IVOA Nov 2021 Interop - DAL 2

Wednesday 3 November 2021 - 20:30 UTC - vconf

Participants: (51) James Dempsey, Grégory Mantelet, Mark Cresitello-Dittmar, Laurent Michel, Patrick Dowler, Pierre Fernique, Marco Molinaro, Tak Lo, Ada Nebot, Séverin Gaudet, Mark Taylor, Marianne Brouty, Janet Evans, Mark Allen, Jesus Salgado, Trey Roby, Tom Donaldson, Yannick Roehlly, Trey Roby, François Bonnarel....

Agenda:

1. DataLink v1.1 WD
2. ProvTAP
3. SODA with Object Storage at CADC

DataLink v1.1 WD

1. Clarification -
2. Links response - link_auth, link_authorized, content_qualifier
3. Service descriptor - contentType, exampleURL params, self-describing service descriptor resource
    1. exampleURL under discussion, see
       https://github.com/ivoa-std/DataLink/issues/55
4. Things delayed to next ver - templated URLs and optional vs required input params in the service descriptor

FB: There was also a discussion about moving service descriptor which has been pushed to the next version
PD: Yes, discussion was moving it from the links to the votable or independant standard

FB: Also dealing with service descriptors for custom services so exampleURL will be very useful for these. DALI examples woudn't be a good match for these
PD: Consensus seems to be going towards keeping exampleURL approach and discussing the name

TR: Question about security content. What happens if the user saves the response?
PD: Yes the saved values would quite likely be out of date in that case. But links table are thought of as transient and to be consumed by the client and not saved.
TR: Security annotations would be useful for our use case

ProvTAP

1. Activity going on for the last year but not visible
2. Why provenance in TAP - can drive discovery by restricting on provenance information
3. The issue - complexity - see ADASS talk and M. Servillat talk in DM
4. ProvTAP status - internal draft on DAL page
5. ProvTAP TAP_SCHEMA
6. ProvHIPS ADQL query examples
    1. requires a few joins so is complex
7. Example of agent and activity which generated entity
8. Issues - denormalised, loop issue, want to create a standardised view
9. Solutions - view with 1 line per used entity - last_step_provenence
10. Solutions - view with 1 line per generated entity - minimum_provenance
11. Examples of these two solutions
12. Success and limitations for view 1
13. Success and limitations for view 2
14. Entity + dataset descriptions
15. Full entity view
16. Reduced query with the full entity view
17. Other views
18. Standard - add some of the views and allow services providing only the views
19. Going further - see ADASS TAP and DM BOF and posters
20. ProvTAP annotation using VODML model_instance attributes
21. ProvTAP - will need clients - either generic or ProvTAP aware
22. Draft will be updated shortly with the new views

OO: Is there any thoughts of having parameters used during generation in the provenance
FB: Yes there is support for parameters, see talk in DM session (13:30 UTC). Will be linked to the activity and will have values

SODA with Object Storage at CADC

1. Background - Storage inventory - including minoc data access service
    1. implemented pixel cutouts ala cfitsio
    2. Have SODA interface to translate from spatial to pixel
2. Storage inventory
    1. libraries - cadc-data-ops-fits and nom-tam-fits in front of storage

adapter
   2. nom-tam-fits - added StandardImageTiler, StreamingImageData classes and RandomAccessDataObject interface
   3. storageadapter has random access adapter
3. SODA implementation (inside minoc)
   1. nom-tam-fits - key is to run cutout operations on the input stream to limit amount of data that is being read
   2. backend is using CEPH object store - using Swift API but S3 could work for this
4. Premilnary results & future work
   1. low level testing - like a slow disc but independant of file size and position in file <10ms to seek and read 32-64 KiB
   2. need to test out segmented files
   3. Adding new params to SODA implementation
5. Collaboration
   1. Implementations in Github on opencadc
   2. nom-tam-fits - currently on a fork but planning to merge to upstream

TR: Very interested in work on upstream nom-tam-fits
PD: Yes keen to do that - have started discussions

FB: Everything that improves SODA is very important, particularly with the emergence of large cubes. See demo tomorrow on Aladin and CASSIS. Accessing large cubes difficult without SODA.

JD: First pass for CASDA has been to download from object storage and then do cutout but that gives large latency for uncached data, particulary for 3.5TB cubes!
PD: The segmenting will impact these large cubes also.

MD: Well, DaCHS has had KIND rather than META for a long time, with values currently either HEADER or DATA. I'm not particularly keen on keeping it that way, but I think planning for more different values might be prudent.
PD: Not settled on the name more introducing the concept

SG: Looking forward to large cubes in SKA and ALMA - in these really need streaming not download to staging

FB: Think this is a good thing that it is appearing now.