheon Company
contract number 80GSFC21CA001



National Aeronautics and
Space Administration

EARTHDATA

SNS Subscription Implementation in the CMR

Presentation Date 1/21/2025

Erich Reiter

CMR Developer

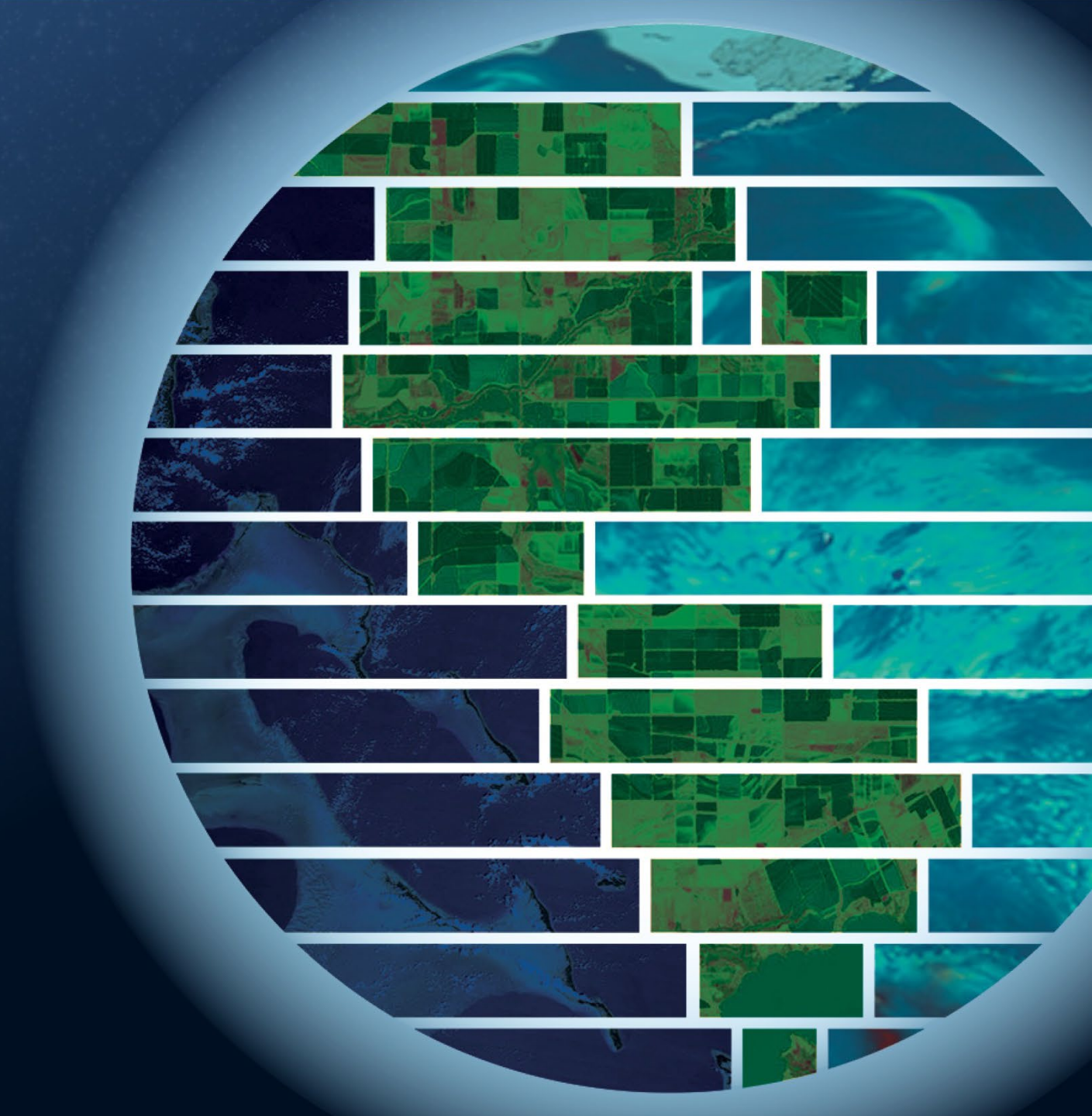


Table of Contents

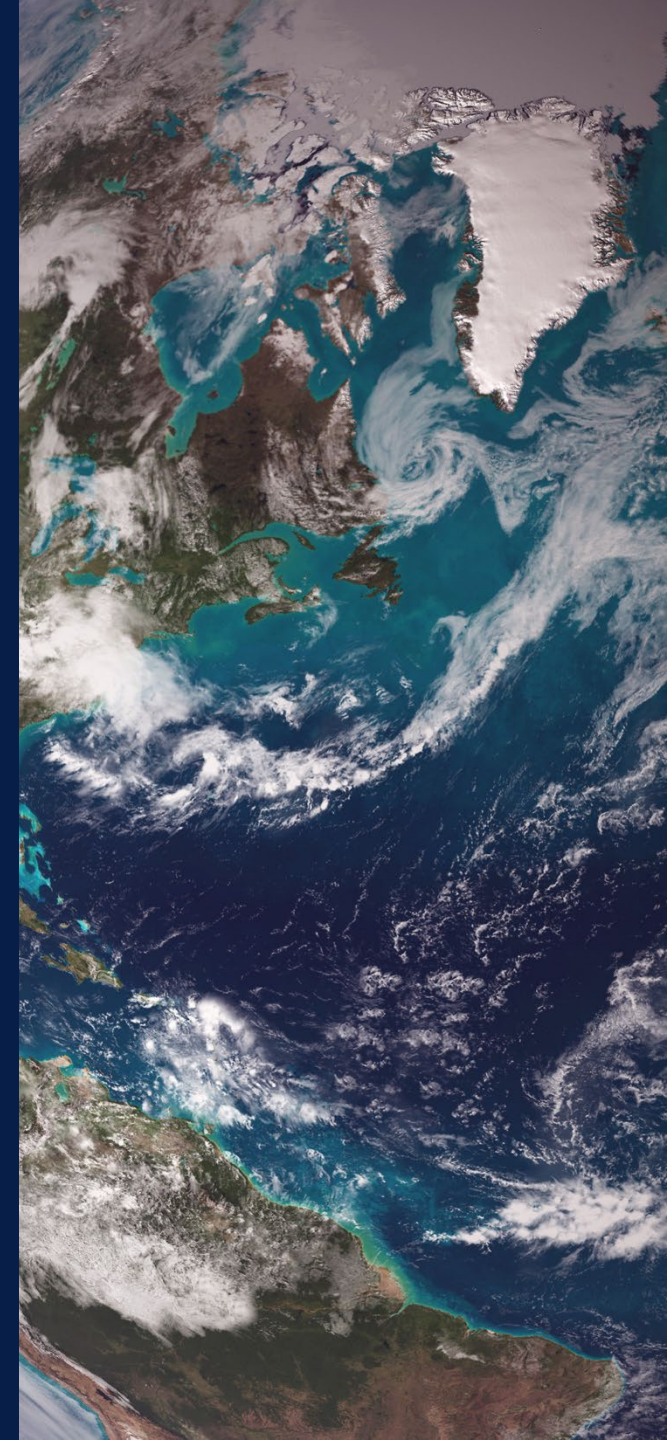
The Problem: Notifications to an AWS Queue

The Solution

The Benefits

The Demo

Resources



The Problem: Notifications to an AWS Queue

- NASA Distributed Active Archive Centers (DAAC) and other Common Metadata Repository (CMR) clients would like a programmable way to subscribe with the CMR to get near real time notifications after observational data files' metadata (granules) have been ingested into the CMR.
- A periodic search subscription solution exists in the CMR that provides the search results to the client in an email.
 - There is a lot of overhead to programmatically triggering processes off of emails
 - Clients do not want to parse emails
 - While it is periodic, the notification may not be as timely as some desire.
- Ideally the DAACs would like to have the notifications delivered to an Amazon Web Service (AWS) Simple Queue Service (SQS) queue so they can pick up the notifications and use them to do more processing.
- The notifications need to contain a few key pieces of information; granule identifiers and the granule's URL.
- A firehose solution where notifications for all ingests are delivered is not desirable nor useful.
- Some DAACs poll the CMR for new data on a periodic basis.



The Solution

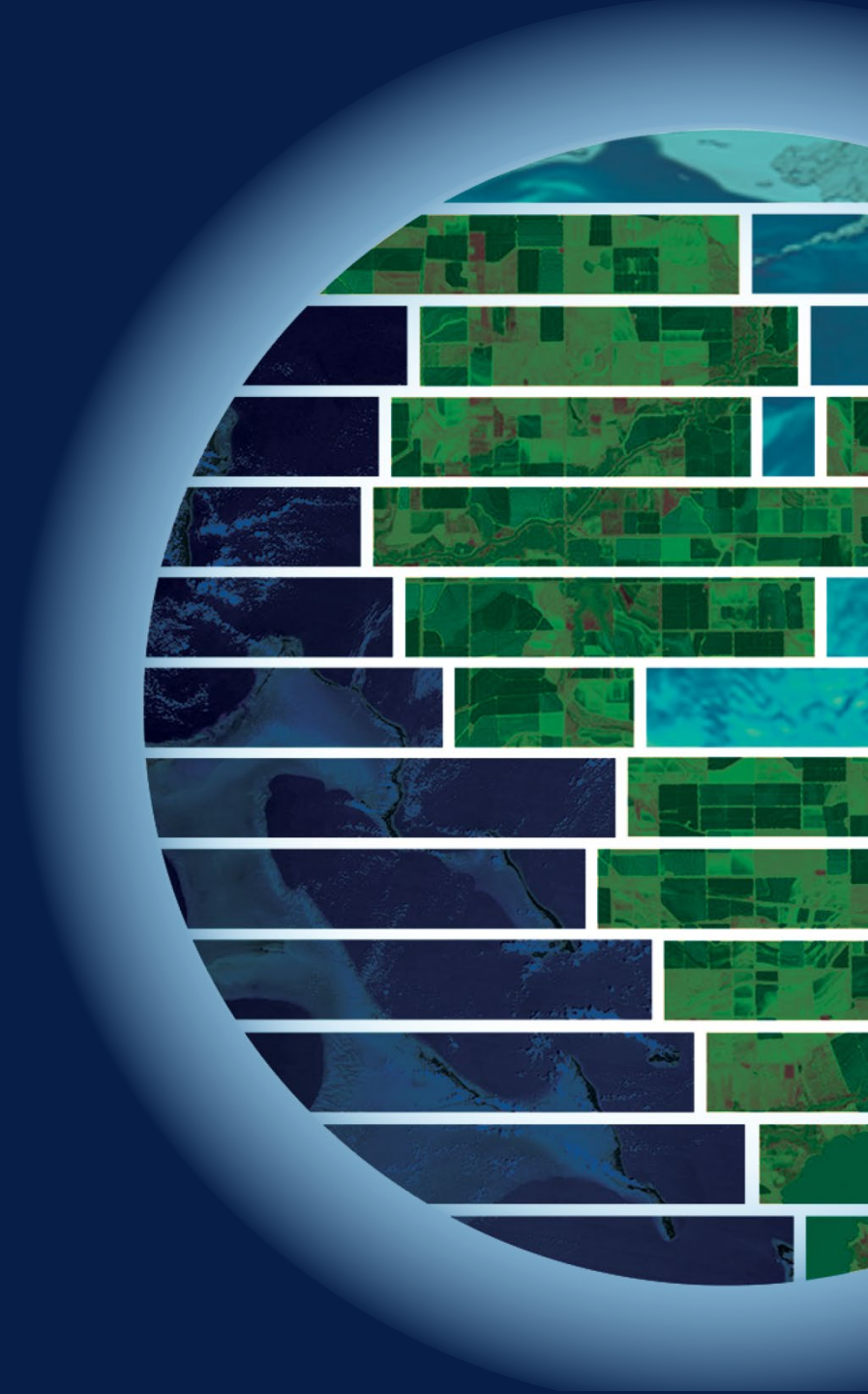
Allow CMR clients to specify an AWS SQS queue Amazon Resource Name (ARN) endpoint when creating a subscription, so that the CMR can send update notifications to the queue.

- Use the current CMR subscription API to allow users to create/update/delete subscriptions.
 - Updated the UMM-Sub schema (version 1.1.1) to add new fields:
 - Method - designates between search (Batch Notification) subscriptions and ingest (Near-Real-Time) subscriptions
 - Mode - describes whether the notification is for New (ingested for the first time into the CMR) granules, Updated granules, or Deleted granules
 - EndPoint - describes where notifications get sent. At this time only AWS SQS ARN's (Amazon Resource Names) are allowed.
 - Schema definition: <https://cdn.earthdata.nasa.gov/umm/subscription/v1.1.1/umm-sub-json-schema.json>
 - CMR Subscription API: <https://cmr.earthdata.nasa.gov/ingest/site/docs/ingest/api.html#subscription>



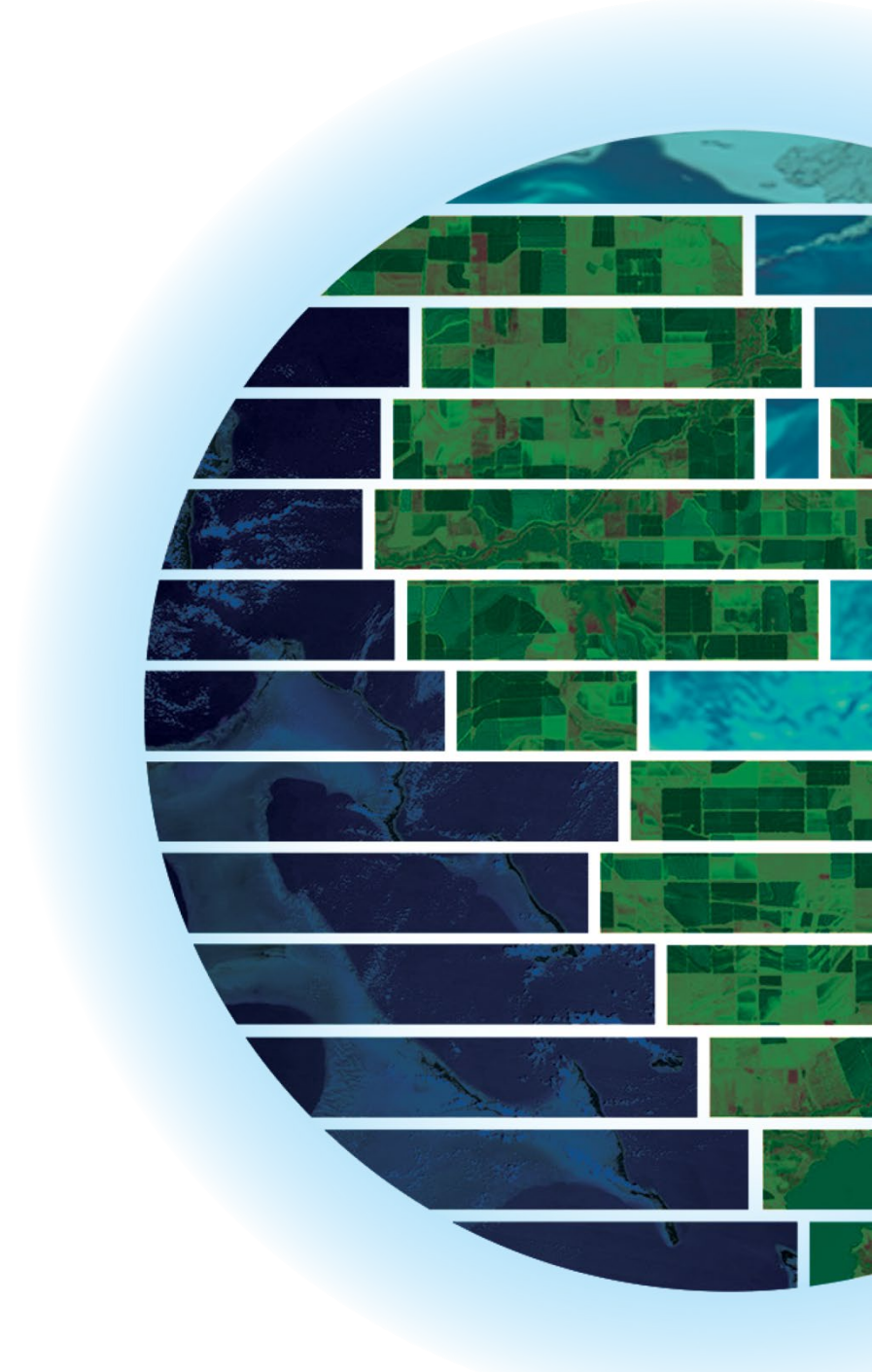
The Benefits

- Reduces load on search infrastructure by reducing reoccurring searches
- Reduces the need for DAACs to create polling software to get the information they need.
- Near real time notification
- No need to setup and parse email messages.
- Uses infrastructure that DAACs currently use, making it easy for them to integrate the notifications into their existing processes.



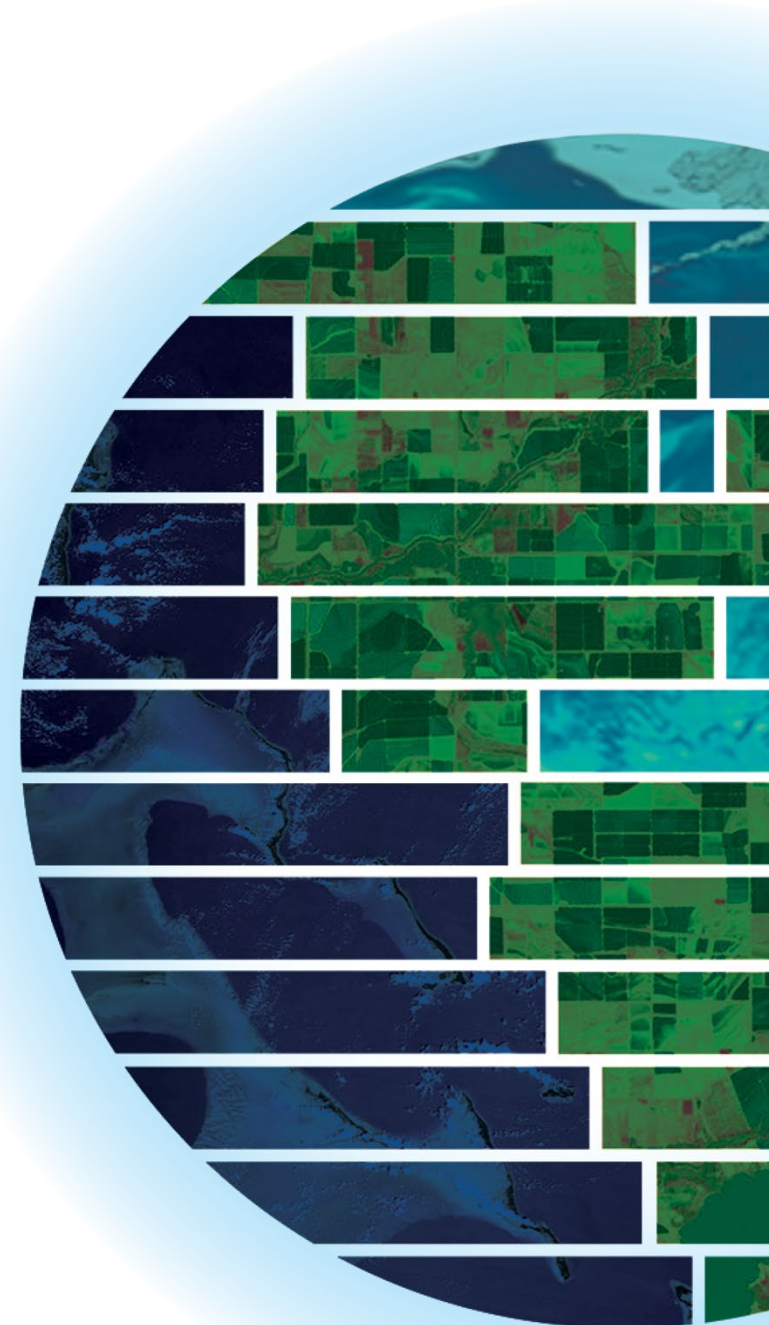
The Demo

- Create Client Queue
 - Fix the Access Policy
- Create subscription record
- Send Subscription to the CMR API
- Ingest a granule to the CMR
- Show the notification message.



Resources

- Schema definition
 - <https://cdn.earthdata.nasa.gov/umm/subscription/v1.1.1/umm-sub-json-schema.json>
- CMR Subscription API
 - <https://cmr.earthdata.nasa.gov/ingest/site/docs/ingest/api.html#subscription>
- Client how-to:
 - <https://wiki.earthdata.nasa.gov/display/CMR/CMR+Ingest+Subscriptions>

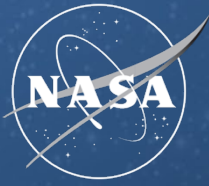




Call to Action

CONTACT

erich.e.reiter@nasa.gov



National Aeronautics and
Space Administration

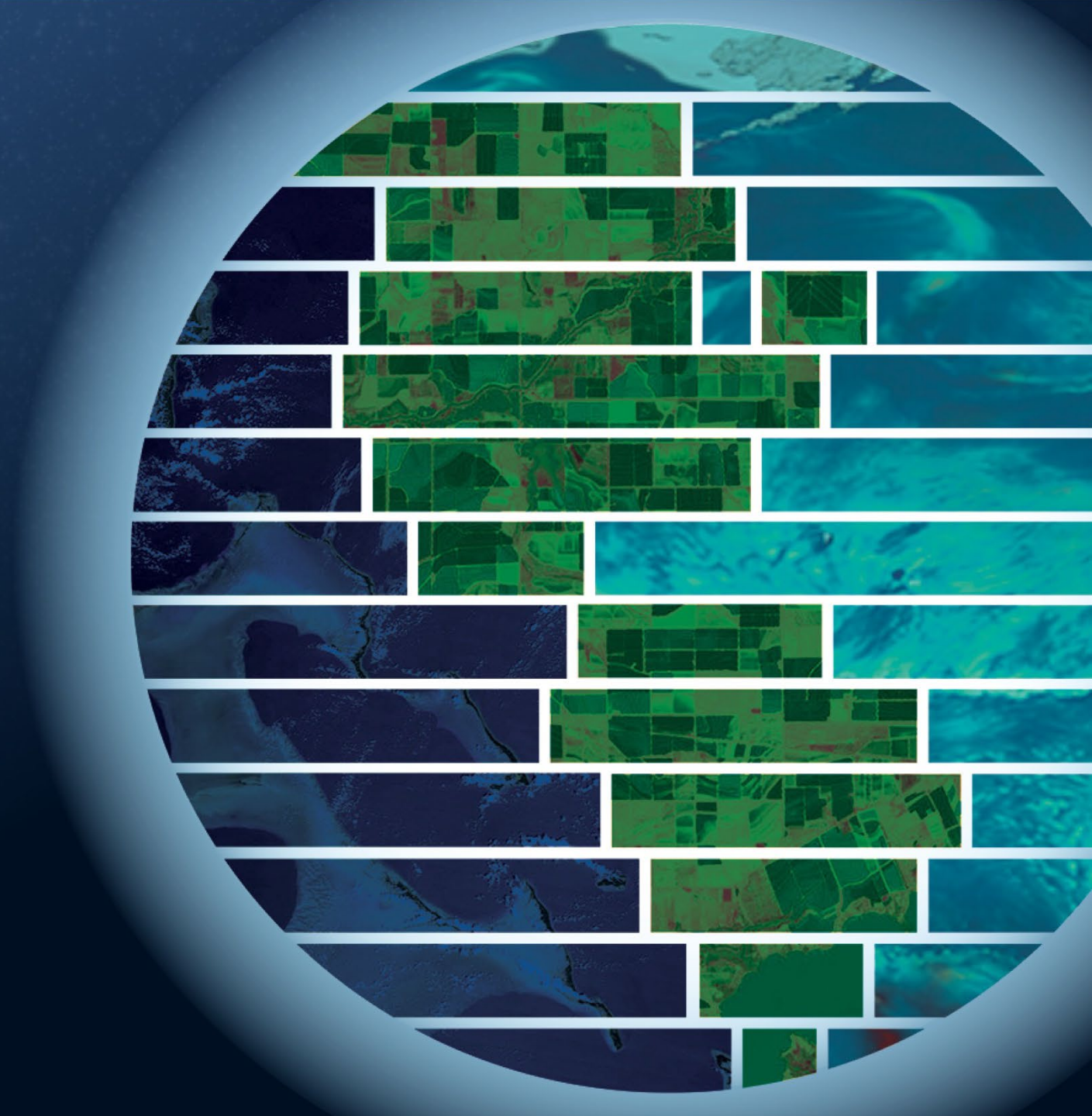
EARTHDATA

Earthdata Search New Features and Improvements

Presentation Date 1/21/2025

Ryan Abbott

Search & Discovery Architect



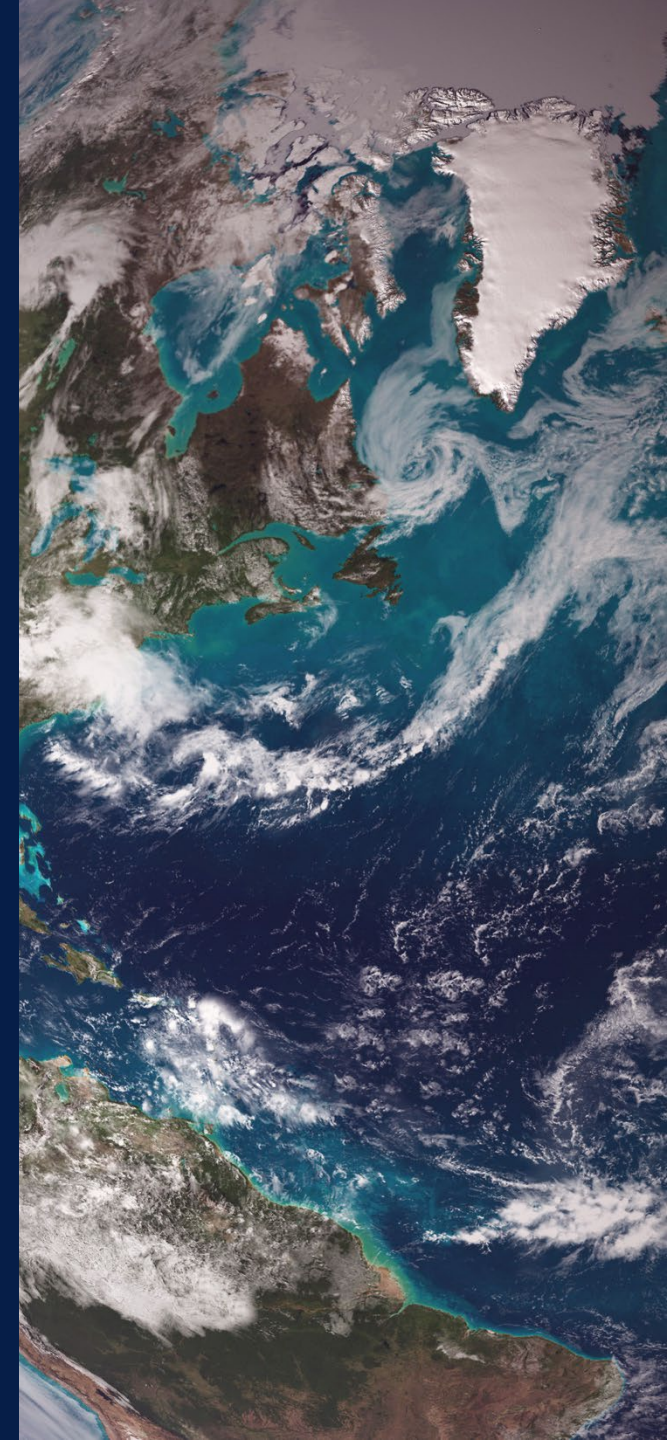
Agenda

Introduction to Earthdata Search

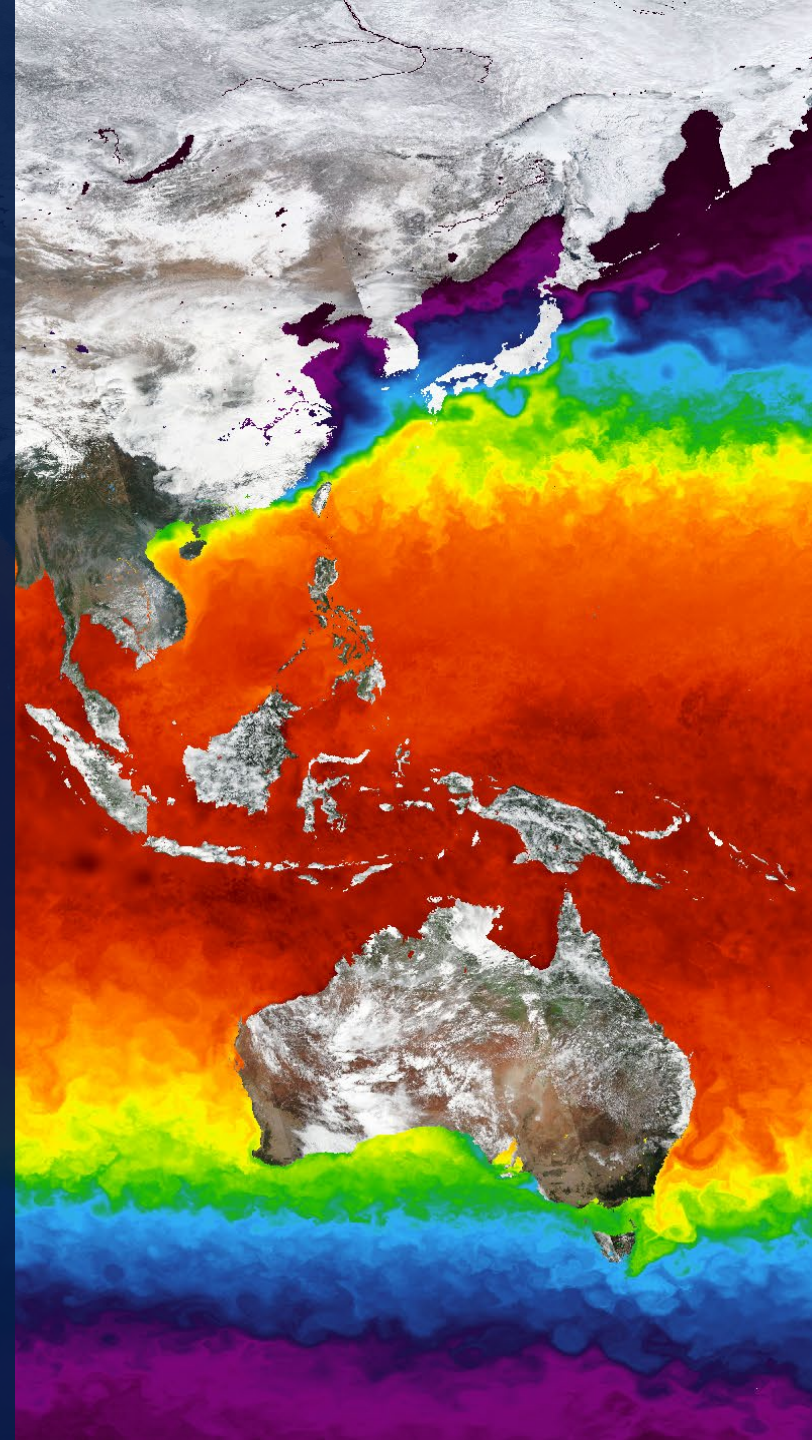
Overview of New Features / Live Demo

More Improvements Coming Soon

Q&A



Introduction to Earthdata Search



Earthdata Search

- Earthdata Search is a web application that enables users to discover and access more than 50,000 Earth Observation data collections from NASA's Earth Observing System Data and Information System (EOSDIS), as well as U.S. and international agencies across the Earth science disciplines.
- Earthdata Search seeks to ease the technical burden on data users by providing a high-quality tool that makes it simple to search for data files by date and spatial area, preview browse images, and download or submit requests for data files, with customization for select data.

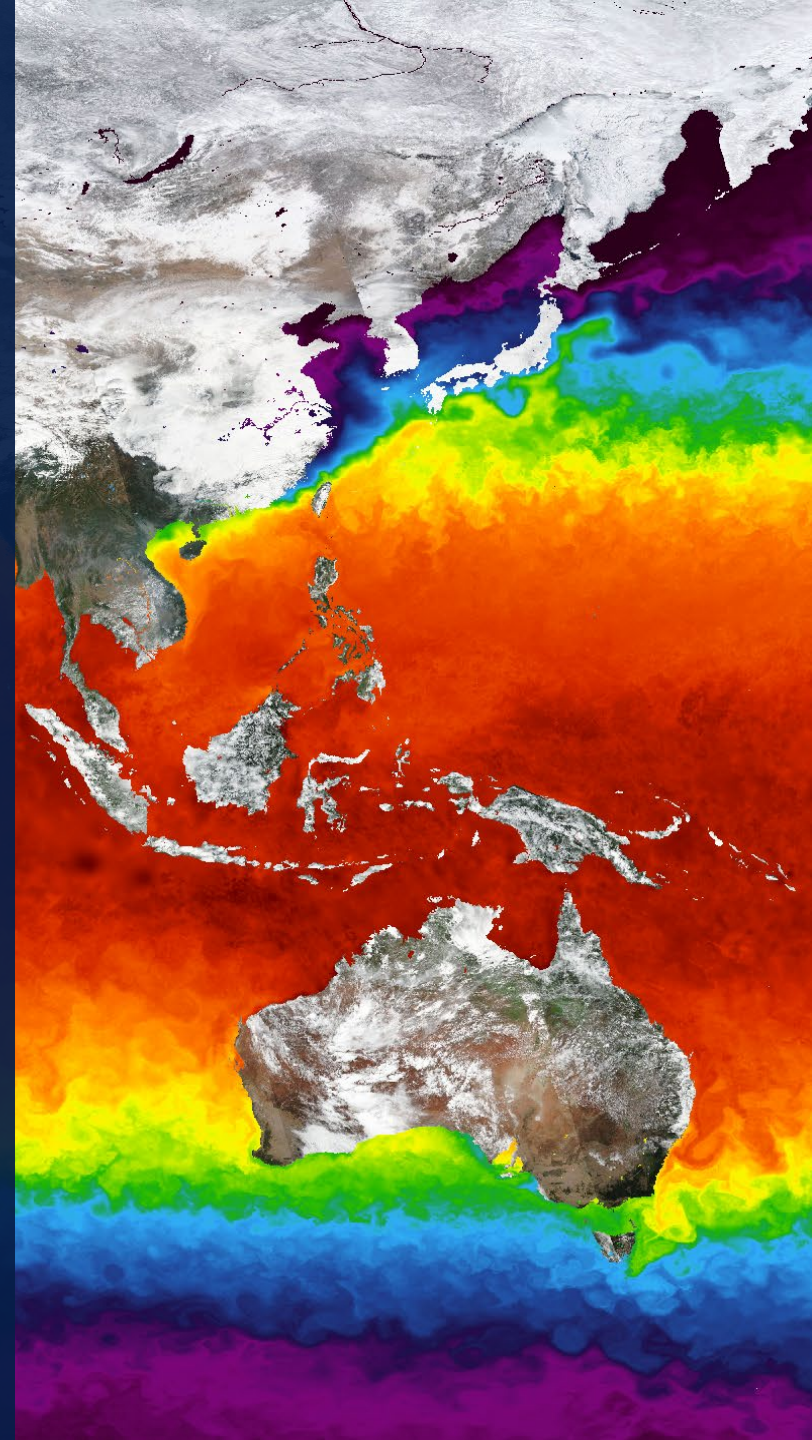
<https://search.earthdata.nasa.gov/search>



National Aeronautics and
Space Administration



Overview of New Features / Live Demo



New Features and Improvements

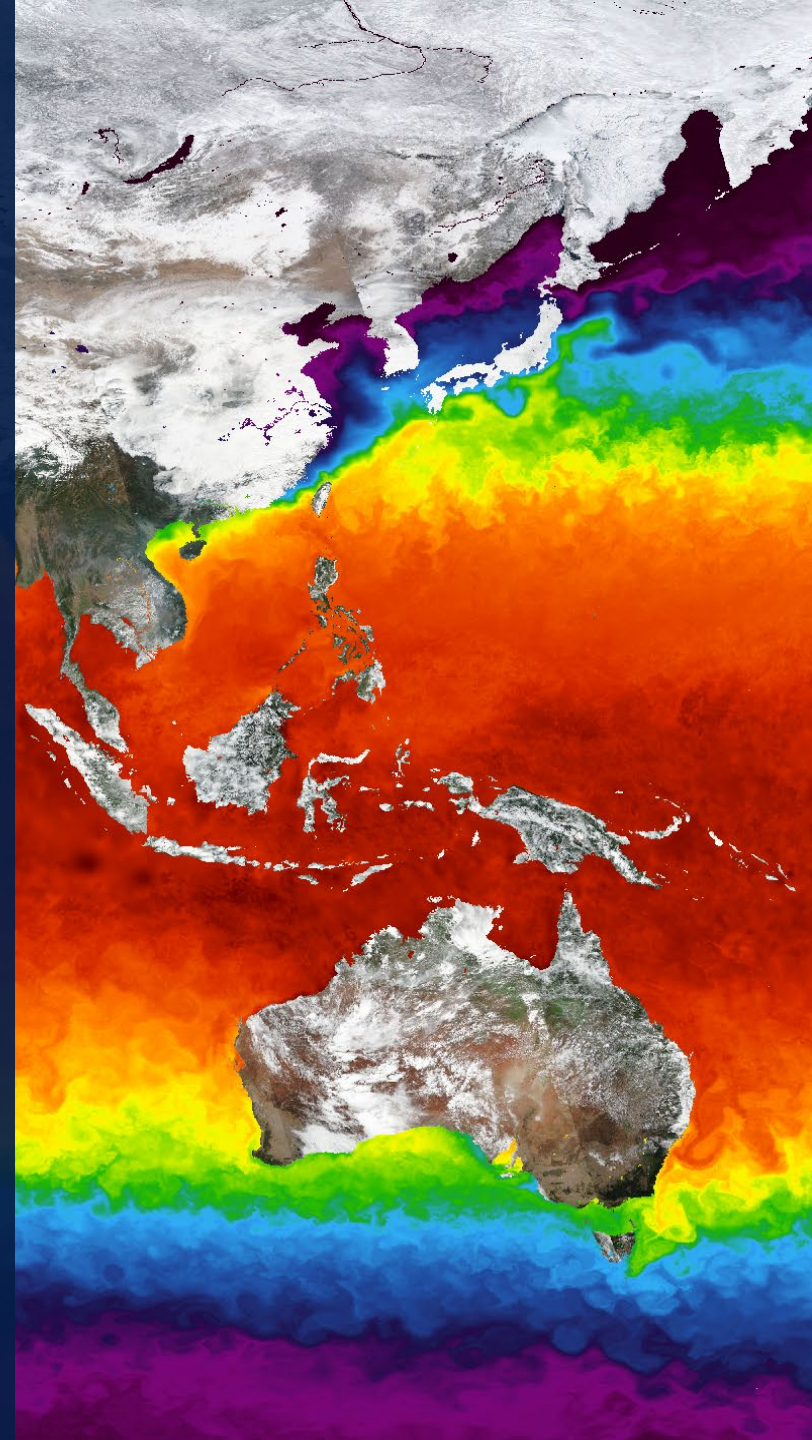
- **Earthdata Search Tour** – context-aware walk-through of core functionality
- **Earthdata Download Application** – enables simplified bulk download of granules
- **Analysis-in-Place** - downloadable Jupyter notebooks for cloud-enabled collections
- **CloudFront Caching** – improved page load performance through edge caching
- **SWODLR Service Integration** - SWOT L2 raster data product generation via SWODLR
- **Concatention Services Integration** – combined data product generation via Harmony concatenation services
- **Portal Browser** – browse, discover, and navigate pre-configured subsets of data
- **Granule Preview Display** – view/select from list of all available previews for a granule



National Aeronautics and
Space Administration



More Improvements
Coming Soon

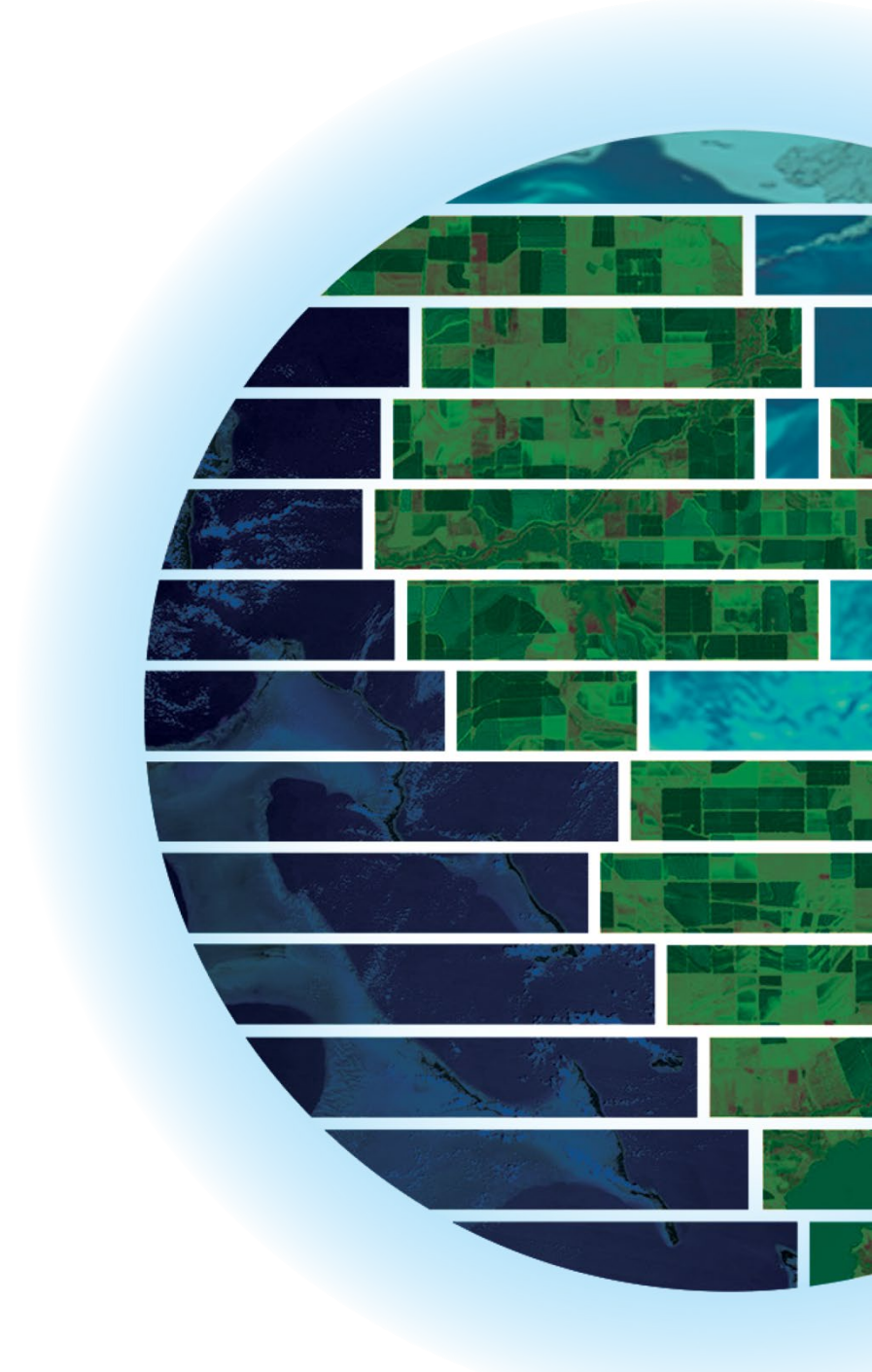


Coming Soon

- **Search and Discovery Tool Convergence**
- **Simplified UI**
- **GIBS Layer Selection**
- **Saved Searches**
- **Increased Zoom Levels on Map**
- **Tutorials**



National Aeronautics and
Space Administration





Please reach out and share your feedback!

Send an email to Earthdata Search Support:
edsc-support@nasa.gov

CONTACT

alicia.m.aleman@nasa.gov

WEBSITE

<https://search.earthdata.nasa.gov/search>

SLACK

@Alicia Aleman / ESIP

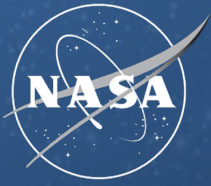
@alicia.m.aleman / ESDIS

EARTHDATA

earthdata.nasa.gov

Thank You for Watching

This work was supported by NASA/GSFC under Raytheon Company
contract number 80GSFC21CA001



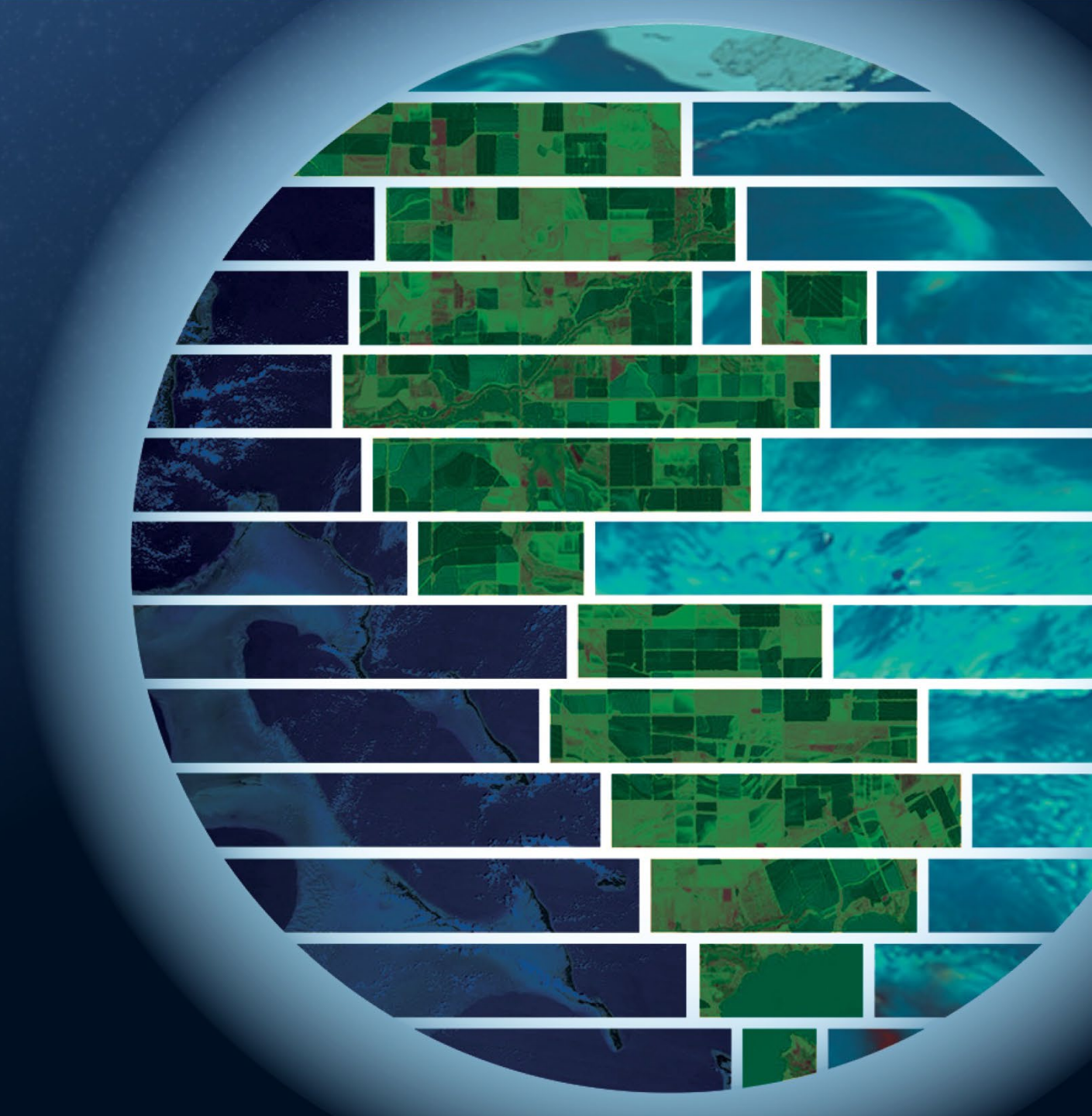
National Aeronautics and
Space Administration

EARTHDATA

Improvements to the Keyword Management System

Presentation Date 1/21/2025

Christopher D. Gokey
Sr. Software Developer



Agenda

What is KMS?

Current State of KMS

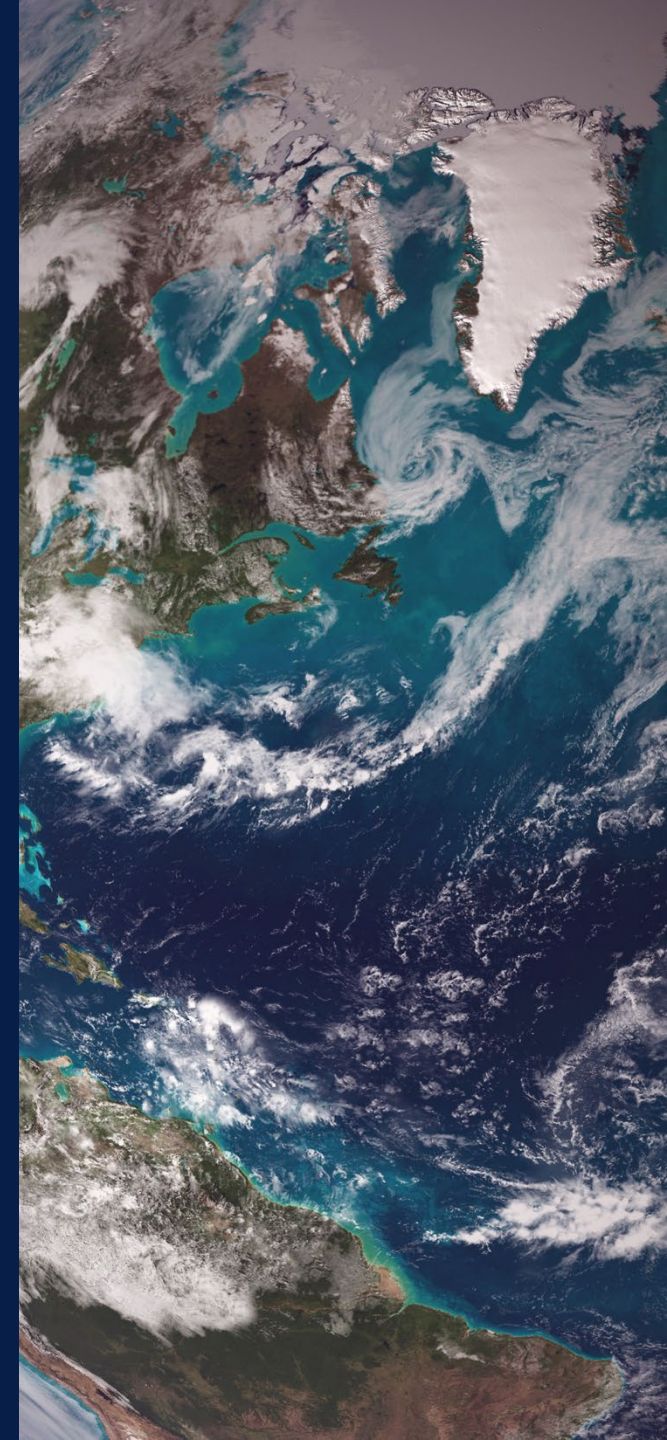
Early Prototype using RDF Database

Building on SKOS

Ontologies/Inference

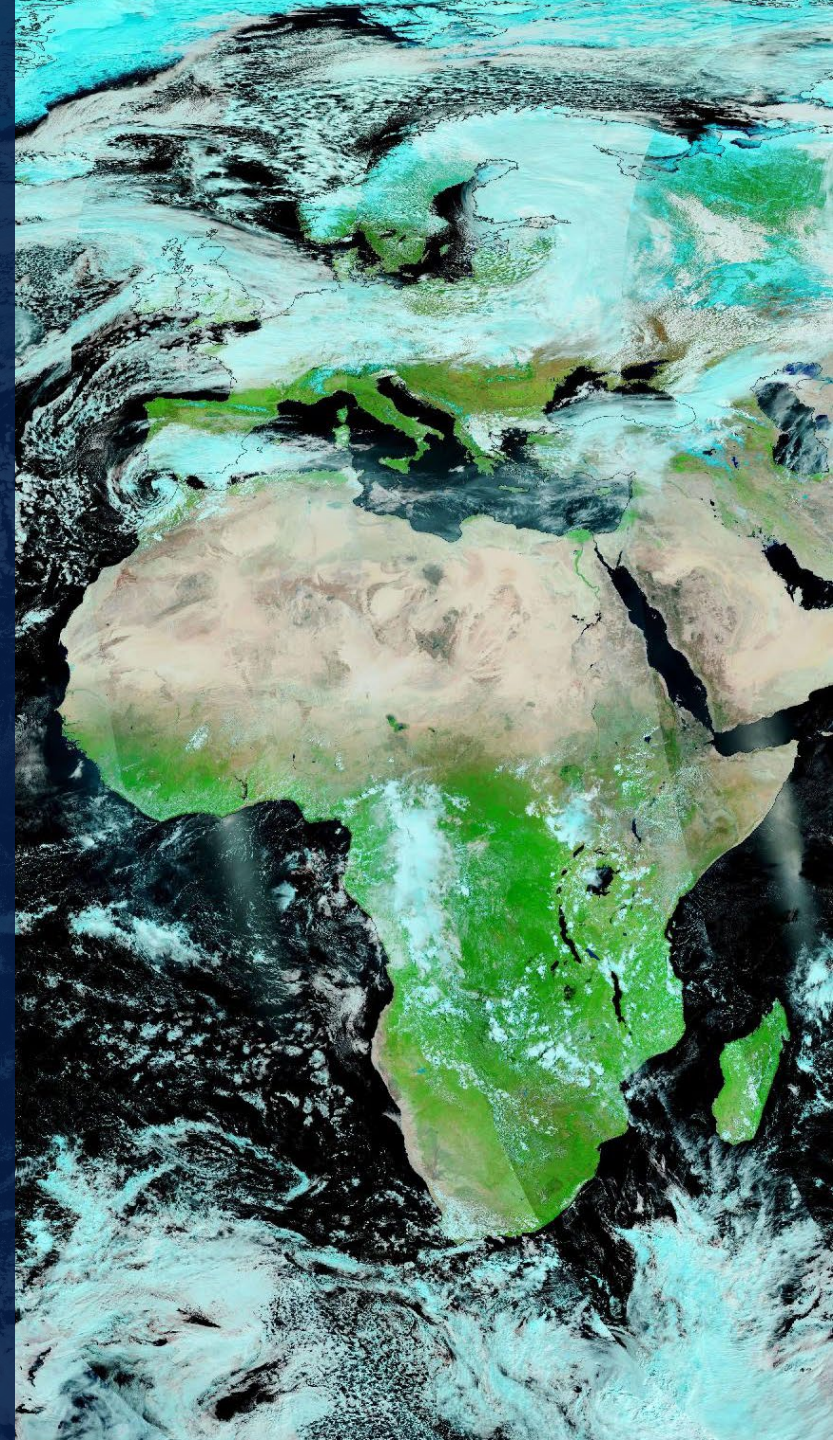
Machine Learning

Demo/Questions



What is KMS?

“KMS” stands for Keyword Management System, which is a set of applications which manages a controlled vocabulary thesaurus used to index and categorize earth science data and services within the CMR database, allowing for consistent searching across different terms related to the same concept; essentially, it is a standardized way to label research topics within CMR system.



Web Service and UI Component

The Global Change Master Directory (gcmd.nasa.gov) Keyword Management Service
Software: Version: 3.1.0 Compiled: 2024-09-29 04:13
Documentation: <https://wiki.earthdata.nasa.gov/display/ED/KMS+User%27s+Guide>

URIs:

- Methods: N/A; URL Path: /kms/status; Parameter Options: ?format=[xml|html];
- Methods: GET; URL Path: /kms/concept_fullpaths/concept_uuid/{conceptId}; Parameter Options: ?None;
- Methods: GET, PUT, DELETE; URL Path: /kms/concept/{conceptId}[.{format}]; Parameter Options: ?format=[rdflxml]&version={label};
- Methods: GET; URL Path: /kms/concept_schemes; Parameter Options: ?version={label};
- Methods: GET, POST; URL Path: /kms/concepts/concept_scheme/{conceptScheme}; Parameter Options: ?format=[rdflxml|json|csv]&version={label}&page_num=&page_size=;
- Methods: GET; URL Path: /kms/concepts/concept_scheme/{conceptScheme}/pattern/{pattern}; Parameter Options: ?format=[rdflxml]&version={label};
- Methods: GET; URL Path: /kms/concepts; Parameter Options: ?format=[rdflxml]&version={label};
- Methods: GET; URL Path: /kms/concepts/root; Parameter Options: ?format=[rdflxml]&version={label};
- Methods: GET; URL Path: /kms/concepts/pattern/{pattern}; Parameter Options: ?format=[rdflxml]&version={label};
- Methods: GET; URL Path: /kms/concept/short_name/{short_name}; Parameter Options: ?version=&scheme=;
- Methods: GET; URL Path: /kms/concept/alt_label/{alt_label}; Parameter Options: ?version=&scheme=;
- Methods: GET; URL Path: /kms/concept_versions/version_type/{versionType}; Parameter Options: ?None;

Capabilities in XML

The screenshot displays the GCMD Keyword Manager interface. At the top, it shows the current released keyword version as 19.8. The main area is the 'Keyword Editor' for a keyword with UUID 77397026-09c9-44e0-b85f-77b2bc9b1630. The keyword is 'Science Keywords>EARTH SCIENCE>ATMOSPHERE' with a preferred label of 'AIR QUALITY'. The definition is 'The study of air pollutants in the atmosphere.' The interface includes a sidebar with a tree view of keyword categories, a top navigation bar with buttons for 'New Selection', 'Generate Report', 'Workflow Document', and 'Publish New Keyword Version', and a bottom status bar indicating the responsible NASA official is Stephen Berrick.

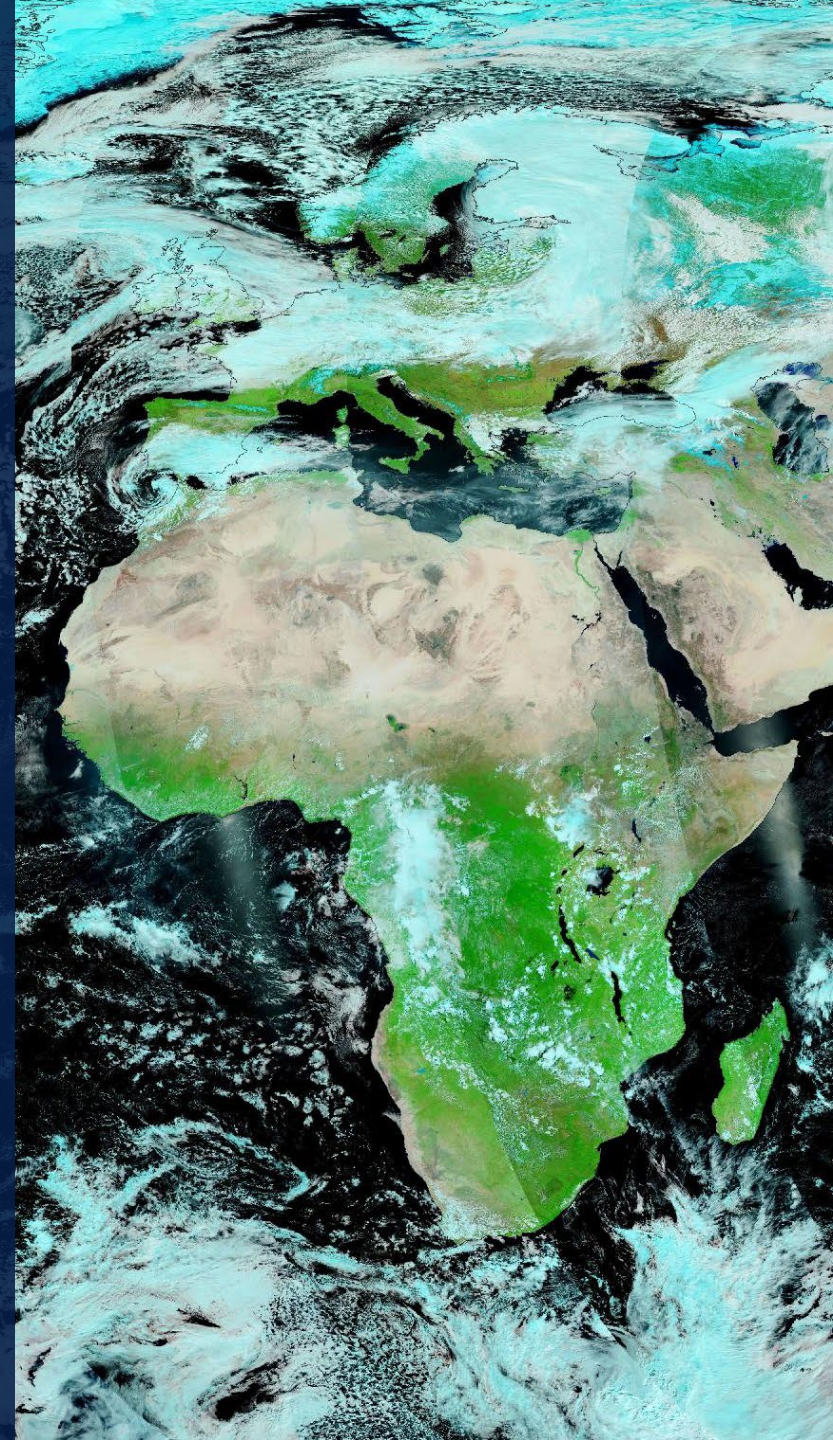


National Aeronautics and
Space Administration

Current State of KMS

Technology Stack

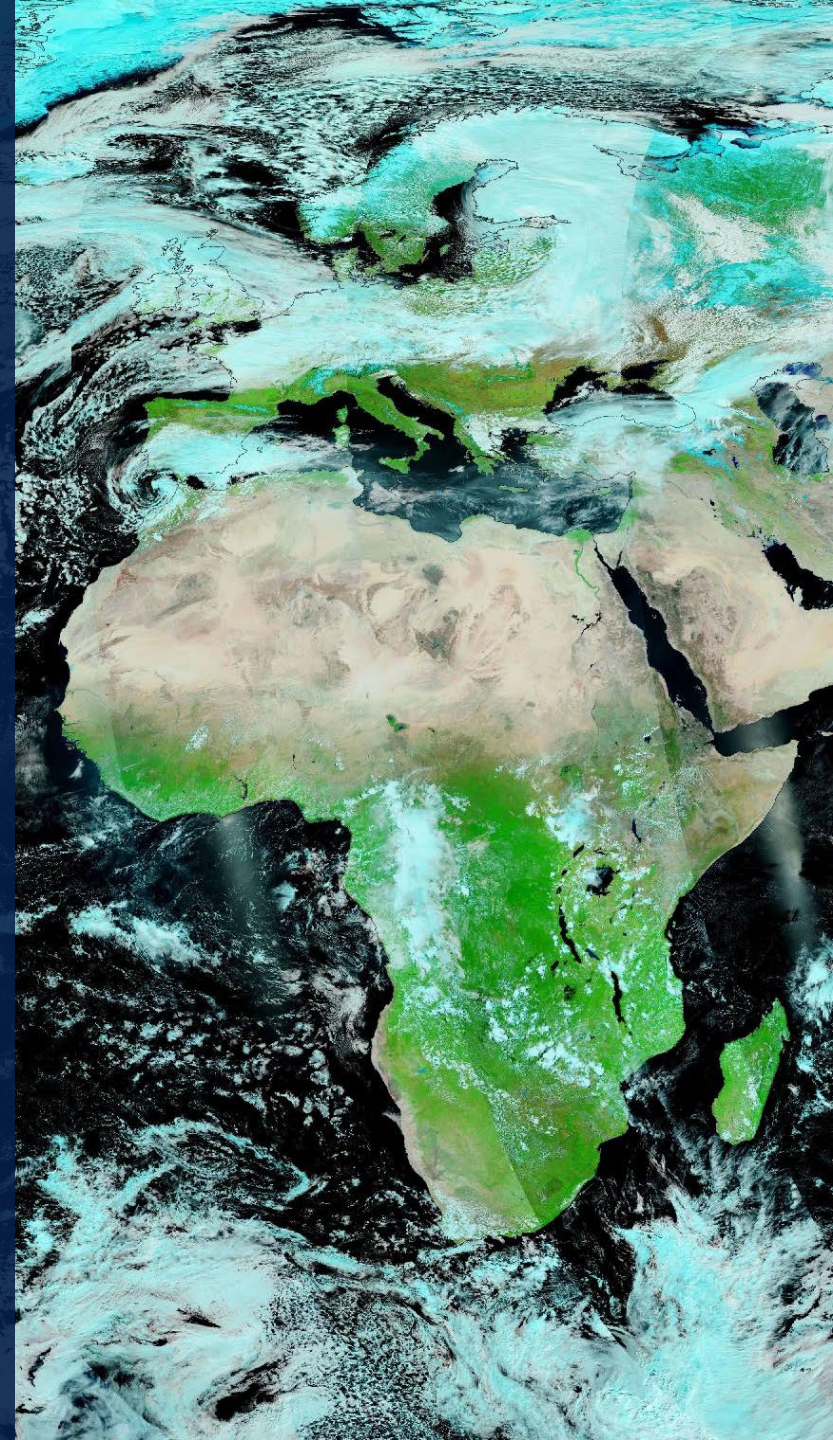
- Written in Java
- Uses MySQL with Hibernate
- GWT frontend
- Jetty 9.4



Early Prototype using RDF Database

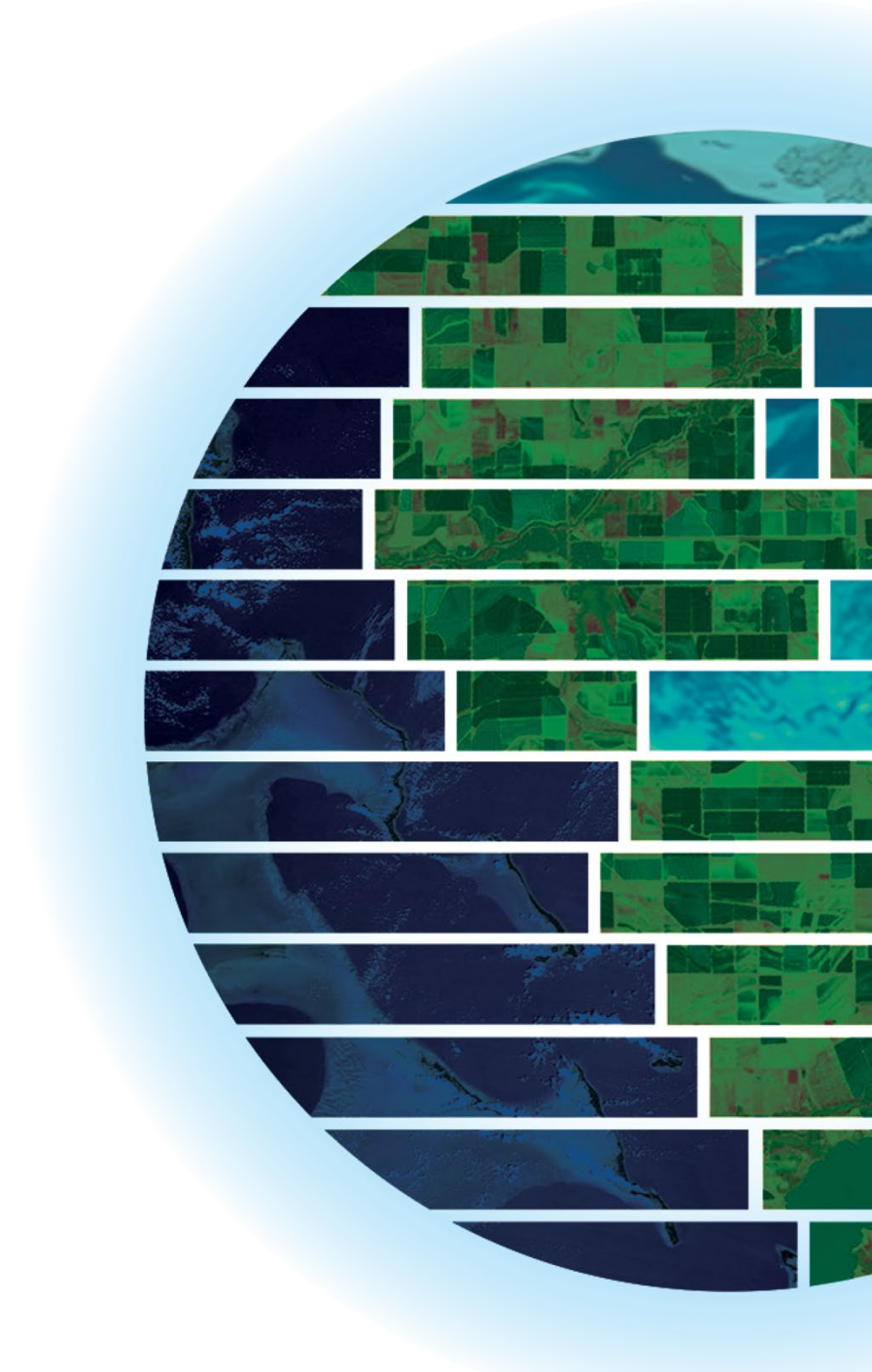
Technology Stack

- Written in nodejs
- Uses RDF compliant database (RDF4j)
- Serverless
- React Frontend with integration with MMT



What did we do?

- Exported our KMS database as RDF
- Imported data into RDF4j
 - 15,152 concepts
 - 160K triples
- Prototyped Key APIs
 - Retrieving a single SKOS concept in RDF
 - Retrieve a batch of SKOS concepts in RDF
 - Finding concepts by pattern (i.e., wind)
 - Exporting all concepts as CSV
- Benchmarked performance.



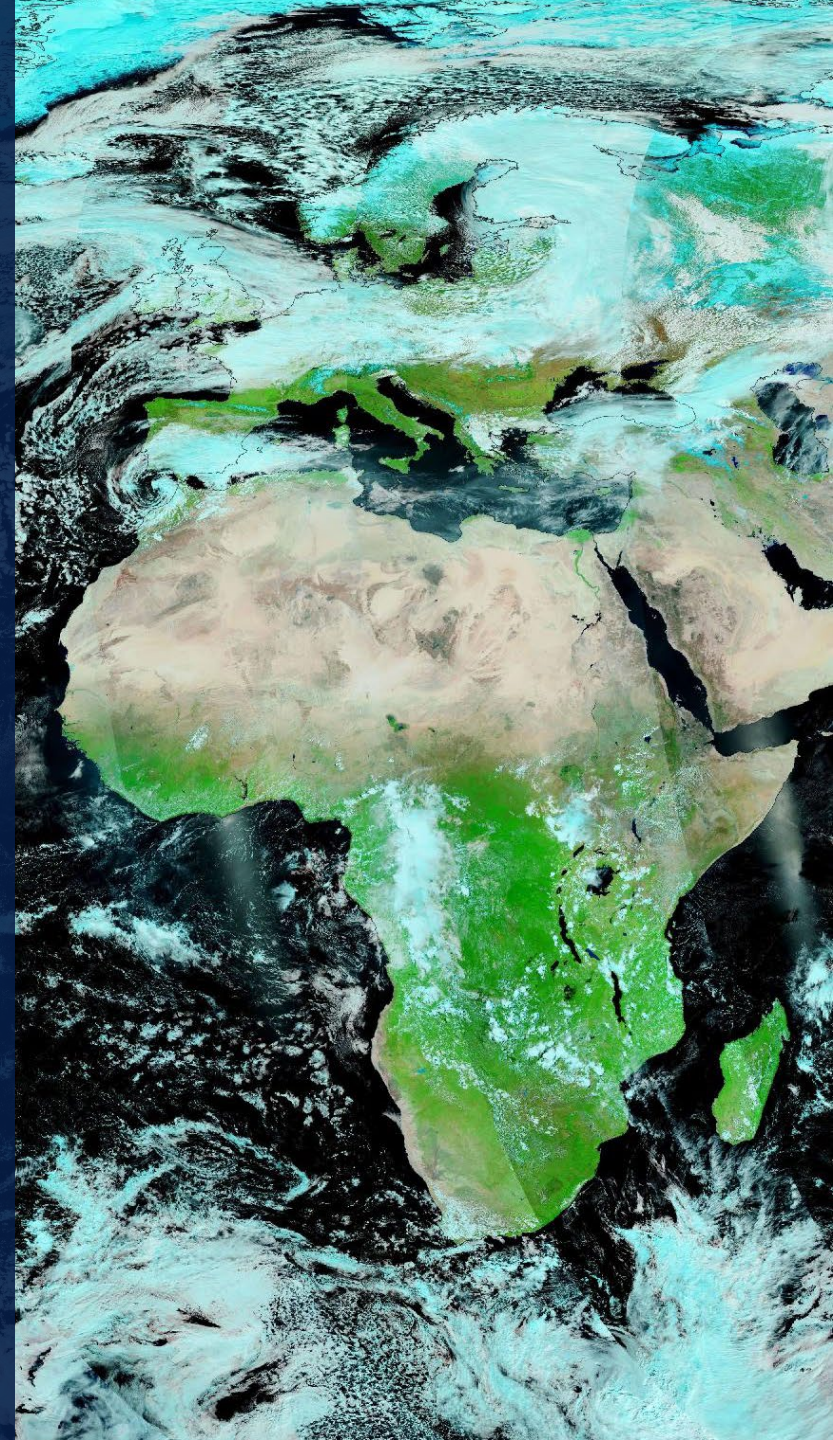
Benchmarks

Operation	MySQL KMS				RDF KMS			
	50th Percentile	95th Percentile	99th Percentile	Maximum	50th Percentile	95th Percentile	99th Percentile	Maximum
Fetch 15K individual concepts and return the RDF concept for each /kms/concept/[identifier]	50	94	260	72595	57	85	98	309
Pattern search through 15K individual terms and return the RDF concept for each /kms/concepts/pattern/[filter]	320	10000	11000	21126	170	240	270	7335
Retrieve CSV for each concept scheme (44 schemes) /kms/concepts/concept_scheme/[scheme].csv	48000	139000	139000	138997	23	400	580	976
Lookup a keyword path from id /kms/concept_fullpaths/concept_uuid/[identifier]	320	10000	11000	21126	30	52	61	312
Bulk download a batch of 2000 concepts at a time /kms/concepts	18000	18000	20000	23195	8700	9500	9900	10441
Parameters: 10 concurrent users, scale up 1 per second								



Building on SKOS

“SKOS” stands Simple Knowledge Organization System, which is a W3C standard for representing and sharing knowledge organization systems (KOS), such as thesauri, taxonomies, ontologies, and classification schemes.

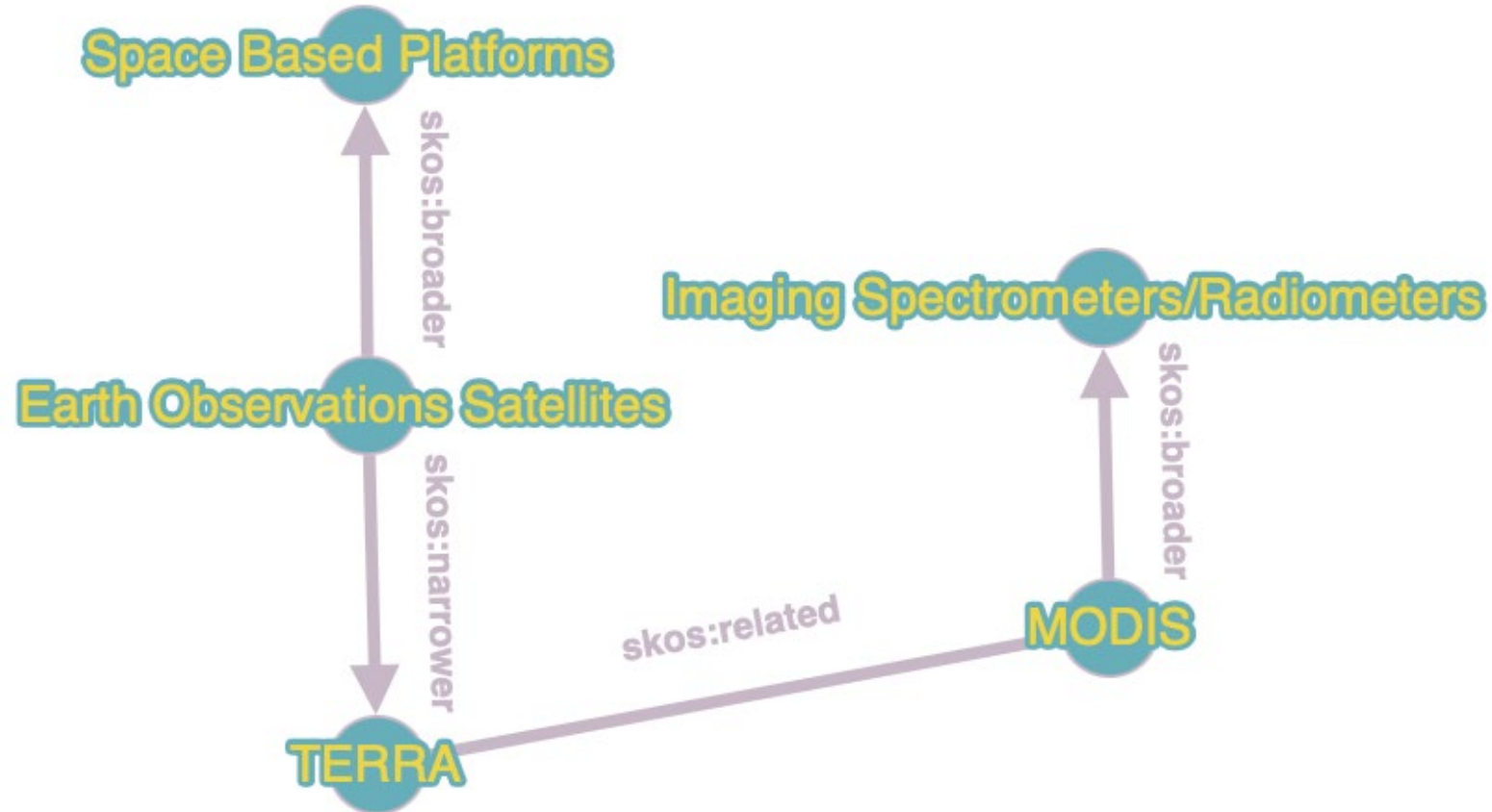


SKOS Example

- Concept: “Earth Observations Satellites”
 - Broader Concept: “Space Based Platforms” (**skos:broader**)
 - Narrower Concept: “TERRA” (**skos:narrower**)
- Concept: “MODIS”
 - Broader Concept: “Imaging Spectrometers/Radiometers” (**skos:broader**)
 - Related Concept: “TERRA” (**skos:related**)



Graphed



Extending SKOS

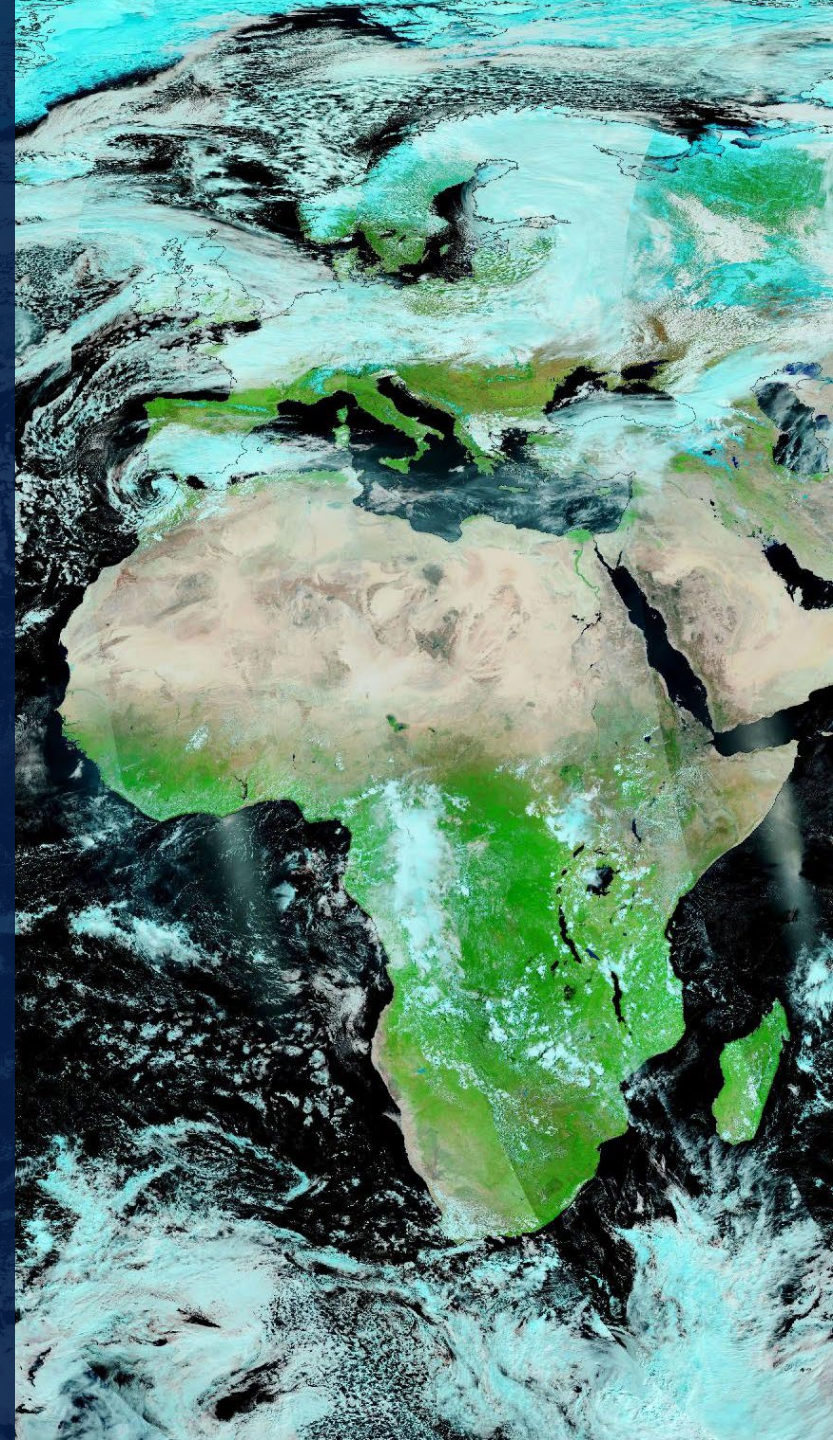
- Adding Custom Properties.

```
:TERRA      skos:prefLabel "TERRA"  
            skos:broader "Earth Observation Satellites"  
            :hasInstrument "ASTER"  
            :hasInstrument "MISR"  
            :hasInstrument "MODIS"  
            :hasInstrument "MOPITT"  
:MODIS      skos:prefLabel "MODIS"  
            skos:broader "Imaging Spectrometers/Radiometers"  
            :isOnPlatform "TERRA"
```



Ontologies/Inference

Leveraging inference engines to improve search



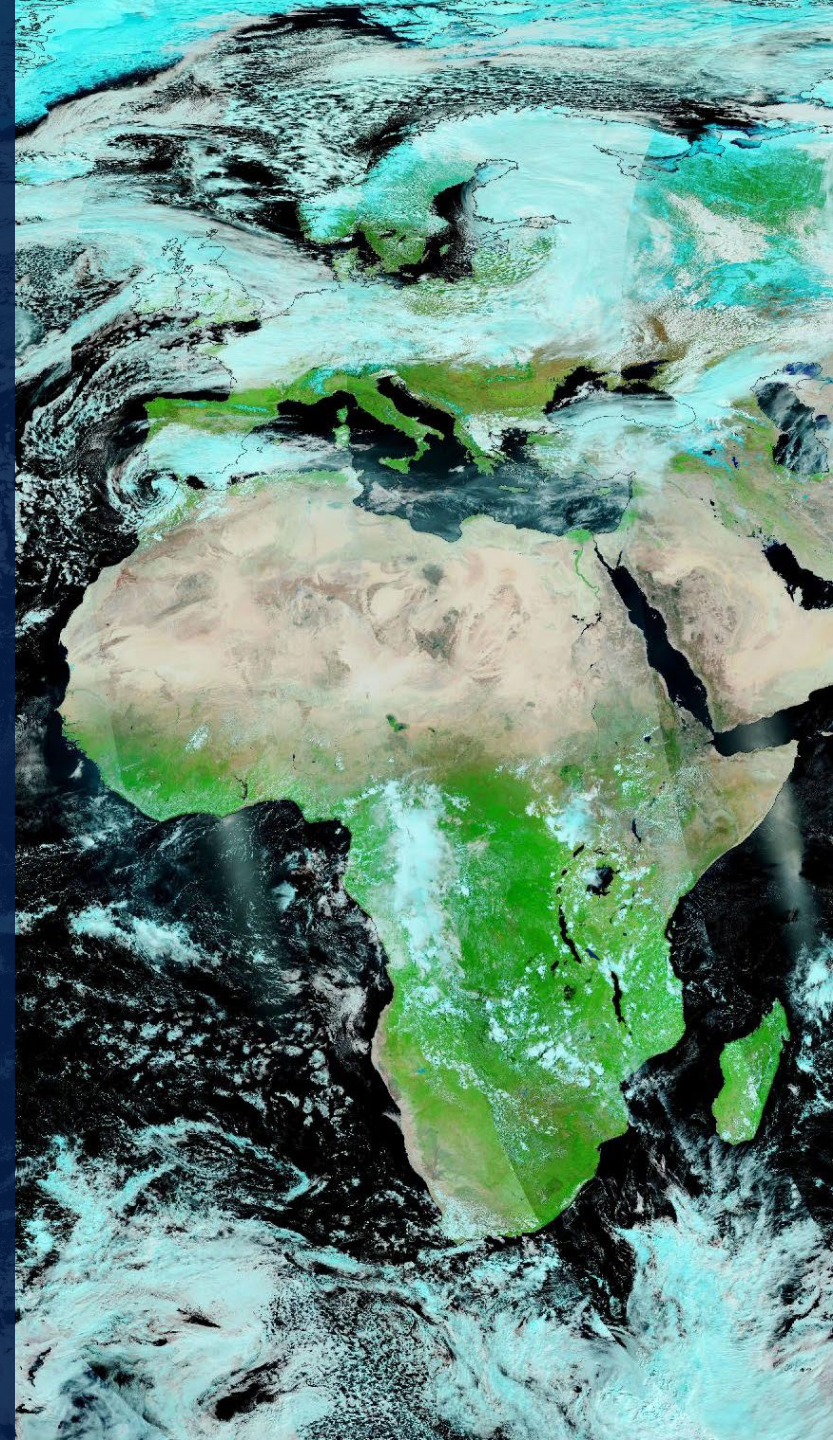
Using OWL with SKOS for Logical Inference:

- **Class Hierarchies and Subclassing**
 - "Sedimentary Rock" is a subclass of "Rock,"
 - "Sandstone" is a subclass of "Sedimentary Rock,"
 - **Inference:** "Sandstone" is a type of "Rock."
- **Is Contained Within (Spatial Relationships)**
 - **State is contained within Country. City is contained within State.**
 - **Inference: City is contained within Country.**
- **Has Influence On (Meteorology and Climatology)**
 - **El Niño has an influence on Global Weather Patterns.**
 - **Global Weather Patterns have an influence on Temperature and Precipitation.**
 - **Inference: El Niño has an influence on Temperature and Precipitation.**
- **Cause-and-Effect Relationships**
 - **Volcanic Eruption causes Earthquakes.**
 - **Earthquakes cause Tsunamis.**
 - **Inference: Volcanic Eruption causes Tsunamis.**



Machine Learning

How the Semantic Web can make machines smarter



Benefits

- Structured way of understanding domain knowledge
- Reasoning can be used to infer missing feature
- Automated Data Labeling

```
prefix spa: <tag:stardog:api:analytics:>

INSERT {
  graph spa:model {
    :myModel a spa:ClassificationModel ;
             spa:arguments (?director ?year ?studio) ;
             spa:predict ?genre .
  }
}
WHERE {
  ?movie :directedBy ?director ;
         :year ?year ;
         :studio ?studio ;
         :genre ?genre .
}
```



GCMD Keyword Recommender

GKR application is a machine learning model that leverages the semantic embeddings ability of Word2Vec models to process metadata and suggest relevant keywords.

Descriptive Keywords

Science Keyword ⓘ

Controlled Science Keywords describing the collection. The controlled vocabulary for Science Keywords is maintained in the Keyword Management System (KMS).

Recommended Keywords
Based on your Abstract, the MMT automatically suggests recommended keywords Recommended for your collection. To associate a recommended keyword to your collection, click the + icon next to the keyword. Once associated to the collection, the keyword will display a green check ✓. To remove a keyword once it's been associated, click the ✖ icon next to the keyword.

- EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC TEMPERATURE > SURFACE TEMPERATURE > AIR TEMPERATURE Recommended +
- EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC WINDS > SURFACE WINDS Recommended +
- EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC WATER VAPOR > WATER VAPOR INDICATORS > HUMIDITY Recommended +
- EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC PRESSURE > ATMOSPHERIC PRESSURE MEASUREMENTS Recommended +
- EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC TEMPERATURE > SURFACE TEMPERATURE > DEW POINT TEMPERATURE Recommended +
- EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC PRESSURE > SURFACE PRESSURE Recommended +
- EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC WATER VAPOR > WATER VAPOR INDICATORS > DEW POINT TEMPERATURE Recommended +
- EARTH SCIENCE > ATMOSPHERE > PRECIPITATION > PRECIPITATION AMOUNT Recommended +
- EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC TEMPERATURE > SURFACE TEMPERATURE Recommended +
- EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC WINDS > UPPER LEVEL WINDS Recommended +
- EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC RADIATION > LONGWAVE RADIATION Recommended +

SCIENCE KEYWORD

Search for Keywords...

EARTH SCIENCE ✖



A satellite view of Earth showing the Americas and Europe. The top half of the image shows North America and Europe, while the bottom half shows South America and the Atlantic Ocean. The background is a dark blue gradient.

Demo/Questions

We are looking for feedback. If you are existing user of KMS, let us know what you like about it, dislike, ideas, requests, etc.

CONTACT

christopher.d.gokey@nasa.gov

EARTHDATA

earthdata.nasa.gov

Thank you!

This work was supported by NASA/GSFC under Raytheon Company
contract number 80GSFC21CA001