# IVOA Provenance Data Model:

# Hints from the CTA provenance prototype

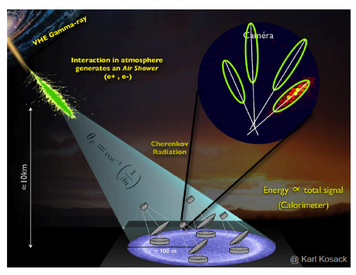
Michele Sanguillon[1], Mathieu Servillat[2], Mireille Louys[3],
François Bonnarel[3], Catherine Boisson[2] ,Johan Bregeon[1]

*[1]LUPM-France, [2]LUTH-France, [3]CDS-France*

## Abstract

We present the last developments on the IVOA Provenance data model, mainly based on the W3C PROV concept. In the context of the Cherenkov astronomy, the data processing stages imply both assumptions and comparison to dedicated simulations. As a consequence, Provenance information is crucial to the end user in order to interpret the high level data products. The Cherenkov Telescope Array (CTA), currently in preparation, is thus a perfect test case for the development of an IVOA standard on Provenance information. We describe general use-cases for the computational Provenance in the CTA production pipeline and explore the proposed W3C notations like PROV-N formats, as well as Provenance access solutions.

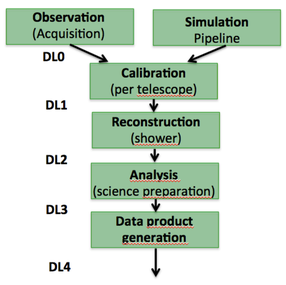
## Cherenkov Astronomy Context: Complex data




- Very high energy gamma ray instrument
- Indirect detection
- Need simulations to compare acquired data to expected ones
  => Complex data :

CTA will be **open to the community**.
High level data (event lists, spectra, sky maps) available through the Virtual Observatory.

| Data Level | Short Name | Description | Data reduction factor |
|---|---|---|---|
| Level 0 (DL0) | DAQ-RAW | Data from the Data Acquisition hardware/software. | |
| Level 1 (DL1) | CALIBRATED | Physical quantities measured in each separate camera: photons, arrival times, etc., and per-telescope parameters derived from those quantities. | 1-0.2 |
| Level 2 (DL2) | RECONSTRUCTED | Reconstructed shower parameters (per event, no longer per-telescope) such as energy, direction, particle ID, and related signal discrimination parameters. | $10^{-1}$ |
| Level 3 (DL3) | REDUCED | Sets of selected (e.g. gamma-ray-candidate) events, along with associated instrumental response characterizations and any technical data needed for science analysis. | $10^{-2}$ |
| Level 4 (DL4) | SCIENCE | High Level binned data products like spectra, sky maps, or light curves. | $10^{-3}$ |
| Level 5 (DL5) | OBSERVATORY | Legacy observatory data, such as CTA survey sky maps or the CTA source catalog. | $10^{-5}$ - $10^{-3}$ |




→**Expose progenitors of science data products**

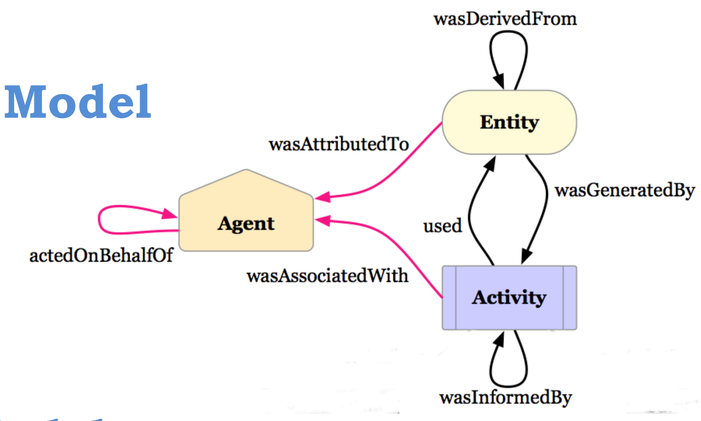
Users need:

- To know what we are talking about : Data Model
- To know how data sets were produced : Provenance description
- To select data on provenance criteria : Query

## Provenance Data Models

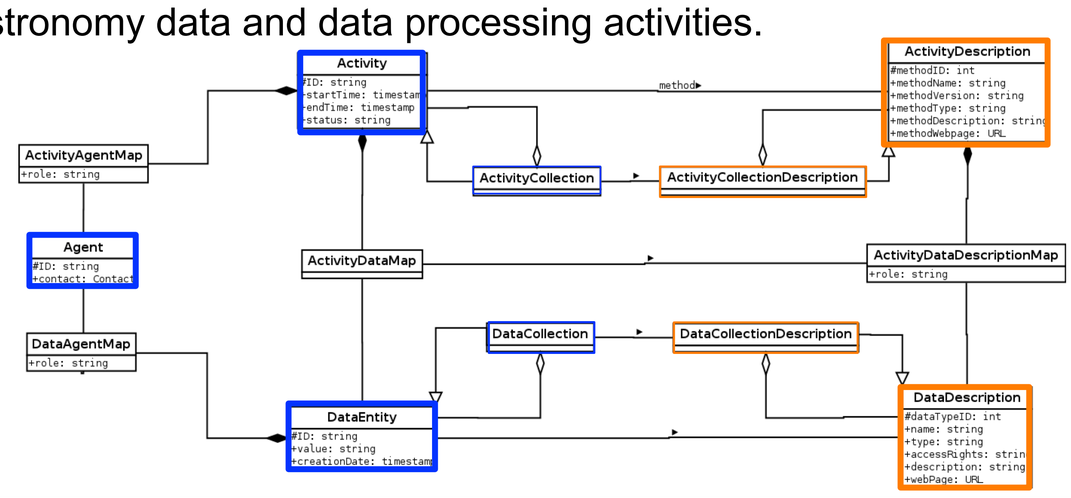
### W3C Provenance Data Model

The model endorsed by the W3C is based on 3 components and relations that connect them to each other.

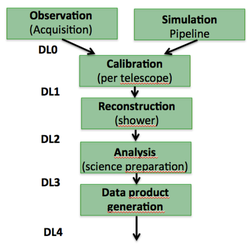



### VO Provenance Data Model

adapted to astronomy data and data processing activities.



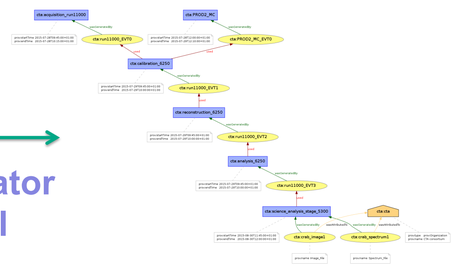

## Provenance description




PROV-N

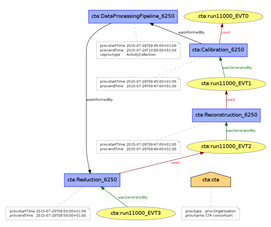
```
// ------------------------------------------------------------------------
// Calibration
activity(cta:calibration_6250, 2015-07-29T09:45:00, 2015-07-29T10:00:00, -)
entity(cta:run11000_EVT1, -)
wasGeneratedBy(cta:run11000_EVT1,cta:calibration_6250,-)
used(cta:calibration_6250,cta:run11000_EVT0, -)
used(cta:calibration_6250, cta:PROD2_MC_EVT0,-)
```

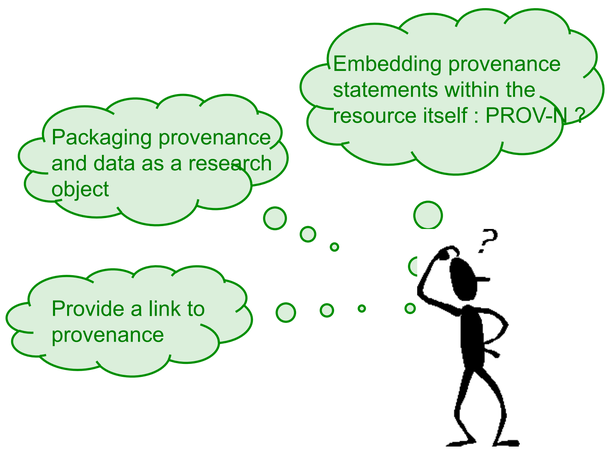
Translator tool



**INTEROPERABILITY : STANDARDIZED DESCRIPTION LANGUAGE PROV-N**

This model is not fully compliant with the W3C model because it does not exist in the latter the opportunity to describe a collection of activities or workflow. We get round the problem by describing on the one hand a workflow and on the second hand independently the activities that compose it. The wasInformedBy relation is used to indicate the beginning and end of the workflow.



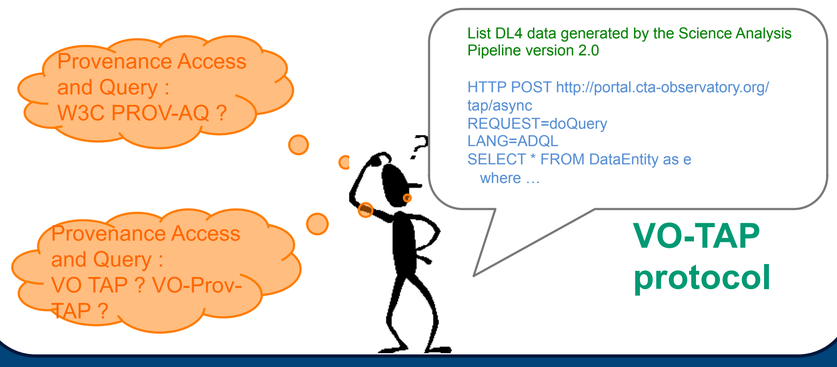



## Provenance query

Selection criteria could be:

- Name of attributes of the provenance data model → **Identified**
- Name of attributes specific to the experiment (run number, ambiant conditions, …) → We need to identify the specific provenance items for each data level.

The VO user need to query which specific attributes could be a criteria




**Acknowledgements:** ASTERICS (http://www.asterics2020.eu/) is a project supported by the European Commission Framework Programme Horizon 2020 Research and Innovation action under grant agreement n. 653477; Additional funding was provided by the INSU (Action Spécifique Observatoire Virtuel, ASOV), the Action Fédératrice CTA at the Observatoire de Paris and the Paris Astronomical Data Centre.