IDN Update

CEOS WGISS-47
Silver Spring, Maryland, USA, April 29-May 2, 2019
Hosted by the National Oceanic and Atmospheric Administration (NOAA)

Michael Morahan
IDN Coordinator
Michael.P.Morahan@nasa.gov
Outline

- IDN Activities (since WGISS-46)
- Registration of services for the IDN
- Setting up CEOS 2-Step OpenSearch in the IDN (CMR)
- Update on efforts to provide consistent Quality Assurance (QA) Rule messaging.
I. IDN Activities  
(since WGISS-46)
Adding FedEO Platform/Instrument relations in KMS

• The relationships show the linkage of Instruments to Platforms.
• Services (Web Services and APIs) used for task:
  - **Skosmos (ESA Thesaurus):** is a web based open source ontology browser.
  - **GCMD Keyword Manager (KMS):** is a web based management system for maintaining and registering Keywords (not available externally).
  - **GCMD KMS API:** is a RESTful web service for maintaining keywords (science keywords, platforms, instruments, data centers, locations, projects, services, resolution, etc.) in the GCMD/IDN system.
    • Documentation URL: [https://wiki.earthdata.nasa.gov/display/gcmdkey/Keyword+Management+Service+Application+Program+Interface](https://wiki.earthdata.nasa.gov/display/gcmdkey/Keyword+Management+Service+Application+Program+Interface)
FedEO Platform/Instrument relations in KMS

FedEO Platform/Instrument relations in KMS

Keyword Editor

Keyword UUID: C729654-77c1-4e27-9957-7153d60213f6
Broader Keyword: Platforms>Earth Observation Satellites>SENTINEL-1
Preferred Label: SENTINEL-1A
Alternate Label(s)
  primary SENTINEL-1A
Add Alternate Label
Resource(s)
  image http://www.esa.int/images/sentinel2_alcblad_M.gif
Add Resource
Definition: The SENTINEL-1 mission is the European Radar Observatory for the Copernicus joint initiative of the European Commission (EC) and the European Space Agency (ESA). Copernicus, previously known as GMES, is a European initiative for the implementation of information services dealing with environment and security. It is based on observation data received from Earth Observation satellites and ground-based information.
Definition Reference: Enter reference for keyword definition
Related Keyword(s)
  C-SAR (instruments)
Add Related Keyword

Edit Keyword Relation

UUID: ed400e7c-229e-48b3-9a93-0412f5d61448
Keyword Title: C-SAR
Relationship Strength: Enter value between 0 to 1.0, 0.5 by default
Resource Type: Platform-Instrument

Helpful Notes
relationship Strength: 0-1, 1 being the strongest, 0.5 by default
Relationship Types:
- Platform-Instrument
- Instrument-Sensor
- Similar: Two keywords in same scheme which are close but different
- Transitive: Keyword in scheme A is the same as keyword in scheme B with x% probability
- Undefined: Default
Platform/Instrument relation using KMS API (JSON)

https://gcmdservices.gsfc.nasa.gov/kms/concept/c7279e54-f7c1-4ee7-a957-719d6021a3f6?format=json
Ingest of NOAA_NCEI (NOAA OneStop) Native ISO

- IDN preformed a comparison between NASA MEND’s ISO metadata and NOAA OneStop ISO metadata.
- 4 issues were discovered that prevented the NOAA OneStop’s ISO from ingesting directly into the CMR for the IDN.
- NOAA, NASA CMR, and IDN worked together to revolve these issues.
- Presently, 70 NOAA_NCEI GHRSST ISO have been ingested into the CMR for the IDN.
New JAXA GCOM-C collection records

• 149 JAXA GCOM-C collection records have been ingested for discovery in the IDN (CMR)

• CEOS WGISS “Discovery and Access” page has information on GCOM-C products:
  - http://ceos.org/ourwork/workinggroups/wgiss/access/

• IDN access to collection metadata:
  - https://idn.ceos.org/search/Titles.do?subset=idn&json={"condition":{"and":{"keyword":"*"},"project":"GCOM-C"}}

Keyword Viewer:
https://gcmd.nasa.gov/search/Keywords.do#keywords
## Transition of DIF-9 to DIF-10

<table>
<thead>
<tr>
<th>Provider</th>
<th># of records</th>
<th># of records transition</th>
<th># of records not transition</th>
<th>Completion Percentage</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCIOPS</td>
<td>15933</td>
<td>15933</td>
<td>0</td>
<td>100%</td>
<td>Completed</td>
</tr>
<tr>
<td>AU_AADC</td>
<td>2819</td>
<td>0</td>
<td>2819</td>
<td>0%</td>
<td>Starting Q2</td>
</tr>
<tr>
<td>CNES</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>100%</td>
<td>Completed</td>
</tr>
<tr>
<td>ESA</td>
<td>112</td>
<td>20</td>
<td>92*</td>
<td>18%</td>
<td>On-going</td>
</tr>
<tr>
<td>EUMETSAT</td>
<td>37</td>
<td>37</td>
<td>0</td>
<td>100%</td>
<td>Completed</td>
</tr>
<tr>
<td>INPE</td>
<td>43</td>
<td>43</td>
<td>0</td>
<td>100%</td>
<td>Completed</td>
</tr>
<tr>
<td>ISRO</td>
<td>34</td>
<td>34</td>
<td>0</td>
<td>100%</td>
<td>Completed</td>
</tr>
<tr>
<td>JAXA</td>
<td>388</td>
<td>388</td>
<td>0</td>
<td>100%</td>
<td>Completed</td>
</tr>
<tr>
<td>NOAA_NCEI</td>
<td>5563</td>
<td>84</td>
<td>5479**</td>
<td>1%</td>
<td>On-going</td>
</tr>
<tr>
<td>USGS_EROS</td>
<td>146</td>
<td>146</td>
<td>0</td>
<td>100%</td>
<td>Completed</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>25095</strong></td>
<td><strong>16705</strong></td>
<td><strong>8390</strong></td>
<td><strong>67%</strong></td>
<td></td>
</tr>
</tbody>
</table>

* New delivery of FedEO ESA records in calendar Q2 will replace existing records.
** Remove and replace NOAA/NCEI records.
Attended Conferences: AGU and ESIP

• ESIP Winter Meeting 2019 (Presentation)
  - **Topic:** GCMD Keyword Management Process and Life-Cycle
  - **Summary:** Show the process for GCMD keywords - fast track and yearly reviews through the ESDIS Standards Office, and how Earth science users can influence keyword additions and modifications. Highlight how keywords facilitate the discovery of EOSDIS data and services. Upcoming keyword reviews.
  - Session Abstract link:

• AGU Winter Meeting 2018 (Poster)
  - **Topic:** Structured Methods for Organization and Discovery of Variables within NASA’s Common Metadata Repository
Structured Methods for Organization and Discovery of Variables within NASA's Common Metadata Repository

NASA's Common Metadata Repository is seeking to support discovery at the level of individual variables, and to support the end-to-end process of data analytics that starts with the discovery of specific variables and ends with analytics-ready data sets. Variable-level metadata forms which organize processing variables with an extensible set of attributes/value pairs, producing CMR-compliant metadata records that can describe the underlying phenomenon and measurement approach represented in the data. Such records will not only help with data discovery, but will also lay the foundation for developing automated data transformation services, which require variable spectrum traits to prepare heterogeneous data for use in multi-variate analysis. A phased approach to processing variables with CMR-compliant metadata is explained, beginning with metadata that will identify the measurement object (the feature of interest) and the quantity being measured. Because the approach to generating variable-level metadata records is to assign descriptive attribute/value pairs to variables (e.g., measurement object/ambient aerosol, quantity/optical depth) it is easily extensible and additional attribute/value pairs can be phased in as needed, based on input from data processing application developers and the science community about what descriptors are needed to support the desired transformation processes.

- 2017/18 goals focused on investigating data transformation use cases and requirements, and found that most use cases involve multi-variable analysis using variables from a number of disparate data sets.
- During the discovery phase, users need to identify suitable variables for their research. Keyword tags are not informative enough to determine suitability.
- Selcting by variable name is one of the most common pre-processing tasks performed by researchers.
- During the analysis phase, users need to pre-process multiple variables from heterogeneous data sets to make them interchangeable: variable level metadata is needed to support these processes.

Class Design

Metadata Examples

<table>
<thead>
<tr>
<th>Measurement Object</th>
<th>Quantity</th>
<th>Vertical Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambient aerosol</td>
<td>Optical Depth</td>
<td>Integrated vertical column</td>
</tr>
<tr>
<td>Aerosol absorption</td>
<td>Mass Concentration</td>
<td>Pressure levels</td>
</tr>
<tr>
<td>Aerosol drift</td>
<td>Optical Depth</td>
<td>Pressure levels</td>
</tr>
<tr>
<td>Cloudiness</td>
<td>Optical Depth</td>
<td>Integrated vertical column</td>
</tr>
<tr>
<td>Cloudiness Bottom</td>
<td>Optical Depth</td>
<td>Top of the atmosphere</td>
</tr>
<tr>
<td>Aerosol optical depth</td>
<td>Temperature</td>
<td></td>
</tr>
<tr>
<td>Air</td>
<td>Pressure</td>
<td></td>
</tr>
<tr>
<td>Ocean</td>
<td>Temperature</td>
<td></td>
</tr>
</tbody>
</table>

Process Workflows:
1) Ingest variable metadata, names, dimensions, standard names, etc. 2) Map Measurement Names to variable records using specification 3) Copy/Download process/copy with community standards without muses for Measurement Names, if necessary.

Issues/Challenges:
- Mapping variables can be challenging, both because of the number of variables that need to be mapped and because semantically consistent maps are needed.
- Allow measurement names to be mapped to variables at the time of ingest.

Maps on variables allow us to sort them into classes and to identify both how two or more variables are similar and how they are different. The modularity of mapping, an approach to assigning fully specific standard names, helps to support different approaches to data discovery and analysis, for example, search for all variables for which the measurement object is aerosol, and discover, e.g., that there are measurements of aerosol vertical optical depth as well as aerosol density.

Similarly, users might search on temperature and discover that there are temperature measurements as well as anemometer temperature measurements. Additional maps, such as vertical depth maps, can further classify or differentiate similar variables. For example, two variables representing measurements of aerosol vertical optical depth might differ in the vertical profiles.

Ultimately, this approach to variable-level metadata records lays the foundation for automating pre-processing tasks required to support data analytics. These processes need to be close to the variable level, e.g. merging, regridding, and automated processing pipelines need access to metadata that is specific to variable level.
II. Registration of services for the IDN
Registering Services into the CMR for IDN

- **(Future)** CEOS Providers will be able to register services using the Draft Metadata Management Tool (MMT)
- **(Present)** providers can send to the IDN the UMM-S Service JSON file or fill out a template of the fields’ values.
Draft Metadata Management Tool (MMT)

User and Provider

Create new draft

Open drafts

Search in CMR

Create Service Record

Create New Record

OR use the search in the top right corner to find published services to clone or edit.

DEV07 Service Drafts

2018-11-07 | DEV07_ESI
DEV07_ESI EOSDIS Service impl...

2018-10-12 | DEV07 SFD Service
DEV07 SFD Service 01

2018-02-22 | <Blank Name>
UNTITLED SERVICE RECORD
## UMM-S Required Fields

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>The name of the service, software, or tool. (Example: &quot;AIRS_L3_OPENDAP&quot;).</td>
</tr>
<tr>
<td>LongName</td>
<td>The long name of the service, software, or tool. It provides a human readable name for the service. (Example: &quot;OPENDAP (Hyrax) framework for AIRS Level 3 data products&quot;).</td>
</tr>
<tr>
<td>Type</td>
<td>The type of the service, software, or tool. (Example: OPeNDAP)</td>
</tr>
<tr>
<td>Version</td>
<td>The edition or version of the service, software, or tool. The version should be defined in the form x, y, and z. where 'x.y.z' means 'major.minor.incremental' version numbers. Typically, 'x' and 'y' are numbers (0 through 9) and 'z' is a number (0 through 99). (Example: 1.1.1)</td>
</tr>
<tr>
<td>Description</td>
<td>A brief description of the service, software, or tool. For example, a description might contain information about what is the service, the purpose of the service, and the parameters (or variables) being invoked, and what are the sources of these data.</td>
</tr>
<tr>
<td>RelatedURLS</td>
<td>This element contains important information about the Uniform Resource Locator (URL) for the service.</td>
</tr>
<tr>
<td>ServiceKeywords</td>
<td>Allows for the specification of Earth Science Service keywords that are representative of the service, software, or tool being described. The controlled vocabulary for Service Keywords is maintained in the Keyword Management System (KMS). (Example: &quot;ServiceCategory: Earth Science Services, ServiceTopic: Data Management/Data Handling, ServiceTerm: Data Search and Retrieval&quot;).</td>
</tr>
<tr>
<td>ServiceOrganizations</td>
<td>The service provider, or organization, or institution responsible for developing, archiving, and/or distributing the service, software, or tool. (Example: &quot;Role: SERVICE PROVIDER, ShortName: INPE, LongName: National Institute for Space Research, Brazil&quot;).</td>
</tr>
</tbody>
</table>
Service records can be searched using the CMR SEARCH API:

**Search for all**
curl 'https://cmr.earthdata.nasa.gov/search/services'

**Search by Keyword**
curl 'https://cmr.earthdata.nasa.gov/search/services?keyword=OPeNDAP&pretty=true'

**Search for specific service**
curl 'https://cmr.earthdata.nasa.gov/search/services.umm_json?name=MUR-JPL-L4-GLOB-v4.1&pretty=true'

**CMR SEARCH API Documentation:**
https://cmr.earthdata.nasa.gov/search/site/docs/search/api.html#searching-for-services
III. Setting up CEOS 2-Step OpenSearch in the IDN (CMR)
CEOS 2-Step OpenSearch “Option 1”

- Send to the IDN a populated spreadsheet with the following values:
  - CMR collection record’s Concept ID (if the record exist in the IDN (CMR) (e.g.: C1532648141-ESA)
  - Collection record’s Entry_ID (ShortName)
  - Granule OSDD link (Example):
    "http://fedeo.esa.int/opensearch/description.xml?parentIdentifier=EOP:ESA:FEDEO:COLLECTIONS:CryoSat.products"

<table>
<thead>
<tr>
<th>Concept_ID</th>
<th>Entry_ID (ShortName)</th>
<th>OpenSearch Granule OSDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1214560189-ESA</td>
<td>SMOS.MIRAS.MIR_OSUDP2</td>
<td><a href="http://fedeo.esa.int/opensearch/description.xml?parentIdentifier=EOP:ESA:FEDEO:COLLECTIONS:SMOS.MIRAS.MIR_OSUDP2">http://fedeo.esa.int/opensearch/description.xml?parentIdentifier=EOP:ESA:FEDEO:COLLECTIONS:SMOS.MIRAS.MIR_OSUDP2</a></td>
</tr>
</tbody>
</table>
CEOS 2-Step OpenSearch “Option 2”

- Populate the DIF-10 Field Related_URL with the following values:
  - Type with “USE SERVICE API”
  - Subtype with “OpenSearch”
  - URL with the granule OSDD link
  - Description with tag_key: opensearch.granule.osdd

- Example of the DIF-10 text:

  ```xml
  <Related_URL>
      <URL_Content_Type>
          <Type>USE SERVICE API</Type>
          <Subtype>OpenSearch</Subtype>
      </URL_Content_Type>
      <Description>tag_key: opensearch.granule.osdd</Description>
      <Mime_Type>application/opensearchdescription+xml</Mime_Type>
  </Related_URL>
  ```
Search for Tags using CMR API

- IDN staff will add the opensearch.granule.osdd tag in the CMR for providers.
- Providers can verify the addition of the tag using the CMR Search API:
  https://cmr.earthdata.nasa.gov/search/collections.json?concept_id=C1532648141-ESA&include_tags=*&pretty=true
- Example of JSON output of the query:

```
...
"tags" : {
  "opensearch.granule.osdd" : {
  "int.esa.fedeo" : { },
  "org.ceos" : { }
},
"id" : "C1532648141-ESA",
...
```
OSDD URL within OpenSearch Metadata

- Discovery: The OpenSearch collection metadata URL is returned to the client user.

  https://cmr.earthdata.nasa.gov/opensearch/collections.atom?uid=C1532648141-ESA

- Within the collection metadata is the OSDD URL used for granule searches:

  ...<link href="http://fedeo.esa.int/opensearch/description.xml?parentIdentifier=EOP:ESA:FEDEO:COLLECTIONS:CryoSat.products" hreflang="en-US" type="application/opensearchdescription+xml" rel="search" title="Non-CMR OpenSearch Provider Granule Open Search Descriptor Document"/>...
IV. Update on efforts to provide consistent Quality Assurance (QA) Rule messaging.
Registering Services into the CMR for IDN

• **New Task:** QA Rule Set Consolidation and Implementation Strategy Across Tools
• **Background:** CMR and IDN Metadata Validation services/tools for providers use different implementations and/or QA Rules.
• **Overall Goal:** Give metadata providers consistent QA results when creating, publishing and updating IDN (CMR) collection metadata base on QA Rule checks.
Questions

Please provide questions/comments to:

michael.p.morahan@nasa.gov (KBR/WYLE)
valerie.dixon@nasa.gov (NASA)
christopher.s.lynnes@nasa.gov (NASA)
Background Information
Example: CERES EBAF Edition 4.0 TOA Cloud Effective Temperature Daynight Monthly.