

Some Thoughts on TAP and Other Ramblings

Ian Evans

Chandra X-ray Center Data System

End-to-End Scientist

All opinions are solely those of the author

Or....

- ... how user-friendly is TAP?

```
public
├── public.flux
│   ├── oidref
│   ├── filter
│   ├── flux
│   ├── flux_prec
│   ├── flux_err
│   ├── flux_err_prec
│   ├── qual
│   └── bibcode
├── public.mesXmm
├── public.mesHgamma
├── public.mesRot
├── public.mesGcrv
├── public.mesposa
├── public.mesMK
├── public.mesDistance
├── public.mesISO
├── public.mesClg
├── public.mesFe_h
├── public.author
├── public.mesIRAS
├── public.ident
├── public.mesCEL
├── public.basic
├── public.otypedef
├── public.mesIRC
```

- TAP presents a **flat** and rather **unfriendly** list of database table and table column names
- The scientist must spend time trying to determine which tables and columns to query
- Sometimes the names are suggestive and that may simplify the determination, but in other cases they are too closely tied to the underlying database structure to be clear
- **In any case, this should be the computer's job and not the scientist's!**
- Sometimes the **list of tables** itself can be overwhelming! (next slide)

- + AtlasOutline
- + DBColumns
- + DBObjects
- + DBViewCols
- + DataConstants
- + Dependency
- + Diagnostics
- + FIRST
- + Field
- + FieldProfile
- + FileGroupMap
- + Frame
- + HalfSpace
- + History
- + IndexMap
- + Inventory
- + LoadHistory
- + Mask
- + MaskedObject
- + PartitionMap
- + PhotoObjAll
- + PhotoObjDR7
- + PhotoPrimaryDR7
- + PhotoProfile
- + Photoz
- + PhotozRF
- + PhotozRFTemplateCoeff
- + PhotozTemplateCoeff

- + Plate2Target
- + PlateX
- + ProfileDefs
- + ProperMotions
- + PubHistory
- + QueryResults
- + RC3
- + ROSAT
- + RecentQueries
- + Region
- + Region2Box
- + RegionArcs
- + RegionPatch
- + Rmatrix
- + Run
- + RunShift
- + SDSSConstants
- + SiteConstants
- + SiteDBs
- + SiteDiagnostics
- + SpecDR7
- + SpecObjAll
- + SpecPhotoAll
- + StripeDefs
- + Target
- + TargetInfo
- + TwoMass
- + TwoMassXSC

- + USNO
- + Versions
- + Zone
- + detectionIndex
- + emissionLinesPort
- + galSpecExtra
- + galSpecIndx
- + galSpecInfo
- + galSpecLine
- + neighbors
- + sdssBestTarget2Sector
- + sdssImagingHalfSpaces
- + sdssPolygon2Field
- + sdssPolygons
- + sdssSector
- + sdssSector2Tile
- + sdssTargetParam
- + sdssTileAll
- + sdssTiledTargetAll
- + sdssTilingGeometry
- + sdssTilingInfo
- + sdssTilingRun
- + segueTargetAll
- + sppLines
- + sppParams
- + sppTargets
- + stellarMassPCAWisc
- + stellarMassPassivePort

OBSID
REVOLUT
MJD_START
MJD_STOP
OBS_CLASS
PN_FILTER
M1_FILTER
M2_FILTER
PN_SUBMODE
M1_SUBMODE
M2_SUBMODE
RA
DEC
POSERR
LII
BII
RADEC_ERR
SYSERR
RA_UNC
DEC_UNC
CX
CY
CZ
HTMID
EP_1_FLUX
EP_1_FLUX_ERR
EP_2_FLUX
EP_2_FLUX_ERR

- Within a table the scientist is often faced with **a flat list of columns without any obvious underlying structure**
- Here for example, position, observation, instrument, and flux properties are all present in a single table
- This may be OK if there are only a few columns but quickly becomes unwieldy as the number increases

Name	Table	Primary Key	Datatype
flux	public.flux	false	real
flux_prec	public.flux	false	smallint

Units	UCD	Description
	instr.precision;phot.flux	flux precision

- Some metadata (units, description, UCD) about the tables and columns may be accessible (and searchable) by the scientist, but **these are not always fully and consistently populated** (see the **flux** column above)
- The **description field** could be a good heuristic to identify columns of interest if concise (e.g., the **flux precision** column above) but not if it provides an extensive description (see the example below)
- But the extended description may be essential for understanding the data content!

Name	Table	Primary Key	Datatype	Units	UCD	Description
SUM_FLAG	twoxmm	false	integer			The summary flag of the source is derived from EP_FLAG. It is 0 if none of the nine flags was set; it is set to 1 if at least one of the warning flags (flag 1, 2, 3, 9) was set but no possible-spurious-detection flag (flag 7, 8); it is set to 2 if at least one of the possible-spurious-detection flags (flag 7, 8) was set but not the manual flag (flag 11); it is set to 3 if the manual flag (flag 11) was set but no possible-spurious-detection flags (flag 7, 8); it is set to 4 if the manual flag (flag 11) as well as one of the possible-spurious-detection flags (flag 7, 8) is set. The meaning is thus: 0 = good, 1 = source parameters may be affected, 2 = possibly spurious, 3 = located in a area where spurious detection may occur, 4 = located in a area where spurious detection may occur and possibly spurious. .

- ▼ Master Sources
 - msid
 - ▶ Source Name
 - ▼ Source Position
 - ▶ ICRS Equatorial Coordinates
 - ▶ Galactic Coordinates
 - ▶ Position Error Ellipse
 - ▶ Source Flux Significance (S/N)
 - ▶ Source Flags
 - ▼ Source Extent
 - ▶ Deconvolved Source Ellipse
 - ▼ Aperture Photometry
 - ▼ Source Region Aperture Fluxes
 - ▶ Photon Fluxes
 - ▶ Energy Fluxes
 - ▶ Spectral Model Energy Fluxes
 - ▶ PSF 90% ECF Aperture Fluxes
 - ▶ Spectral Hardness Ratios
 - ▼ Model Spectral Fits
 - ▶ Power-Law Model Spectral Fit
 - ▶ Black-Body Model Spectral Fit
 - ▶ Galactic Neutral Hydrogen Column Density
 - ▼ Temporal Variability
 - ▶ Intra-Observation Variability
 - ▶ Inter-Observation Variability
 - ▼ Observation Summary
 - ▶ ACIS Observations
 - ▶ HRC Observations

- For the *Chandra* Source Catalog query tool (CSCview) we provide a **hierarchy** of properties with **human-readable** titles
- The property titles are kept short so they can be easily displayed, but are also (hopefully) **intuitive**
- The hierarchies are kept to a few levels, so that the scientist can drill down the list of property titles quickly, e.g.

Master Sources →

Aperture Photometry →

Source Region Aperture Fluxes →

Energy Fluxes

- ▼ Source Observations
 - posid
 - ▶ Observation-Specific Information
 - ▼ Detected Source Properties
 - ▶ Observation-Specific Source Identification
 - ▶ Source Position
 - ▶ Source Significance
 - ▶ Source Codes and Flags
 - ▶ Source Extent
 - ▼ Aperture Photometry
 - ▶ Aperture ICRS Equatorial Coordinates
 - ▶ Source Region Aperture
 - ▶ PSF 90% ECF Aperture
 - ▶ PSF Aperture Fractions
 - ▼ Source Region Aperture Fluxes
 - ▶ Total Counts
 - ▶ Net Counts
 - ▶ Net Count Rates
 - ▶ Photon Fluxes
 - ▼ Energy Fluxes
 - ▼ ACIS Broad Energy Band
 - flux_aper_b
 - flux_aper_lolim_b
 - flux_aper_hilim_b
 - ▼ ACIS Hard Energy Band
 - flux_aper_h
 - flux_aper_lolim_h
 - flux_aper_hilim_h

- Only at the very bottom level does the scientist see the actual database column names
- We try to make the columns names reasonably **self-descriptive** so that a property title is not required for each column
- Metadata (datatype, units, description) are provided for each data column
- The descriptions provide the detail necessary to interpret the column data
- Because of the large number of data columns (~900) multiplexed over 6 energy bands, with most properties having separate lower and upper confidence limits we provide **multiple views of the property hierarchy**
- In this display, the hierarchy is Property → Energy Band

- ▼ Source Observations
 - ▼ Energy Band Independent Properties
 - posid
 - ▶ Observation-Specific Information
 - ▶ Detected Source Properties
 - ▼ Energy Band Dependent Properties
 - ▼ ACIS Broad Energy Band
 - ▶ Source Position
 - ▶ Source Significance
 - ▶ Source Extent
 - ▼ Aperture Photometry
 - ▶ PSF 90% ECF Aperture
 - ▶ PSF Aperture Fractions
 - ▼ Source Region Aperture Fluxes
 - ▶ Total Counts
 - ▶ Net Counts
 - ▶ Net Count Rates
 - ▶ Photon Fluxes
 - ▼ Energy Fluxes
 - flux_aper_b
 - flux_aper_lolim_b
 - flux_aper_hilim_b
 - ▼ Spectral Model Energy Fluxes
 - ▼ Power-Law Model Energy Fluxes
 - flux_powlaw_aper_b
 - flux_powlaw_aper_lolim_b
 - flux_powlaw_aper_hilim_b
 - ▶ Black-Body Model Energy Fluxes

- Alternatively, a scientist who is primarily interested in a single energy band (e.g., the broad band) can select
Energy Band → Property
as the hierarchy

On Position Queries ...

- A significant majority of the *Chandra* Source Catalog queries are **position searches surrounding a point**, where the user wants to **select** and **retrieve** properties for a possible X-ray source at a predefined location, e.g.,

POINT('ICRS', ra, dec), CIRCLE('ICRS', 40.669879, -0.013289, 0.016666667)

- However, a large fraction of TAP services **do not** support this kind of position query!
 - Out of 50 TAP services queried, only 8 (**16%**) have implemented this capability
 - Could this vitally important capability be **too hard to implement** ?

On Domain-Specific Information ...

- TAP also needs to support **domain-specific representations as determined by the data provider**

- For example, the *Chandra* Source Catalog uses the following energy bands:

Ultra-soft	u	0.2–0.5 keV	Soft	s	0.5–1.2 keV
Medium	m	1.2–2.0 keV	Hard	h	2.0–7.0 keV
Broad	b	0.5–7.0 keV	Wide	w	0.1–10 keV

- The existing UCD's, which define `em.X-ray.soft` as “Soft X-ray (0.12–2 keV)” and `em.X-ray.medium` as “Medium X-ray (2–12 keV)” are clearly not adequate!
- Properties such as these are certainly domain-specific and are often **facility-specific**
- **All properties should be representable using domain-specific units**
 - For example, ObsCore requires representing the spectral limits `em_min` and `em_max` in units of meters
 - For *Chandra*, that's roughly 0.0000000002 and 0.0000000006
 - That may be OK for a machine to interpret but is clearly not very useful for an X-ray astronomer (not to mention the data provider) !

The Bottom Line:

- ... not very