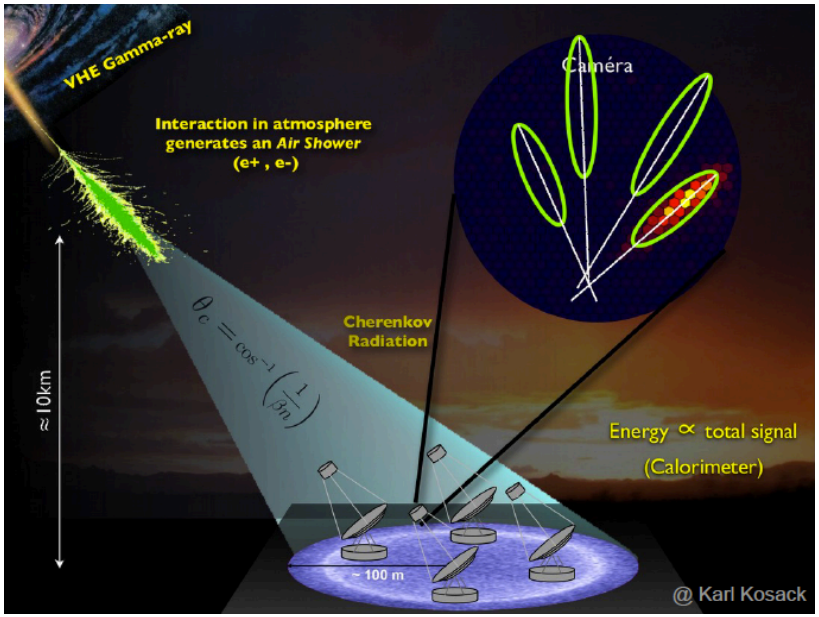
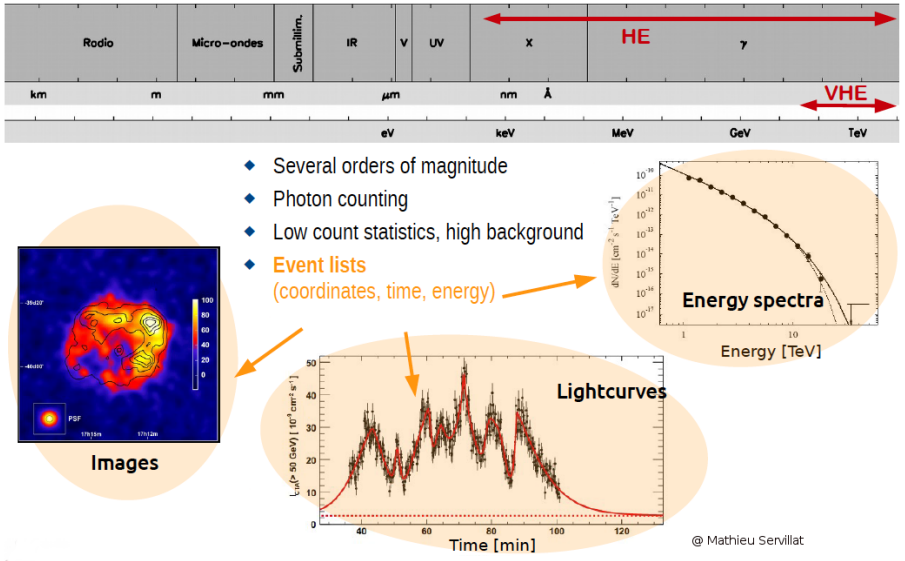
# IVOA PROVENANCE DATA MODEL: Hints from the CTA provenance prototype


**Michèle Sanguillon, Johan Bregeon** – **LUPM, Montpellier, France**
**Mathieu Servillat, Catherine Boisson** – **LUTH, Meudon, France**
**Mireille Louys, François Bonnarel** – **CDS, Strasbourg, France**

# Cherenkov Astronomy Context




- Very high energy gamma ray instrument
- 3 types of telescopes in CTA
- Complex data :
  - Indirect detection
  - Need simulations to compare acquired data to expected ones
- Final products data available on the VO

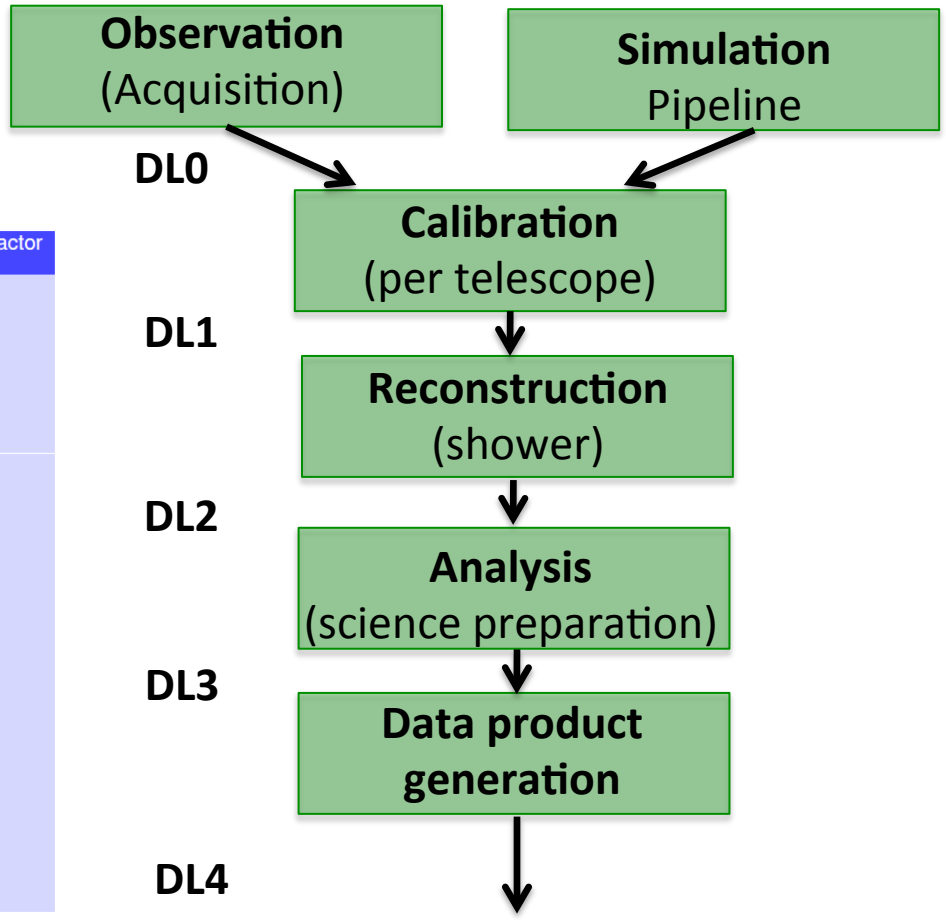



CTA will be the first Cherenkov Observatory providing its high level data (event lists, spectra, sky maps) on the Virtual Observatory

# Data and workflows

Different levels of data : DL0 to DL5.
DL3, DL4 and DL5 data available on the VO

| 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. | $10^{-5}$ - $10^{-3}$ |



# Provenance : what do VO users need ?

- **To know what we are talking about : Data Model**

- **To know how data sets were produced : Provenance description**
  (Observing process and conditions, Data reduction, selection and extraction methods applied to raw measures to build up science-ready data products

  => To establish quality, relevance

  => To track information through complex transformation

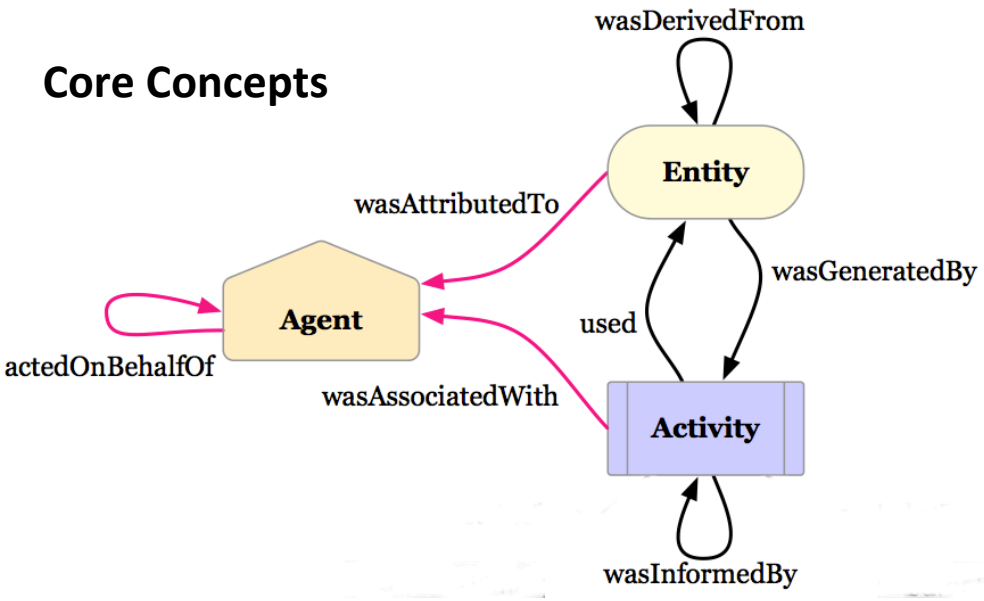
  => To describe the experiment for understanding

  => To provide evidence in support of scientific claims

- **To select data on provenance criteria : Query**
  - Derive selection criteria to filter out suitable data for their scientific needs
  - Estimate better which data release fits the best for their needs

**→Expose progenitors of science data products**

# Provenance in the W3C




**Benefits :**

- 4 recommendations and a number of non-prescriptive notes
- Tools to validate and translate a description format in another
- Possible to define our own attributes

**4 recommendations (30/04/2013)**

PROV-DM : the PROV data model

PROV-O : the PROV ontology

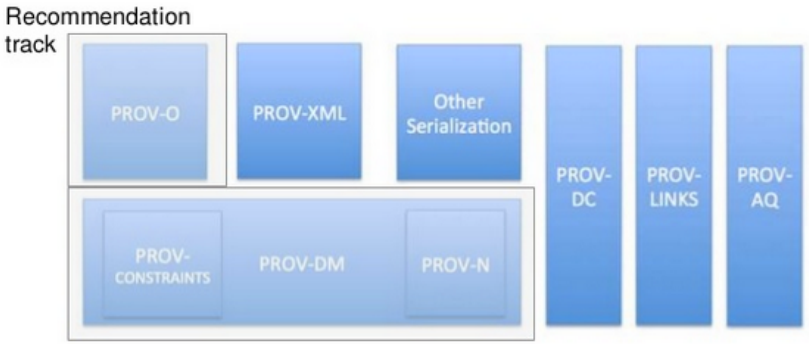
PROV-Constraints : Constraints of the PROV Data Model

PROV-N : a notation for provenance aimed at human consumption

**and a number of non-prescriptive notes**

PROV-XML : an XML schema for the PROV data model

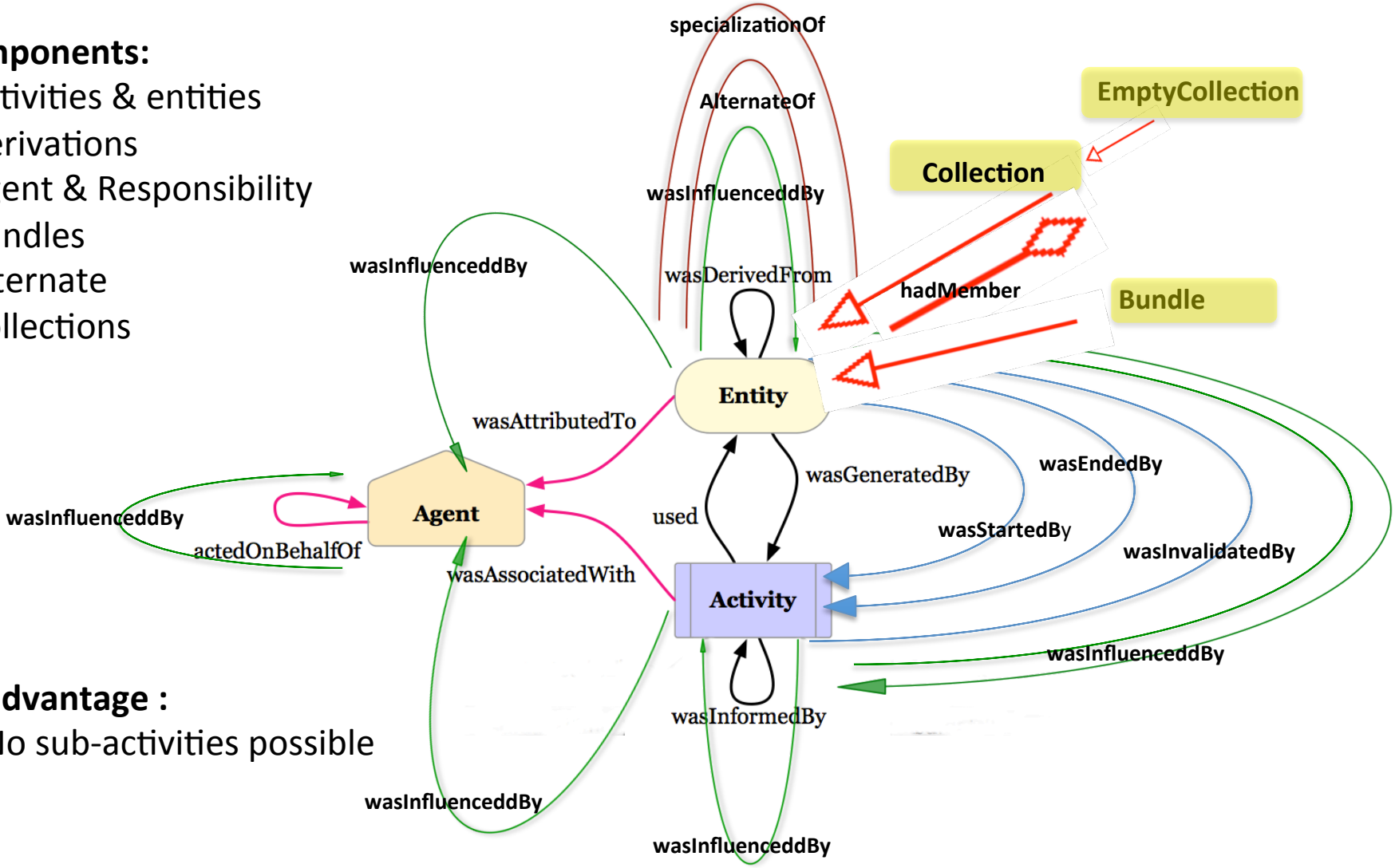
PROV-AQ : Provenance access and query



# Provenance in the W3C

**6 components:**

- Activities & entities
- Derivations
- Agent & Responsibility
- Bundles
- Alternate
- Collections

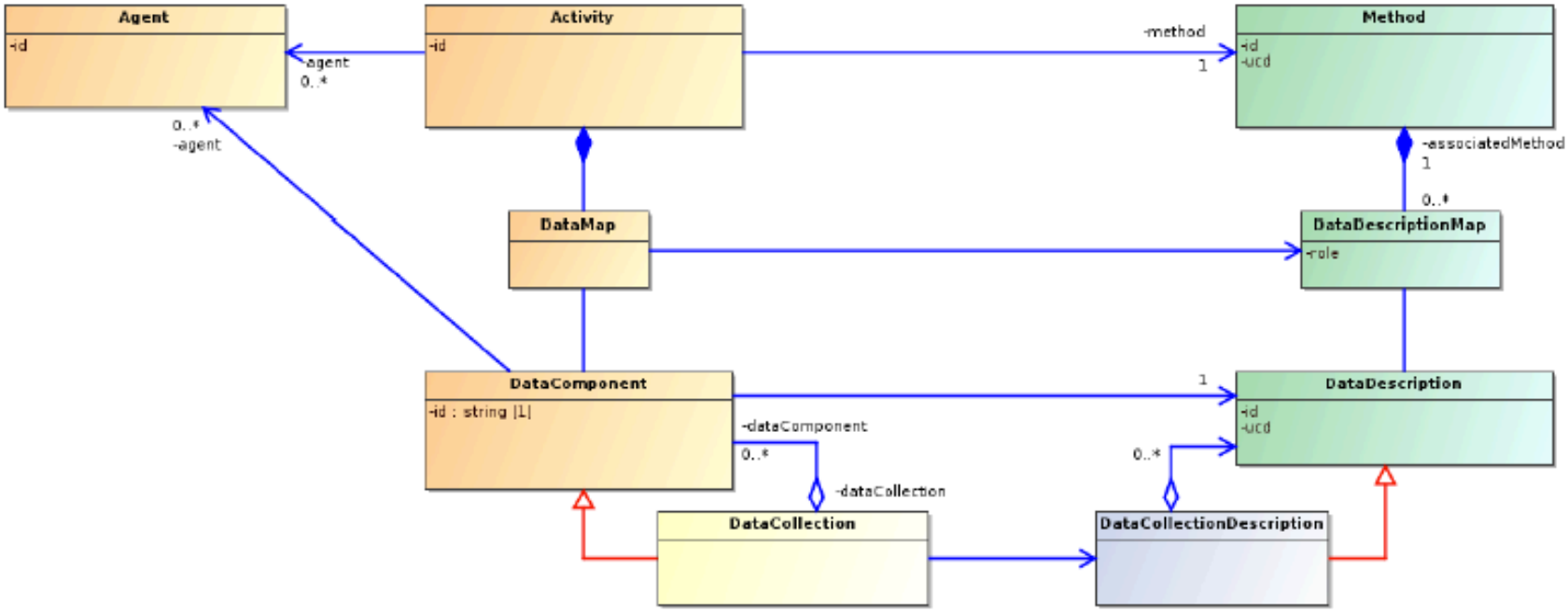



**Disadvantage :**

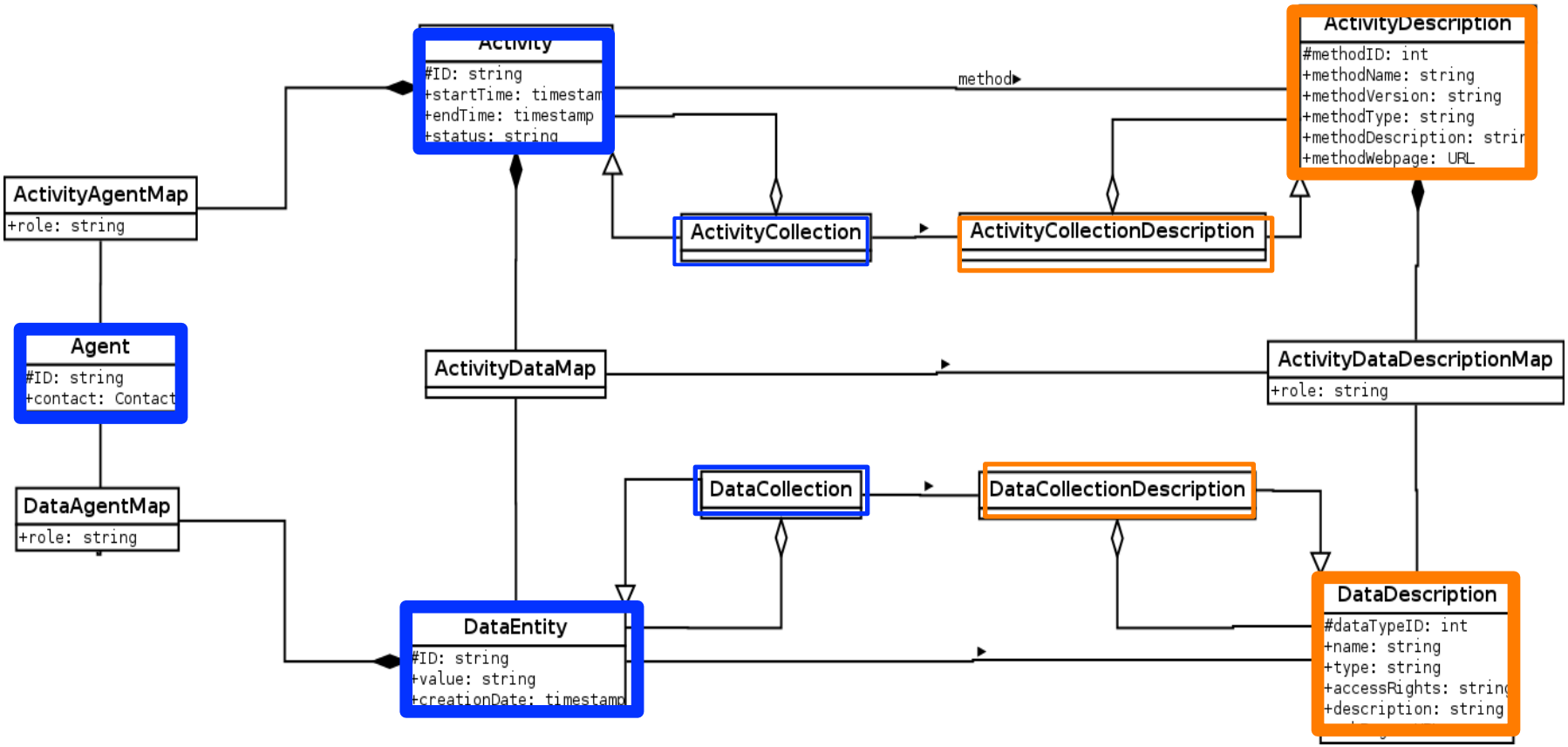
- No sub-activities possible

# Provenance in the VO : current DM

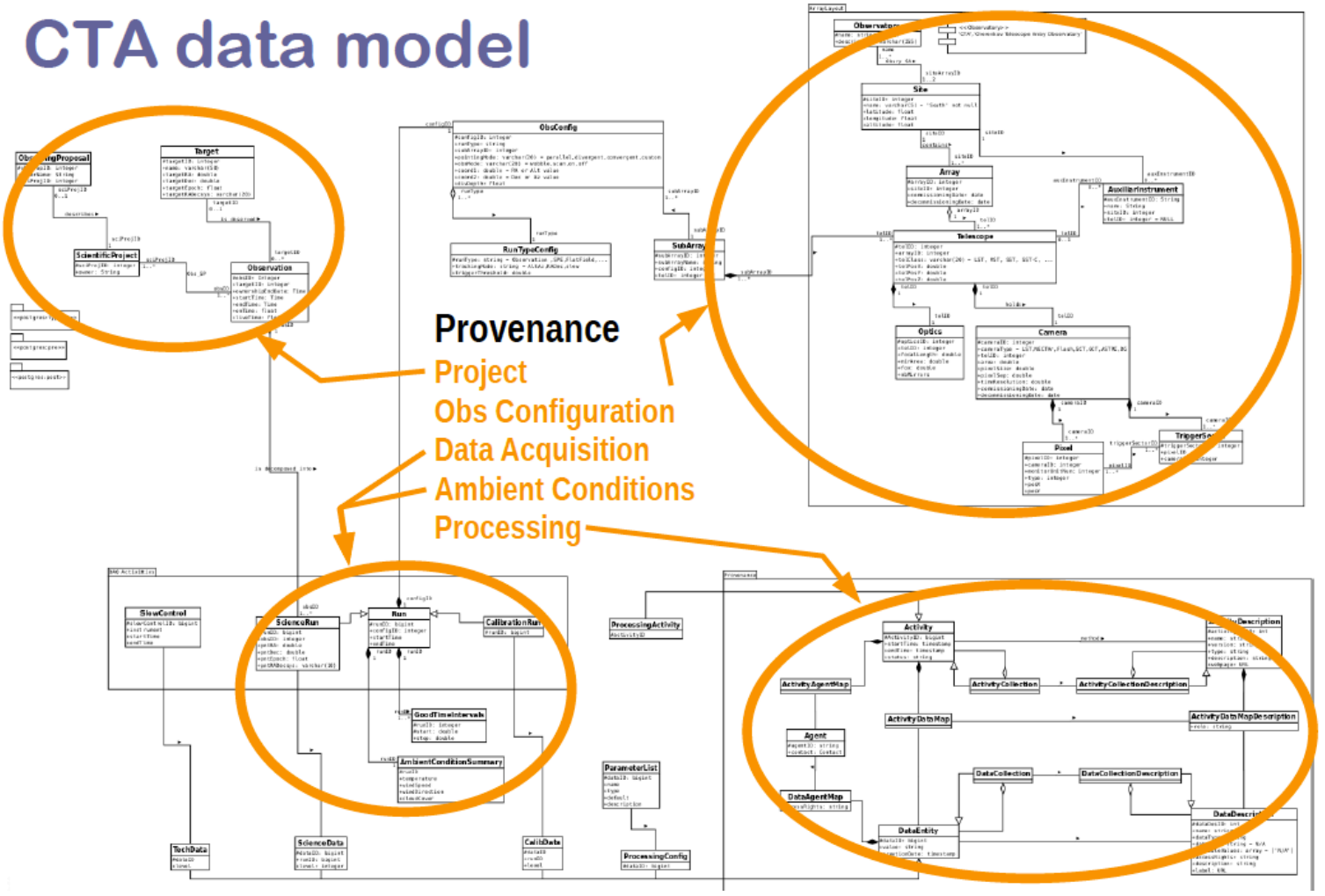
**IVOA Provenance Data Model**
**Version 0.1**
**IVOA Working Draft 2015-05-18**



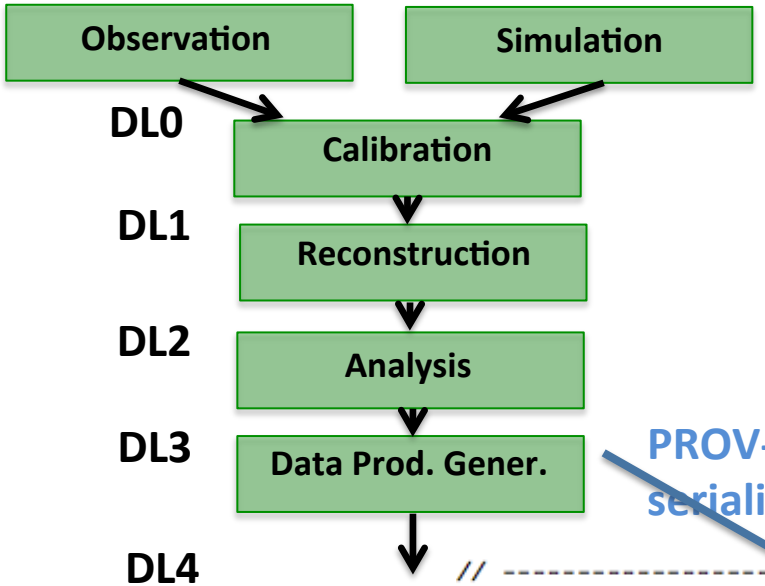
# Provenance in the VO : a DM proposal



# Provenance in CTA

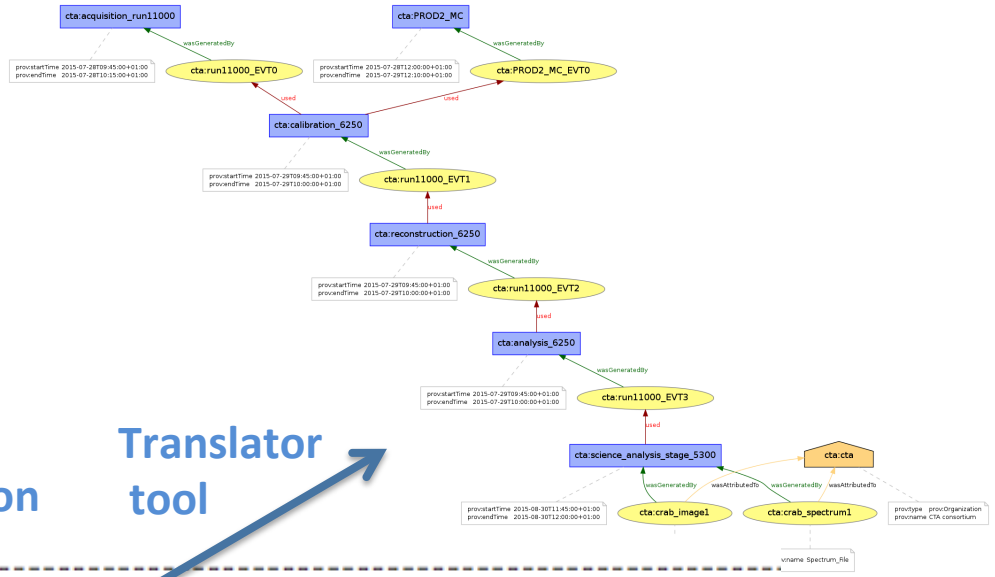


# CTA Provenance description with W3C




PROV-N serialization

Translator tool




```
// ------------------------------------------------------------------------------
// 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,-)
```

PROV-XML → VOTable

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

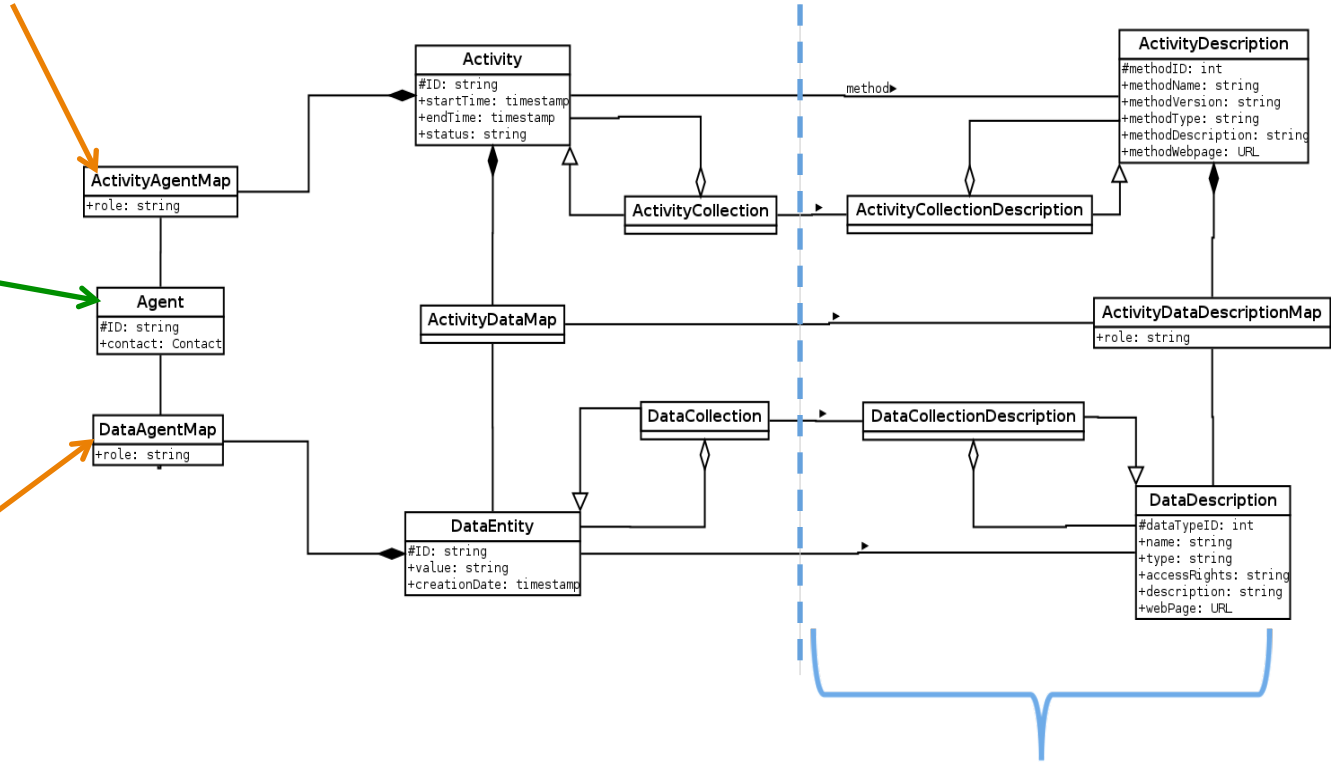
# Provenance in the VO : notation

## wasAttributedTo relation with attributes

wasAttributedTo(cta:run11000_EVT0, cta:cta, [ voprov:role= "Owner"])

## Agent with attributes

```
agent(cta:cta,
  [
    prov:type="prov:Organization",
    prov:name="CTA consortium",
    voprov:contact = ""
  ]
)
```

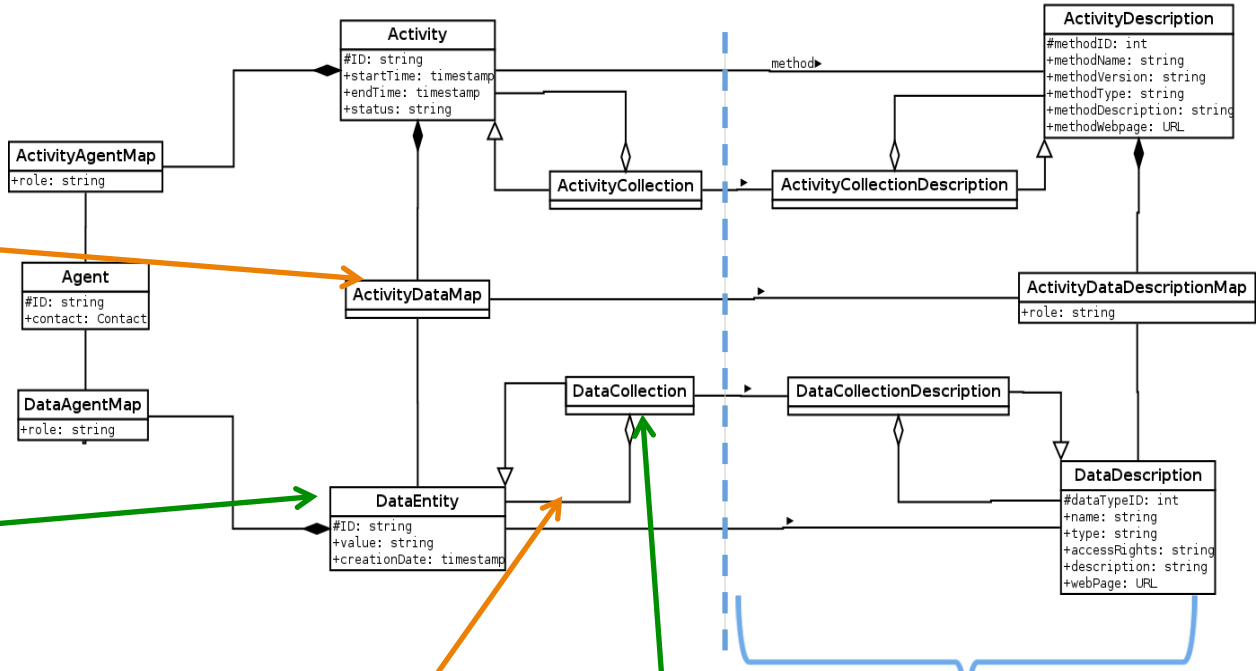



## wasAssociatedWith relation with attributes

Attributes of the left-hand side components

# Provenance in the VO : notation

**used and wasGeneratedBy relations**




**Entity with attributes**

```
entity(cta:run11000_EVT0,
 [
  prov:label = "EVT0 file",
  voprov:name= »EVT0_File »,
  voprov:type = "File",
  voprov:accessRights= "private »,
  cta:runNumber=11000,
  cta:telescope="MST21"
 ]
)
```

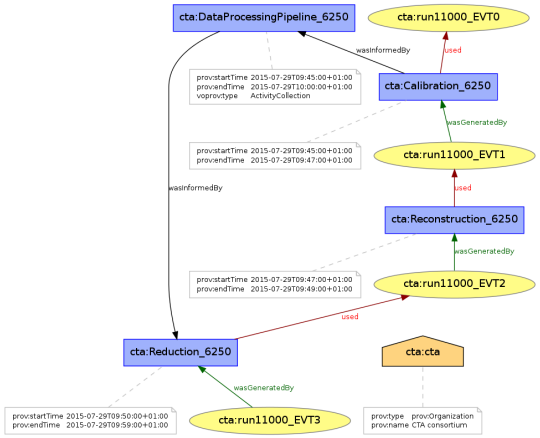
**hadMember relation**

**Entity with attributes**

entity(cta:ObsConfig11, [prov:type = "prov:Collection"])

Attributes of the left-hand side components

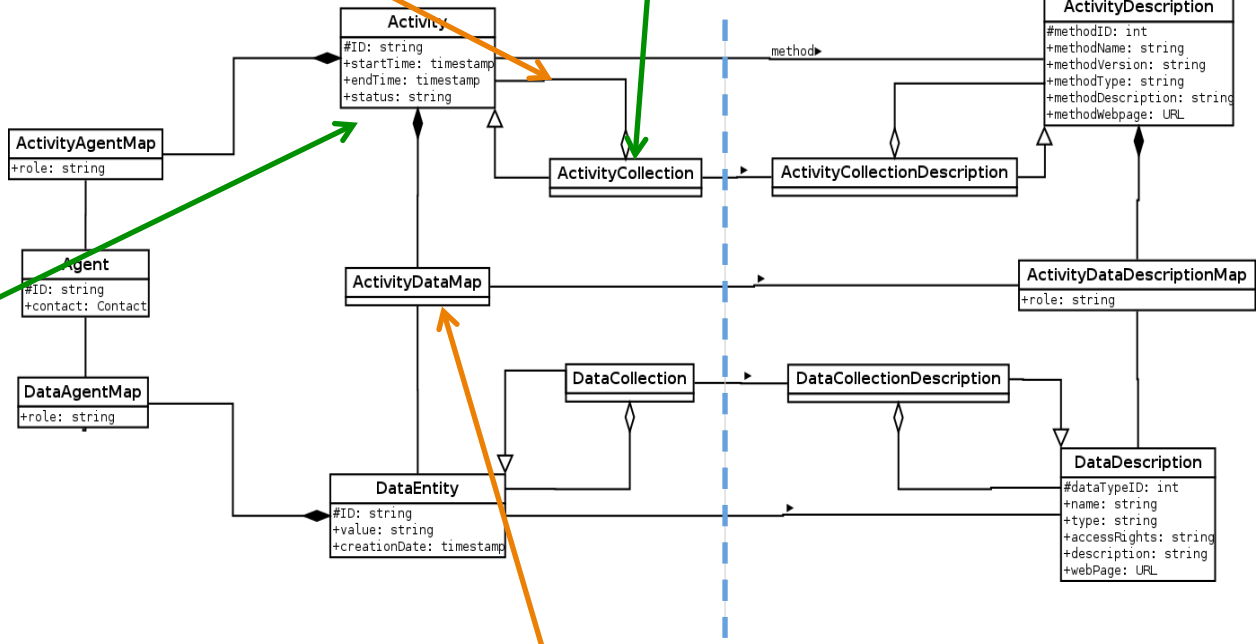
# Provenance in the VO : notation




**wasInformedBy relation ???**

**Activity with attributes**

activity(cta:acquisition_run11000,
2015-07-28T09:45:00, 2015-07-28T10:15:00,
[voprov:type = "ActivityCollection"] )




**Activity with attributes**

activity(cta:acquisition_run11000,
2015-07-28T09:45:00, 2015-07-28T10:15:00,
[
voprov:status: »OK »",
prov:type = "obs:observation",
voprov:method_name= »DAQ",
voprov:method_version="1.0",
voprov:description= »Data Acquisition"
cta:runNumber= »13000"
cta:runType="Observation"
]
)

**used and wasGeneratedBy relations**

Attributes of the left-hand side components

# CTA Provenance : query

**Selection criteria could be :**

- **Name of attributes of the W3C or IVOA Provenance data model**
- **Name of the attributes specific to the CTA context (run number, ambiant conditions, …)**

**Identified**

**WE NEED TO IDENTIFY THE SPECIFIC PROVENANCE ITEMS FOR EACH DATA LEVEL.**

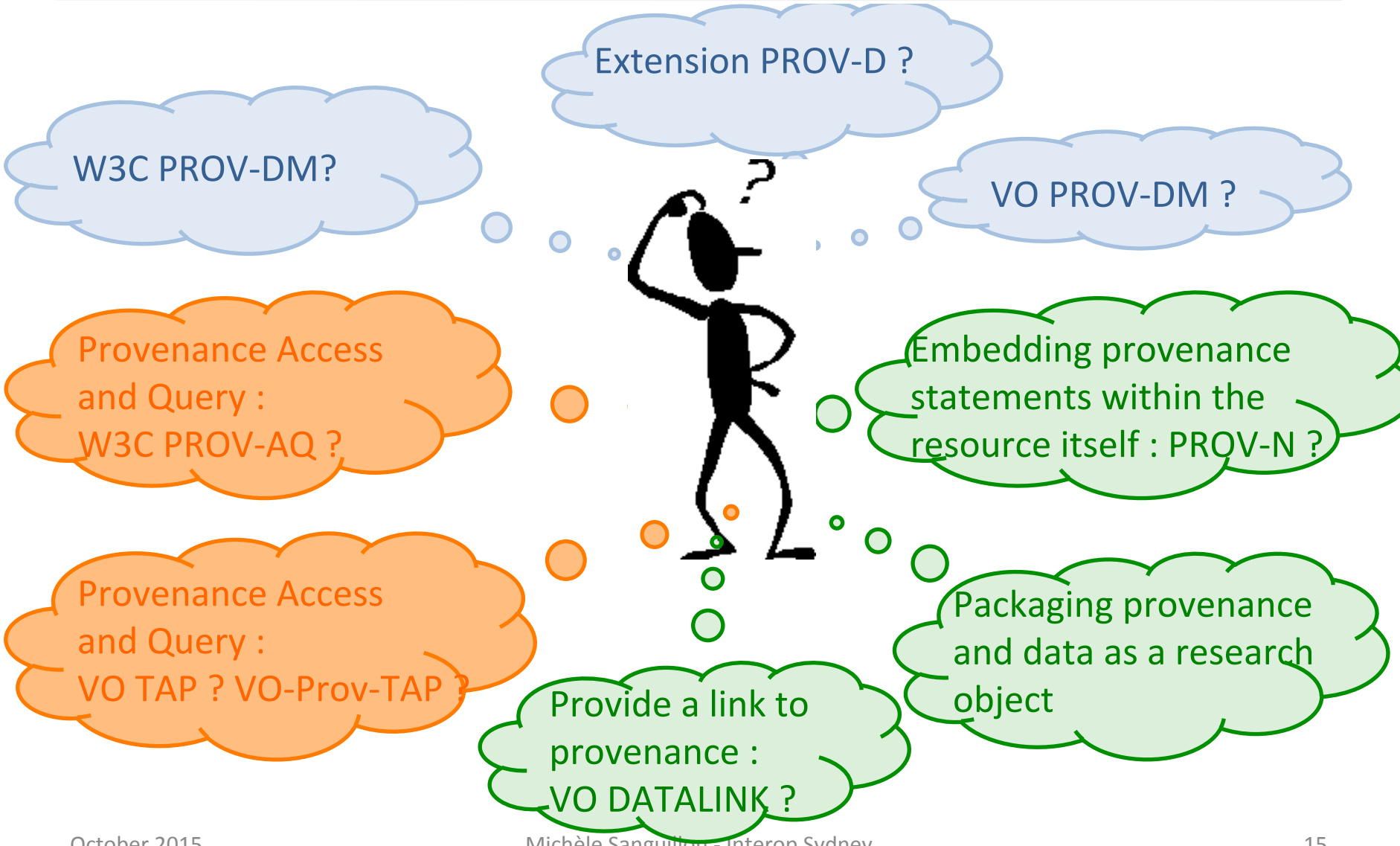
**THE VO USER NEED TO QUERY WHICH SPECIFIC ATTRIBUTES COULD BE A CRITERIA**

List DL4 data generated by the Science Analysis Pipeline version 2.0

```
HTTP POST http://portal.cta-observatory.org/tap/async
REQUEST=doQuery
LANG=ADQL
SELECT * FROM DataEntity as e
  where …
```

**VO-TAP protocol**

# Provenance : a lot of questions to solve for implementation



# THANK YOU
# FOR YOUR ATTENTION