EOSDIS
NASA'S EARTH OBSERVING SYSTEM DATA AND INFORMATION SYSTEM

New Metadata Capabilities Within NASA's Common Metadata Repository (CMR)

ESIP Summer Meeting 2018

Simon Cantrell
EED-2 Senior Principal Systems Engineer
simon.cantrell@nasa.gov

Tyler Stevens
EED-2 Senior Discipline Engineer
Tyler.B.Stevens@nasa.gov

Joe Rincione
EED-2 CMR Product Owner
joe@element84.com

Kathleen Carr
EED-2 Senior Principal Systems Engineer
Kathleen_Carr@raytheon.com

Valerie Dixon
NASA ESDIS
valerie.dixon@nasa.gov

Outline

I. Overview of the Common Metadata Repository (CMR)

II. Introduction To UMM-Var

III. Introduction To UMM-S

IV. Migration of SERF to UMM-S

V. Community Engagement of the Models

VI. Questions/Discussion

VII. Backup

EOSDIS

2

I. Overview of the CMR

Common Metadata Repository (CMR)

What is the Common Metadata Repository (CMR)?

The CMR is a high-performance, high-quality metadata repository for earth science metadata records. CMR manages the evolution of NASA Earth Science metadata by providing a central storage and access capability that streamlines current workflows while anticipating future capabilities.

What Does the CMR do Today?

CMR acts as a source of truth for earth science metadata. Data partners can manually or programmatically use the CMR to ingest new data records or update existing records. CMR helps power Earthdata Search by returning the right data elements searched by the user.

What Will the CMR do in the Future?

Our development team is committed to increasing the metadata capacity of the CMR, as well as optimizing our ingest, indexing, and search performance. We'll be expanding to support newer UMM Models (UMM-G) and new versions of existing UMM models (UMM-S, UMM-Var, UMM-C). We're also searching for ways to support a better user experience in Earthdata Search by providing relevant search results and powerful filtering tools.

EOSDIS

Unified Metadata Model (UMM)

What is The Purpose of the Unified Metadata Model (UMM)?

The UMM provides a common metadata model to unify legacy systems (i.e. GCMD, ECHO) with new systems (i.e. CMR).

How Does the UMM Support Interoperability?

In the context of the CMR, the UMM enables legacy systems to interoperate to enable continued support to user groups who depend on these systems.

The UMM is Subdivided Into Several Sub-Models:
- UMM-Collection
- UMM-Granule
- UMM-Variable
- UMM-Services

Access the Models:
https://wiki.earthdata.nasa.gov/display/CMR/CMR+Documents

EOSDIS

II. Introduction to UMM-Var

What is the UMM-Var Model?

What is a Variable?

The motivation for UMM-Var is to provide a variables model applicable to CMR that:

- Enables services at the variable level, as an extension to the UMM-C, UMM-Common and UMM-G metadata models, and

- Permits users to search and discover variables via well known geophysical terms (a.k.a. measurement terms), as well as using existing Science Keywords
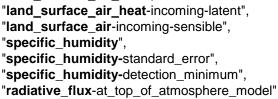
Note: Because key terms "Variable" and "Measurement" are frequently used within the Earth Science community, often with different meanings, operative definitions are provided below.

Measurement: An observable property, usually geophysical. For models, it is a simulated observable property. Measurement Names are of the form: <object>__<quantity>.
Names shall contain only lowercase letters and numbers along with the Standard Names separator characters (_, -, ~, __).

Examples:
"**land_surface_air_flow**",
"**land_surface_air_heat**-incoming-latent",
"**land_surface_air**-incoming-sensible",
"**specific_humidity**",
"**specific_humidity**-standard_error",
"**specific_humidity**-detection_minimum",
"**radiative_flux**-at_top_of_atmosphere_model"

Variable: A measured geophysical quantity associated with a particular data product, typically a point, swath or grid of values nested within the data file. One or more variables may be associated with a given collection.

Examples:
psl (mean sea level pressure)
O3_ppbv (ozone mixing ratio reported in parts per billion by volume)
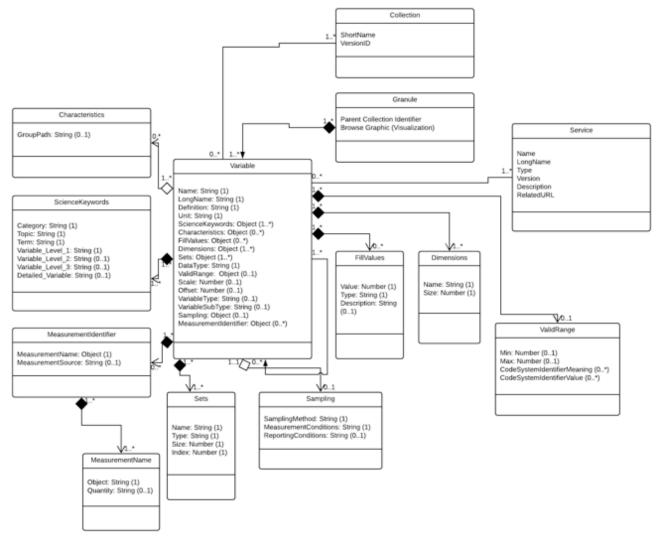**Scat_550** (total dry aerosol scattering coefficient at 550 nm)
LST_1KM_Day (daily daytime 1km grid land surface temperature)
Sur_Refl_b01 (surface reflectance band 1)
WDB_L3MCA10 (Aerosol Optical Depth 550nm (Land Only))

EOSDIS

UMM-Var Design (v1.2)

New Capabilities and Features

- Measurement Names with <"Object"_"Quantity"> structure
- Sampling class
- Addition of Code System Identifier valids
- GroupPath
- Variable Sub Type

EOSDIS

Challenges in Populating UMM-Var Records

- One of the biggest challenges in populating UMM-Var records is the large number of variables expected and the wide variety in naming and structure

- We are now prototyping methods of auto-population, with sources of variable metadata coming from OPeNDAP, but also Data Set User Guides and HDF/netCDF and other community sources

- Scalability of generalized variable handling code due to the variability of naming and dimensionality (1D, 2D, 3D etc.)

EOSDIS

10

Use Cases and Examples (summary only)

- Browse Variables of a Collection

- Faceted Browse

- Update Variable Associations

- Search Relevancy Ranking

- Cross-Site Data Subsetting

- Integrating GIBS

- Measurement comparison of two in-situ measurements of Species X (where X is e.g. Nitrous Oxide)

EOSDIS

III. Introduction to UMM-S

What is the UMM-S Model?

What is a Service?

We need to recognize SERF legacy for services, but extend NASA web services to include future support for access of NASA data from the cloud. These services provide a method of transforming the data (e.g. subsetting, reprojection or reformatting, or a combination of these).

How to I Discover and Utilize a Service?

In order to locate web services or portals, software or tools, we can use the GCMD client to search and return metadata for services. https://gcmd.nasa.gov/

In order to discover and transform data, we can use the EDSC client to search and return *transformed* data via services. https://earthdata.nasa.gov/

How to I Describe a Service?

Typically, services are described by their attributes, e.g.

Service Name: "SERVIR"

Service LongName: "Mesoamerican Visualization and Monitoring System (SERVIR)"

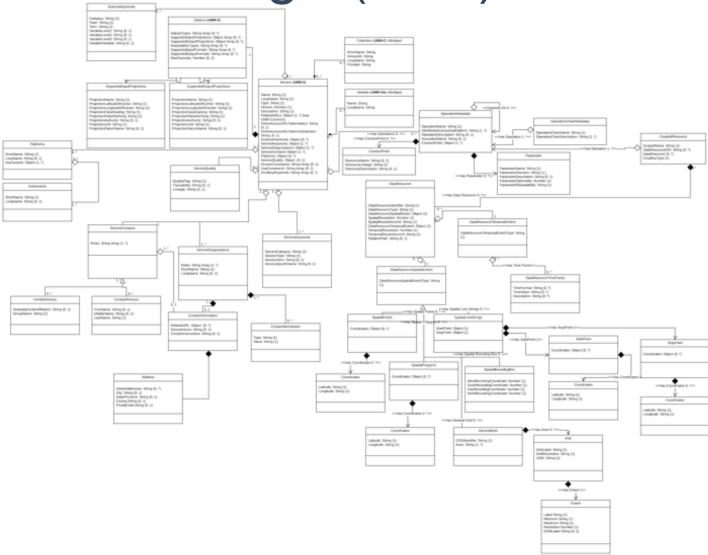**Service Type:** "WEB PORTAL"

Service Version: "1.8"

Service RelatedURL: "https://www.servirglobal.net/default.aspx"

> **Services in general can be characterized by their: Name, LongName, Type, Version, RelatedURL, etc..**

EOSDIS

13

UMM-S Design (v1.2)

New Capabilities and Features

- RelatedURLs increased cardinality to 1 to N
- Options class expanded to include supported Input and Output projections, Input and Output formats.
- ServiceQuality class added
- Platforms and Instruments classes added
- Addition of OperationMetadata class to support a wide variety of server-side operations (e.g. GetCapabilities, GetMap, GetCoverage)

EOSDIS

Challenges in Populating UMM-S Records

- One of the biggest challenges in populating services is the naming and scope of the service, in terms of options available

- Associating Services to Collections and Variables in the CMR

- Coupling between services and variables requires human intervention
    - Services like OPeNDAP have tightly coupled data sources
    - Others, i.e. WMS, WCS, have loosely coupled data sources
    - Methods of auto-population are being explored

EOSDIS

IV. Migration of SERF to UMM-S

What is the SERF (Service Entry Resource Format)

- Metadata standard used to describe Earth science tools, software, and models (e.g. EDSC, Giovanni, Panoply, USGS Model Viewer)

- Supports the discovery and access of these tools, software, and models

- Originated (in year 2000) as a prototype within NASA's Earth Science Technology Office (ESTO) and supported by NASA and the Committee on Earth Observation Satellites (CEOS)

NASA GCMD Data Service Types

DATA ANALYSIS AND VISUALIZATION (680)
calibration/validation, geographic information systems, global positioning systems, statistical applications, visualization/image processing show all...

DATA MANAGEMENT/DATA HANDLING (349)
archiving, cataloging, data compression, data delivery, data interoperability show all...

EDUCATION/OUTREACH (58)
curriculum support, exhibit materials, interactive programs show all...

ENVIRONMENTAL ADVISORIES (110)
agricultural advisories, fire advisories, geological advisories, health advisories, hydrological advisories show all...

HAZARDS MANAGEMENT (49)
disaster recovery/relief, disaster response, hazards mitigation, hazards planning show all...

METADATA HANDLING (87)
authoring tools, data discovery, metadata transformation/conversion, service discovery show all...

MODELS (340)
atmospheric chemistry models, atmospheric general circulation models, carbon cycle/carbon budget models, climate change impact assessment models, component process models show all...

REFERENCE AND INFORMATION SERVICES (93)
bibliographic, digital/virtual reference desks, gazetteer, identification/classification systems, knowledge/decision systems show all...

WEB SERVICES (35)
data application services, data processing services, information management services show all...

EOSDIS

18

Why Migrate to UMM-S in CMR

- Use the Expanded Model and Infrastructure To Make Services More Useful
  - Extended service capabilities (service invocation, subsetting, reprojection, time aggregation, etc...)
- Link Services With Applicable Data Sets
  - Allows for service capabilities and specific tools to be accessible to/from their applicable data sets, which renders the services discoverable, available, and useful to users that might not already be aware of them
- Search Across a Common Repository for Services
  - Centralized catalogue of all services, tools, software, models across U.S. and international agencies that work with Earth science data
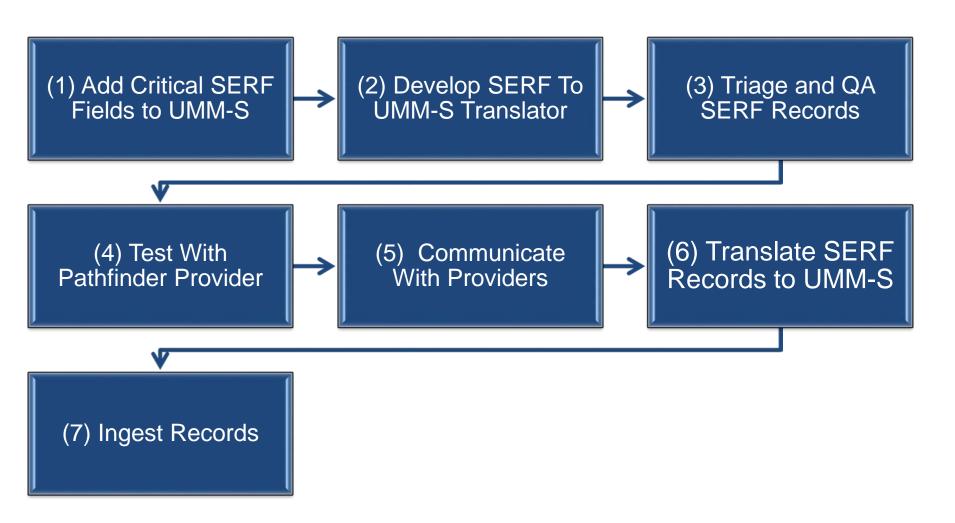
EOSDIS

What's Being Migrated

- There are ~1,200 SERF records in GCMD
- Should I stay or should I go
  - Does the service describe a tool or software
  - Is the service still viable?
  - Is the service being deprecated
  - Do the links still work?
  - Is the service from a commercial provider?
  - Are the other attributes still accurate?

Migration Process

```
(1) Add Critical SERF      →    (2) Develop SERF To    →    (3) Triage and QA
Fields to UMM-S                 UMM-S Translator            SERF Records
```

```
(4) Test With              →    (5)  Communicate       →    (6) Translate SERF
Pathfinder Provider             With Providers              Records to UMM-S
```

```
(7) Ingest Records
```

EOSDIS

What Users Need to Know

- Providers will be contacted to assess and update their SERFs prior to UMM-S migration

- SERF will be deprecated after migration is complete

- Service records will be discoverable in the CMR API, Earthdata Search Client (via collection associations) and the Global Change Master Directory (GCMD)

- Providers will continue to be able to submit UMM-S records for review and ingest

EOSDIS

V. Community Engagement of the Models Through the ESDIS Standard Office

Role of ESO In Community Review

- The ESDIS Standards Office (ESO)
  - Assists the Earth Science Data and Information System (ESDIS) project in formulating standards policy for NASA Earth Science Data Systems (ESDS)
  - Coordinates standards activities within ESDIS
  - Recruits members for review groups from NASA and the broader Earth science community as appropriate
  - Posts the status and results of reviews online
- If you would like to participate in reviews, please contact the ESO at eso-staff@lists.nasa.gov
- See https://earthdata.nasa.gov/about/esdis-project/esdis-standards-office-eso for more information
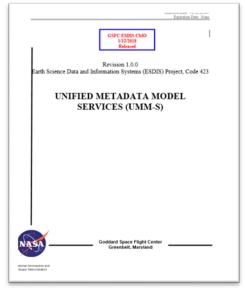
EOSDIS

24

ESO Review Process

Initial Screening

Community Review

Triage Feedback

Publish Document

Published UMM Document

JAMA Requirements Management System

EOSDIS

Issues and Lessons Learned from Reviews

- ## Issues from Conducting Full Model Reviews
  - Expensive because the reviews involve a large number of people reviewing almost the exact same large model all over again
  - Comments tend to be duplicates of previous years questions and some of them are contradictory of past comments and decisions
  - Comments of the actual metadata providers are hard to distinguish from the philosophical ideals of other reviewers
  - Large number of comments take months to triage and implement

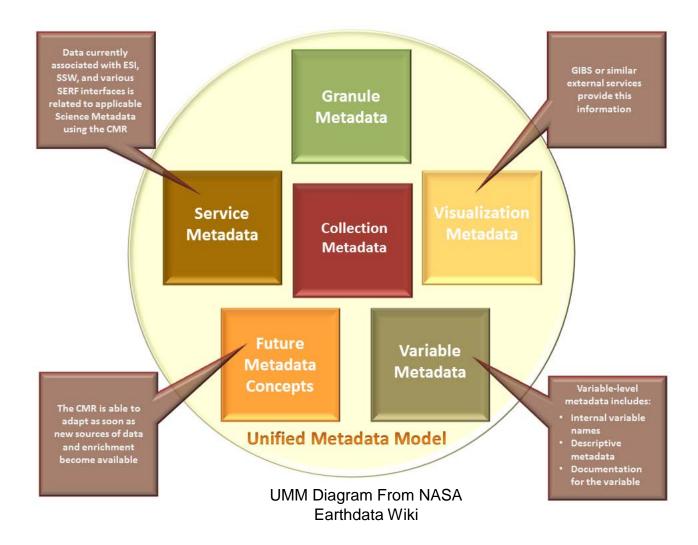- ## Lessons Learned for Future Reviews
  - More frequent targeted reviews of a specific issue or problem based on end-user and client needs
  - Wiki page documenting the issue, background information, recommended solution, and an example of the change
  - Faster turn-around time for triage and implementation
  - Reviewers can still do an full review of the model if they want

EOSDIS

VI. Questions/Discussion

VII. Backup

Unified Metadata Model

UMM Diagram From NASA
Earthdata Wiki

EOSDIS

ESIP Summer 2018

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C.

in partnership with