Notes

Tuesday Morning

Morning attendees: Ada, Marco, Jiri, Dave, FX, François, Sébastien,Pierre LS, Baptiste, Laurent

Discussion on Time Series Data and Minimal Requirements

- Presentation to set context and questions.
- Should we start from scratch or extend stds?
    - François promotes not developing a new DAL/model standard but extending existing ones.
    - Sebastien warns that new standards could lead to more confusions.
    - Baptiste adding time to existing models/DAL: starting from Obstap would be good
    - Jiri Obscore natural way to discover data even time data
    - Marco TS can contain any type of observable: if described in obscore.subtype by free text: difficult to discover
    - EPN does not use subtypes but UCD lists to describe observables
    - Time axis. JD in double can be better than a microsec: Obsore limits could be fixed with ISO time.
- What is time series and what other things are not time domain specific?
    - Ada: Statistics on time data not part of the discovery
    - François: Phased data can be described by subtypes
    - No vocabulary for subtypes: difficult to use for discovery
    - Baptiste : one must be able to discover TS having one observation at a given time.
    - Ada : we need the exposure time range
    - Marco: unique exposure time not relevant for multi WL data. MM observations: tricky point which can be skipped.
    - Baptiste: can be worked around with min/max values without regards to the WL attached to each exptime.
    - Ada: important to characterise time between observations (variability discovering) (time spam)
    - t_exp/MIN/MAX and t_delta/MIN/MAX seem to be enough to describe time sampling
    - Baptiste: keeping the same names as EPNCore while quantities and format match
    - Obscore t_xel/resolution can be kept

Important fields for discovery purposes.

- **sampling / cadence**: delta_t_min, delta_t_max
- **exposure time**: t_min, t_max, t_mean, t_exp_total
- **date of observation** (min, max): t_min, t_max
- **time resolution/precision**: t_resolution
- number of observations: t_xel
- EPNCore covers most of these fields and so an extension of Obscore should be easy to implement.

Obscore extension: making a difference between analysis data (not in Obscore) and discovery data. Separating char data and discovery data. Obscore extension could provide a vocabulary for subtypes.

Tuesday afternoon

Attendees: Ada, Marco, Jiri, Dave, FX, François, Sébastien,Pierre LS, Baptiste, Laurent, Mireille, Gilles,Thomas

Discussion on data model and serialization:

- Jiri summarized the status of the Note:
  - highlighting the difference between cutouts of dices and slices
  - The model only works with tables thanks to colIDs in the model.
  - The model holds standard statistics for dependant axis.
- Laurent is concerned about column refs in the model.
- Jiri's reply: Axis domain models points to the column refs and not the other way around.
- Emphasis on the ability of joining filters from different tables.
- François mentions the difficulty of extending this to others use cases.
- Mireille presented her UML. This sets the context for models in the IVOA.
- Laurent presented the VO-dml approach.

Wednesday morning

Attendees: Mireille, Gilles, Jiri, François, Sébastien, Ada,Marco, Thomas, Laurent.

Data examples and serialisation.

- Sebastian took some example **data** from VizieR, representing different science cases and Francois presented the **serialisation** as 4 tables to map correctly different parts of the proposed UML by Mireille.
- Laurent presented **data** and **serialization** of XMM-Newton data. 52 observations for one source, so it can be taken as a time series of time series data. Images, spectra and light curves for each observation in the form of a preview (PDF format).
- Marco presented **data** from GAPS: RV against time for discovering planets around stars. Two different velocities, one in the infrared another in the optical, different exposure times in the spectra. There is the need from the user point of view of having information on physical parameters (e.g. activity of the host star). TAPHandle can deal with joins / Obscore customised
- Ada: another example of light curves, spectra and RV **data**: https://sdss-wdms.org/parameters.php#

To combine data coming from different surveys or missions, that is to combine the "times" it is very important to know have certain information. For a good practice see A&A 574, 36 (2015): https://www.aanda.org/articles/aa/abs/2015/02/aa24653-14/aa24653-14.html

Mapping to a pivot format

For each table we need to describe:

- **Time Scale** (time_scale): Can be TT, TDB, TAI, GPS, ET, UTC, TCG, TCB. If we don't know then fill in with "UNKOWN".
- **Time Frame Position** (time_frame_position): place where the time is measured. Observatory (latitude, longitude, altitude), satellite (which?). Are times barycentre corrected already. If we don't know then fill in with "UNKOWN".
- **Time Uncertainty** (time_uncertainty): Resolution of the time
- **Time Systematic Error** (time_sys_error): We can use default values, if time_scale is not known then **100s** as DEFAULT value, if time_scale and time_frame_position are not known then use **1000s** as DEFAULT value. Approximately 100s is good for the time_scale since that's related to changes in the clock in space/earth; 1000s is good if we don't know if times are corrected for the position of the Earth/satellite on its orbit around the Sun since that's approximately twice the time it takes the light to travel in the Sun-Earth orbit.

- **Time Stamp**: the measurement itself
- **Format/unit**: days, hours, seconds,…
- **Offset**: time after GRB that happened on YYYYMMHHMMSS.SS or random number the authors have subtracted from data to provide the data.
- **Description**: "Photometric variability in filter V", "Radial velocity curve in HJD".

In VizieR the recommendation might be to give time in **JD(TDB;BARYCENTRE)**, and then there are easy ways to convert from one system to another.

Time as (offset since reference) is common. In this case, metadata for the time column would need to include the reference time.

Wednesday afternoon

Hands on

We worked in two groups:
1. serialization
   a. Time offsets and origin. We have time scale and ref position. ISO easy to distinguish, but MJD, JD not easy to distinguish.
   b. Compared the three serialization, took same file three annotations and show to the community for discussion. It's in volute:
2. discovery questions
   a. Give me all the time series data around a certain radius of M31 and with at least 5ksec of exposure time: This kind of question makes sense at least for X-ray data. In this case the question is whether with the duration of the observation time be the parameter to check (t_exptime) or would it be it needed to add exposure time information on each element of the block of observations (time_exp_min, time_exp_max). Check with EPNCore team if:
      - time_exp_min == min(t_exptime)
      - time_exp_max == max(t_exptime)

DAL summary

Consensus so far:
- Keep *multi-D DAL framework* as a basis (ObsCore/TAP, SIA2, DataLink,SODA)
- TimeSeries Extensions for ObsCore, SIAV2, SODA
- TimeSeries DataModel and serialization is a specification

How to proceed for these extensions?
- Generic and specific extensions mechanisms as Light spec new versions or endorsed notes?
- A *TimeSeries discovery and access* specification is created. Must rely on all other specs.

Discovery: Obscore
- Set a new TimeSeries extension table of the ivoa TAP schema
  - More columns
  - Restriction/extensions on existing columns

Access: Data Representation requires modelling and serialization → DM task

SODA: TimeSeries generation:
- Add a *DataProductType attribute* to SODA to generate TimeSeries instead of Cubes.
- Add resampling parameter(s) to SODA interface.

SIAV2:
- Reflect new Obscore-like attributes in the SIAV2 query parameters
- Virtual data discovery capability
  - *access reference* is no more a *retrieval* URL but a *SODA* URL.

Data Model summary

Consensus so far:
- Time Series data model is
  - Based on CubeDM.
  - VO-DML-XML document due
  - Time as independent axis
- Points to be discussed:
  - Which dependant axes ?
  - How to describe them
  - Cube model / axis agnosticity ?
- TimeSeries representations / DataModel serializations:
  - Data organization: Main data tables + additional Tables/GROUPS of PARAMS (for metadata).
  - Which DataModel Mapping? Several proposals to be discussed:
    - Utypes (all role and meaning information conveyed at the column level)
      - Classical one (long composed utypes on FIELDS/columns)
      - GROUP/FIELD separation (utypes on both on GROUP, FIELD)
    - VO-DML mapping (rebuild model objects from VOTable)
      - Light (L.Michel)
      - Full mapping (Cresitello)

Summary and conclusions

1. Minimal requirements: Discovery metadata proposals to be exposed and discussed.
2. We have agreed on having another face to face meeting in March.
3. Documents to be written:
   - DM document: Mireille and Jiri have agreed to continue online on DM. Model details to be discussed further before writing a common draft.
   - DAL document: Marco and Francois with suggestions on how to modify what (Obscore, ObsTAP, SODA). DAL chair/vice-chair to propose a DAL guideline as an IVOA note → IVOA discussion to be driven.
4. Three serializations were presented: François, Jiri and Laurent will store data samples annotated with each of these three proposals under:
   - https://volute.g-vo.org/svn/trunk/projects/time-domain/time-series/data_sample/
   - https://volute.g-vo.org/svn/trunk/projects/time-domain/time-series/standardized_votables/
   - For each example dataset:
     - Construct an example votable serialization.
     - Check with the client developers that their application would be able to understand it.
     - Describe how a user would discover this data (what params would find this data).
     - Check with service providers that they will provide these selection criteria.
4. For discovery purposes, inherit keyword vocabulary of source classification from the registry. More accurate classification based on VizeR experience. To propose in later version.
5. Timeline:
   - Dec-Jan collaborate on votable serializations.
   - Mid Jan for teleconference regarding votable serialization(s), unified if possible.
   - Jan-Mar collaborate with client developers (Aladin and Topcat)
   - Mar - aim to have final serialization, with sign-off by client developers. (DADI tech forum Edinburgh?)
   - May - Drafts may appear around the interop in Victoria.