CTA Data Provenance
*The Cherenkov Telescope Array*

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

RDA 6th meeting, Paris, 24 Sept. 2015
- Two arrays of **100 (South)** and **20 (North)** Cherenkov telescopes (4, 12 et 24 m in diamètre)
- July 2015: **Site Selection**, Chile (ESO) and La Palma
- 2016: **Construction phase**
- Current experiments: H.E.S.S., MAGIC, VERITAS
  H.E.S.S.: experiment with 4+1 telescopes (4 x 12 m + 1 x 28 m)
- Event Reconstruction: photon, particle shower, Cherenkov light (faint, few nanoseconds)
- Atmosphere = calorimetre
  Simulations, assumptions
- Complex Metadata, need to be structured
Very high energy data

<table>
<thead>
<tr>
<th>Radio</th>
<th>Micro-ondes</th>
<th>Submill.</th>
<th>IR</th>
<th>V</th>
<th>UV</th>
<th>X</th>
<th>HE</th>
<th>γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>km</td>
<td>m</td>
<td>mm</td>
<td>μm</td>
<td>nm</td>
<td>Å</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

**Images**

**Lightcurves**

**Energy spectra**

<table>
<thead>
<tr>
<th>10^{-30}</th>
<th>10^{-29}</th>
<th>10^{-28}</th>
<th>10^{-27}</th>
<th>10^{-26}</th>
<th>10^{-25}</th>
<th>10^{-24}</th>
<th>10^{-23}</th>
</tr>
</thead>
</table>
| Energy [TeV] | dN/dE [cm^{-2} TeV^{-1}] |}

RDA 6th – 24 Sept. 2015
CTA requirements for data diffusion

- **Diffusion** of high level data products via Virtual Observatory protocols (event lists, images, spectra, lightcurves)

- **Use cases:**
  - Provide data products that include some provenance information useful to the end user
  - Filter data using provenance selection criteria
  - Check the production of data (ensure quality)

- **Provenance categories:**
  - Data acquisition, observing configuration
  - Data processing, reduction
## CTA data levels and workflow

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

Workflow description

Data Level description
CTA data model

Provenance
- Project
- Obs Configuration
- Data Acquisition
- Ambient Conditions
- Processing
Some **Custom** Activities and Entities (Run, ObsConfig, ...) 

- Link to **Provenance data model** for Data Processing
Examples:
Number of telescopes involved
Field of view
Pointing direction

---
Different aspects

- **Filling** the Provenance information
  - Structured database based on *data model*
  - Interface (API) for ingestion

- **Restitution**
  - Metadata attached to a *data file*
  - PROV-N, the W3C Provenance Notation
    → use of existing tools!

- **Direct Access**
  - IVOA *protocols* (Table Access Protocol, DataLink)
  - But Provenance data are project dependant
    → Provenance *Profiles* customized for CTA
entity(rave:0645m522l0049.fits, [prov:type = 'std:fits'])
entity(rave:0645m522l0049.wav.fits, [prov:type = 'std:fits'])

agent(aao:Paul_Cass, [prov:type='prov:Person'])
agent(rave:Alessandro_Siviero, [prov:type='prov:Person'])

activity(rave:act_observation, 2008-02-16T13:25:24, -,
[ prov:type = 'obs:Observation' ])
activity(rave:act_irafReduction, 2008-03-04T09:46:57, -,
[ prov:type = 'std:reduction' ])

wasAssociatedWith(rave:act_observation, aao:Paul_Cass, -,
[ prov:role = 'obs:Observer' ])
wasAssociatedWith(rave:act_irafReduction, rave:Alessandro_Siviero, -)

wasGeneratedBy(rave:0645m522l0049.fits, rave:act_observation, -)
used(rave:act_irafReduction, rave:0645m522l0049.fits, -)
wasGeneratedBy(rave:0645m522l0049.wav.fts, rave:act_irafReduction, -)
wasDerivedFrom(rave:0645m522l0049.wav.fts, rave:0645m522l0049.fits)