ObsTAP Mandatoriness

To Null or Not to Null?

2010.05.18

A.Micol

Why ObsTAP?

- CSP (Schade): Maintain a focus on building usage of VO services by astronomers

- ObsTAP focus: Astronomers, not Data Providers!
  - Aim is to ease SCIENCE not to ease Data Publishing

- Science cases:
  - 1.1. Show me a list of all data that satisfies:
    - 1.1.1. DataType=Image
    - 1.1.2. Wavelength=V or I or Z
    - 1.1.3. Spatial Resolution < 0.7 arcseconds FWHM
    - 1.1.4. Exposure Time > 1000 Seconds
    - 1.1.5. Data Quality=Fully Calibrated

ObsTAP tables

- Core part: mandatory columns
  - Table (or view) with mandatory columns
  - with fixed names, units, and reference frames.
- Extended part: optional columns
  - Table with optional columns that an obstap service might include
- Universal queries:
  - One and same query can be sent to all obscore

To NULL or NOT To NULL?

- Full compliance to selected science cases obtained only through complete mandatoriness
- If NULL are NOT allowed:
  + All science cases are satisfied
  + No record excluded on the basis of “I do not know and didn’t bother to ask”
  + Result sets will always be complete
  + Result Set analysis not compromised:
    - Sorting, visualising important for refining queries
  – More work for data providers

To NULL or NOT To NULL?

- If NULL are allowed:
  - some science cases won't be satisfied;
  - laziness might (and will) prevail;
  + more collections will be available sooner;
  - Result sets won't be complete, hampering analysis of results
    - No sorting, no visualisation, incomplete metadata
  - Astronomers will not have full overview/control

To NULL or NOT To NULL?

- Quantity versus Quality: Is more better?
- Personal view:
  - ObsCore = La crème de la crème
    - Only fully characterised data products
    - A strong service
  - For the rest: regular TAP, or the extended ObsTAP, can be used.
    - A weak service
- Including 90% of all data is “good enough”
- Including non-characterisable data (<<10%) is not good at all.

ESO ObsTAP choices for ESO + HST data

s_region	all	Bounding box, not footprint (lazy)
- s_fov	image	Longest diagonal of box
- s_fov	spectrum	Shortest size of box
s_resolution	spectrum	= s_fov
t_span	all	= t_stop - t_start
t_resolution	all	= t_span
t_exptime	all	Total exposure time
em_domain	all	Function(em_min, em_max)
em_res_power	image	(em_min+em_max)/2/(em_max-em_min)

To NULL or NOT To NULL?

- More work for data providers:
  - In which case is it really IMPOSSIBLE to assign approximated (good enough) values?
  - On which of the few ObsCore attributes?
  - So far, no compelling answers!
  - E.g., t_exptime (wrong name?) depth indicator valid only within one collection, useful to avoid e.g. shallow products. Does not need to be the REAL exposure time, it is not going to be used in data analysis, only for data discovery.
- Difference between: “I do not know” and “I didn’t have the time”