2017-10-26 ARC-CMR All Day Meeting notes

Date
26 Oct 2017

Attendees
- Erich Reiter
- Jon Pals
- Kathleen Carr
- Kaylin Bugbee
- Simon Cantrell
- Jeanne' le Roux
- Adam Sisco
- Betzy Hernandez Sandoval
- Carlee Loeser
- Dana Ostrenga
- Tyler Stevens
- Scott Ritz

Goals
1. Work off previous action items
2. DIF 9/10 Distribution is not being translated anymore.
3. Schema changes (requested by ARC or newly implemented by CMR)
4. Mapping changes (requested by ARC or newly implemented by CMR)
5. Validation rule changes (by either team)
6. Hot topic areas requiring further discussion
7. GCMD Items
8. Open floor

Related URLs Document from ARC Team
https://docs.google.com/document/d/1oMqUGH2c9wxmV0ykRJM4-qnxKKwNKnM1CdZrKPx3Q/edit?usp=sharing

Discussion Items
1. Previous still open action items
   a. 6-29
      Erich Reiter - Pending discussions in section 2 removed Price and DataFormat from the JAMA document - Didn't remove it because I only talked about what I think should happen, but there was no discussion and therefore I don't want to assume agreement.
      Erich Reiter - Pending discussions in section 2 write an issue to remove Price, DataFormat from the ECHO 10 schema. - Didn't write an issue for the same reason as the previous action.
   b. 7-13
      Erich Reiter - Look at code to determine if OnlineResourceURLs/OnlineResourceURL/MimeType values are being converted to UMM Format enum. (Either need to drop it, convert it correctly or enhance ECHO10 format)
   c. 8-24
      Kaylin Bugbee and Jeanne’ le Roux Provide a list of data formats to Erich to possibly add to enum list Plan on sending in the next few weeks
      Kaylin Bugbee ARC to review GCMD platform and instrument list for missing long names Plan to complete this between phase 1 and phase 2 interim (Dec 2017)
   d. 8-31
      Dana Shum - talk to GCMD about validation of Instruments to Platforms
      Dana Shum - write-up change cycle
   e. 9/21
      Erich Reiter - Start documenting how the CMR can auto-calculate the temporal ranges - make it optional. - For now GES-DISC will use end-at-present flag. Counter for spatial - you can validate the granules based on the collections spatial boundaries.
      Kaylin Bugbee - Check the spreadsheet against the the ISO findings to see where the CMR may be reading from more than one place. Code changes and spreadsheet changes may be needed. There appear to be a few duplicate mappings, probably need a separate meeting to talk through this.
   f. 10/12
      Erich Reiter Add the recommendation about collection browse to the decisions made that ARC will disseminate.
1. Erich Reiter create examples from DIF 10 -> UMM that will take the place of Distribution. How does the data from distribution make its way to related url in UMM - DIF 10 data needs to move - be able to describe how to add it to DIF 10.

2. Related URL
   a. DIF 9/10 Distribution section is not being translated - check this - either way we are going to stop because it is being removed from the UMM.
   b. MMT demo of Related URL
   c. We need a solution for the zipped collection level data download URL. ORNL wants to provide a link to download all the collection level data in a zip bundle. Recommendation would be to leverage the description and type elements to describe it. Thoughts?
   d. Get translation examples.

3. Go through RelatedURLs section of the UMM-Common in detail. ARC has prepared a document containing discussion topics which includes a breakdown of each RelatedURL element, along with issues/ discussion topics for each pertaining to both the concept overall as well as mapping issues. Mapping issues are separated out into their own discussion section. This is a google doc currently and we fully expect comments and additions to the document as the meeting occurs. The link can be found here: https://docs.google.com/document/d/1oMqUGH2c9wxmV0ykRUM4-qxnKwKzKwNKe1CdZKXa7Q/edit

4. ARC is not recommending that collections have browse data - there are some tickets about use a granule browse when a collection browse doesn't exist. Bring this recommendation to the DAAC's with the decisions from the All day meeting.

5. Query for empty elements in the CMR.

6. Bulk update demo if time allows
   a. Bulk update changes - ECSE-224
   b. Discuss changing Data Center so that it includes only 0 or 1 home page. Today the CMR model looks like the following:

7. Change Cycles - Unified Metadata Model Change Management

GCMD Items

Proposed GCMD Related_URL Type Keywords (Version 8.6 Draft)
Proposed Unified EOSDIS MimeType List (Draft)
Action Items

- Kathleen Carr, Erich Reiter - Include a warning in the MMT and at ingest when no Related URL/Description is provided.
  - CMR-4620
  - MMT-1245

- Dana Shum - Support the display of the Related URL/Description in the EDSC.
- Kaylin Bugbee, Jeanne’ le Roux - ARC to strongly encourage the inclusion of RelatedURL/Description in reviews.
- Kaylin Bugbee, Jeanne’ le Roux - ARC to review proposed URL Type/Subtype list from GCMD and provide feedback.
- Kaylin Bugbee, Jeanne’ le Roux - ARC to leverage GCMD KMS URL Type/Subtype list as authoritative list once ESO review is complete.
- Erich Reiter, Kathleen Carr - UMM documentation and all other metadata related curation documentation to include a link to GCMD URL Type/Subtype list as a reference.
- Scott Ritz - Add a URL type for ‘Data Citation Policy’ to the URL Type/Subtype list.
- Erich Reiter, Kathleen Carr - CMR team to review proposed URL Type/Subtype list from GCMD and provide feedback.
- Christopher Lynnes, Valerie Dixon - Create a curation mailing list ‘EOSDIS CMR Curation.’
- Kaylin Bugbee - Provide Chris and Valerie with a list of DAAC PoCs to be added to the curation mailing list.
- Tyler Stevens, Scott Ritz - Provide Chris Lynnes with a list of EOSDIS DAAC PoCs to be added to the curation mailing list if applicable.
- Tyler Stevens, Scott Ritz - Incorporate proposed changes to ‘Get Data’ URL type/subtype into GCMD list. Deprecate types as needed.
- Tyler Stevens, Scott Ritz - Incorporate new ‘GET SERVICE’ URL Types in the GCMD URL Type proposed Version 8.6 KMS. Ensure URL Subtypes are properly assigned in relation to the new URL Types.
- Tyler Stevens, Scott Ritz - Create proposed KMS list of data formats.
  - ECSE-256
- Kaylin Bugbee, Jeanne’ le Roux - ARC to share proposed data format list from ORNL and ARC reviews to GCMD for inclusion in data format list.
- Dana Shum - Let EDSC know that we need to display data format information on the collection level page.
- Erich Reiter, Kathleen Carr - Move data format from UMM schema to KMS (once KMS is in place).
  - ECSE-256
- Tyler Stevens, Scott Ritz - Create proposed KMS list for mime type.
- Erich Reiter, Kathleen Carr - Move new element to UMM for ‘native data format’ (exact nomenclature can be discussed as needed).
- Valerie Dixon, Christopher Lynnes - When ready, share data format and mime type KMS lists via communication channels (Slack, mailing list).
- Erich Reiter, Kathleen Carr - Make Get Data/Size and Get Data/Unit not required elements at the collection level.
- Erich Reiter, Kathleen Carr - Provide two separate optional elements for Get Data/Size. One would be called something like ‘AVERAGE FILE SIZE’ and the other would be called something like ‘FILE SIZE’ (See notes for more details).
- Erich Reiter, Kathleen Carr - Update and clarify definition of new ‘Size’ elements under RelatedURLs in the UMM schema.
- Erich Reiter, Kathleen Carr, Simon Cantrell - Update UMM-C Related URL/Services information to align with UMM-Service including writing new definitions for the elements in UMM-C/Common.
- Erich Reiter - Share updated definitions for the Related URL/Services elements to ARC team.
- Kaylin Bugbee, Jeanne’ le Roux - ARC to review updated Services elements definitions.
- Dana Shum, Erich Reiter - Prioritize EDSC ticket 1550 concerning link issues in the abstract in the umm-json view of records.
- Kaylin Bugbee, Jeanne’ le Roux - Reproduce Earthdata login issue & share with Dana & Chris.
- Dana Shum, Erich Reiter, Christopher Lynnes - Complete a test run on Earthdata Drive with ORNL and/or PODAAC (whichever is ready first).
- Kaylin Bugbee, Jeanne’ le Roux - ARC to consider alternate use cases for UMM-Var during UMM-Var Jama review.
- Stephen Berrick - Figure out how to provide DAACs with a list of users who have downloaded specific data products. Also need to provide DAACs URS metrics on a regular basis.

Pending Decisions

After reviewing the meeting notes, from ARC’s perspective, the following items need further clarification and a final decision needs to be made:
1. Native Data Format
   a. Did we decide to only support one native data format for display on the collection level page? Or do we want to support multiple formats?
   b. We want to avoid comma separated data formats in a single element
   c. There are two possible approaches here:
      i. Provide guidance through ARC reviews and documentation that only the primary data format information should be provided here
      ii. Support multiple data format elements so that DAACs can provide multiple formats
   Decision: We will support one element at the collection level for the native data format. The data format information included in the element will be the most important file of the collection.

2. Data Format with URLs
   a. Are we still going to include data format information with Get Data URLs?
      Was still planning to do this

3. Distribution section in UMM
   a. Did we decide to bring back the 'Distribution' section in the UMM?
      If we are only going to bring back only native data format, may not need 'Distribution' section
   b. If so, several actions will need to be added in order to support this (see https://docs.google.com/document/d/1oMqUGH2c9wxmV0ykRJ4M-qnxKwKne1CdZrKPSXx7Q/edit?usp=sharing for more information).
      If sizes are provided by data providers, schema will need to change. If computed, will not change.
   c. From the completed action item that I didn't want to loose, Erich Reiter - mention to ARC and Bruce that we are not translating DIF 10 distribution anymore. The UMM is getting rid of this instead use RelatedURL. Find and let them know what to use instead and give an example. This will change based on the outcomes of the meeting. It has been proposed that the Distribution/Size information be retained (and split into 2 fields; Average granule size and Total collection size) as well as the Distribution/Format (to be mapped to a 'new' UMM element; Native data format which is not tied to a RelatedURL).

Notes from 11/9/17 Discussion:
Valerie votes to compute sizes for granule/collection. If we compute, would not have to add to UMM since it would be computed on the CMR side.
Categorizing format and size information into Distribution helps provide categorization and organization so a Distribution section will help with this.

☐ Erich Reiter - Complete an analysis of strengths and weaknesses of adding Distribution.

Decisions Made:

☐ Type your task here, using "$@" to assign to a user and "$/" to select a due date