



Data Model: DAQ outlook

E. Lyard

ISDC – Univ. Geneva

FACULTÉ DES SCIENCES
Section d'Astronomie



**UNIVERSITÉ
DE GENÈVE**

The DAQ Outlook

- 7 different cameras
 - Formats to be unified “as much as possible”
- Data will be compressed in storage
 - IO layer does the bits-shuffling
- Reflection (aka Introspection) would allow us to keep several format during commissioning
 - Will we select the file-format before finalizing the data model ?

7 different cameras

- » ASTRI
- » DIGI
- » FLASH
- » GCT
- » LST
- » NECTAR
- » SCT

- Accommodate optional fields
 - And avoid piles of IFs in the code
- Accommodate versioning
 - Without having to change the code at 20+ locations
 - Until hardware is finalized and unlikely to change ever again
 - By then we can get rid of extra features and optimize for performances
 - This may introduce a fair amount of IFs and/or meta-programming in the framework

Compressed in storage

- Compression is expensive
 - Difficult to vectorize
 - Take the opportunity to make the IO-layer deliver data in the desired format
 - Should be cheap compared to compression
 - Could be used to get data “as we like it”, e.g. following a given pixels ordering, calibrated,...
- Calibrate events on-the-fly ?
 - Implies to store calibration constants close to the data

Introspection

- Need to be able to store data
 - Requires a file format / IO layer
 - Do not think twice when reading old data
 - We must select the format asap
- But....
 - New requirements may appear in the future
 - Introspection would allow to read/write to/from different formats, with new fields
 - Transcode automatically all data sets in the final, best format once the CTA instrument is better understood
 - Validate transcoding by re-running analysis on new format, and compare results
 - Possibly get rid of introspection at that time, low-level optimization

Hillas test in DLO data challenge

- Algorithms performance closely related to layout of data in memory
 - Compare Hillas perfs. of each of the proposed formats
 - Compare extracted parameters to validate results (i.e. must be close within a few percents)
 - Overall ranking takes the full chain into account
- Should we measure number of CPU cycles along with overall time ?