



EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

Making Metadata Better with the CMR and MMT

Summer ESIP 2016

Jason Gilman

Dana Shum

The material is based upon work supported by the National Aeronautics and Space
Administration under Contract Number **NNG15HZ39C**

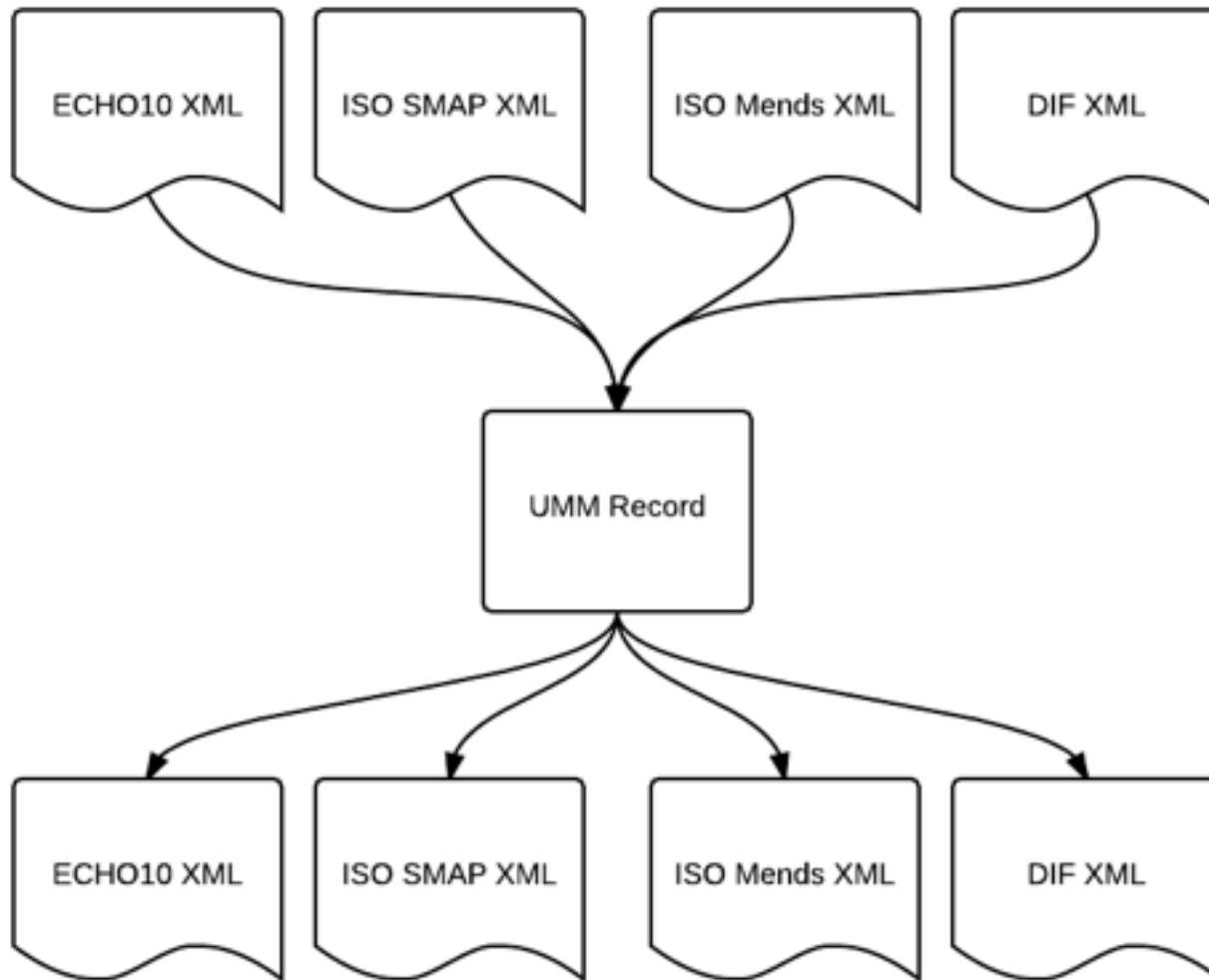
METADATA PROBLEMS

Completeness and Consistency

- Misspellings: “Bioosphere”
- Legacy Terms: “AM-1” instead of Terra
- Inconsistent Names: Processing levels “Level 1”, “1”
- Whitespace around element values
- Missing elements

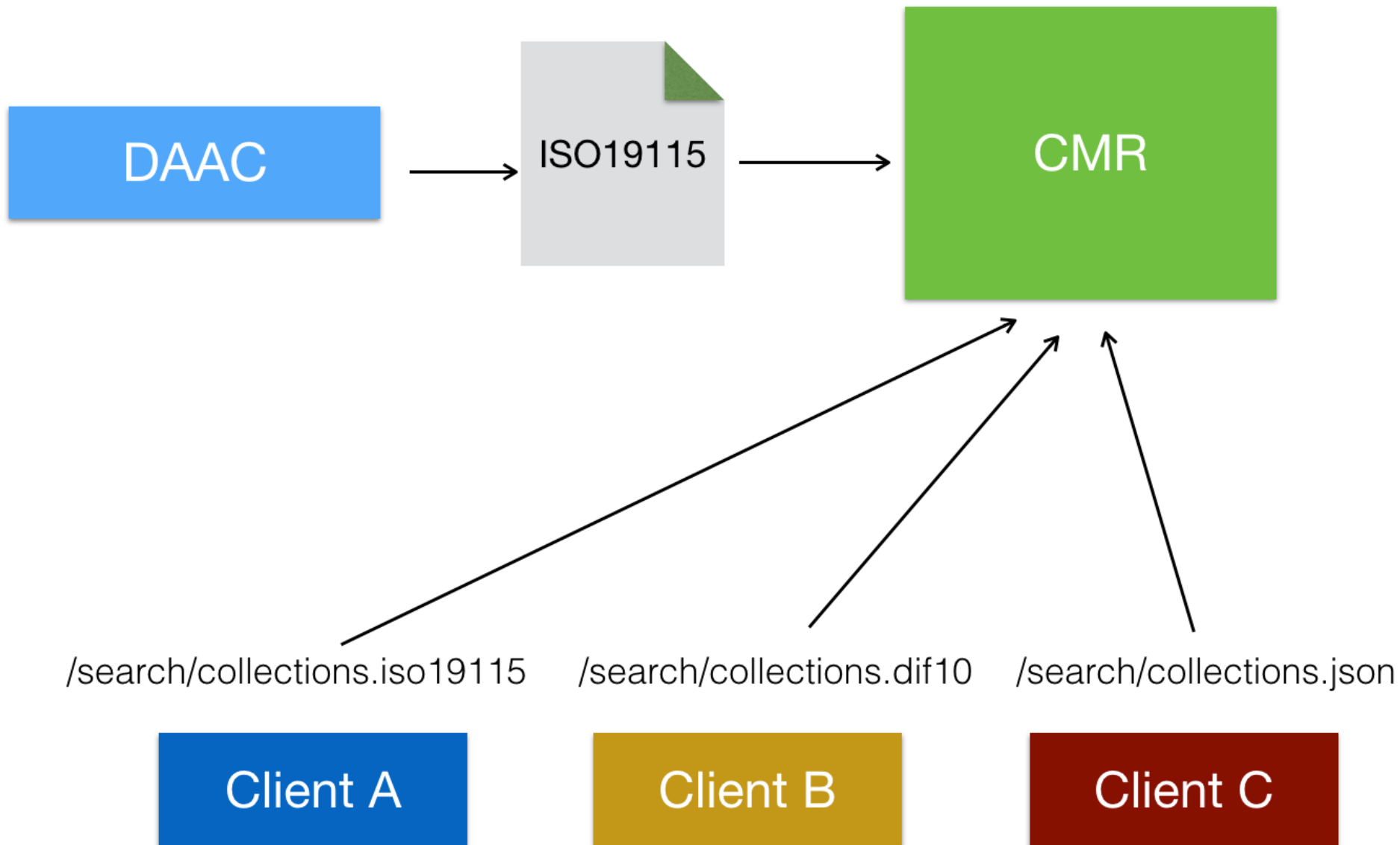
Overview and Benefits

UNIFIED METADATA MODEL



UMM Benefits

- Common validations across all metadata dialects
- Consistent features across all dialects.
- Allows conversion from any dialect to any other dialect.



MMT



{UMM
JSON}

CMR

/search/collections.iso19115

/search/collections.dif10

/search/collections.json

Client A

Client B

Client C

COMPLETENESS AND CONSISTENCY ENABLERS

MMT Pre-Ingest Validation

The screenshot displays the 'testRecord1_001' draft record page. At the top, it shows 'Manage Metadata', 'Drafts', and 'testRecord1_001'. The record title is 'testRecord1_001' with the description 'This is test record 1'. A 'DRAFT RECORD' label is present. On the right, it indicates 'VERSION 001' and a 'Quality Score: 20' with a note 'Required fields not complete'. Below the title are 'Publish Draft' and 'Delete Draft' buttons. The 'Metadata Fields' section lists several categories with progress indicators: Collection Information (4 required, 2 optional), Descriptive Keywords (1 required, 2 optional), Spatial Information (1 required, 3 optional), Organizations (1 required, 1 optional, 1 does not pass validation), Personnel (1 required, 1 optional), and Collection Citations (1 required, 1 optional). A legend explains the icons: a green circle with a checkmark for 'Required and complete', a green circle with an 'r' for 'Required but not complete', a white circle with an 'o' for 'Optional and not complete', a grey circle with an 'o' for 'Optional and complete', and a red circle with a minus sign for 'Does not pass validation'. A 'Close' button is located at the bottom right of the legend box.

Manage Metadata Drafts testRecord1_001

testRecord1_001
This is test record 1

DRAFT RECORD

Publish Draft Delete Draft

Metadata Fields ⓘ

- Collection Information (4 required, 2 optional)
- Descriptive Keywords (1 required, 2 optional)
- Spatial Information (1 required, 3 optional)
- Organizations (1 required, 1 optional, 1 does not pass validation)
- Personnel (1 required, 1 optional)
- Collection Citations (1 required, 1 optional)
- Metadata Information (1 required, 2 optional)

VERSION 001

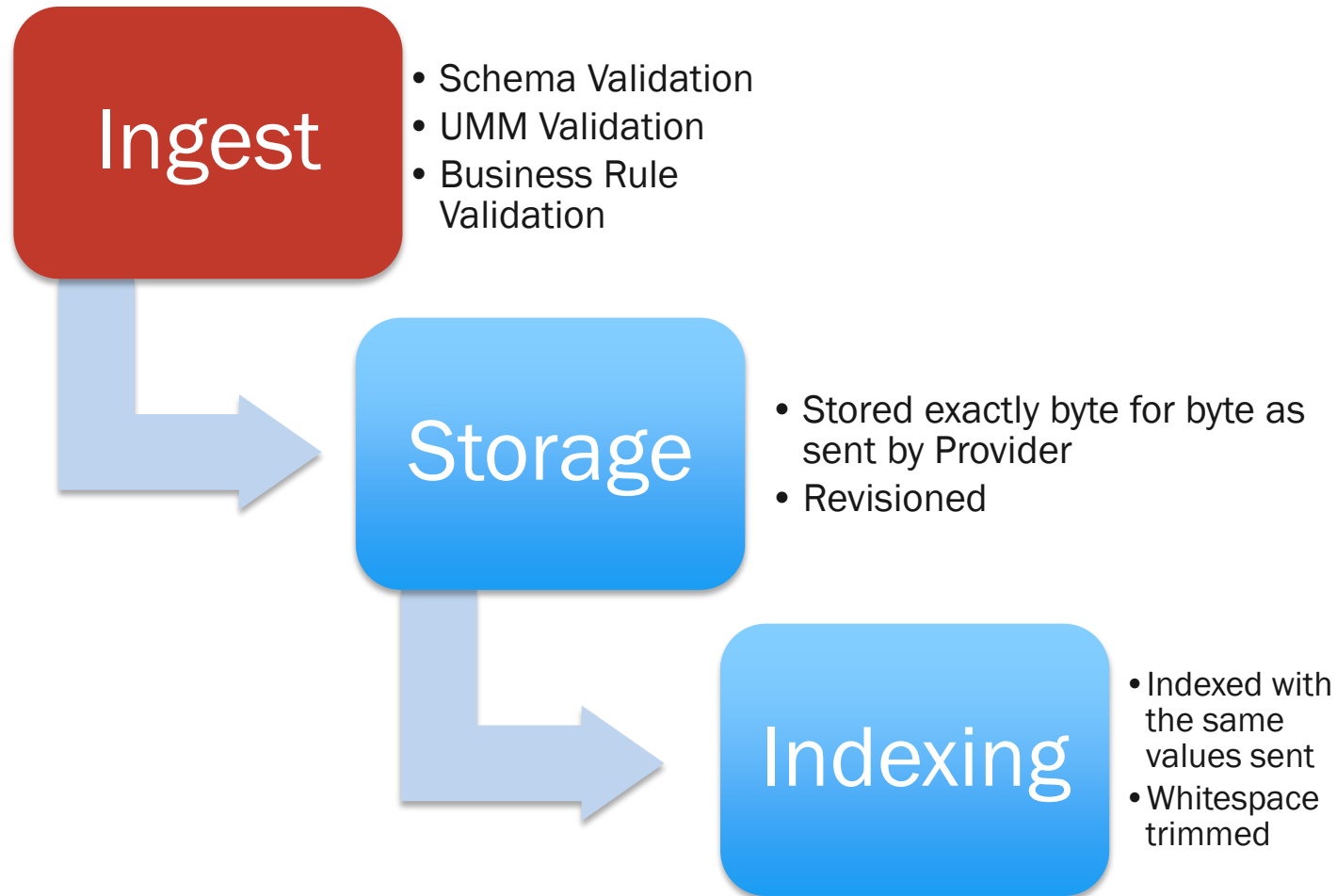
Quality Score: 20
Required fields not complete

The icons below each form name indicate progress toward completion. See below for information on what each icon represents.

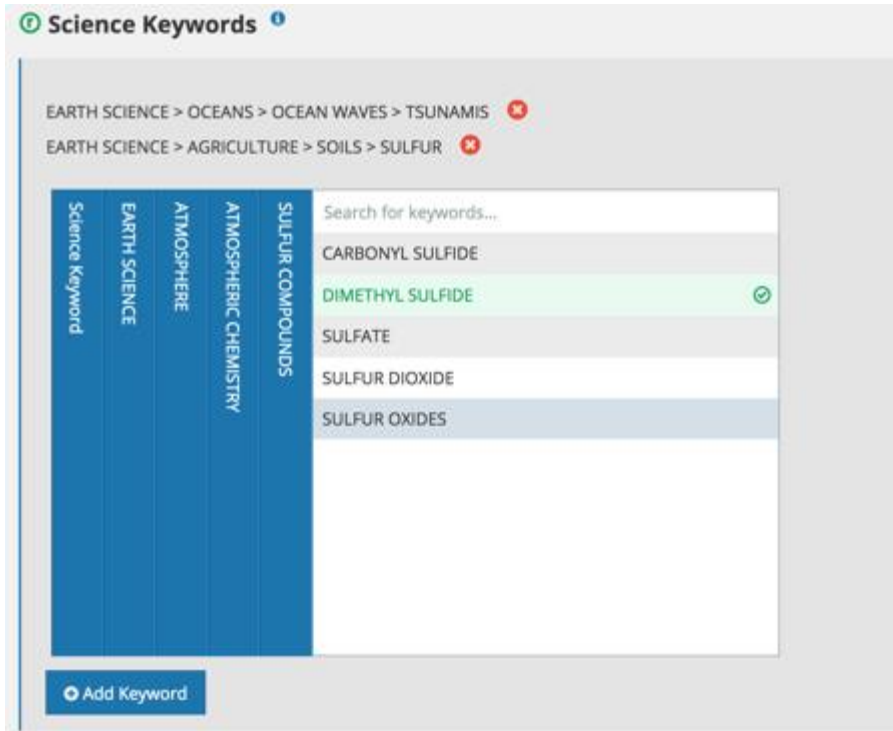
- Required but not complete
- Required and complete
- Optional and not complete
- Optional and complete
- Does not pass validation

Close

CMR Ingest Validation



MMT and CMR Keyword Validation



```
curl -i -XPOST ... -H "Cmr-Validate-  
Keywords: true" \  
https://cmr.earthdata.nasa.gov/ingest/pro  
viders/LARC_ASDC/validate/collection/some  
NativeId -d \  

```

```
"<Collection>  
...  
<ScienceKeywords>  
  <ScienceKeyword>  
    <CategoryKeyword>EARTH  
SCIENCE</CategoryKeyword>  
  
<TopicKeyword>BIOOSPHERE</TopicKeyword>  
  <TermKeyword>SOILS</TermKeyword>  
</ScienceKeyword>  
</ScienceKeywords>  
</Collection>"
```

HTTP/1.1 422 Unprocessable Entity

```
<path>ScienceKeywords/0</path>  
<error> Science keyword Category [EARTH  
SCIENCE], Topic [BIOOSPHERE], and Term
```

Preview Panel / Collection HTML



Common Metadata Repository

https://cmr.earthdata.nasa.gov/search/concepts/C197265171-LPDAAC_ECS

ASTGTM_002
ASTER Global Digital Elevation Model V002

PUBLISHED RECORD

VERSION 002

☆☆☆☆☆
Quality Score: 20
Required fields not complete

Abstract

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) was developed jointly by the U.S. National Aeronautics and Space Administration (NASA) and Japan's Ministry of Economy, Trade, and Industry (METI). ASTER is capable of collecting in-track stereo using nadir- and aft-looking near infrared cameras. Since 2001, these stereo pairs have been used to produce single-scene (60- x 60-kilometer (km)) digital elevation models (DEM) having vertical (root-mean-squared-error) accuracies generally between 10- and 25-meters (m). The methodology used by Japan's Sensor Information Laboratory Corporation (SILC) to produce the ASTER GDEM involves automated processing of the entire ASTER Level-1A archive. Stereo-correlation is used to produce over one million individual scene-based ASTER DEMs, to which cloud masking is applied to remove cloudy pixels. All cloud-screened DEMs are stacked and residual bad values and outliers are removed. Selected data are averaged to create final pixel values, and residual anomalies are corrected before partitioning the data into 1 degree (°) x 1° tiles. The ASTER GDEM covers land surfaces between 83°N and 83°S and is comprised of 22,702 tiles. Tiles that contain at least 0.01% land area are included. The ASTER GDEM is distributed as Geographic Tagged Image File Format (GeoTIFF) files with geographic coordinates (latitude, longitude). The data are posted on a 1 arc-second (approximately 30-m at the equator) grid and referenced to the 1984 World Geodetic System (WGS84)/ 1996 Earth Gravitational Model (EGM96) geoid. While the ASTER GDEM 2 benefits from substantial improvements over GDEM 1, users are nonetheless advised that the products still may contain anomalies and artifacts that will reduce its usability for certain applications, because they can introduce large elevation errors on local scales. The data are provided "as is" and neither NASA nor METI/ERSDAC will be responsible for any damages resulting from use of the data. V002 data set release date: 2009-06-28 Data Set Characteristics: Geographic Extent: Global between 83° latitude Scene Coverage: 1° x 1° tiles Image Dimensions: 3601 x 3601 Total Number Tiles: V001: 22,604; V002: 22,702 Tile Volume: ~25MB, 6.4 MB compressed Compression Type: zip File Format: GEOTIFF Map Projection: Geographic Lat/Lon Datum: WGS84/EGM96 Resolution: 1 arcsecond (30-m horizontal posting at equator)



Spatial Coordinates

Bounding Rectangle

N: 82.0 S: -83.0 E: 180.0 W: -180.0

Location Keywords

No Spatial Keywords found

Temporal Coverages

DateTime Ranges

1999-12-18 to 2011-02-28

Help Text

The screenshot shows a web browser window with the URL https://mmt.uat.earthdata.nasa.gov/drafts/147/edit/acquisition_information. The page title is "Metadata Management Tool". The main content area is titled "Platforms" and shows a form for editing "Platform 1". The form includes fields for "Type" (set to "Earth Observation Satellites"), "ShortName" (set to "terra"), and "LongName". A "Characteristics" section is visible, containing a table with columns for "Name", "Unit", and "Data Type". A help dialog box is open over the "Characteristics" section, titled "Characteristics", with a close button (X) in the top right corner. The dialog box contains the following text:

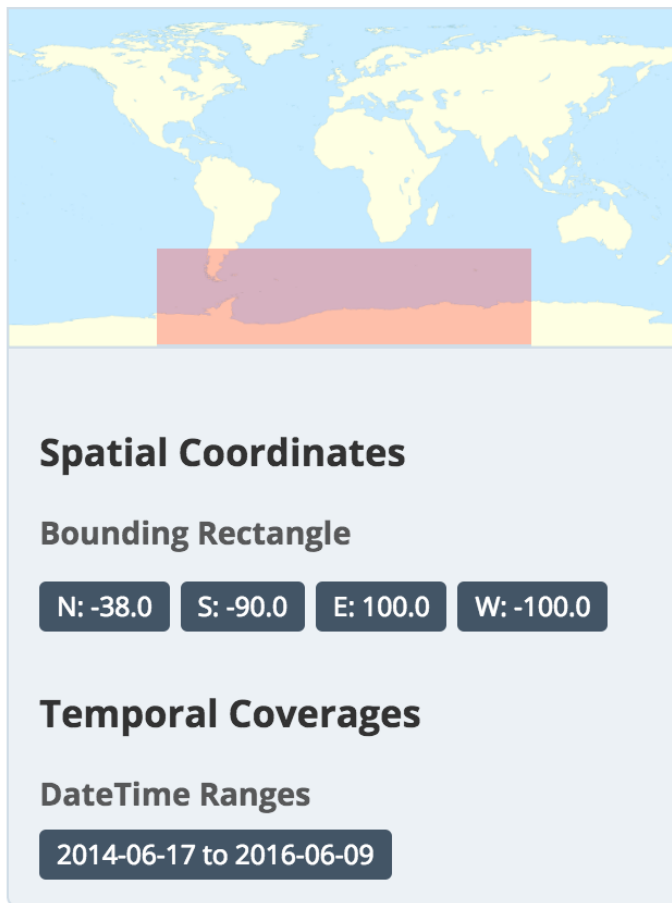
Characteristics

Platform-specific characteristics, e.g., Equator Crossing Time, Inclination Angle, Orbital Period. The characteristic names must be unique on this platform; however the names do not have to be unique across platforms.

Validation

- Minimum Items: 0

A "Close" button is located at the bottom right of the dialog box.




METADATA QUALITY RUBRICS

Metadata Quality Rubrics

- Provide indication of completeness of Metadata based on different recommendations (UMM-C, CSW, DataCite...)
- Automatically evaluated and tagged in CMR
- Badges and guidance displayed in MMT.
 - Encourages better quality metadata
- Future
 - Reports, aggregations, search relevancy.

Metadata Management Tool x Mark

← → ↻ <https://mmt.uat.earthdata.nasa.gov/drafts/147> ☆ 🛑 ❤️ 🗨️ 🗑️ 🗨️ 📷 🌐 🌐 ☰



EARTHDATA
Metadata Management Tool BETA

Quick Find Find

🔍 Full Metadata Record Search

MMT_1

Mark Reese [Change Provider](#) [Logout](#)

Manage Metadata [Drafts](#) [Test-Record_001](#)

Test-Record_001

Kathy's 101 test record - Cloned

DRAFT RECORD

VERSION 001

☆☆☆☆☆

Quality Score: 20

Required fields not complete

Publish Draft

Delete Draft

Metadata Fields ?

<p><input checked="" type="checkbox"/> Collection Information r r r r ○ ○</p> <p><input checked="" type="checkbox"/> Descriptive Keywords r ○ ○</p> <p><input checked="" type="checkbox"/> Spatial Information r ○ ○ ●</p> <p><input checked="" type="checkbox"/> Collection Citations ○</p>	<p><input checked="" type="checkbox"/> Data Identification r ○ r ○ ○ ○ ○ ○ ○ ○ ○</p> <p><input checked="" type="checkbox"/> Acquisition Information r ○</p> <p><input type="checkbox"/> Organizations - ○</p> <p><input checked="" type="checkbox"/> Metadata Information ● ○ ○</p>	<p><input checked="" type="checkbox"/> Distribution Information r ○</p> <p><input checked="" type="checkbox"/> Temporal Information r ○ ○</p> <p><input checked="" type="checkbox"/> Personnel ○</p>
--	---	--

HUMANIZERS

Humanizers Prototype

- Give clients clean, consistent facet data without having to change the underlying metadata.
- Temporary solution to help users while metadata is cleaned up.
- Humanizers can inform changes that are needed to metadata.

Fixing Problems in Facets

- Misspellings: “Bioosphere”
- Legacy Terms: “AM-1” instead of Terra
- Inconsistent Names: Processing levels “Level 1”, “1”
- Whitespace around element values
- Use normal case when appropriate

Before Humanizers

The image displays two screenshots of the NASA EarthData search interface, illustrating the state of the search results before humanization.

Top Screenshot: Search for 'Terra'

- Search bar: Terra
- Results: 630 Matching Collections
- Search Time: 1.1s
- Featured Collection: ASTER Global Digital Elevation Model V002 (ASTGTM v002 - LPDAAC, 1999-12-18 to 2011-02-28, 22702 Granules)
- Left sidebar: Browse Collections list with 'TERRA' highlighted (630 collections).

Bottom Screenshot: Search for 'AM-1'

- Search bar: AM-1
- Results: 48 Matching Collections
- Search Time: 9.4s
- Featured Collection: ASTER L1A Reconstructed Unprocessed Instrument Data V003 (AST_L1A v003 - LPDAAC, 1999-12-18 ongoing, 2823846 Granules)
- Left sidebar: Browse Collections list with 'AM-1' highlighted (48 collections).

After Humanizers

The screenshot shows the NASA EarthData Search interface. At the top, there are navigation links for 'Data Discovery', 'DAACs', 'Community', and 'Science Disciplines'. The search bar contains the text 'Terra'. Below the search bar, there is a 'Browse Collections' sidebar on the left with a table of collections:

Collection Name	Count
SPOT-4	17
SPOT-5	15
TERRA	x678
THEOS	12
UK-DMC	12

The main content area displays '630 Matching Collections' and includes a 'Learn More' button. Below this, there is a 'Recent and Featured' section with a card for 'ASTER Global Digital Elevation Model V002'. The card includes the text 'No image available', 'ASTGTM v002 - LPDAAC', and the date range '1999-12-18 to 2011-02-28 | 22702'. At the bottom of the page, there is a footer with the version 'v 1.19.3' and links to 'NASA Official: Andrew Mitchell', 'FOIA', 'NASA Privacy Policy', and 'USA.gov'.

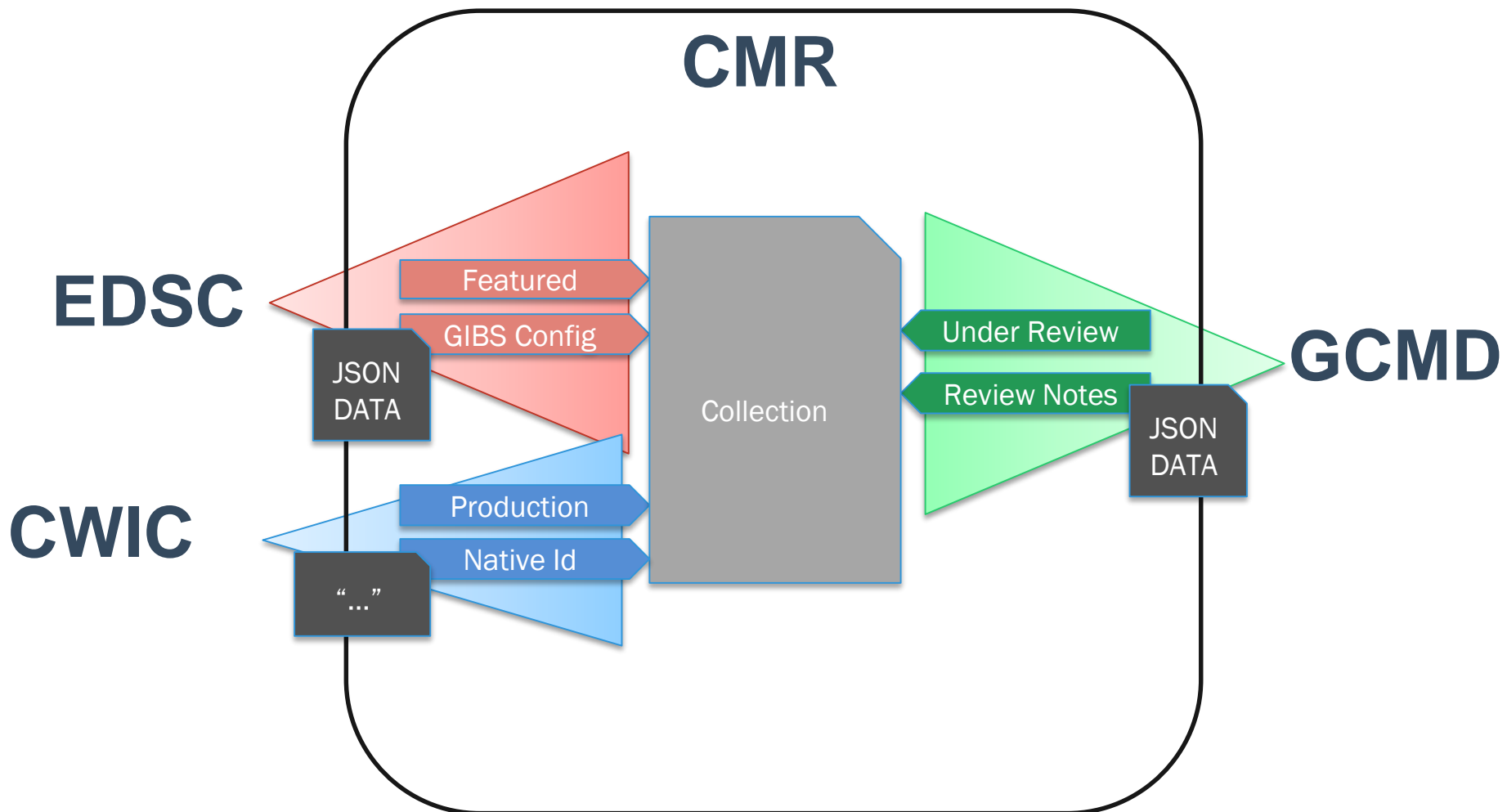
How it works

1. Admin creates humanizer instructions in CMR
2. CMR indexes impacted collections with additional humanized fields
3. Clients optionally request facets with humanized values.

Underlying metadata and existing fields are **not modified**.

TAGS

Tags
=
Enhanced Metadata



What is a Tag?

- Key

- “edsc.extra.gibs”

- Description

- “Used to associate GIBS configuration for a collection”

- Category (optional)

- “EDSC_display”

What is a Tag Association?

- Connects a Tag with a Collection
- Tag Key
 - “edsc.extra.gibs”
- Collection Id
 - “C1000000560-NSIDC_ECS”
- Data
 - Any arbitrary JSON up to 32K. Single strings will be made searchable.

Tag Data can be Retrieved

```
curl "https://cmr.../search/collections.json?\n      tag_key=edsc.extra.gibs\
```

...

```
"tags" : {\n  "edsc.extra.gibs" : {\n    "data" : [ {\n      "maxNativeZoom" : 5,\n      "antarctic" : false,\n      "product" : "GHRSSST_L4_G1SST_Sea_Surface_Temperature",\n      "geo" : true,\n      "arctic" : false,\n      "title" : "Sea Surface Temperature (L4, G1SST)",\n      "source" : "Multi-mission / GHRSSST",\n      "match" : {\n        "time_start" : ">=2010-06-21"\n      },\n      "format" : "png",\n      "resolution" : "1km"\n    } ]\n  }\n}
```

Tag Data can be Searched

Find all collections:

- tagged with “org.ceos.wgiss.cwic.granules.native_id”
- with data “JPL-L2P-MODIS_A”

```
curl "https://cmr.../search/collections.json?\
```

```
tag_data[org.ceos.wgiss.cwic.granules.native_id]=JPL-L2P-MODIS_A"
```

Tags Enable Many Use Cases

- Allows layering on additional information to each collection.
- Examples
 - Mark “featured” collections.
 - Categorize collections (Reviewed, approved, needs work, etc).
 - Store visualization information
 - Add a client specific id to collections.

This material is based upon work supported by the National Aeronautics and Space Administration under Contract Number **NNG15HZ39C**.

Raytheon