DAL architecture for Time Series

F.Bonnarel (CDS as DAL chair)

Acknowledgments :
M.Molinaro, A.Nebot, M.louys and L.Michel for discussions

CENTRE DE DONNÉES ASTRONOMIQUES DE STRASBOURG

Astronomy ESFRI & Research Infrastructure Cluster

IVOA

TimeSeries discovery

- 3 discovery modes
  - Source driven (direct or via DataLink)
  - ObsCore/SIAV2-like driven (are extensions needed ?)
  - Physical Content driven (project specific?)

Source driven (Use case : GAIA)

- We retrieve sources via a TAP or an SCS service
- For each source an URL retrieves TimeSeries
- How do we put a standard tag on this URL ?
  - Utype on a single FIELD ?
  - LINK feature with new « content » attribute ?
  - Service descriptor (DataLink) towards a TimeSeries retriever ?
  - Links to TimeSeries in a {link resource} (DataLink) associated to the source ?

Obscore-like driven (use case :SVO, planets, GAIA, all)

- CoordSystem is ICRS, TT, BARYCENTER
- Obscore allows discovery of « data_product=TimeSeries » datasets with other constraints
- What should be added (if any -→ see Adas' talk)? Eg :
  - Time Support tag (when do we have significant observation?)
  - Time Support summary tags (min/max of « parts »)
  - Time sampling frequency, or frequency bounds tags
  - Time sample width bounds tags
  - Extend o_ucd value domain : what is varying with time
- Close to previous SSA-like approach (SVO)

Physical Content- driven (INAF exoplanets, ESA missions)

- List of metadata
  - Signal periodicity
  - Periods
  - Object type candidate (exoplanet, variable star, etc..)
  - Transiancy
  - Artefacts
  - Etc…
- Requires specific analysis
  - Project specific
  - Additional physical content metadata table.
  - Joints to Obscore-like table

TimeSeries representation and access

- Proper TimeFrame and Time representation definition (stc)
  - Time Scale
  - Reference position
  - Representation : JD, MJD, Offset (no ISO, no HJD…)
- TimeSeries data model :
  - based on CubeDM ? VO-DML-XML document due
  - Which dependant axes ?
  - Which Mapping/serialization (from VO-DML or whatever)
  - -->Tommorrow session

TimeSeries representation and access (2)

- Single URL-based retrieval to TimeSeries insufficient
  - Extraction from a database : SODA-like approach to generate the actual TimeSeries
  - Usefull to discover which TimeSeries can be generated by a service : virtual data discovery.

How does DAL can tackle all this ?

- This is only a proposal
- See also the two last slides of FB talk in Shanghai : http://wiki.ivoa.net/internal/IVOA/InterOpMay2017-TD/TimeSeries.pdf

How does DAL can tackle all this ?
1) Obscore

- Pick up the required characterization of the Time axis (see slide 3) from TimeSeries datamodel
- set a new TimeSeries extension table of the ivoa TAP schema. Create this :
  - As an endorsed note ?
  - As a real new standard ?

How does DAL can tackle all this ?
2) representation

- Requires modelling and serialization
- ---> It's a DM task (see tommorrow)

How does DAL can tackle all this ?
3) TimeSeries generation

- Add a « DataProductType attribute » to SODA (to generate TimeSeries instead of Cubes)
- Add resampling parameter(s) to SODA interface
- Again :
  - new version ?
  - Endorsed note ?

How does DAL can tackle all this ?
4) SIAV2 (PARAM-based) discovery

- Reflect new Obscore-like attributes in the SIAV2 query parameters
- Virtual data discovery capability
  - (access reference is not a retrieval URL but a SODA URL)
- Again:
  - endorsed note ?
  - or new version of the standard ?

How does DAL can tackle all this ? Provisional conclusion

- TimeSeries DataModel and serialization is a spec
- 3 endorsed notes = generic (and agnostic) mechanism to complement Obscore with datasettype-sepcific attributes and extend SIAV2 Parameters and capabilities and SODA parameters and capabilities
- A « TimeSeries discovery and access » specification is created
  - Is based on DataModel Spec
  - Is base on the 3 endorsed notes described above
  - Contains all TimeSeries specific capabilities and parameters definition
  - Provides a mechanism for discovery of TimeSeries in source-driven context