

DAL Running Meeting #18

18th April 2024 - 05:00am UTC

Participants: François Bonnarel (FB), James Dempsey (JD), Grégory Mantelet (GM), Mark Taylor (MT), James Tocknell (JT), Markus Demleitner (MD), Renaud Savalle (RS), Robert Butora (RB), Marco Molinaro (MM)

SODA

What do we want to standardize?

- SODA-1.0 only selecting data samples from datasets
- Other UWS services are non standard apart from the process management and can provide ad-hoc functionalities
- Increased data rates from instruments such as SKA drive the need for processing services like SODA
- Data extraction close to the data is still something interesting especially for large data (e.g. Jupyter Notebook)

Metadata delivery

- Needed to support cutouts using WCS
- MD: DaCHS has had the KIND parameter which lets folks retrieve the header only (e.g., http://dc.g-vo.org/flare_survey/q/mdl/dlmeta?ID=ivo%3A//org.gavo.dc/~%3Fflare_survey/data/plates/ESO040_004362.fits). Of course, that's *a bit* FITS specific (cutting out the metadata certainly is more complex in, say, HDF5).
- MT,FB: LOFAR might use HDF5
- FB: CADC also have SODA which retrieves metadata

SODA extension

- HiPS generation on the fly
 - JT: Brent created some code to do this for a survey DC is supporting where the output would be greater than our whole storage capacity. It would worth chatting with Brent as he experimented with how to do this quickly (as the HiPS were being passed to aladin-lite)
 - MT: "application/HiPS" is not legal MIME (unless registered), but you could have "application/x-hips"
 - FB: Agreed
 - MD: or we do it properly this time and just register a media type? This would also be a useful prototype to finally get a media type for VOTable...
- pixel cutouts (extraction of 1 pixel upon n):

- MD: is all for keeping it simple: PIXEL_1 .. PIXEL_n for each axis. It's not clear how we'd deliver multi-array in a useful way anyway. And attaching metadata (length of each axis) is also simple and carefree with PIXEL_n.
- Conversion using RESPONSEFORMAT (e.g. for images)
- Rebinning/resampling
- data product type transformation. MD: This is basically a way to sum up along an axis, right? FB's cube example would suggest that, at least. Perhaps we should do this by axis rather than abstracting this into a data product type?
- Cutout by MOC MD: Does that make sense? It would basically be shifting the bbox calculation from the client to the server, and that seems a rather minor win; if people do things of that complexity, they perhaps shouldn't do that with just a shell script.

MD: Important to provide ranges to users so that they can work out what values to supply to the SODA transformation calls

FB: Agreed, can use call without params to get back the service descriptor (e.g. CFitsIO)

MD: Values in the VOTable should show valid ranges for each axis

MD: Data product type - sounds like summing up along a certain axis. Maybe better to have parameter to sum up along specific axis, so can sum along Dec without having new DPTYPE

FB: Suggest opening a GitHub issue or make comments on PR

MD: What is the use case driving a MOC cutout - wouldn't this just provide a square array with nulled out values outside MOC. Couldn't client provide bounding box and then null out moc itself?

FB: Could be useful for a script to pass the MOC from a discovery service to the SODA service

MD: Better to do this client side and not complicate server protocols and implementations - note this does make sense for discovery services which return a table of matching results.

MM: Agree

FB: One driver is to be able to use the same parameters in both discovery and access

MM: MOC handling is different between discovery and cutout though and seems to add a lot of complexity

JT: If you have the multi-extension fits solved, HDF5 should work

MD: Not sure about that, not only because there are about 2^n ways to encode metadata in HDF5...

JT: If you have the path to the group/dataset, then the metadata is key/value

MD: Well, if only it were that simple; at least for tabular data, every programme distributes it into arrays in slightly different ways, and quite a few put the metadata into arrays, too. But I give you that's probably worse for tables than for

actual arrays... Still HDF5 is hierarchical, so there **is** an additional level of complexity over flat FITS.

FB: Should it become a WD to have more visibility?

JD: Would a side session at Sydney to compare/harmonise SODA extensions be useful?

Many: Yes but MT, MD not attending in person. Could do an afternoon session after the last session

JT: May be a closing time, particularly for hybrid, will check with Simon O'Toole

JT: R.e. meetings after 5:30pm AEST at the interop, we'll need to check with uni security, but we'll see what we can do

Mark Cresitello Dittmar: Many of the new services are directly connected to the original motivation for working on the Cube model and its subcomponents.

They are supposed to give the data model support behind the services:

- image cutouts in WCS and/or Pixels => Coords and Transforms
- dataset metadata transfer => Dataset DM

I would like to make sure there is a component to this project which relates the services and implementations back to the data models to make sure this goal is satisfied.

It also ties into the discussion for the Joint session at the interop...

- how will the implementations tag the information in the query and response to the application.. associating the interface arguments with the model elements.?

FB:

- For the "metadata" feature we can indeed imagine that the serialization of datamodels such as datasetDM, transform, CAOM, ProvDM could be released (for example using MIVOT on top of a VOTable)
- For the data extraction itself it was already the case in SODA1.0 and will be more in SODA1.1 (with resampling for example) that what SODA does is actually forcing the response to match some ObsCore characterization features.
- In other words you are demanding the service to build from the original dataset a new dataset which will have `s_ra`, `s_dec`, `s_fov`, `su_resolution`, `em_min`, `em_max`, etc.... such and such.
- I don't know yet if other datamodels than ObsCore could play the same role

There is a PR on GitHub which I already modified after feedback from Pat.

Includes proposal for :

- Pixel cutouts (instead of world coordinates) on all axes are missing (*SODA GitHub issue #3* - <https://github.com/ivoa-std/SODA/issues/3>)
- No possibility to control output WCS by regridding/rebinning exists in SODA.

(SODA GitHub issue #4 - <https://github.com/ivoa-std/SODA/issues/4>)

- It is not possible to query SODA services by MOC

(SODA GitHub issue #5 - <https://github.com/ivoa-std/SODA/issues/5>)

- SODA spec doesn't tell us how to provide dataproduct_type transformation

(SODA GitHub issue #7 - <https://github.com/ivoa-std/SODA/issues/7>)

- Format transformation : FITS to png/jpeg, FITS to HiPS, HiPS to FITS etc

(SODA GitHub issue #14 - <https://github.com/ivoa-std/SODA/issues/14>)

- Extracting metadata from the dataset

(SODA GitHub issue #15 - <https://github.com/ivoa-std/SODA/issues/15>)

See *GitHub* or also : https://wiki.ivoa.net/twiki/bin/view/IVOA/SODA-1_0-Next

SIA-extended or DAP

See <https://github.com/ivoa-std/SIA> for most issues below
and <https://github.com/ivoa-std/DAP> for the DAP

DAP is the extension of SIA to other dataproducts than images and cubes.

- Input PARAMETERS with limited list of values : better description

(GitHub SIA issue #1 - <https://github.com/ivoa-std/SIA/issues/1>)

- No input PARAMETER exists to select the RELEASE DATE

(GitHub SIA issue #2 - <https://github.com/ivoa-std/SIA/issues/2>)

- No Wild-carding of the input PARAMETERS values exists

(GitHub SIA issue #3 - <https://github.com/ivoa-std/SIA/issues/3>)

- Input PARAMETERS values are case sensitive

(GitHub SIA issue #4 - <https://github.com/ivoa-std/SIA/issues/4>)

- 1 shot discovery (and then access) to cutouts was possible in SIA1 but no more in SIA2

(GitHub SIA issue #6 - <https://github.com/ivoa-std/SIA/issues/6>)

- SIA2 cannot discover rebinned data like SIA1 was able to do

(GitHub SIA issue #8 - <https://github.com/ivoa-std/SIA/issues/8>)

- It is not possible to query SIA services by MOC

(GitHub SIA issue #9 - <https://github.com/ivoa-std/SIA/issues/9>)

- Update matching in-progress DataLink update in examples

- Extension of SIA-style protocol usage outside the image/cube "camp"

(GitHub SIA issue #10 - <https://github.com/ivoa-std/SIA/issues/10>)
-----> See DAP PR #3 - <https://github.com/ivoa-std/DAP/pull/3>

Apart from GitHub SIA and DAP repositories the discussion can alternatively be read there : https://wiki.ivoa.net/twiki/bin/view/IVOA/SIAP-2_0-Next

DataLink

Version 1.1 has been released.

There are still points which have been discussed and not integrated.

See : https://wiki.ivoa.net/twiki/bin/view/IVOA/DataLink-1_1-Next
and GitHub repository - <https://github.com/ivoa-std/DataLink>

- *Service descriptor*: URL could benefit go be templated (for example to pick up part of the path from the user or from the table and not only parameters). Proposals have been made but it was delayed to next version until some service implement it. Probably DataLink 2.0
- (GitHub issue DataLink #27 - <https://github.com/ivoa-std/DataLink/issues/27>)
- *DataLink underlying data model*: is content_type, content_qualifier, semantics triplet enough? Should we distinguish conveyed "information" and conveyed "relationship" (issue DataLink #44 - <https://github.com/ivoa-std/DataLink/issues/44>)? This is a discussion for DataLink 2.0.
- *Service descriptor inputParams* enhancing distinction between required and optional parameters. No obvious solution has been found at the moment.
- (GitHub issue DataLink #51 - <https://github.com/ivoa-std/DataLink/issues/51>)
- MD: If we touch this, let's really have a hard look at translating/simplifying PDL into VO-DML and then declare information like this using MIVOT.
- *Service Descriptor to be removed* from DataLink and pushed to VOTable. This is a major revision of DataLink but also of VOTable. Version 2.0. This may also be completed by introduction of the templating mechanism or even more by the use of PDL inside inputParams.
- (GitHub issue DataLink #53 - <https://github.com/ivoa-std/DataLink/issues/53>)
- *Content parameter in DataLink MIME type*. This is apparently to be solved in VOTable before DataLink-1.1 becomes a REC
- (GitHub issue DataLink #82 - <https://github.com/ivoa-std/DataLink/issues/82>)
- *UCD for ID column in {links} service output*. Is not the same as the UCD of ID parameter in any SODA service referring to the ID column in the {links} table. Is that an issue ? apparently not
- (GitHub issue DataLink #89 - <https://github.com/ivoa-std/DataLink/issues/89>)

And a new one (which may be solved by an erratum?)

- *inputParams* could refer to *PARAM* as alternative to *FIELD*
(<https://github.com/ivoa-std/DataLink/issues/115>)

DataLink implementation note

While finishing the DataLink recommendation process we decided to have some non-normalized proposals for recognition in an Implementation Note.

This is now here:

<https://github.com/ivoa/DataLinkRecImplNote>

And the draft can be read here:

<https://github.com/ivoa/DataLinkRecImplNote/releases/download/auto-pdf-preview/DataLinkImp-draft.pdf>

FB: Comments on the note welcomed and would like to push to document repository soon.

MT: Section 3 overlaps a lot with the DataLink spec

FB: Was trying to explain it in a bit simpler terms to help implementors

MT: Will take a closer look