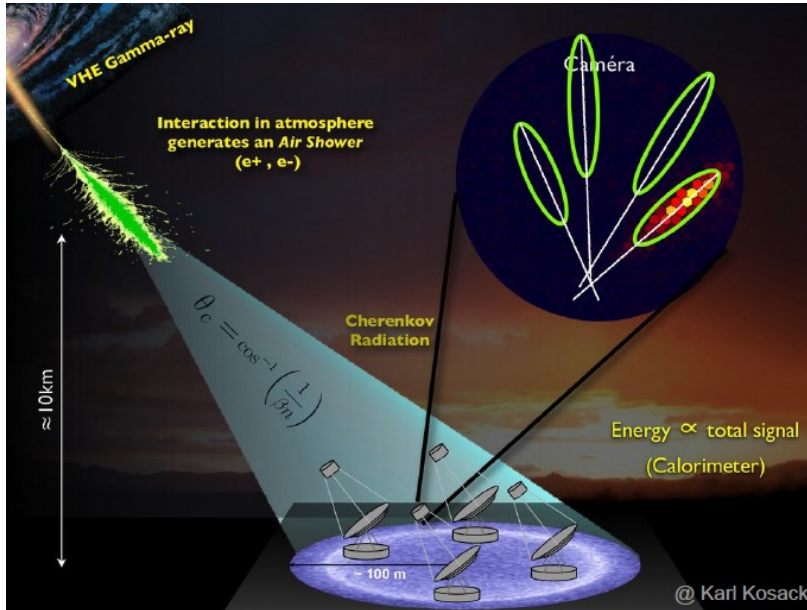# IVOA PROVENANCE METADATA
## LINKS TO EXISTING MODELS
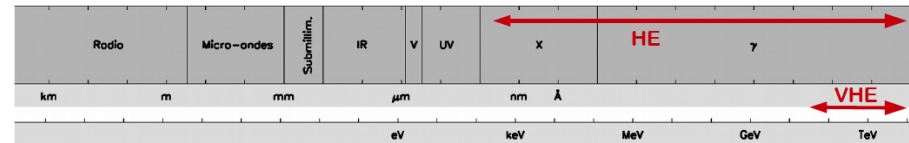
**Michèle Sanguillon**[(*)(1)(2)], **Mathieu Servillat**[(**)(1)], **Julien Le Faucheur,** [(**)(1)], **Catherine Boisson**[(**)(1)], **Johan Bregeon**[(*)(1)], **Ana Palacios**[(*)(2)], **Mireille Louys**[(***)(3)], **François Bonnarel**[(***)(3),] **Pierre Le Sidaner**[(****)(3)]

(*) LUPM, Montpellier (**) LUTH, Meudon (***) CDS, Strasbourg, (****) DIO, Paris
(1) CTA, (2) Pollux, (3) VO Experts

# CTA 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
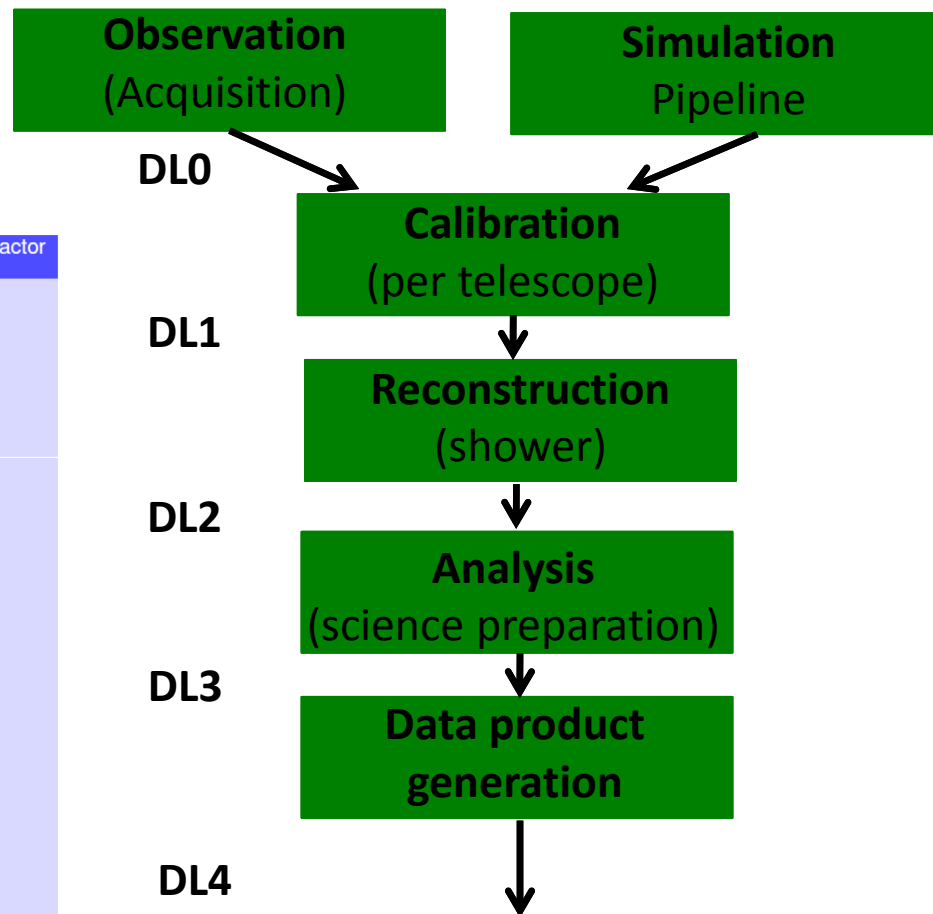- Final products data available on the VO



- Several orders of magnitude
- Photon counting
- Low count statistics, high background
- Event lists (coordinates, time, energy)

Energy spectra

Images

Lightcurves

@ Mathieu Servillat

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

Mireille Louys -  Provenance- Asterics meeting  March 7-8 , 2016

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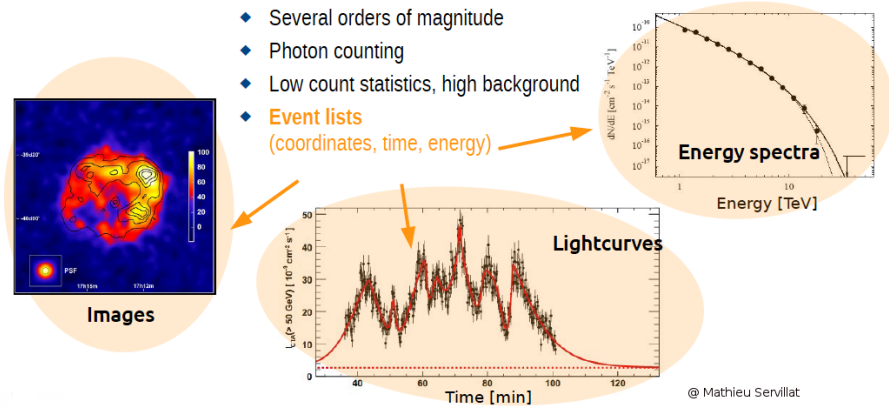
Different levels of data : DL0 to DL5.
DL3, DL4 and DL5 data available on the VO

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

**Observation** (Acquisition)

**Simulation** Pipeline

DL0

**Calibration** (per telescope)

DL1

**Reconstruction** (shower)

DL2

**Analysis** (science preparation)

DL3

**Data product generation**

DL4

# CTA search cases

- **Use case 1:**
  Search data available for a given Target at a given time interval.

  → ObsCore
  Any protocol

- **Use case 2:**
  Search public data for all blazars

  → Extended ObsCore
  TAP

- **Use case 3:**
  Search data that include LST (Large Size Telescope).

  → CTA ObsConfig
  TAP

- **Use case 4:**
  Search data produced using a given reconstruction method

  → No data model
  No current protocol

- **Use case 5:**
  Search data for a given target produced with loose cuts

  → No data model
  No current protocol

# Provenance in the IVOA

- **Explains how data sets were produced:**
  - Observing process and conditions
  - Data reduction, selection and extraction methods applied to raw measures to build up science-ready data products (source lists, spectra, light curves, images, …)
  - Worflows to build theorical data (spectra, images, …)
- **Helps VO users to:**
  - Derive selection criteria to filter out suitable data for his/her scientific needs
  - Estimate better which data release fits the best for their needs
  - Run his/her own reduction method on intermediate data products in order to refine data analysis

# Provenance in the W3C

**W3C Provenance definition:**

« Provenance is information about entities, activities and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability and trustworthiness. PROV-DM is the conceptual data model that forms a basis for the W3C provenance (PROV) familiy of specfications. »

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

PROV-DM: the PROV data model

PROV-O: the PROV ontology

PROV-Constraint: Constraints of the PROV Data Model

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

**and a number of non-prescriptive notes**
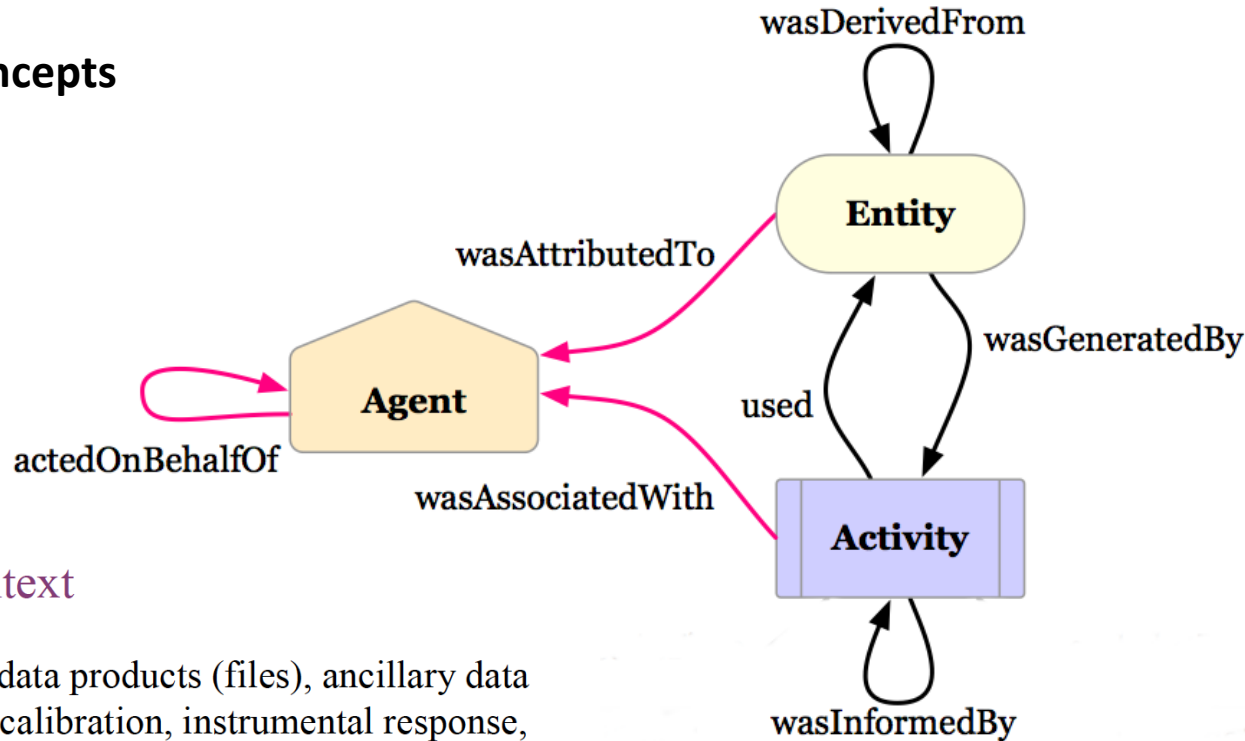
PROV-XML: an XML schema for the PROV data model

PROV-AQ: Provenance access and query

**Benefits:**
- Four recommendations and a number of non-prescriptive notes
- Tools to validate and translate a description format in another one
- Possible to define our own attributes

# W3C Provenance Data Model

**Core Concepts**



In our context

| | |
|---|---|
| Entity | ■ data products (files), ancillary data (calibration, instrumental response, etc.), processing parameter files |
| Activity | ■ data acquisition, mosaicing, regridding, fusion, calibration, …, transformation |
| Agent | ■ Telescope astronomer, pipeline operator, principal investigator, etc. |

# IVOA current DM

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

# IVOA proposal DM



Ivoa:ObsDataset
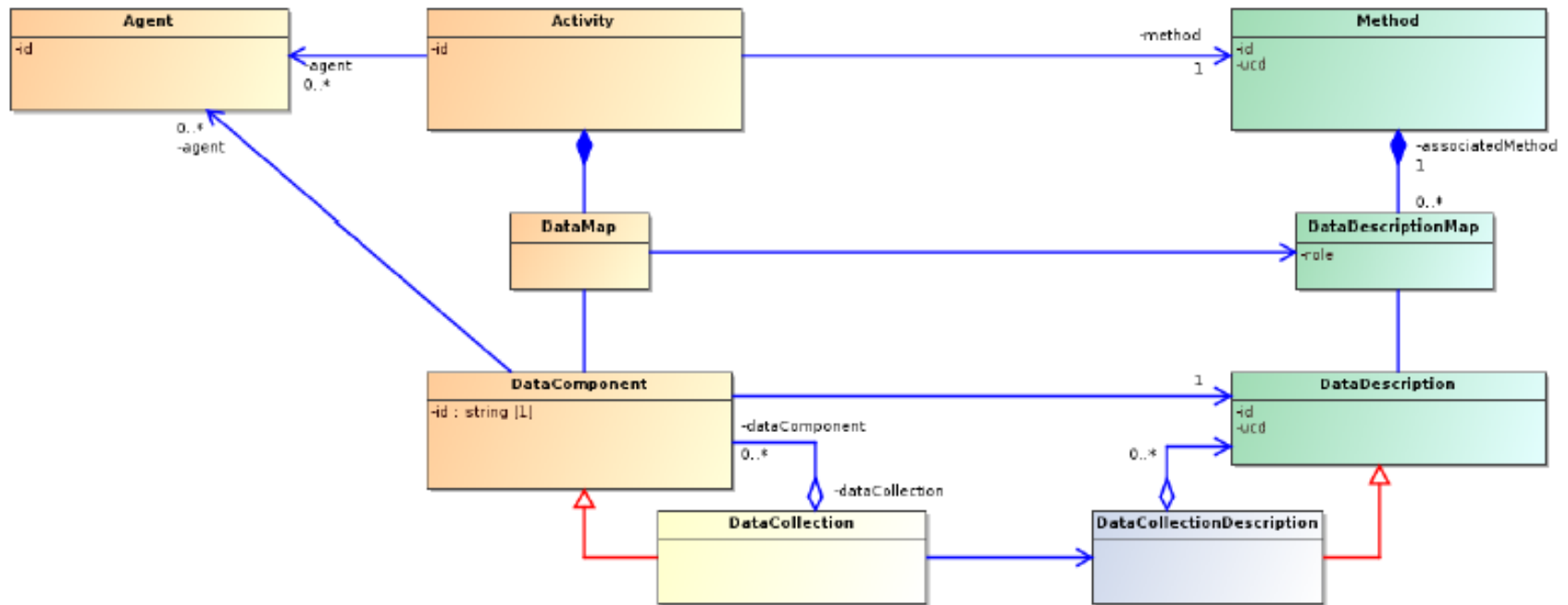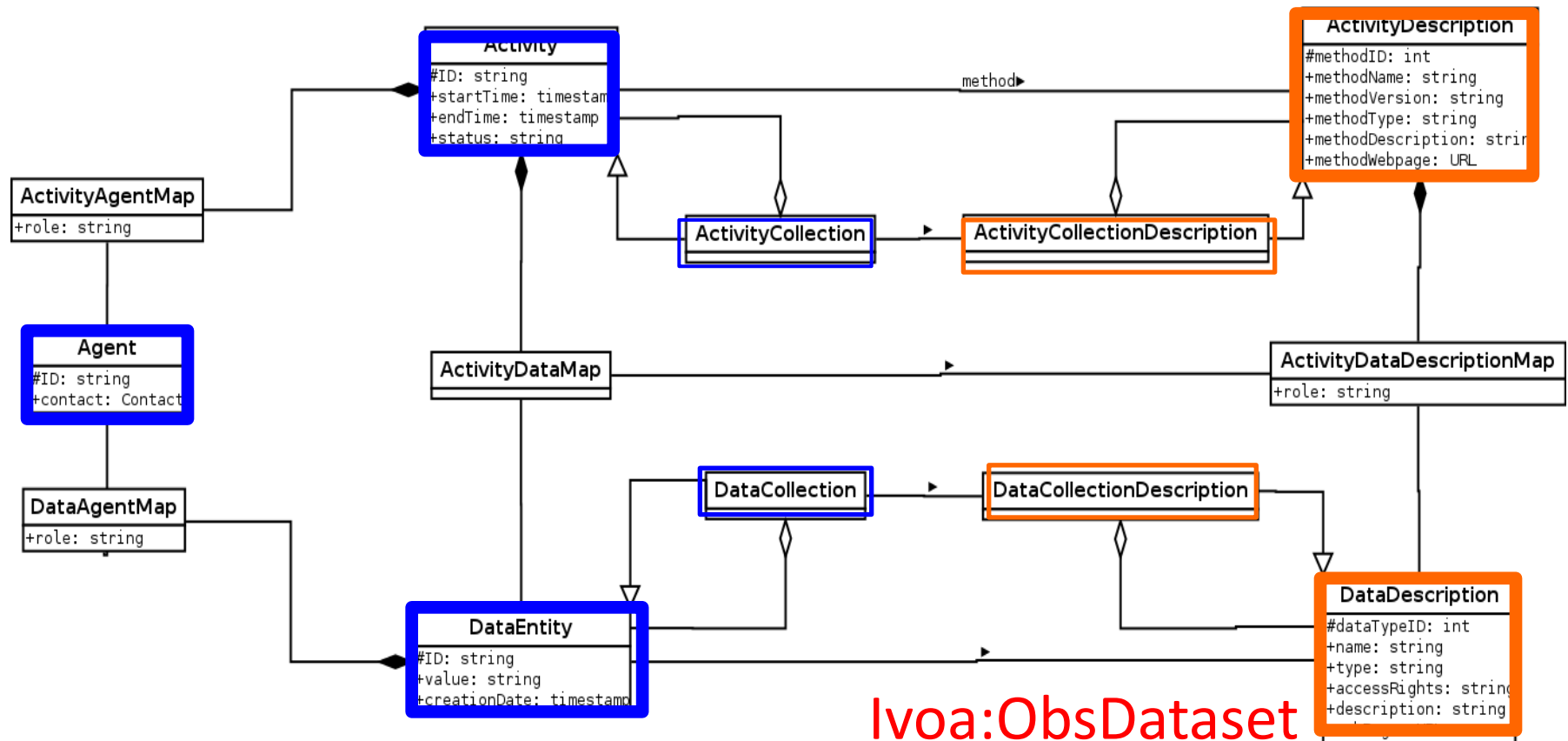
Kristin Riebe, Leibnitz Institut für Astrophysics Postdam

http://wiki.ivoa.net/internal/IVOA/InterOpJune2015DM/Provenance.pdf

- All W3C PROV formats tried out

- To continue in the IVOA DM WG

# Pollux stellar database

temperature = 5000°K
gravity = …
…

http://pollux.graal.univ-montp2.fr/
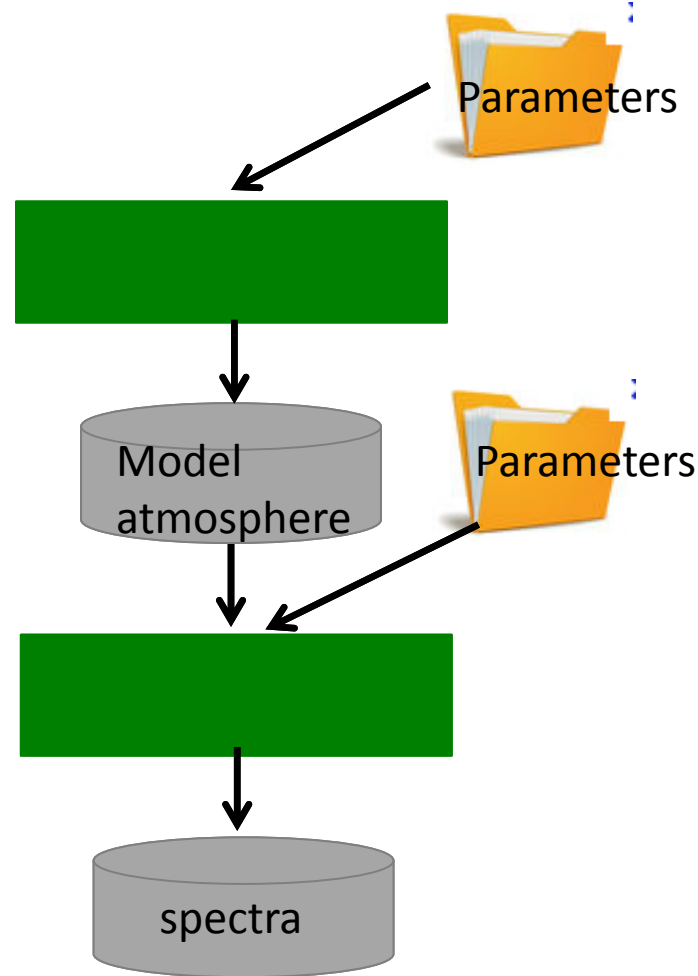
POLLUX database collects and presents synthetic spectra computed* at high resolution.



A_p10200g3.9z−1.0t2.0_a0.00c0.00n0.00o0.00r0.00s0.00_VIS.spec
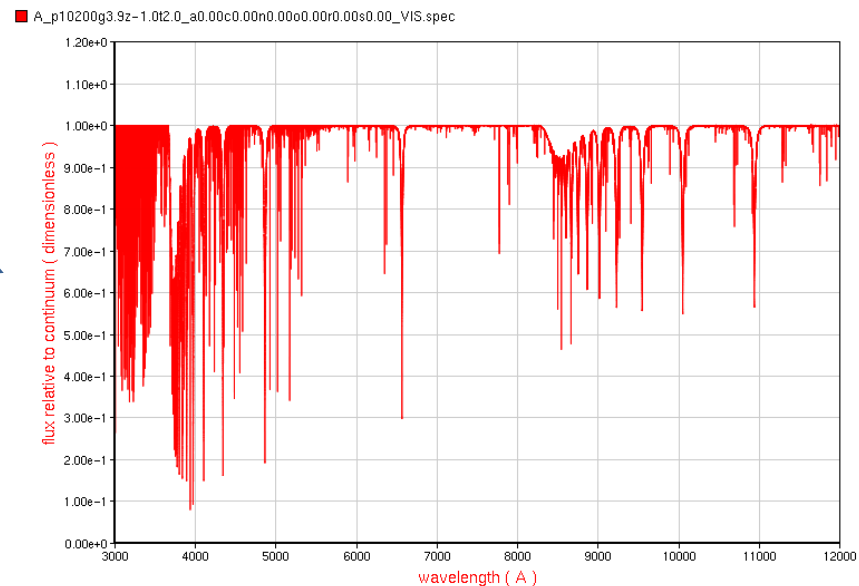
* The computing of the spectra is done by the producers and not done on the fly.

# Pollux data and workflow



Parameters

Model atmosphere

Parameters

Only the spectra are available on the VO

spectra

Mireille Louys -  Provenance- Asterics meeting  March 7-8 , 2016

# Pollux use cases

- **Use case 1:**
  Show me a list of synthetic spectra satisfying :
  - domain of wavelength = visible
  - domain of effective temperature = [4000, 5000]

- **Use case 2:**
  Show me a list of synthetic spectra satisfying :
  - code for model atmosphere = MARCS
  - type of model atmosphere = spherical

- **Use case 3:**
  Show me a list of synthetic spectra satisfying :
  - code for spectral synthesis = turbospectrm
  - version of this code = 2008.1

Data Model:
- Obs* could not be applied
- SimDM implements only a simulator and a PostProcessor

Protocol
- SSA protocol with format = METADATA but only few criteria are currently available.
- SimDAL not done for this use

# Conclusion for IVOA Provenance

**Rich and diverse projects explored for Usecases**

- CTA, RAVE, Pollux DB, SVOM(http://www.svom.fr/svom.html)

- A Core Provenance Model on the way

- Existing modeling to reuse for data products
  - ObsCore/ Dataset Metadata DM

- Protocols available  TAP?

- Translation tools in W3C tools

*Discussion to happen here?*