The use case of the CTA data model

A test bed for a Provenance Data model

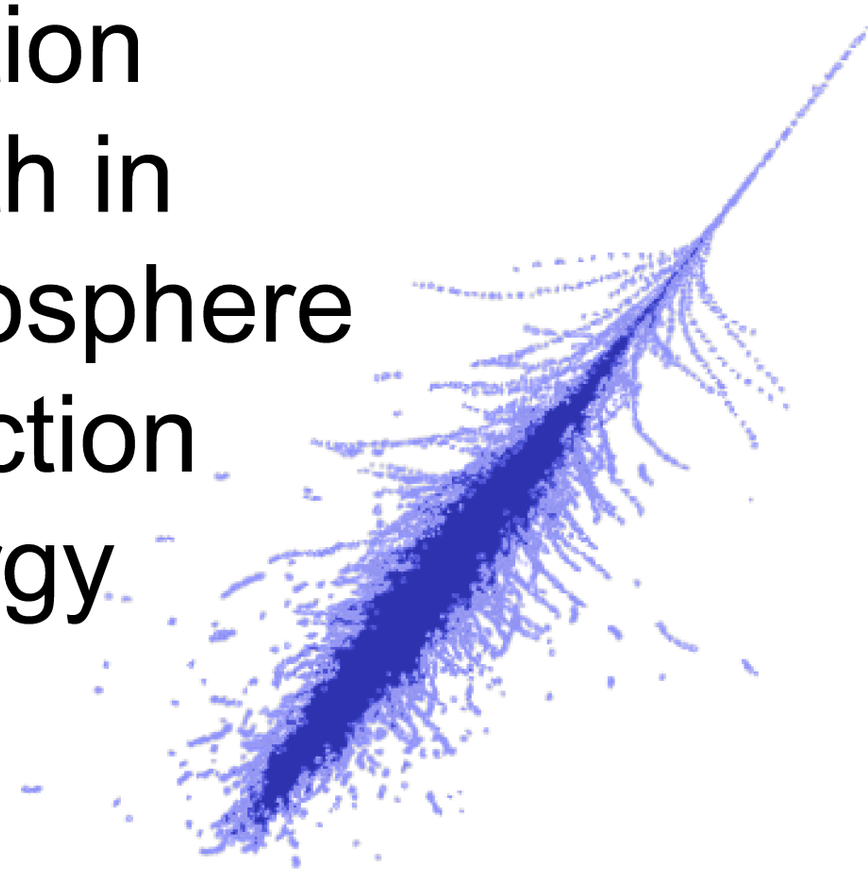
From C. Boisson (on behalf of CTA VO WG)
LUTh, Observatoire de Paris
Heidelberg Cosadie 'Data Center Forum'

C. Boisson, C. Lavalley, Karl Kosack,
J. Salgado, F. Bonnarel , M. Louys

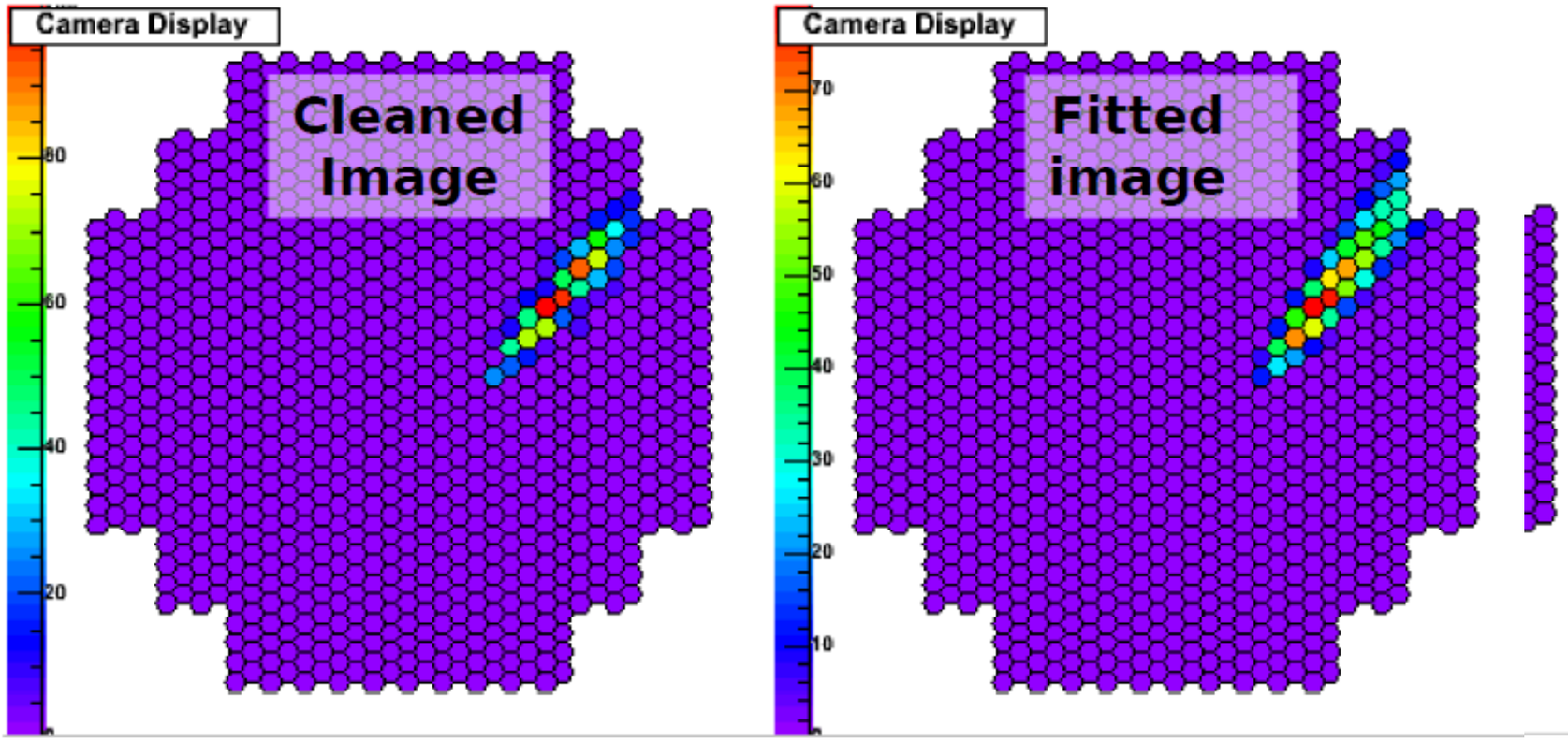
Laboratoire de l'Univers et de ses Théories

- Given:
  - (cleaned) **image** of the shower in **N telescopes**
  - Geometry and pointing direction of telescopes
  - Current event time

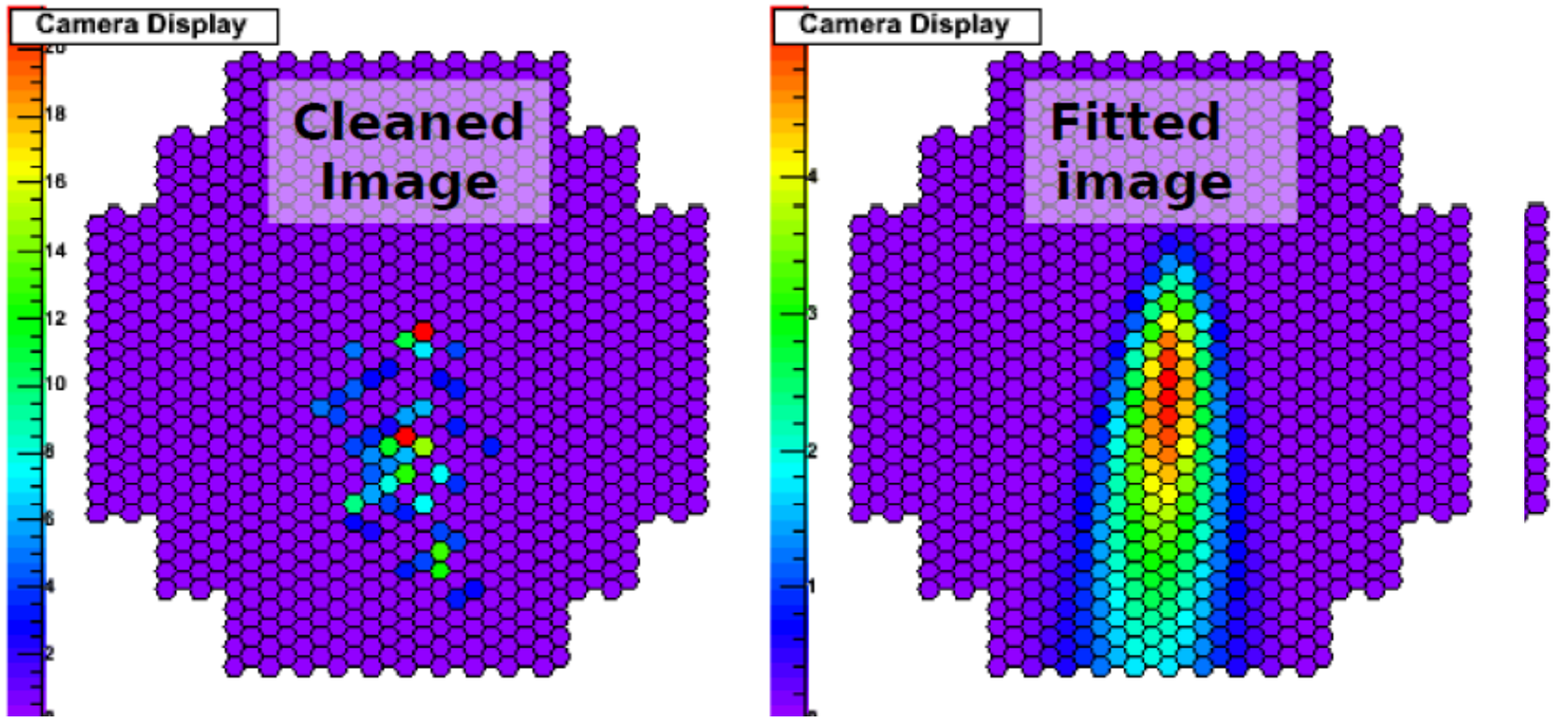
- Determine physical shower parameters:
  - Impact parameter/Core location
  - Depth in Atmosphere
  - Direction
  - Energy



Gamma Ray:

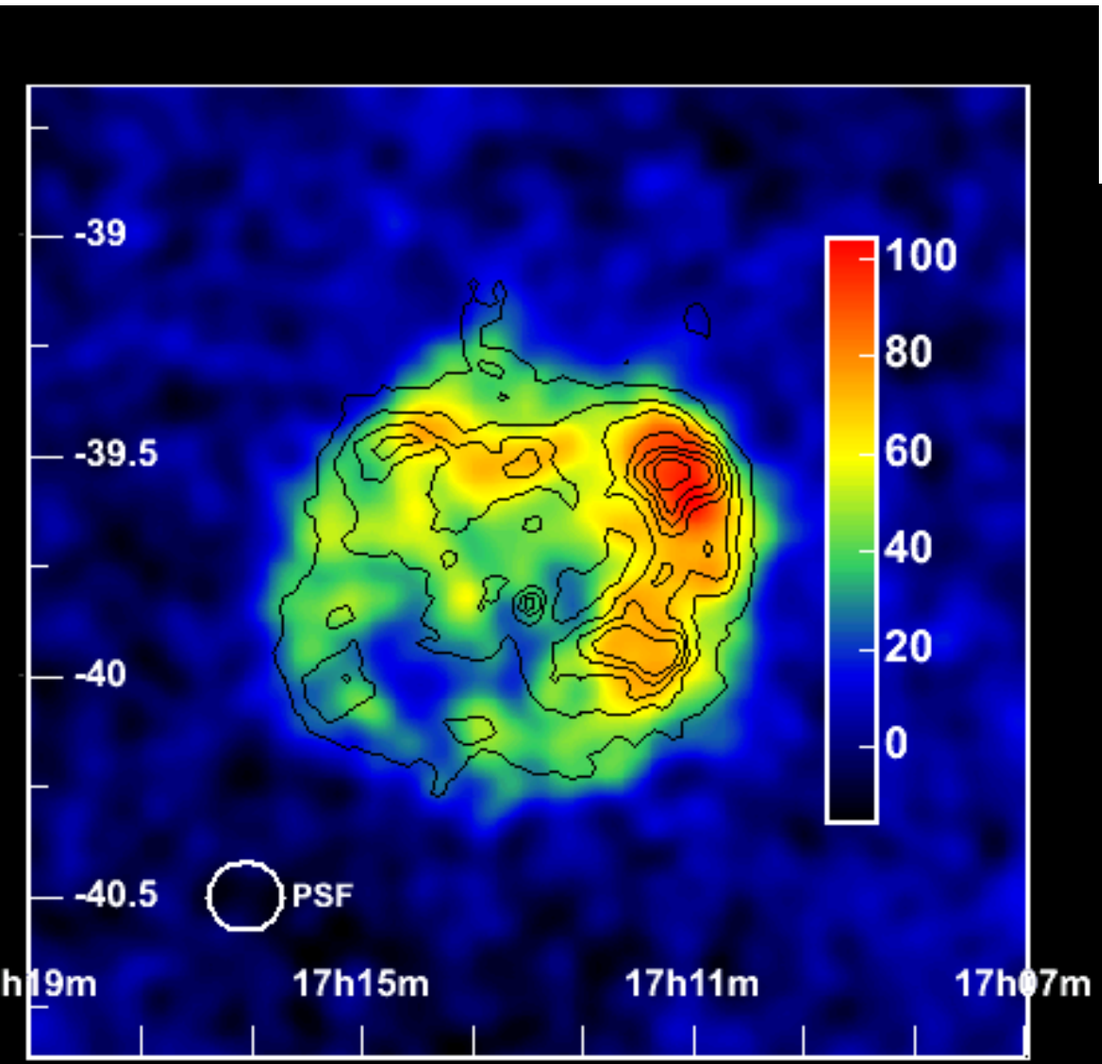


Hadron:

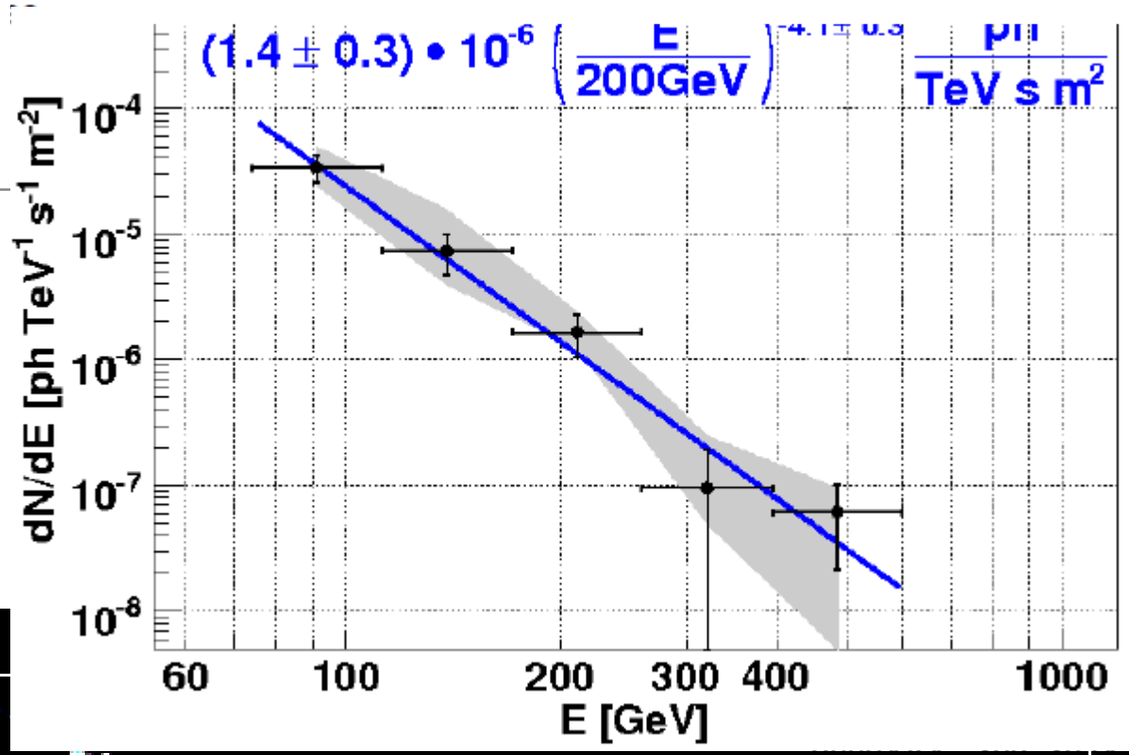


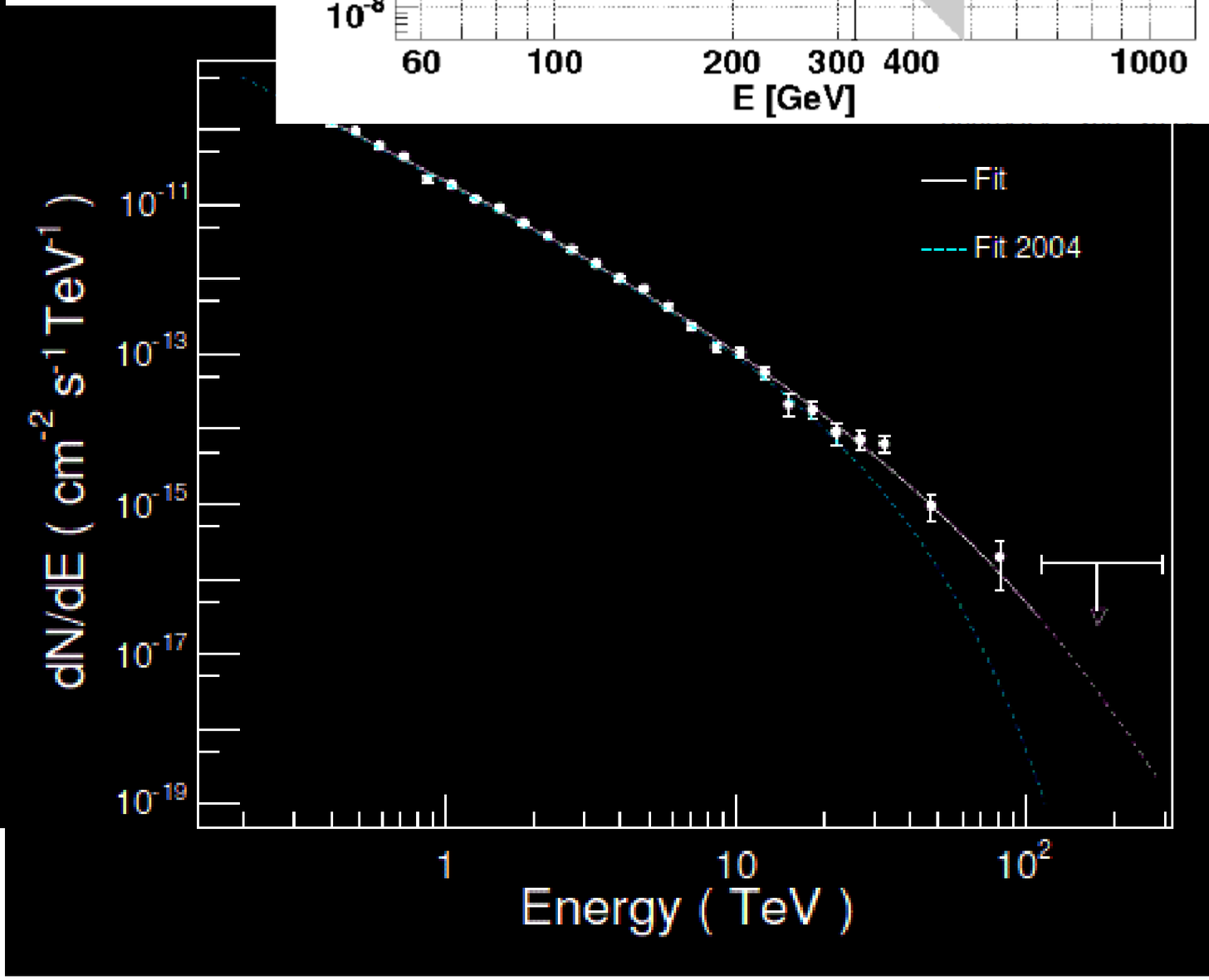
- Simulate processes in atmosphere
- Simulate telescope
- Compare shower images

→ reconstruct energy →

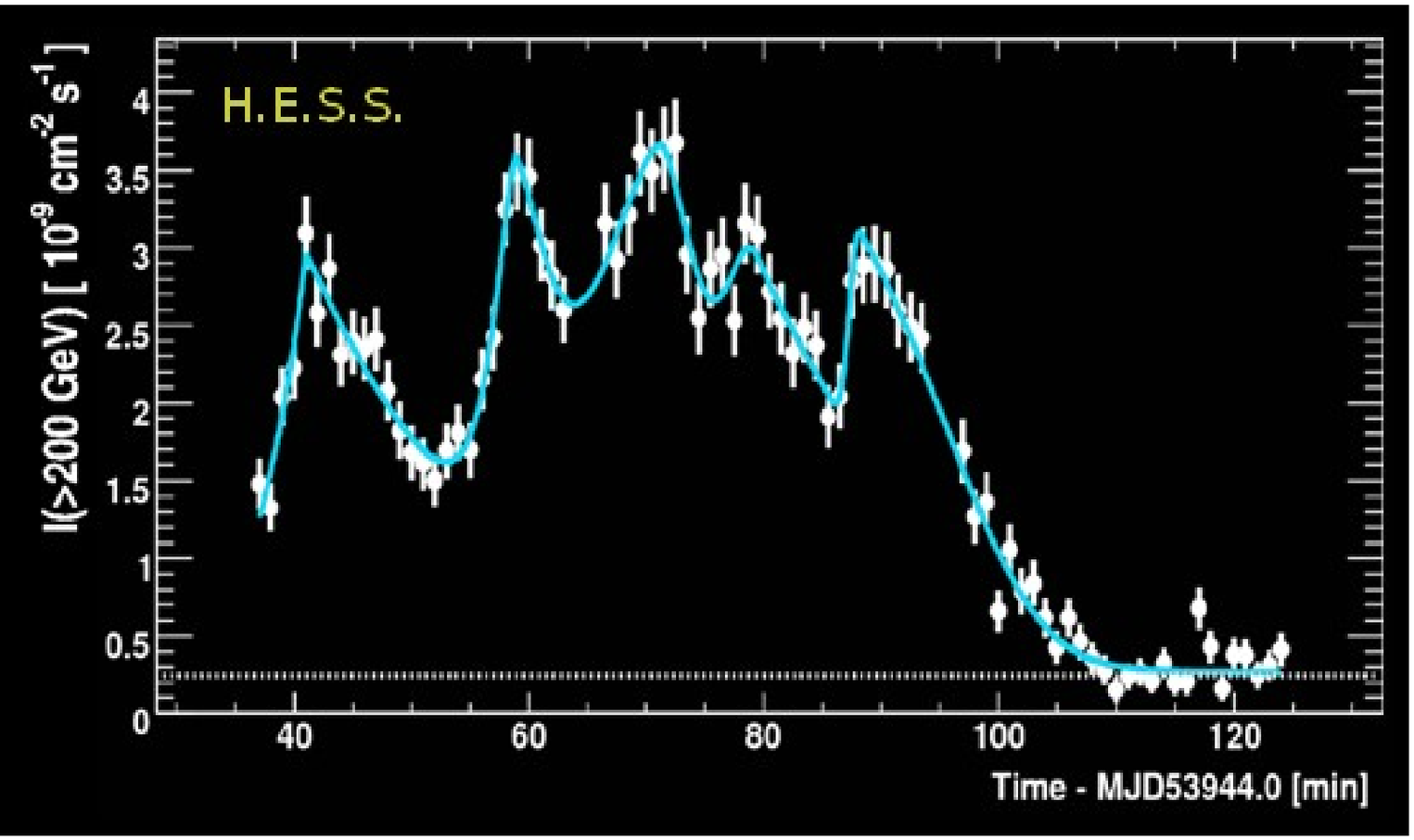


Aharonian et al., 2004, *Nature*, 432, 75



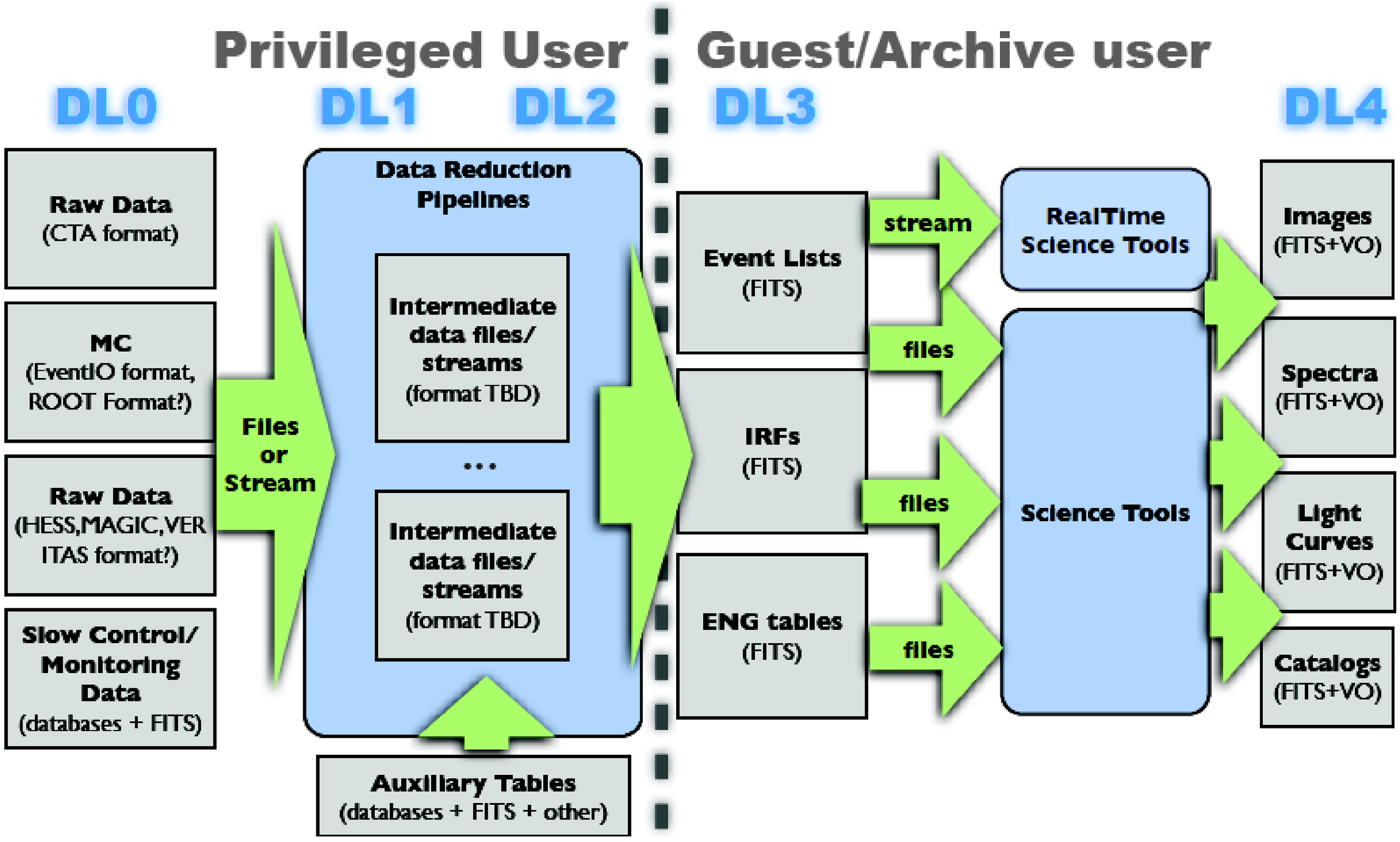


Resolved extended TeV source
Correspond with X-Ray morphology



- First major AGN flare for a source
- Variability on 5 min (or less) scales

Light curves at different timescale from months down to minutes


Privileged User
Guest/Archive user
DL0
DL1
DL2
DL3
DL4
Raw Data
(CTA format)
MC
(EventIO format,
ROOT Format?)
Raw Data
(HESS,MAGIC,VER
ITAS format?)
Slow Control/
Monitoring
Data
(databases + FITS)
Files
or
Stream
Data Reduction
Pipelines
Intermediate
data files/
streams
(format TBD)
...
Intermediate
data files/
streams
(format TBD)
Auxiliary Tables
(databases + FITS + other)
Event Lists
(FITS)
IRFs
(FITS)
ENG tables
(FITS)
stream
files
files
files
RealTime
Science Tools
Science Tools
Images
(FITS+VO)
Spectra
(FITS+VO)
Light
Curves
(FITS+VO)
Catalogs
(FITS+VO)

Kosack 2012

ICTA – Data

An observation – instrumental chain

Observations are “reconstructed” to get a list of event:
e(x,y,Ra,Dec,E,t, ambient conditions)

Observing time is not continuous → series of GTI exposures : run
Observation usually sum of 5 to 10hrs on source

Reconstructed position of the source → ROI, may be different from pointed
Ra, Dec
Background(s), image response functions

Wide field → more than one source In the field

ICTA – Data

Products

- Excess maps
- Significance maps
- Exposure maps
- PSF variation (function of t, E, position across image)

- Skymaps in different energy bands
- Integrated spectra
- Light curves
- Photometry point when too low statistics

CTA – an Observatory

Data products

Data obtained will be systematically pipeline-processed to generate a range of **scientific and auxiliary data products**

Minimal data product to provide is *a list of events with the relevant quality informations, engineering data, required image response functions, as well as science data (images, spectra, light curves)*

But recall that products are only one **representation of the data**.
One may want to select another set of guess (hadron-gamma separation, bkg,...)

VO

No standard yet to archive high energy astronomical data

- SSAP protocol defines a uniform interface to remotely discover and access simple 1-dimensional spectra → OGIP standard PHA format of X-ray spectra not accepted, so difficult fo VHE

- Spectral Data Model does not describe completely the HE data

- HE spectra are not physical units but in instrument counts - calibration needed and a model should be assumed to obtain a spectrum in physical units

- ObsTAP - makes it possible to discover and access the whole dataset of the observation, but doesn’t access the calibration files needed for the analysis

- Units not adequate (e.g. meter) : problem of precision

VO

No standard yet to archive high energy astronomical data

- Missing keywords (Utypes) to the Spectral Data Model to describe High Energy astronomical data: e.g. calibration version, model used to extract spectrum, PSF instead of aperture model, time boundaries of observation together with live time

- Calibration DM : useful for x-calibration, changes flux value but here we have a fully new analysis (bkg, bin sizes, …)

- Partial data cubes : single exposure data cube access n different observations for better S/N and look to final results

- Light curves : need to talk to « time series » WG ?
  Not properly tackled yet.

VO – publishing

Two kind of data to publish

- Highest data products and catalogs where modeling is predefined
  → just define interface: DataLink? But still need to tell the history of the data set : Provenance ?

- Some more data with calibrations (on tool side: final post-processing, workflow system rather than a final product)
  → a complete DM

We have identified the need of an extension or combination of existent IVOA Data Models such as ObsCore & Characterisation (data products), SimDM & PDL (pipelines), DataLink (protocol), to take account of the particularities of HE Data Products. Provenance is another important block.

Provenance

How the observation is taken : ObservingConfiguration

- Ambient Conditions :altitude, weather, wind, pressure versus time, aerosol, moon phase...
  Create atmospheric sets ? Keep all metadata?
- Observing Configurations : How was the telescope configured ?
  How many telescopes involved ?

How the observation is processed: Processing workflow descriptions

- Data Processing : calibration, reconstruction and analysis pipeline

Overview of the CTA Model

See the UML diagram in the PNG file linked here.

UML diagram for CTA DM

Conclusions

Several aspects of provenance involved

Observing configuration

Processing

Multiple types for data products