Running meeting notes

Workshop goals

As you may have noticed, the discussions around the data models are not always consensual.
We have some data models that work fine, such as Obscore, PhotDM, Provenance or specDM.
In fact, disagreements appear when we try to match models with scientific quantities in query responses.
This is a recurrent issue in the VO for years.
The fact is that in 2021, there is still no standard mechanism to annotate, or to completely describe data one can find in catalog data.

- Physical quantities
- Computed quantities
- Coupled quantities

One can do a lot with <GROUP>, but the way that <GROUP> are built is ad-hoc.
The link between groups and DM is rather weak, and this is bad for interoperability.
This has been discussed many time but no solution have been accepted right now

We had a lot of discussion about DM for year, often tense, but:
- There is no consensus about the scope of datamodels
- There is no consensus on the way to use them
- There is no consensus on the way to annotate data

That is the situation now.

However a lot of actions have been undertaken to sort this out:
- 2014: STC annotation note
- 2015: Catalog DM proposal
- 2016: Source DM proposal
- 2016: VODML REC
- 2017: VODML mapping proposal
- 2018: Hands-on at Victoria
- 2019: Source DM session Paris
- 2020: Data Models in the VO (Virtual interop May)
- 2020: Mango proposal (Virtual interop Fall + ADASS)
- 2020: Discussion in the extended DM session (Virtual interop Fall)
- 2021: Meas/Coords RFC
- In addition: topic discussed several times in the TCG

There are multiple reasons to get stuck with this situation, but the major one is likely what can be named as the DM deadlock

- Client developers are waiting for datamodel compliant data before to undertake developpements
- Data providers are waiting on datamodel aware clients before to provide model compliant data
- The fact is that working with datamodel-compliant data is a big step forward and everyone is waiting for the other to dive first.

2 important initiatives have been triggered to break the loop
- Proposal of a component based model (MANGO/ModeInstanceInVot LM 2020) that could soften the dive
- Running a workshop trying to gain some sort of consensus (Tom 2020)
  - Here we are!

The WG has been committed by the TCG to run a workshop trying to fix it, or **the** workshop that will fix it instead.

Not believing in fairy tales, I consider that a simple workshop could not solve something pending for so long.

We believe it will be more efficient to take a bit of time to give the opportunity for all involved people to get all facets of the issues, to contribute and to exercise different solutions. On various use-cases.

After this we will be able to make the summary of this collaborative work in the workshop and then to present the results at the next interop.

After this long introduction, I would like to propose a 3 steps roadmap and to ask for your agreement.

1. Propose an open list of concrete use-cases
2. Give people time to contribute
   a. New use cases
   b. Solution for specific use cases
   c. Comment on any proposals
   d. Questions to the authors
3. Running a workshop where :
   a. involved people could present their contributions
   b. A final discussion will lead to large consensus.

We propose to do this with a Github repository.
We prefer working on Github and then to report the conclusions on the IVOA Wiki.
The reason is that Github has very convenient features for our purpose (code repository, Issue manager, PR, versioning … markdown Wiki).

This repository is 2 folds:
1. A Wiki page where people can feed with free text a poll on the following topics.
   - What is the role of the datamodels in the VO?
   - Do we need another way to map data on datamodels?

- Should data annotations be considered part of the data model specification or should they be model-agnostic?
- Does the VO need one common annotation syntax?
- What are the annotation expectations from a client point of view?
- How to make the data annotation process affordable for data providers?
- Do we expect TAP servers to be able to generate annotations in the future?
    -

Most of these topics are related to the data annotation because this is the major concern for either users, data providers and client developers.

2. A set of concrete use cases with data, documentation and discussion threads.
    - Allow people to see how VOTables can be annotated.
    - Shows examples of code using annotations
    - Interested people can publish there their own examples
    - This is the place for questions responses

The dm-usecases repository

The repository is partially set up.
We have 10 use cases taken from the source DM proposal and also from the Mango use cases.
Each use cas is hosted by a subdirectory containing one or more VOTable with raw data plus one subdirectory for each proposal.
The content of the proposal directories is free. However we suggest to put README files everywhere and to avoid storing too much code there, pointers are better.

7 on 10 use cases already contain proposals based on Mango. This is due to the fact that both Mango and ModelInstanceInVot have been designed and tested upon these use cases

- Questions or remarks on specific points can be addressed with Github issues.
- General remark on a proposal must addresses on a specific Wiki page

Workshop conduct

It is a bit early to setup the workshop running, but assuming we could get a larger audience, we can sketch something in 4 steps
- Remind of goal of the exercice
- Reports on each use cases
- Discussion
- Wrap up.

Date
End of April expected.