Astronomy ESFRI & Research Infrastructure Cluster
ASTERICS - 653477

CTA Data Provenance

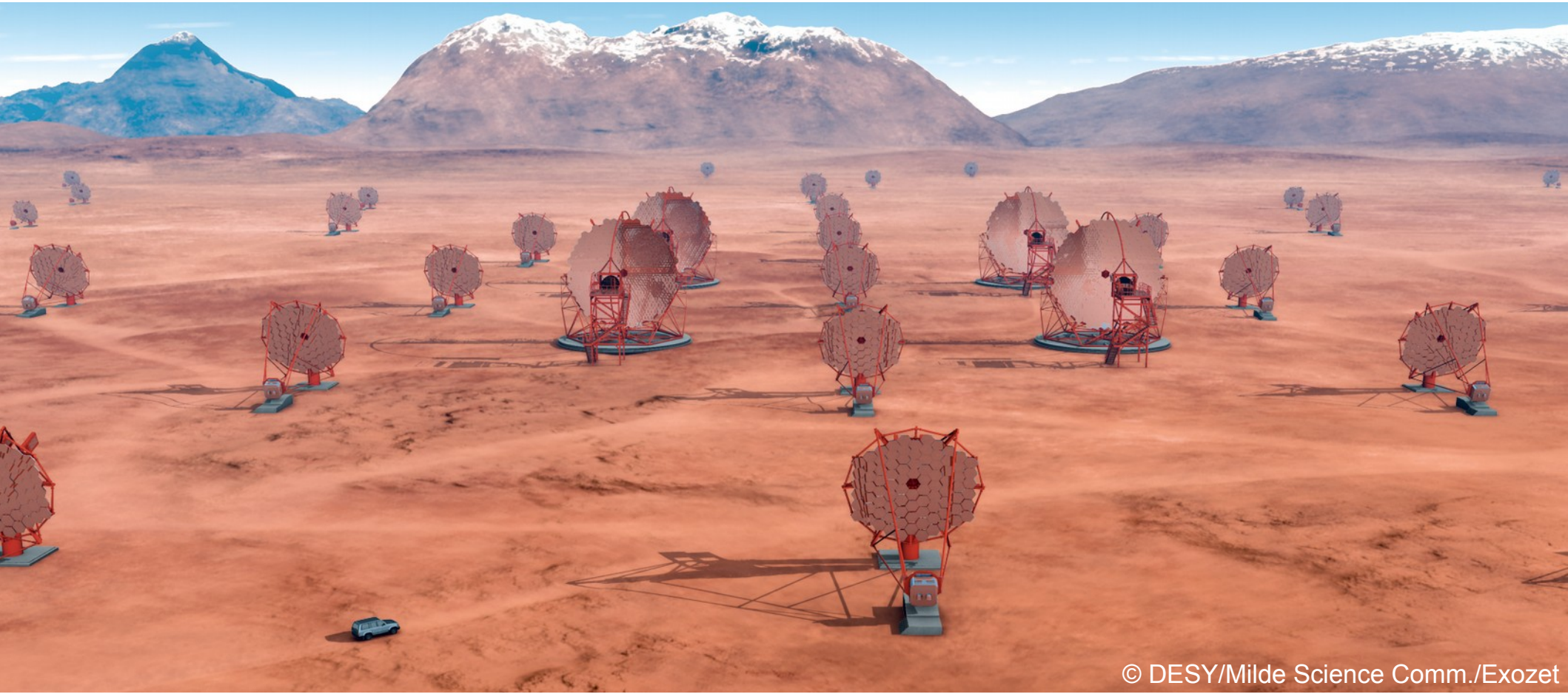
The Cherenkov Telescope Array

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

Observatory

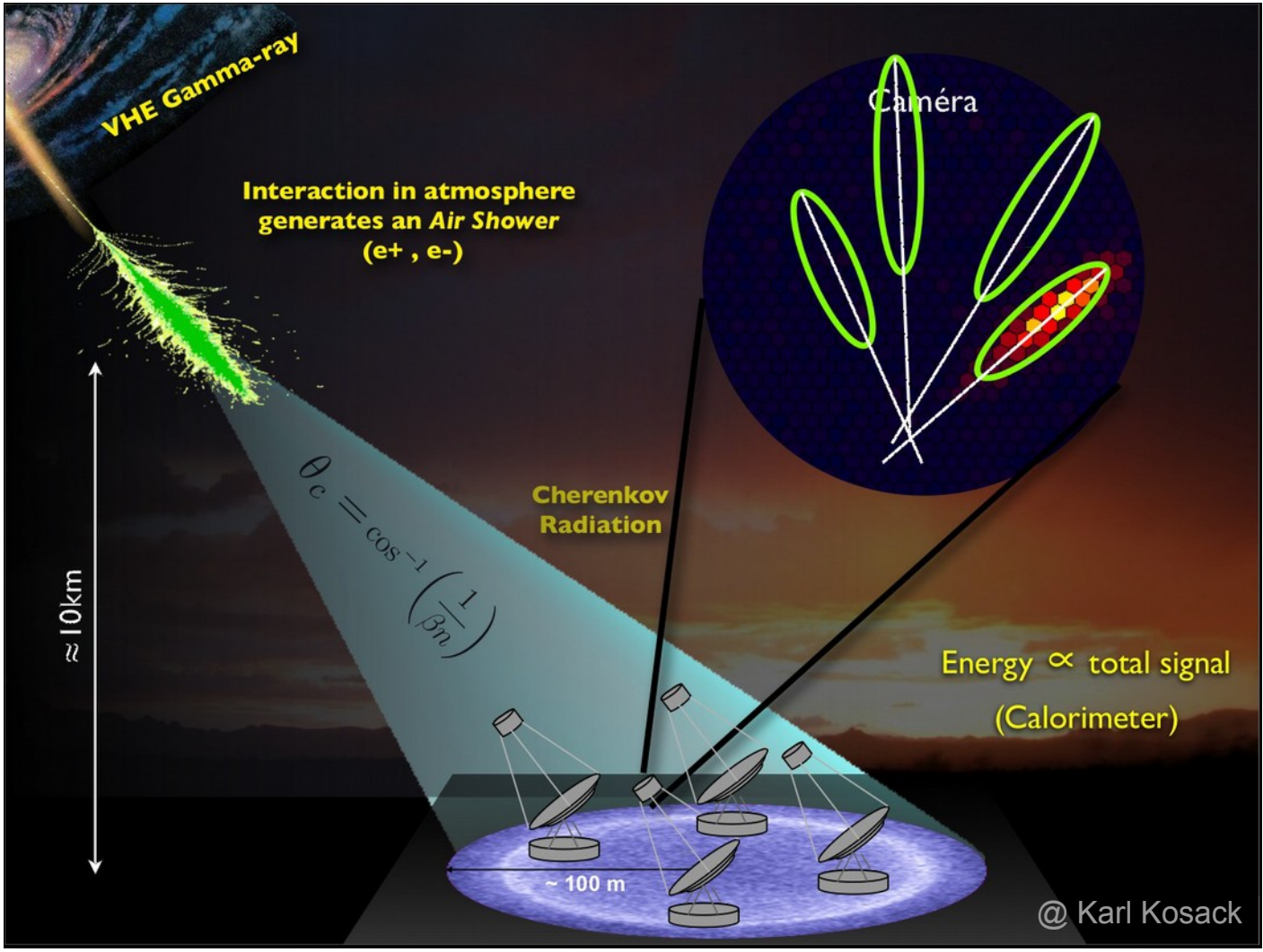
- Two arrays of **100 (South)** and **20 (North)** Cherenkov telescopes (4, 12 et 24 m in diametre)
- July 2015: Site Selection, Chile (ESO) and La Palma
- 2016: Construction phase
- Current experiments: H.E.S.S., MAGIC, VERITAS
  H.E.S.S.: experiment with 4+1 telescopes (4 x 12 m + 1 x 28 m)

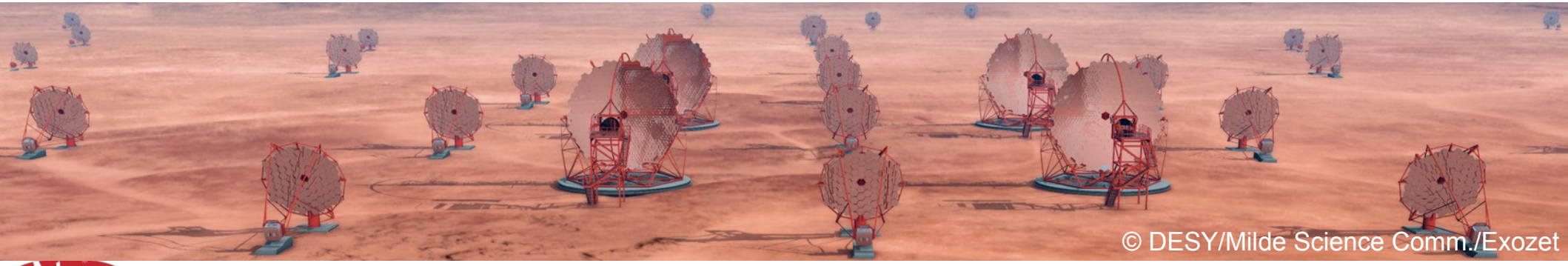


cherenkov telescope array

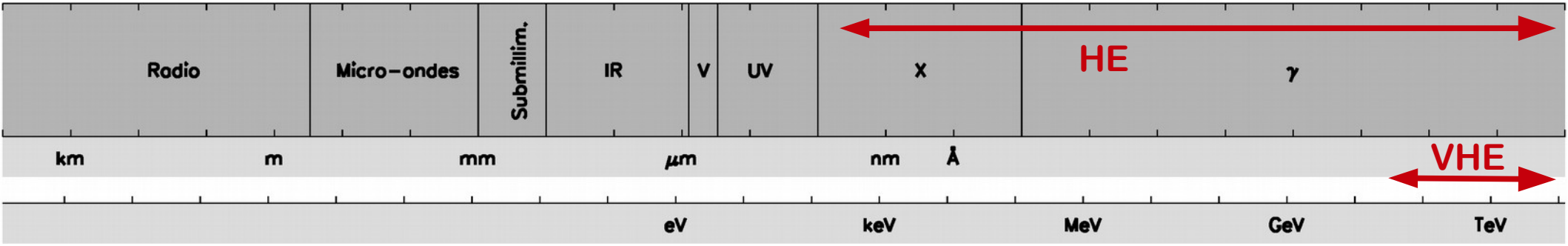
Observatory

- Event Reconstruction: photon, particle shower, Cherenkov light (faint, few nanoseconds)
- Atmosphere = calorimetre Simulations, assumptions
- Complex Metada, need to be structured

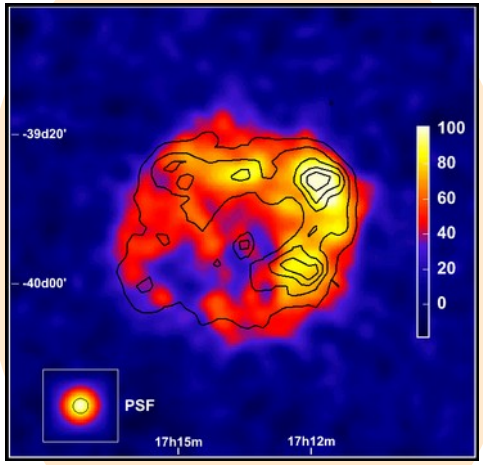




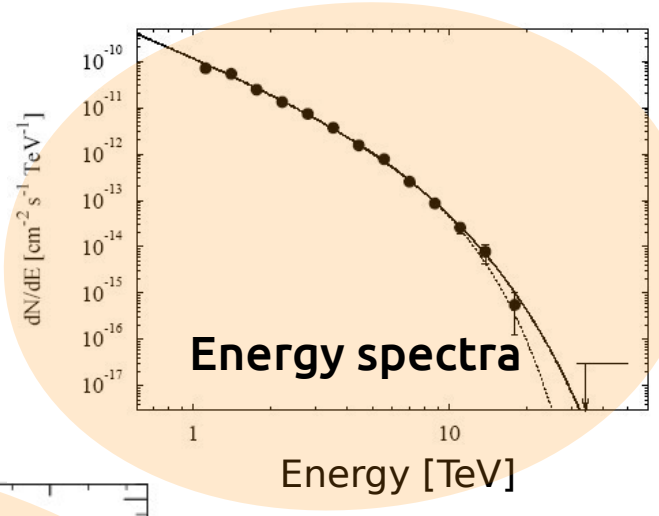
Very high energy data

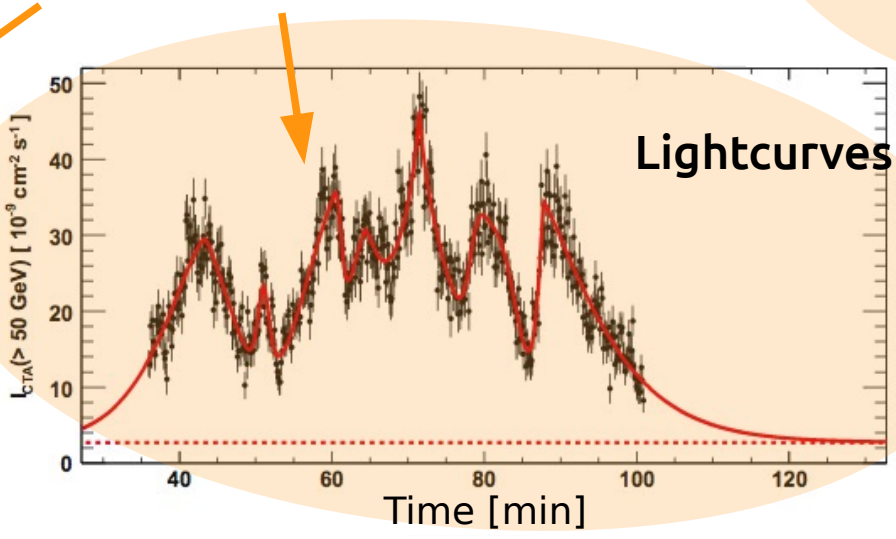


- Several orders of magnitude
- Photon counting
- Low count statistics, high background
- **Event lists**
  (coordinates, time, energy)



Images



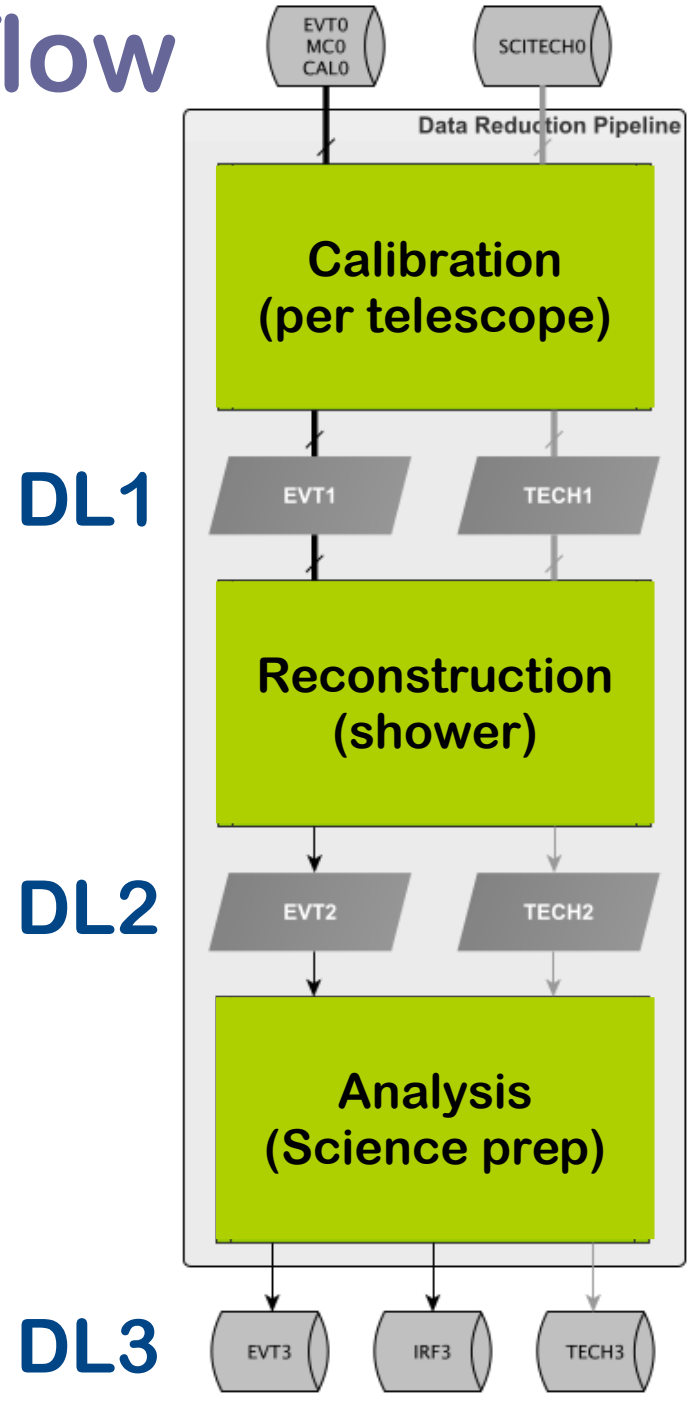


CTA requirements for data diffusion

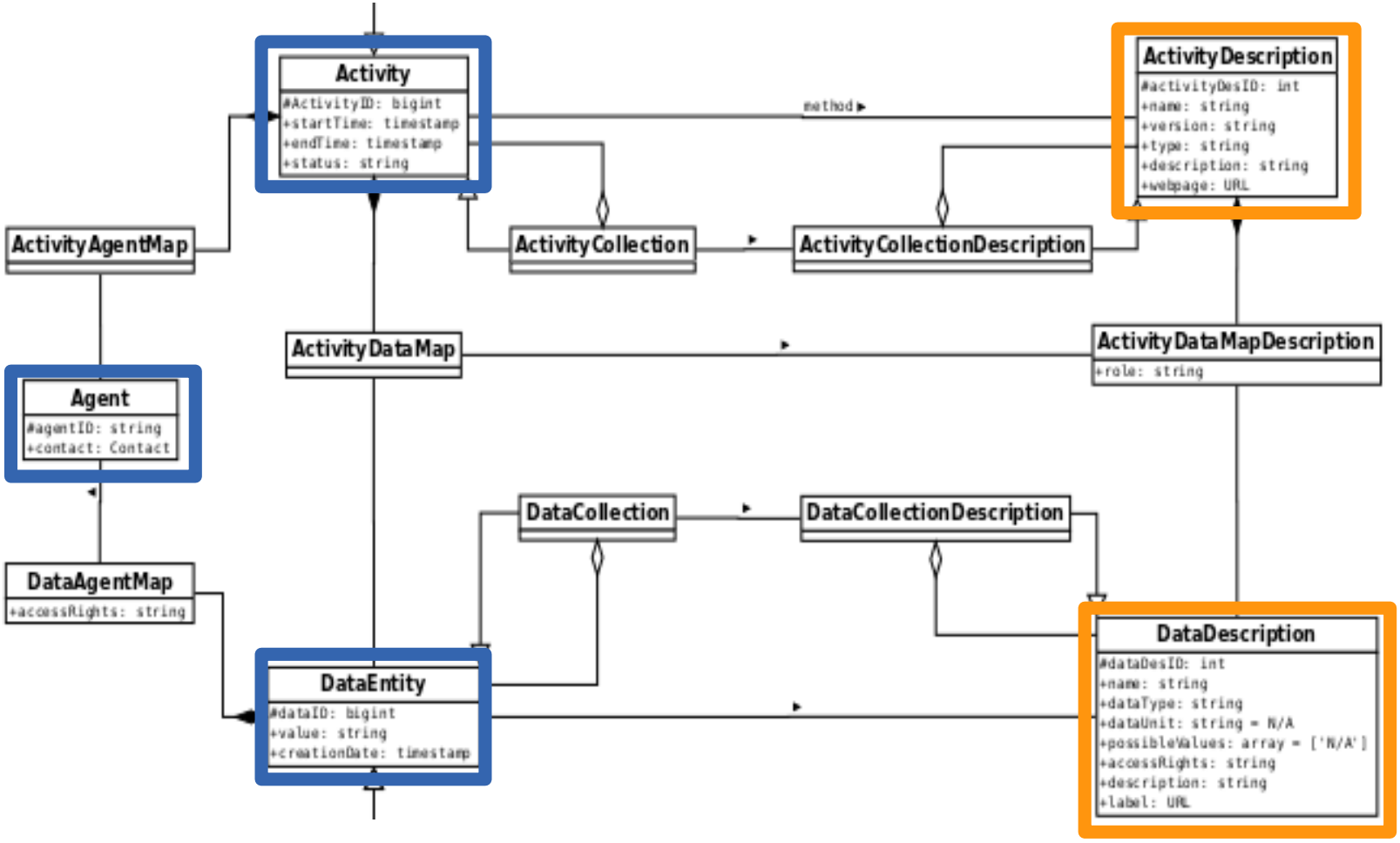
- Diffusion of high level data products via Virtual Observatory protocols
  (event listsimages, spectra, lightcurves)
- Use cases:
  - Provide data products that include some provenance information useful to the end user
  - Filter data using provenance selection criteria
  - Check the production of data (ensure quality)
- Provenance categories:
  - Data acquisition, observing configuration
  - Data processing, reduction

CTA data levels and workflow

Data Level	Short Name	Description
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.
Level 2 (DL2)	RECONSTRUCTED	Reconstructed shower parameters (per event, no longer per-telescope) such as energy, direction, particle ID, and related signal discrimination parameters.
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.
Level 4 (DL4)	SCIENCE	High Level binned data products like spectra, sky maps, or light curves.
Level 5 (DL5)	OBSERVATORY	Legacy observatory data, such as CTA survey sky maps or the CTA source catalog.



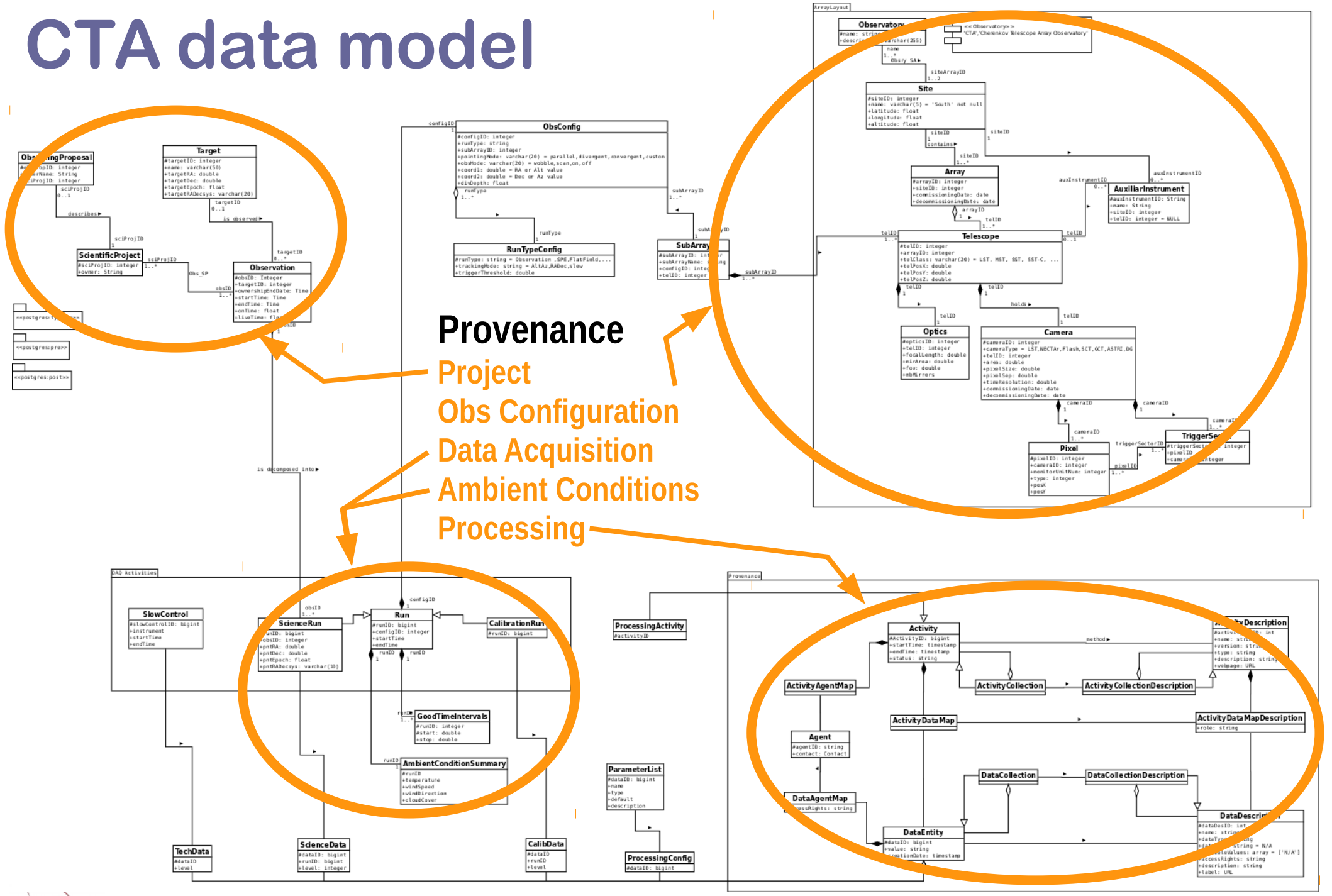
Provenance IVOA data model



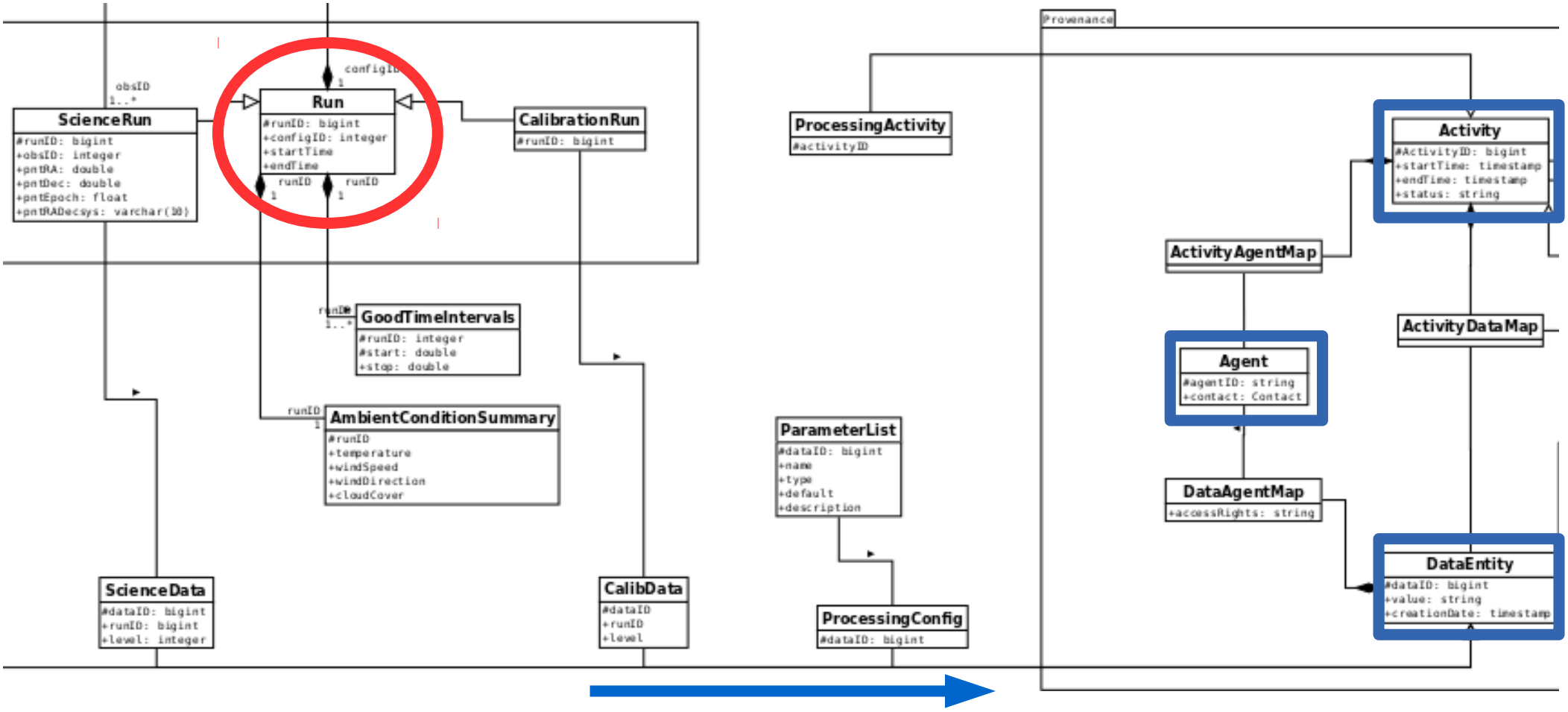
Workflow description

Data Level description

CTA data model

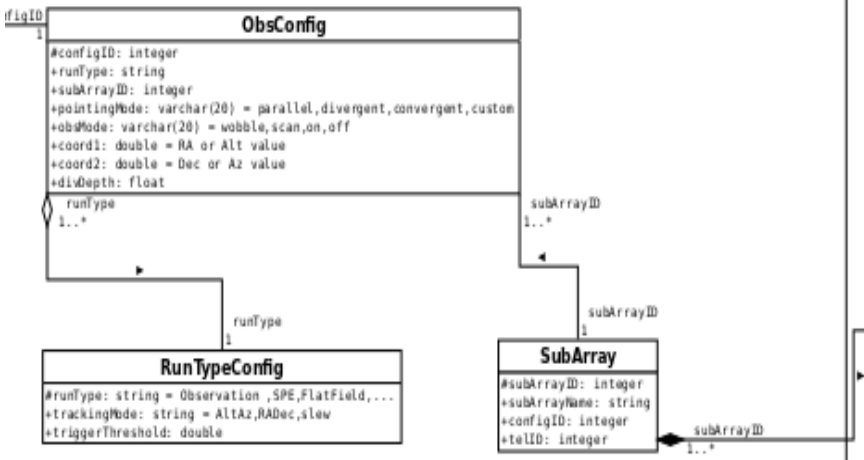


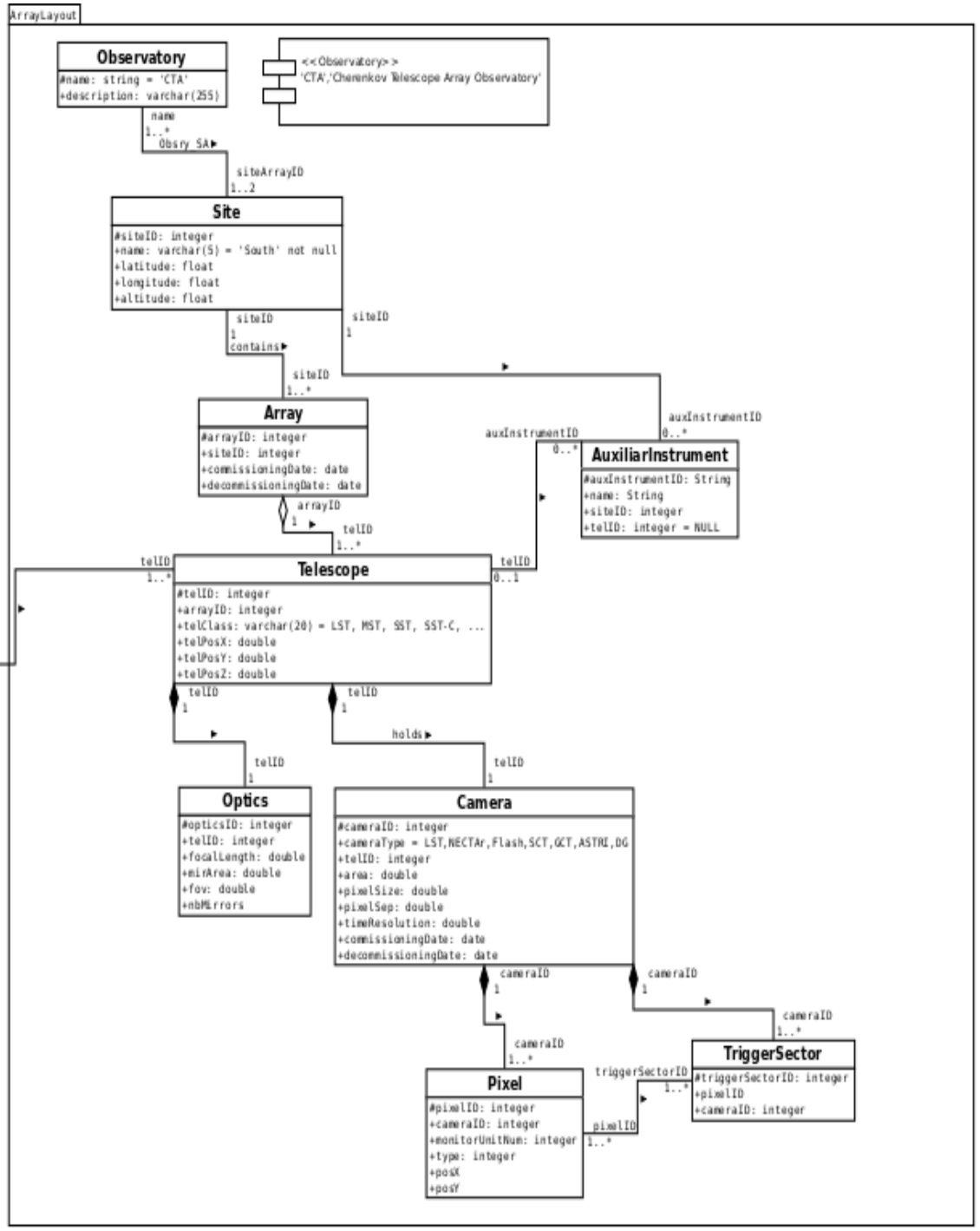
CTA data model – Data Processing



- Some Custom Activities and Entities (Run, ObsConfig, ...)
- Link to Provenance data model for Data Processing

CTA data model ObsConfig





Examples :

Number of telescopes involved

Field of view

Pointing direction

...

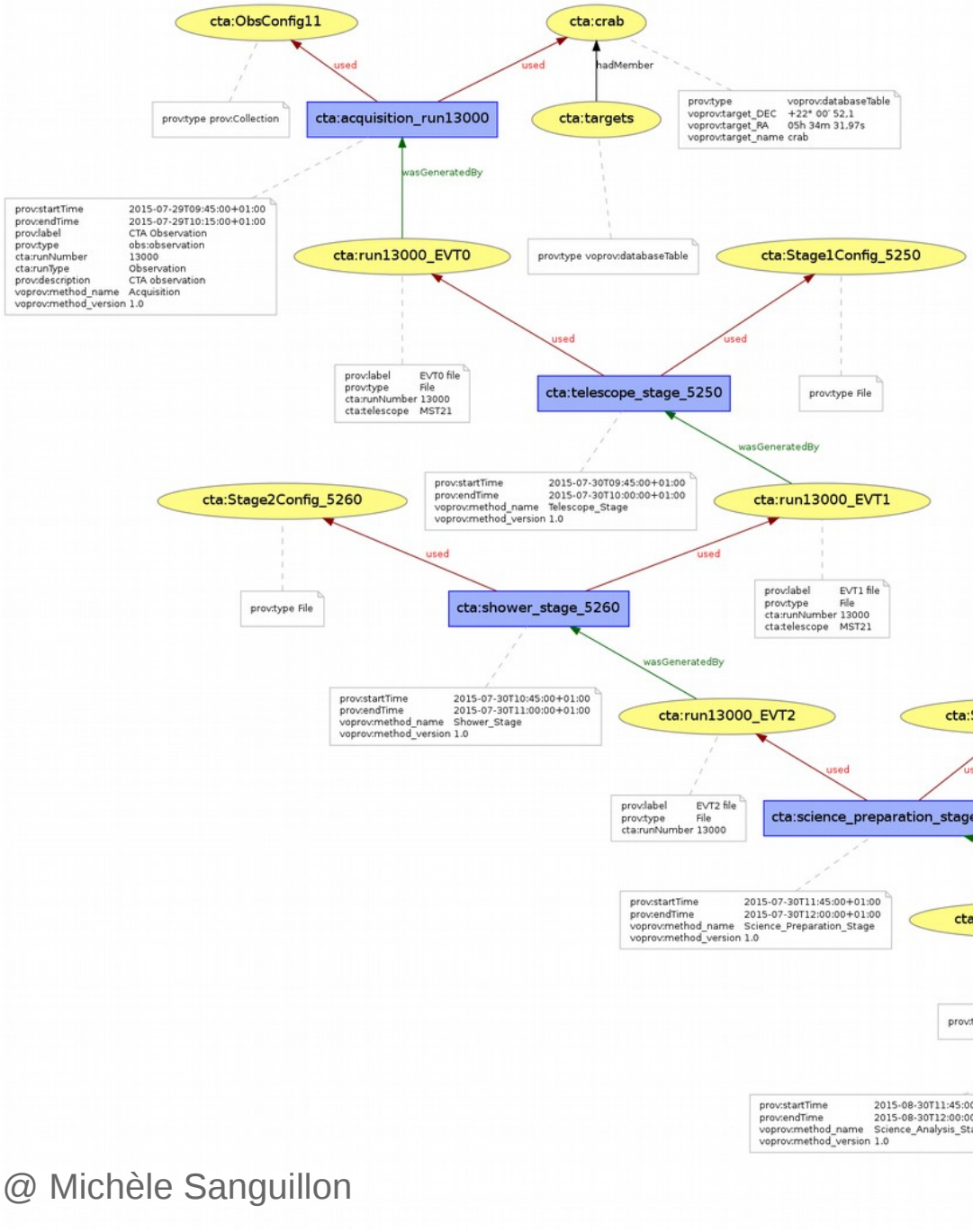
Different aspects

- **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

PROV-N

```
entity(rave:0645m522I0049.fits, [prov:type = 'std:fits']
entity(rave:0645m522I0049.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:0645m522I0049.fits, rave:act_observation, -)
used(rave:act_irafReduction, rave:0645m522I0049.fits, -)
wasGeneratedBy(rave:0645m522I0049.wav.fts, rave:act_irafReduction, -)
wasDerivedFrom(rave:0645m522I0049.wav.fts, rave:0645m522I0049.fits)
```

@ Kristin Riebe



PROV-N file:

```
// Stage 0 : Acquisition
activity(cta:acquisition_run13000, 2015-07-29T09:45:00, 2015-07-29T10:15:00,
    [
    prov:label = "CTA Observation",
    prov:type = "obs:observation",
    prov:description = "CTA observation",
    voprov:method_name="Acquisition",
    voprov:method_version="1.0",
    cta:runNumber=13000,
    cta:runType="Observation"
  ]
)

// Entities Description
entity(cta:ObsConfig11, [prov:type = "prov:Collection"])
entity(cta:run13000_EVT0, [prov:label = "EVT0 file", prov:type = "File",
cta:runNumber=13000, cta:telescope="MST21" ])

// Relations
wasGeneratedBy(cta:run13000_EVT0,cta:acquisition_run13000,-)
used(cta:acquisition_run13000, cta:ObsConfig11,-)
used(cta:acquisition_run13000, cta:crab,-)

//
-------------------------------------------------------------------------------
// Stage 1 : Teslescope Stage (Calibration)
activity(cta:telescope_stage_5250, 2015-07-30T09:45:00, 2015-07-30T10:00:00,
  [
    voprov:method_name="Telescope_Stage",
    voprov:method_version="1.0"
  ]
)
```

Turned into a diagram

@ Michèle Sanguillon