Strasbourg Time Series Data Model
17 - 20 July 2018

Mark Cresitello-Dittmar, Francois Bonnarel, Mireille Louys, Laurent Michel, Ada Nebot

Tuesday 17 July
Morning

- **Revision of the scope and agenda of the meeting**
- **Teleconference with Gaia**

    Present + Alcione Mora, Jesus Salgado, Juan Gonzalez, and more tech team.

    - **Time Series access and content in DR2**
        - Light-curves for each source_id derived from the three cameras and all the transits:
            - For each source there is a light-curve in three bands
            - http://gea.esac.esa.int/archive-help/index.html
            - At the serialisation information on time, fluxes, magnitudes,… are mixed, that is there is one column for band which can take values "G, BP, RP" and one column for time, flux, mag,…
        - Access through datalink: Preferred datalink over TAP due to large amounts (10^8 light-curves) of data in DR3: two step access "TAP+Datalink"
        - Feedback on their implementation of datalink was given by Francois Bonnarel. Sent to Alcione Mora, who will send an internal Gaia ticket.
    - **Time Series access and content in DR3 / DR4**
        - DR3 will be distributed on 2020
        - DR3 & DR4 content still to be defined:
            - DR3 could be similar to DR2
            - DR2-like but adding per-CCD data time / magnitudes. Right now values are average times and magnitudes on all CCDs. This could be included in DR4 but probably not in DR3.
            - Time Series of spectra (see more later)
            - Epoch astrometry (see more later)
        - Meta-data seemingly useless for end users but not for VO third parts can be added if required by the TDIG (e.g. PubID, curation,... fields DatasetMetadata Characterisation)
        - Query mixing **space** (Gaia main table) and **time** (only in lightcurves) is at the moment not possible and there's no obvious solution.
            - Obscore-like table/access could solve this issue.
            - There is a table with statistics on the time series. This could be the source for populating Obscore-like table.
        - Searching which sources have been observed at a given time
            - TMoc might be a solution
            - Obscore-like too searching for a time series dataproduct for a specific Target
    - **Time-Series of epoch spectra**
        - spectra with time data => Time-Series of spectra
        - Fluxes and wavelength are not externally calibrated.
        - User needs to know transformations/conversions.
        - Two options: add metadata vs give externally calibrated spectra. Provide not calibrated data as for X-ray data?
        - Sparse cube model good for multiD view of data but the annotation is too verbose. Note this is the first direct feedback item from a data provider, that the annotated examples are more verbose than they would like to include in their products.
        - ESAC agrees to implement a VO compliant serialization as long as it exists
        - Time scale: need a prototype first soon, not in 2 years.

- **Mean spectra**
  - Not a standard type of spectra (wavelenght, flux) but a set of parameters. Possibly need something like STC transform (Basis Function in spectrum comps) to characterize how spectra has been built (e.g. calib function) (no VO standard).
  - Francois Bonnarel proposes to expose this through a specific service. This data can be exposed to the user without being a VO standard.
  - Ada highlights this is not a TimeSeries related problem.
- **Epoch astrometry**
  - Epoch astrometry per object (there is no VO standard): parallax, proper motions, …
  - There is no defined data at the moment, but it could be residuals with respect to a model,...
  - Will be included in DR4
- **Physical parameters**
  - Providing a set of computed physical parameters with a solution id (there is no VO standard). Monte-Carlo parameter determination, so instead of a value a distribution of possible values. This is not related to a problem to be tackled by or with the TDIG
  - ESAC ask which metadata can be queried
- **PhotDM - VODML compliant version**
  - Action on Jesus with support of Mark Cresitello-Dittmar and Laurent Michel.

Afternoon
- **Telecon notes checking and comment**
  - Gaia needs a framework to support time oriented queries. They are ready to implement a VO/TDIG oriented prototype.
  - FB proposed a fix for Datalink responses
    - Using the semantic proposed in Victoria
    - Setting the file sizes
- **Demo of the GAIA samples in Topcat**
  - We see the description is just text in ascii, with all information mixed in simple sentences.
  - First priority: promote a template for describing simple metadata and propose it to Gaia team. It seems Gaia team is ok to consider what we can propose
  - How to do it: take the example → write metadata in a parameter list → check what the model offers → validate
- **How to proceed for the model?**
  - How to provide something stable enough to rise uptake by Gaia or other projects?
  - Make them (data providers) experiment and validate in following loops.

Wednesday 18 July
Morning
- **How to proceed for the model? (cont.)**
  - Examined the Gaia serialisation with respect to the TD Model
    - Two cases:
      - 1 source 1 LC (Light Curve)
      - Catalog of light Curves (Wise, Gaia, others..),
- **Comparison with classes**
  - Time: It has obviously an offset 2266.43 — this is written in the text

- Band / Filter information: Just a name or a reference to the Filter → need to resolve the filter id with a filter profile service or a photDM instance.
- Flux, mag, error mag, Solution Id
- Signal to noise. There is no UCD - check under http://cdsweb.u-strasbg.fr/UCD/cgi-bin/descr2ucd
- *time*:
  - UCD: should be replaced to time.epoch
  - Units: fraction of days
  - UType see latert
  - Time reference is the same: time ref and offset
    - Refpos: barycenter → coords: TimeFrame.refPosition
    - Timescale TCB → coords: TimeFrame.timescale
    - Time0: 2455…. in JD   the shift to be added to any time value (2010-01-01T00:00:00)   → Fixed for all the survey? yes
    - Time representation JD: which is a DMType
  - How can we define the exposure time in Gaia? Can be done by adding classes to the model: no need to break the actual time stamp (see measurement model). Needed to estimate the signal to noise inside the observation.
- Observable *flux*:
  - Belongs to the Timeseries Model
  - In MCD time series DM : spec package
    - Luminosity relates to a distance of the object → not appropriate here
  - How to bind the flux to the related Filter:
    - In the model : use PhotCal
    - In the column of the VOTable: how to bind it in a column, i.e how to go to the Filter information?
    - Need for a class describing mag and flux
    - Question: should both be bound in the same class: possibly if that class include a photometric system.
    - When fluxes are not calibrated: can be modelled by spectral meas including the error
    - Moving to mag needs a filter definition (photometric measure)
- Observable *mag*:
  - Reminder on the PhotDM and connection to Spectral DM
  - PR model by Mark for a IVOA spectral DM Flux Frame (see photo)
  - FluxFrame connected PhotDM: Photcal
    - No longer on the agenda. Will be a long work to restart Spectral DM now
  - Agreed to work out a solution for the current TD data model, try and prototype the PhotometricMeasure class and add it to *meas* STC package.
  - Mark can add a new class :
    - PhotometricMeasure
    - Attributes
      - Units
      - UCD: to label what kind of photometry: phot.flux, phot.mag, and associated subclasses, etc.
- *SpectralMeasure*
  - Mean spectral value
  - Error on spectral value

Afternoon

- **How to bind filter info to the TimeSeries data VOTable**
  - Different strategies proposed by Francois:
  - ***Strategy 1.a :*** Add a table for filter description into the TS-Votable document. Reference it from the detection row stored in Gaia TS table.
    - See the filter description of PhotDM page 36 to be filled in the Filter table.
    - NB ML: this was used in the example of a SED in PhotDM by Sebastien Derriere
  - ***Strategy 1.b:*** Resolve the filter ID by accessing SVO Filter profile service.
    - It means Filter Id are standard.
    - There is a parametrized URL to get the Filter description as VOtable excerpt.
  - ***Strategy 2:*** Use the VOTable <LINK> construction
    FIELD name=Band value =G etc …
    <LINK Href = 'svo filterprofile description''>
    - This relies on the possibility to query a filter service and to grab the info you need.
  - http://svo2.cab.inta-csic.es/svo/theory/fps/fps.php?Instrument=GAIA&?FORMAT=METADATA will get the description of all filters from the Gaia instrument
    - This supposes we are able to build the proper URL from the band value: data providers and filter provider must agree on this
    - This feature is not supported by VOMDL syntax: a <subset> block must be added for this <SUBSET columnID="band" value="G"> instance annotations </SUBSET>
    - S2 and VOMDL can co-exist since S2 do not impose a specific data arrangement
    - One can get lots of metadata from the spectrum attached to the photDM record: one has to check that this is not specific to Gaia
  - http://svo2.cab.inta-csic.es/svo/theory/fps/index.php?mode=voservice shows how you can get metadata about those filters
  - Questions:
  - Can PhotCal be considered as a frame? Yes
    - This was hooked before by adding Flux frame in spectralDM, but it had nothing more that PhotCal, so in fact FluxFrame is indeed replaced by PhotCal from PhotDM.
  - **ACTION** make sure SVO urls can be templated with FilterIDs, etc…

- **VODML**
  - See section on TS DM Note:
  - For Simple Time series, if we use Cube, then all classes will be be named according to Cube prefix name
  - The cube paradigm: helps to tackle all dimensions
  - Cube: NDpoints of 3 observables
  - TS annotated with band names:
    - timeMeasure required
    - Sorting /splitting by band (TBD)
  - Spectral data (Spectrum or SED)
    - Band (wavelength coords) required

- Sort /group by time
- Possibly aggregate time
- Here we can have SED, SED depending on time (slice by time)
  - Time series multiband: slicing by lambda
  - All axes are discrete/sampled.
- Time series of Spectra
  - Spectral axis = continuous, Time sampled, Flux sampled,...
- Gaia (Jesus) interested in keeping the generality of the Cube axes, and sort, select each depending on the dataproduct axes organisation.
- CoordSpace has axis class :
  - How can we advertise the use of the STC leaf classes ?

- **STC discussion**
  - Francois presented questions about STC Coords / classes attributes
  - Lots of abstracts classes and in general level of classes.
  - ANSW: yes it is complex but it's the price for flexibility
  - Lots of cross refs between abstract classes which must be specialized at mapping time

Thursday
Morning
- Question LM: How can we translate subsettings in the VODML xml document?
- ANSW it's well formalized and then parsable (see simple TS model on Volute).

- **Update of the model - discussion**
  - Progress in the Modelio version by Mark
  - Mark report on photDM in our TS model
    - photDM grey box now has PhotCal, Photfilter mag system zeropoint, transmission curve, etc
  - All the S_ xx classes are the SpectralDM v2.0 candidates classes
  - As SDM2.0 does not exist , its classes are (re)defined in photDM instead of being imported
  - Agreed: Transmission curve to be represented as a spectrum
  - PhotDM classes must be revamped to be VODMlized: GL worked on it
  - All classes of PhotDM transcripted
  - TS model only needs to include S_bounds (start, stop, extent and eff wl) + url on the full instance. The rest of photDM stuff is provided by SVO e.g.
  - How can we be consistent in the serialisation content if the service we point to e.g. SVO FPS does not hold all the filter and PhotCal information?
  - Does the SVO FPS know all the PhotCal objects or do we have to provide it into the serialisation for each dataset?
    - Laurent : Can we just add the needed info in a second table: Lambda eff, url to transmission, f0, system as Ada showed in her regular usage of the filter information?
    - Should we ask the SVO to provide a PhotDM compatible output? We can ask and see.

Afternoon
- **Update of the model - discussion (cont.)**
  - Integration of TS of spectra on the way with spectra as DataProduct
  - Discussion on the roles of DataSet vs DataProduct

- Mireille rose the issue of char in mutiband dataset.
  - Char is degenerated for the lambda axis because there is one point
  - Discussion about one dataproduct per band or one for 3 bands
  - The current model supports both approaches: its an implementation choice
- Ada requests to always provide users and readers with UML basics!!!
- Model updated to support TS of values, of DataProducts and even of TS of not time stamped data (radio visibilities)
- Next question : making STC specific classes for flux/mag.
  - We probably have to define a phot axis which is not in STC.
  - Can be done by constraining the type of PhotometricMeasure
  - Mireille: Here we can rely on the ucd phot.mag or phot.flux
  - Can be done by extended coords with a PhotCoords class which can be extended
- If we have Flux and Mag derived from PhotometricMeasure, then each class as the appropriate methods to handle the representation of values.
- Invoque Flux object or Mag object when needed.
- **Model looks now simple and consistent as well.**
- Just need to know where to put Flux/Mag classes?
- Can be in TS, where we now need it
- Can be with other
- We can have a try to generate the XMI and generate the VODML-TS.xml
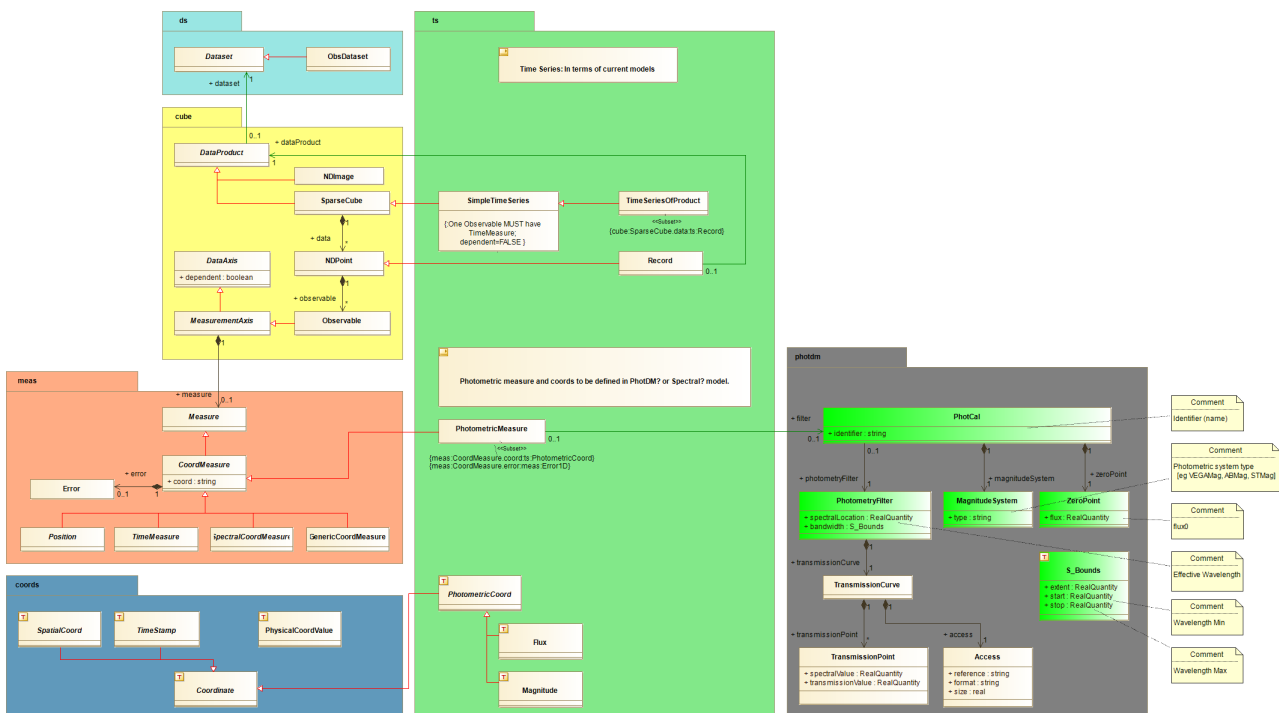
- **Mapping**
  - Four different mapping approaches
    - VOdml
    - VODML lite
    - Utypes approach
    - Jiri approach (utype up to a certain point)
  - One current format for serialisation: VOTABLE
    - Needs a mapping technique to translate instances of the model

- We can imagine other formats: JSON, XML (tree) with XML schema.
    - To be considered later.
- Purpose of TD note was to compare the various mappings and see their pros and cons.
    - Clients must be able to read, interpret, and use the mapping: splat, topcat, vizier, Aladin.
    - Get their feedback is important.
    - What is understandable to the human, what is understandable for programs?
    - The way to convince scientists to use annotated data is to show to client developers the same code can read different datasets as long as they implement the model.
    - We also have to show that the annotation process is affordable for data providers.
- Laurent: there are indeed 3 ways to annotate:
    - the field and provide your definition (like provide the readme for each field)
    - tags on various fields (utypes)
    - Convey the model description inside the serialisation file:
        - lite description
        - full description
- **Why we are limited with utypes?**
    - LM: grouping with several intrication levels makes it difficult for utypes
    - You need to understand the model to interpret the utypes
    - Utypes focused on access (protocols)
    - ML, all agreed: New mapping techno was introduced to enhance the development of applications and develop code easily by instantiating classes from the UML classes.
    - Utypes are easy to use when you are looking for one specific notion, when the path is direct from a root element in the model.
- Ada: annotation must be scriptable
    - The data mapper template must be done once for each data collection . When data sets are very heterogeneous, we must provide an assisted way to the documentalist to define the mapping template (which may be adapted for every dataset entered in Vizier)
    - Example of VizieR Metatime tables
        - Prototype annotation tool for documentalists for time / follows the STC principle
        - We could check to make it completely consistent
        - Example Gaia in vizier
            - Vizier has a version of the Gaia data collection stored at ESA
            - 3 TS coded as 3 RESOURCE in Votable

Friday 20 July
Morning
- **How can we meet the deadlines?**
    - Laurent raised the question how to meet the deadlines, in particular for GAIA.
    - Need feedback soon on the mapping strategy (VODML mapping doc will be discussed mainly when it enters the REC period)
    - How can we get more feedback soon and enter into the mapping discussion at the VO level?
    - Ada: keep a stable version of the DM before people use it

- Francois: should we propose data organisation first and then see how we can do the mapping for it
- Mark: we can suggest data organisation but not mandate it. No archive will be happy with it.
- Ada: reminder use cases :
    - Simple lightcurve
    - Multi filter light curve
    - Multi source multi filter light curve (Catalog)
        - GROUP by source, then by filter
- Catalog of timeseries?
    - Do we need to solve it now?
    - Can start with a good def of multiple filter time series, groups are done by applications (sort by source, then filter)
    - Properties of the sources can be expressed in a different source catalog (and modelled via the Source Catalog DM eventually)
    - Groups are feasible in both mappings LightVODML and Full-VODML
    - Coordinate system is a couple of attributes → should be easy to represent
    - Francois: In catalogs you may want to match a CoordSysid value e.g. ICRS with a record buildup from the mapping.
    - **Mapping of a time frame — Time char — needs TimeSys**
    - **Ref position is missing in the SkyCoordinate**
    - Seems simple for one ICRS system. More tricky if you have several position with diff coordinate system eg. galactic coords/icrs coords
    - See the VOtable errata : http://wiki.ivoa.net/twiki/bin/view/IVOA/VOTable-1_3-Err-1
    - Mireille: I see that when we try to understand the mapping, we need also to understand the Model used as example.
    - Mapping of a time frame — Time char — needs TimeSys

Addition of Timesys
- Example based on http://wiki.ivoa.net/twiki/bin/view/IVOA/VOTable-1_3-Err-1

```
<RESOURCE name="myFavouriteGalaxies">
<COOSYS ID="sys" equinox="J2000" epoch="J2000" system="eq_FK5"/>
<Timeframe id='TF' timescale='TS'    refFramepos='Barycenter'
time0='2,567 000' timerepresentation='TimeOffset'>
%time0 optional
%Given in JD
%TimeOffset will have to be in JD.
<TABLE name="results">
<DESCRIPTION>Velocities and Distance estimations</DESCRIPTION>
<PARAM name="Telescope" datatype="float" ucd="phys.size;instr.tel"
     unit="m" value="3.6"/>
<FIELD name="RA"   ID="col1" ucd="pos.eq.ra;meta.main"
   datatype="float" width="6" precision="2" unit="deg" ref="sys"/>
```

- **This COOSYS/ TIMESYS solution as shown here is in fact a mapping procedure using attributes in specific VOtable ELEMENTS.**
- **TS: note that the actual value would be a valid coords:domain.time.TimeScale instance. (one among TT, TAI, …)**
- It covers the cases of standard frames.
- It needs methods (library) to parse this in application, in addition to the usual parsing methods usually applied to interpret FIELDS PARAMS and Model annotation.
- TimeSYS can include the time0 element as an attribute in order to factorize t0 for all time coord column.
- This allows to reuse the knowledge of the STC:Coords model, and present it in the shape CDS implementers, dealing with large catalogs Vizier, Aladin, etc
- Pb with Time origin : should be always given in the same representation
  - Pbs with the looping question …
  - Solution: fix TimeOffset to JD to avoid loop

- **Data Samples to send to mapper guys:**
  - GAIA help page http://gea.esac.esa.int/archive-help/index.html
  - Vizier   http://vizier.u-strasbg.fr/viz-bin/VizieR-3?-source=I/345/transits&-out.max=50&-out.form=HTML%20Table&-out.add=_r&-out.add=_RAJ,_DEJ&-sort=_r&-oc.form=sexa
- Example of simple VOTABLE annotated by Omar:
  - https://github.com/olaurino/jovial/blob/new-mapping/src/test/resources/votable/example.votable.xml
  - Ada: although in appearance it is not so complicated, it refers to a model which details are to be checked and that needs to be stable.
- Francois : feedback on Utype serialisation:
  - Have to know more about association direction.
  - Took the utypes of the 2013 PhotDM REC and not the one of the FilterProfile service .
  - Should be complete in the photDM Modelio or Magicdraw version.
  - A new question by FB: how to finish the def of PhotometricMeasurement
  - Discussion:
    - Finalising PhotometricMeasure Class
    - Which is the nature of the PhotometricMeasure usually :
      - Ada: can be very different quantities:
        - Flux, mag, differential mag, colors (e.g. ZTF , LSST ?), flux ratio
        - Diff. magnitude: m1 in time 1 - m2 in time 2 → Record m1-m2
        - Color: m1 in band 1 - m2 in bbnd 2 → Record m1-m2
      - Classes flux mag diffmag
      - The mapping will say which nature is the Phot right away
      - If you have more flavors of Phot measure, you can specify by a ucd the kind of quantity : arith.ario; phot.flux
      - Photcal works as a Photometric Frame, we should reference from PhotometricMeasure the PhotCal
      - Should Flux and Mag be under PhotometricMeasure, and not inside coords?

Afternoon: Closing Session
- **Writing a document based on the note to be quickly circulated (* see below)**

- An **email** to interop@ivoa.net summarising the work progress of the week :
    - What we discussed here
    - We have converged towards one unique model
    - We started from Gaia requirements, examples serialisation & feedback
    - Assessment to provide a working prototype by the end of 2019
        - VOTable content from ESAC Datalink responses
        - Model
        - Annotations
- **Action** to involve data providers and client developers in serialization design loop
    - Data provider: Gavo (MD), CDS (GL), Heasarc (TMcG), Ipac (S Groom)
    - Client dev: Splat (MD and Margarida), Topcat (MT), Aladin (PF), Vizier frontend and widget (TB), MAST/IRIS/CIAO/TapHandle already in the loop, Edinburg U (D Morris)
- TSModel (ts namespace)
    - PhotDM model:
        - Relevant fields which must be used in TSDM and serialization
        - URL to the photCal instance
        - Filter id, photometric system, 0-point , Effect wave length, [min max] and band name
        - Action on Jesus with DM (vice)chairs: VODMLization
    - To be complete (MCD)
    - Process VODML (LM)
    - Circulating the diagram
    - Putting the diagram on the Wiki TDIG page
- Data sample
    - Ada will publish the data somewhere (TDIG wiki or volute…)
    - Annotate and send
    - Initiate a loop with data providers and Apps.
- Deadlines
    - PhotDM expected in VODML and wished before October, let's see
    - TS Model for beginning of September
    - Annotation for mid September
    - Starting the loop with data provider and apps dev. by the end of September
    - Running the loop until Interop
    - Showing a summary of our work at Interop

- **TODO on the Note**
    - **Action list**
        - Complete the DM
        - Add documentation to classes inside Modelio
        - Generate xmi → check imports STC:Coords, PhotDM → process to VODML-TS.xml
    - **Rewrite the TS note**: various parts distributed among the authors:
        - Section 1 Intro: to do
        - Section 2 Time Series: rename
        - Make a better separation between discovery model and TS model
        - Remove EPNCore
        - Section 3 goes out
        - Merge sections 4 and 5 and update with the new DM (Mireille & Mark)
        - Section 6 mapping with TIMESYS (Ada&Francois)
    - **Deadline** for recompilation of all parts: end of september.