« Time Series » HackaThon report
21/03/2017

Participants and institutes:
APC: Eric Chassande-Mottin
ARI/UHEI, Heidelberg: Hendrik Heinl, Markus Hundertmark, Grégory Mantelet
CDS: François Bonnarel, Sebastian Derriere, Mireille Louys (Icube, Université de Strasbourg), Ada Nebot, Jenny Sorce, Thomas Boch, Madjid Bouchair, Ibrahim Yapici, Pierre Fernique, Gilles Landais
Czech Technical University in Prague: Jiří Nádvorník
ESAC: Jesus Salgado, Raúl Gutiérrez
INAF/OATs: Marco Molinaro
INTA: Jose Manuel Alacid
JIVE: Mark Kettenis
KM3NeT/CPPM: Vladimir Kulikovskiy
Observatoire Astronomique de Strasbourg: Laurent Michel
SVO: Carlos Rodrigo
UEDIN: Dave Morris

During the hackathon, it was decided to postpone the discussion on data models and serialisation attempts or prototyping. We focused on two topics:

1. Metadata needed to describe TimeSeries

Gathering experience from SVO, High Energy groups (XMM archive and SVOM project), VizieR, Planetary science, Prague group, and GAVO. The question is what metadata are needed for discovery purposes.

- Spatial coordinate system
- Time coordinate frame including time scale (TT, UTC, etc..), reference position, representation (JD, MJD, etc..)
- Time, spectral, space and polarisation characterisation and statistics:
    - raw or mean position
    - raw bounding limits
    - standard deviation

- What are the dependant and independent axes ? Describe the dependant axes quantities.
- Time sampling characterisation and statistics:
    - mean sampling step
    - sampling steps limits
    - sampling step standard deviation
    - total exposure time
    - mean exposure time per step
    - min, max and standard deviation of exposure time per step

- Characterisation on the time frequency axis:
    - Periodograms are another representation of data.
    - We can have period(s) for periodic data or variability
    - We can proceed to frequency analysis and provide coefficient and frequencies following this formula:

$$X(t) = \sum_{i=0}^{n} A_j \times e^{i\omega_j t} + \epsilon(t)$$

We can store the vector: $[A_j, \omega_j]$,
We can also give the phase ($\phi \in [0{:}1]$), defined as: $\quad \phi = \dfrac{(t-t_o)\%P}{P}$
and the zero point (t_0).

- Which mode are the data? variable, or transient? this can be seen on periodogram or by the Target class.
- Target name, class, subclass, are needed (e.g. SN, eclipsing binary, spectroscopic binary,…). This also gives an hint of the variability type. Reuse of standard vocabulary suggested.
- Questions such as "have we more observations on Wednesdays or every day between one and two o'clock?" have to be possible to track artefacts.

2. What do the data we call "TimeSeries" encompass for your group?
- It's a temporal sequence of "measurement points" containing:
  - a time coordinate and either
  - one or several flux(es), with errors, resolution, etc.. or a derivative (mag, mag diff etc..)
  - a radial velocity (double stars, exoplanets)
  - a position (solar activity)

  but also:
  - spectra
  - images ?

  In the latter 2 cases is that a cube (with a sparse axis) or more like an event list (i.e. a catalog of measurements).

- Should we recommend a time representation? Probably yes, and this will be MJD.
- What about relative time (for theoretical data)?

The discussion will continue on the TDIG DAL and DM list and during the TDIG/DAL/DM common session in Shanghai.