Data Model for VHE gamma-ray data and relations with IVOA standards

B. Khélifi[†] (APC, Paris)

With inputs from the CTAO data model group

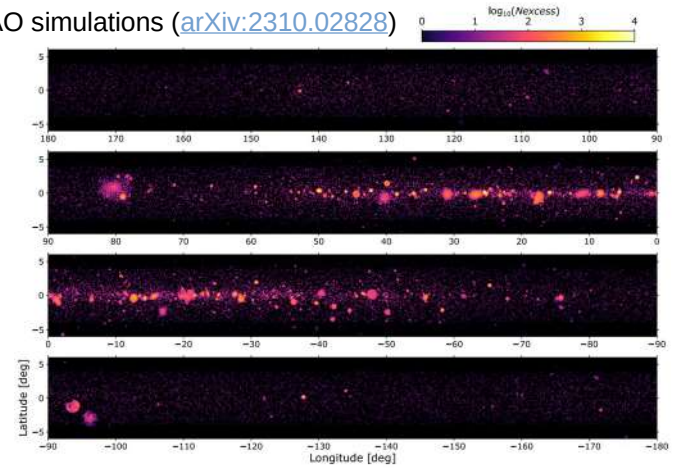
IVOA Interop meeting (La Valletta, Malta)

[†]: Gammapy Project Manager, VODF convener, one of the HESS DL3 release responsibles, SWH ambassador, etc

VHE γ-ray observations

Acceleration processes: relativistic particles

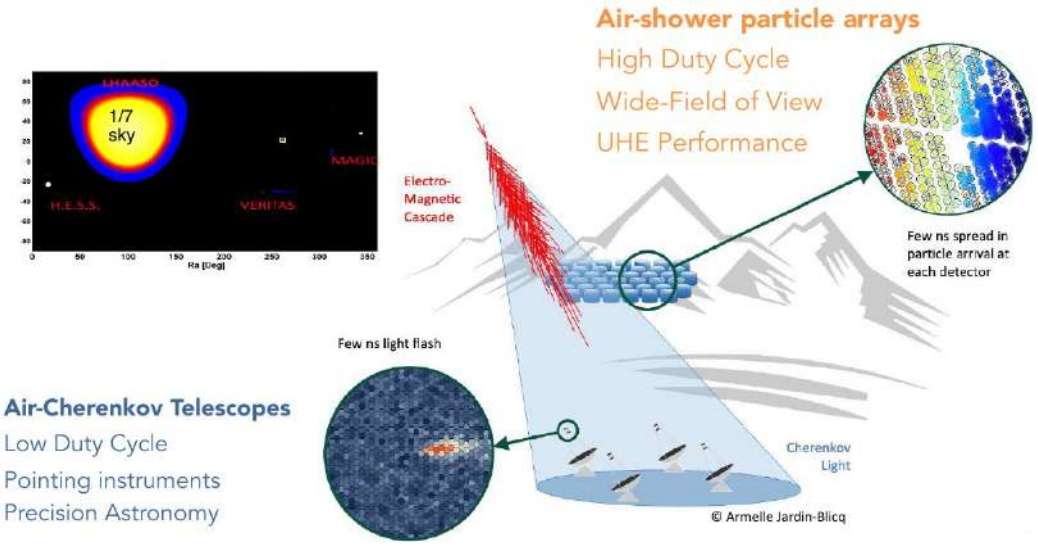
CTAO simulations (arXiv:2310.02828)



- Pulsar Wind Nebula, Supernova Remnants, Stellar Clusters, etc
- Active Galactic Nuclei
- Gamma-ray Bursts
- Etc

VHE γ-ray detection

Atmospheric showers



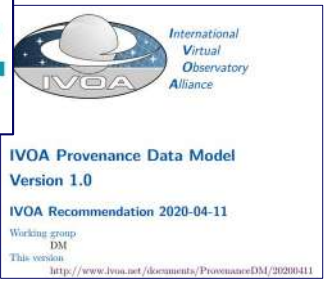
→ Counting experiments

- One event = One shower

VHE data

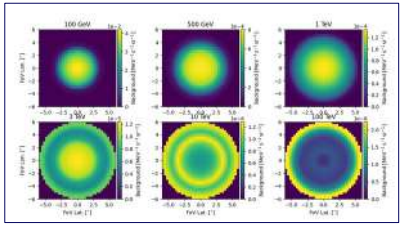
Future experiments: observatories

- CTAO, SWGO
- Open data and their dissemination
- Some observations are proposal-driven
- Need FAIR data and Provenance information



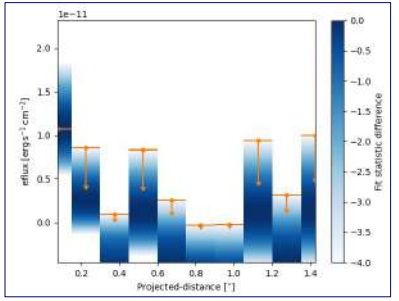
Particularity of imaging Cherenkov telescopes

- Instrument Response Files (IRFs) are time-dependant
  - One set of IRFs per “Good Time Interval” (GTI)
- IRFs are multi-dimensional: 3D to 6D



VHE Astrophysical products

- Complex (as results of fits) → multi-dimensional
- To be published → FAIR



Data modelling

The past achievements

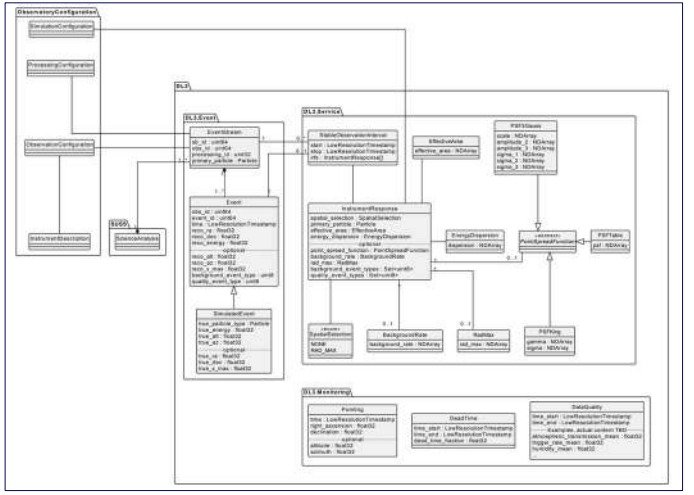
OGIP X-ray format (90^s) → Fermi-LAT GeV format (00^s) → GADF TeV format (10^s)

- **GADF** is from a community-driven initiative and is in use for running TeV experiments (H.E.S.S., MAGIC, VERITAS, HAWC)

CTAO work

See K. Kosack talk, yesterday

- In-depth study of data modelling
- Community is ready to create a data format respecting the FAIR principles and following as much as possible the IVOA recommendations
- Some of them might need some upgrades

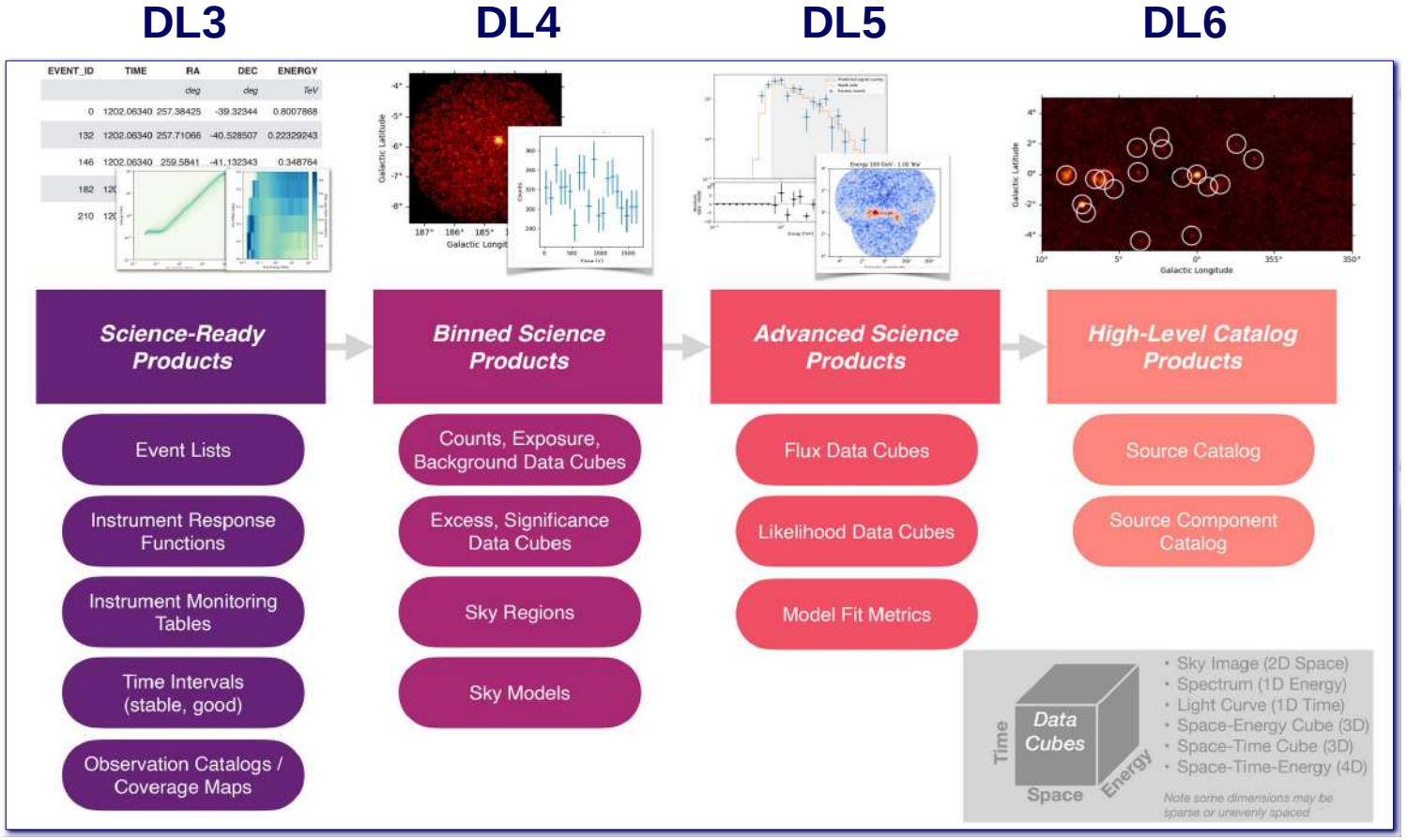


IVOA High Energy Interest Group

Talk content

1. Data levels
2. Data characterisation with metadata
3. Events and IRFs
4. Higher level products

1. Data levels



Khélifi, B., *et al.*, Proc. of 38th ICRC (2023)

These data will be public

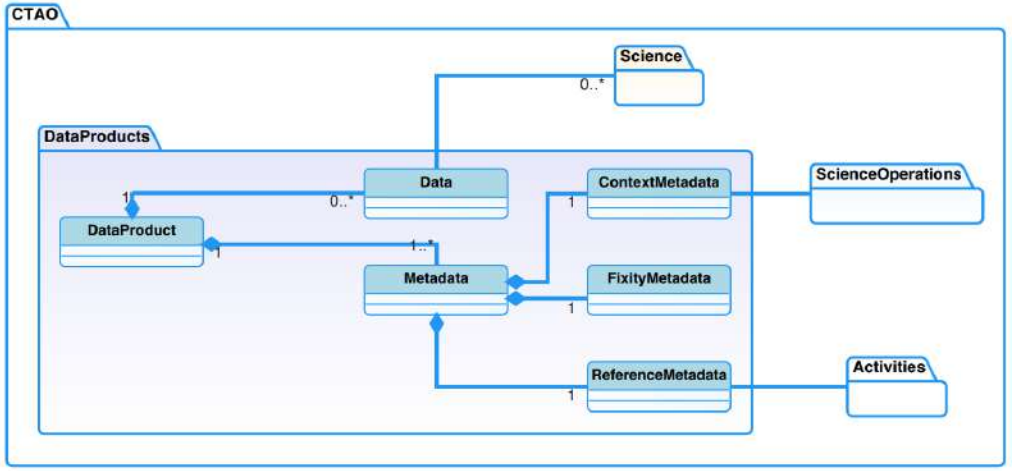
- There are 6 other lower levels

Correspondence with X-rays

- L1 w/o IRFs ⇔ DL3
- L2 ⇔ DL4
- L3 ⇔ DL5+DL6

2. Data characterisation

For each data product



CTAO top-level data model release_2a

Reference metadata

Permits unique identification associated to an activity. Contains also licence, copyrights, contact information, “release”

Contains the `Provenance` information of the last activity step

See M. Servillat talk, IVOA Sydney (05/2024)

Context metadata

Activity-specific. For an observation, link to the SchedulingBlock

Direct mapping with `ObsCore`

Fixity metadata

Ensure data integrity (e.g. checksum)

2. Data characterisation

The DL3 data and ObsCore

- VHE data are list of events with associated IRFs (by the event timestamp)
- How to characterise them?
  - `ObsCore` “dataproduct_type” ?==? “event”
  - Should the notion of “event-list” dataset be more described?
  - How to link the complex IRFs? With `DataLink`?
- Check whether the `Observation Model` draft, the `Dataset Model` and the `Dataset Metadata Model` draft are compliant for the VHE data releases

Some prototyping has been made

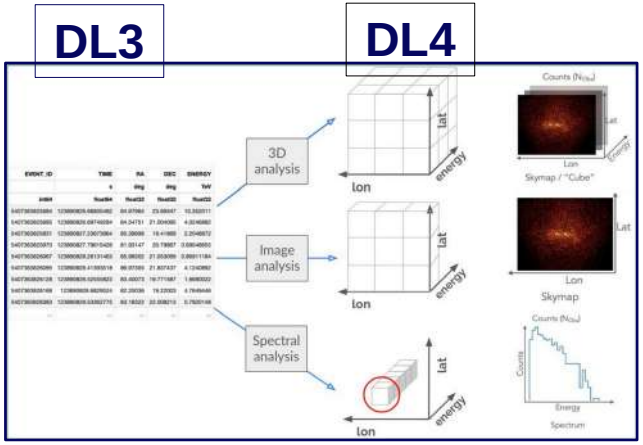
See M. Servillat talk, IVOA Sydney (05/2024)

- VO registry (`TAP` server) of H.E.S.S. test public data (DOI 10.5281/zenodo.1421099)
- Mapping between the “Observation” of the Gammapy library and `ObsCore` (`gammapy.data.ivoa.to_obscore_table()`)

3. Events and IRFs

Event list into N-dimensional maps

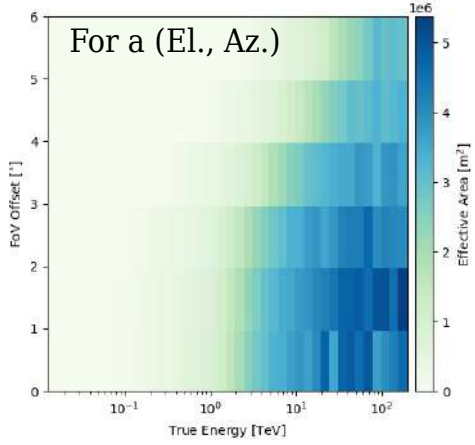
- Binned analysis makes projections of events (and IRFs): DL3 → DL4

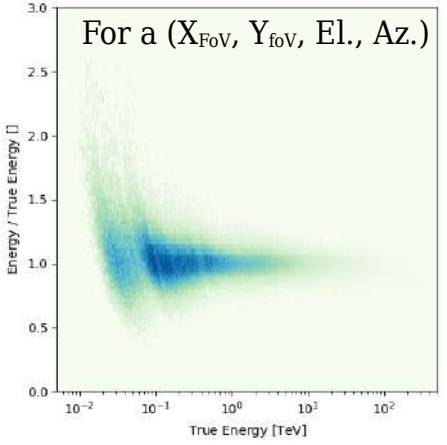


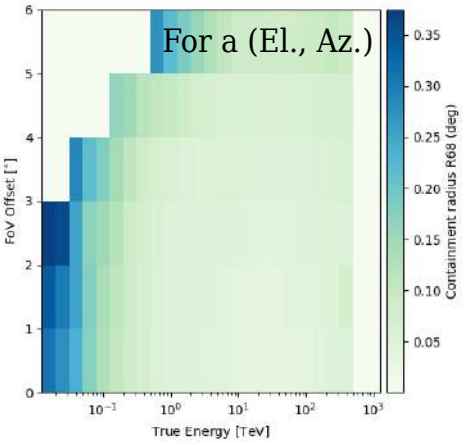
From the Gammapy documentation

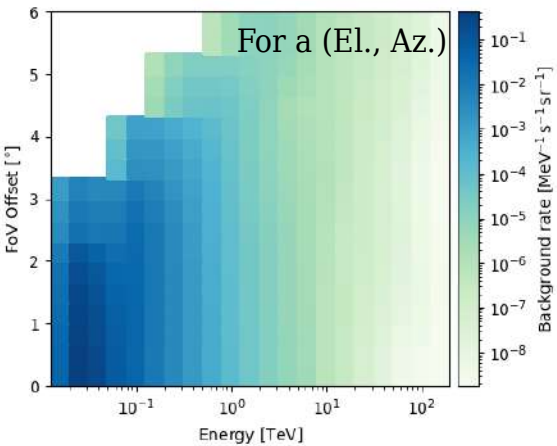
IRFs are stored into ND-cubes

- The four response functions , for CTAO South:









3. Events and IRFs

VHE needs

Axis

- Any physical dimension (many units for a given dimension) or None (e.g. a label)
- Defined per node or bin (linear, sparse, log, sqrt, etc is then a consequence)
  - Be able to support WCS and HEALPix coordinate systems
  - Can be cyclic (e.g. periodogram)
- Knowledge of the interpolation scheme (lin, log, sqrt)

Data Content

- w or w/o dimension (e.g. probability)

PS: for a data model, speaking about dimension seems more general than handling units

Voxel

- defined by the N axes, creating the ND-cubes

Metadata and Provenance

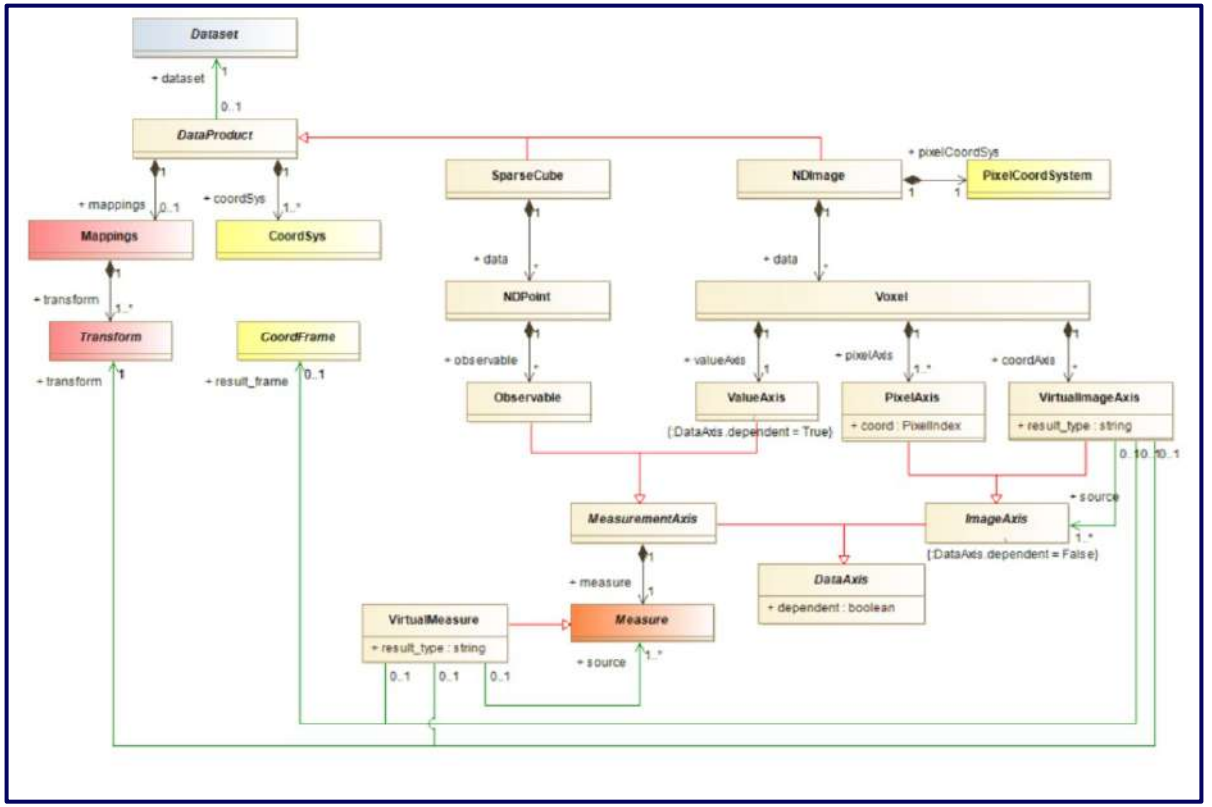
- Aim to publish of FAIR ND-cubes (e.g. counts or exposure cubes as MOC)

3. Events and IRFs

Are the VHE needs compliant with the [N-Dimensional Cube Model] draft??

Why separating SparseCube and NDImage? Why not EventList and ND-Cube?

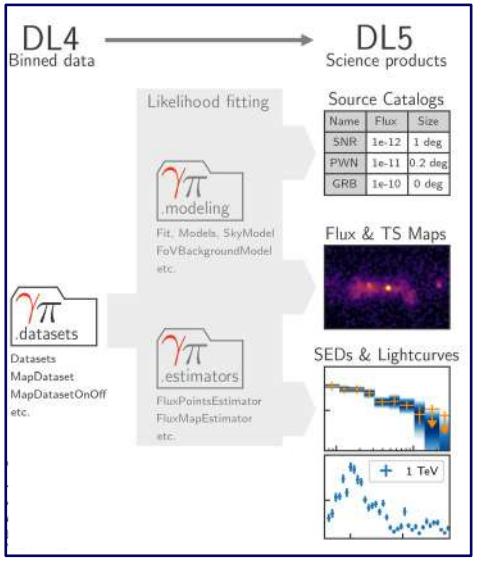
Where is the provenance? Is there no specific metadata?



4. Higher level products

Modelling and fitting

- Use of the forward-folding technique to extract astrophysical final products

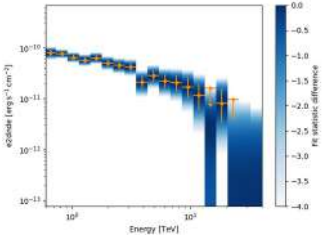


Some products

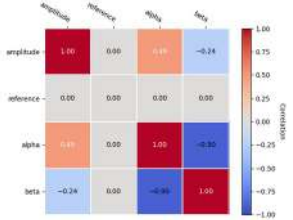
Adjusted data model

```
model type    name     value         unit      ... min max frozen link prior
----- ---- --------- ---------- -------------- ... --- --- ------ ---- -----
 crab          index 2.2727e+00                ... nan nan  False
 crab      amplitude 4.7913e-11 cm-2 s-1 TeV-1 ... nan nan  False
 crab      reference 1.0000e+00            TeV ... nan nan   True
 crab         lambda_ 1.2097e-01          TeV-1 ... nan nan  False
 crab          alpha 1.0000e+00                ... nan nan   True
```

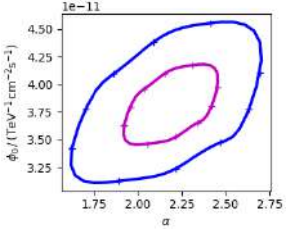
Flux points and their likelihood



And the parameter correlation

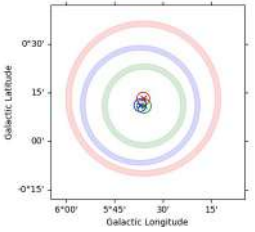


Two-fold parameter correlation

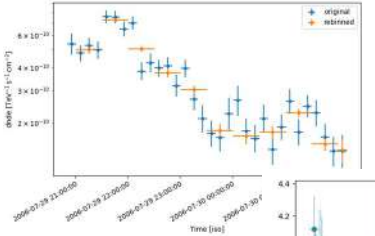


From the Gammapy documentation

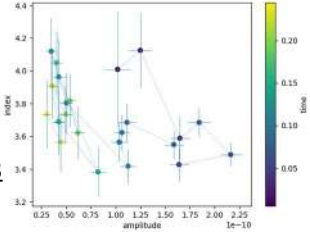
Fitted spatial model with errors



Light curves at different energies



Time-dependency of spectral parameters



4. Higher level products

Are the VHE needs compliant with the [N-Dimensional Cube Model] draft?

This draft also defines

- `DataProduct`
  - Has to contain a `CoordSys`, independently to the axes definition!?
- Surprising to mix the ND-Cube and DataProduct models into the same document…

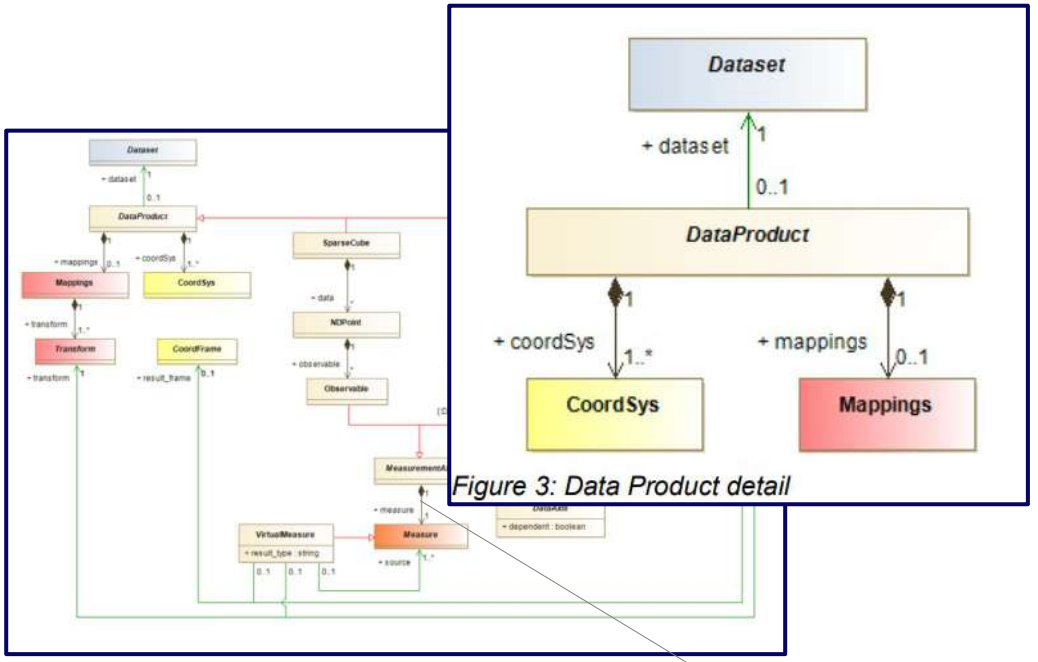


Figure 3: Data Product detail

→ Not (yet) compliant with all VHE use cases

- Large variety of data products
- Need to also handle “model parameters list”, a core concept to create catalogue of sources (which is a `DataProduct` by itself)
- Are the metadata enough for publication and VO dissemination?

PS: notion of `Measure` for a fitted value than is model and prior dependant is not general enough

Summary and personal conclusions

VHE gamma-ray observatories are counting detectors

- These open data will soon populate the VO data bases (CTAO, SWGO) – and maybe the archives of current experiments...
- Counting detectors → notion of *event* & *event-list*
- *Instrument Response Files* are time-dependent → to be coupled to the event-list, forming together a *Dataset*

Many commonalities with other high-energy experiments

- Astrophysical VHE **neutrino observatories** (KM3NeT, IceCube)
  - The IRFs can be factorised in the same manner than the ones of gamma-ray detectors
  - Maybe less time-dependent

See J. Schnabel talk, yesterday

- **X-ray and GeV observatories** (XMM-Newton, Chandra, Fermi-LAT)
  - The IRFs can be factorised in the same manner than the ones of gamma-ray detectors
  - Same concept of data levels (DL_*X* = L_*X*+*2*)

Summary and personal conclusions

Astrophysics usage

- Many **open Science Analysis Tools**
  - They should follow FAIR4RS principles to handle FAIR data → Need of updated IVOA concepts
- **Multi-wavelength and multi-messenger astrophysics**
  - Libraries now permit real joint fit of multi-instrument datasets → new use cases for the models
  - Final astrophysics products should contain Provenance and precise metadata → for journals or VO

High Energy Interest Group

Will revisit the current models associated to our astrophysical use cases

Will permit to publish event lists (DL3) up to final results (DL5)

See the TD/HEIG session (yesterday): link

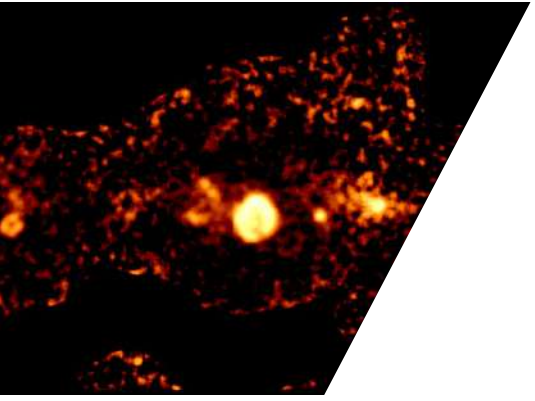
HEIG wiki: link

International Virtual Observatory Alliance

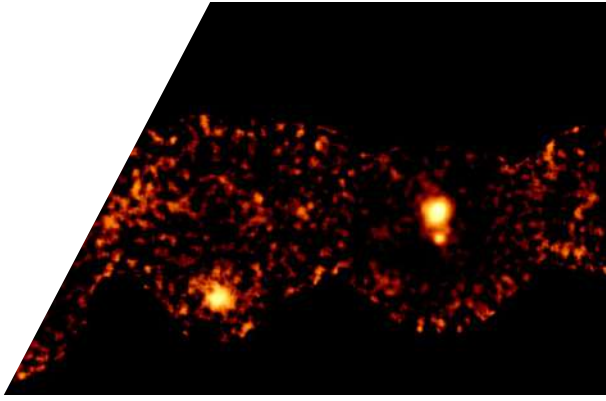
Virtual Observatory and High Energy Astrophysics

Version 1.0

IVOA Note 2024-11-12

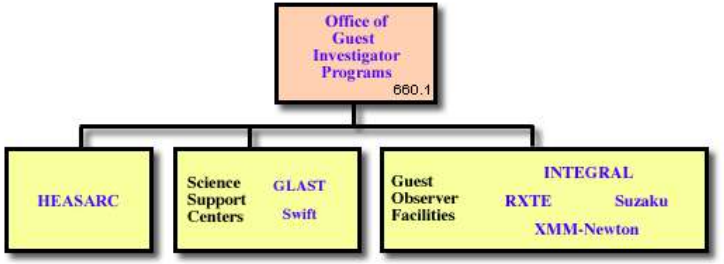


Annexes



X-rays → GeV: OGIP

Office of Guest Investigator Programs conventions



15 specific format recommendations (link) Corcoran *et al*., 1995 **ADASS**

- None on the final astrophysical products

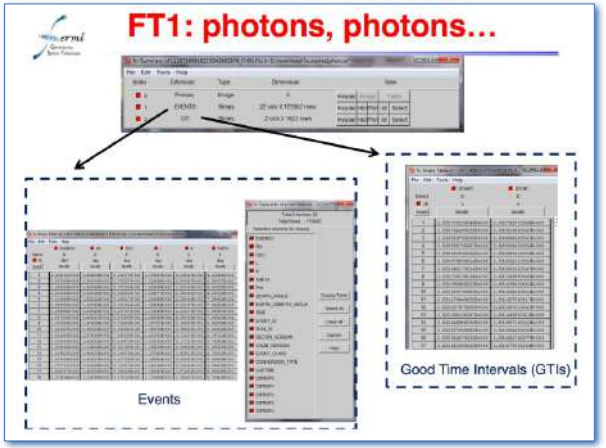
- **Event list:** well described, except errors, no detailed metadata
- **Instrument Response Files** (IRFs):
  - PHA (XSPEC compatible): ie RMF+ARF, ie Edisp+Area, includes statistical errors and systematics, link to the background file, GTI, metadata (e.g. creation, history)
  - PSF (radial or 2D): includes statistical errors, metadata (e.g. creation, history)

Name	Units	Description
TIME	`s' or `d'	The time associated with the event
RAWX	`pixel'	Raw telemetry X position of the event
RAWY	`pixel'	Raw telemetry Y position of the event
DETX	`pixel'	Linearized X position of the event on the detector
DETY	`pixel'	Linearized Y position of the event on the detector
X	`pixel'	Projected X position of the event on the sky
Y	`pixel'	Projected Y position of the event on the sky
PHA	`chan'	Pulse height analyzer' energy channel

HE γ-rays: Fermi-LAT

Formats regulated very early, and afterwards updated

- Project Data Management Plan (PDMP), 2007 (link)
- Science Data Products File Format Document (FFD), 2019 (link)



Definition of Data levels (L0 → L3)

- like for X-rays (HEASARC continuity)
- (raw data)
- **L0**: 'cleaned raw data'
- **L1**: starting point for scientific analyses
- **L2**: result of science analysis tools
- **L3**: catalogs or compendia of DL2 data, including e.g. flux history, sce identification

In addition, **ancillary data**: diffuse galactic interstellar, extragalactic emission models, pulsar ephemerides

Data format

- based on OGIP + specificities (e.g. interstellar emission model, BAT data, LAT LLE)
- specific format for L2 (LC, spectrum) and L3 (catalogs)
- Metadata: s/w name and version, but not the release name, no provenance in data (but in web pages)

Almost FAIR
Light compliance to IVOA standards

VHE γ-rays: GADF

Up to mid-10's, VHE community worked in a totally competitive and closed mode

- All was private
- Except few MoUs around scientific projects

Better results
Interoperability between instruments
Respect of the FAIR principles

Some 'dreamers' worked towards the opening of the VHE astrophysics

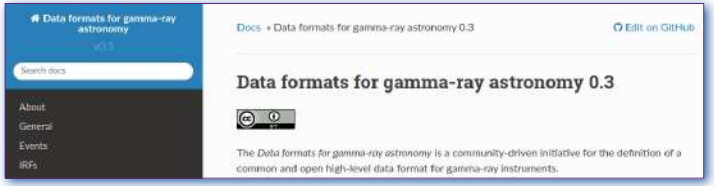
Deil, C., *et al.*, ASTERICS 2016 (link)

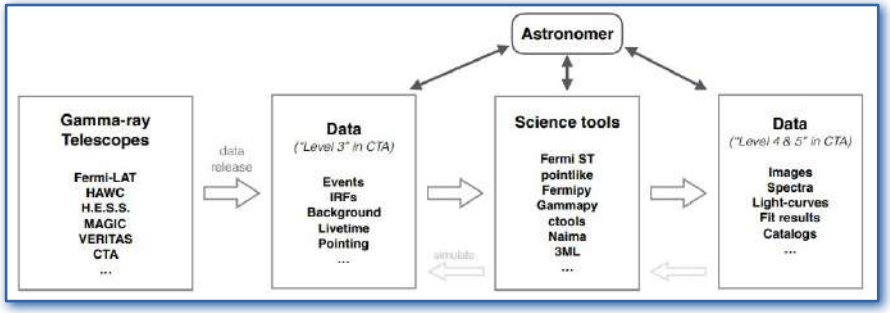
- Data format standardization: open initiative 'Open Gamma-Ray Astro'
- Open Science Tools: Gammapy

The Gamma Astro Data Format

DOI 10.5281/zenodo.7304668

- Strongly influenced by the Fermi-LAT format (and OGIP) and serialization into FITS
- Same type of data levels: DL3, DL4, DL5
- Full description of the DL3: event list and IRFs





Deil, C., *et al.*, Proc. of Gamma 2016

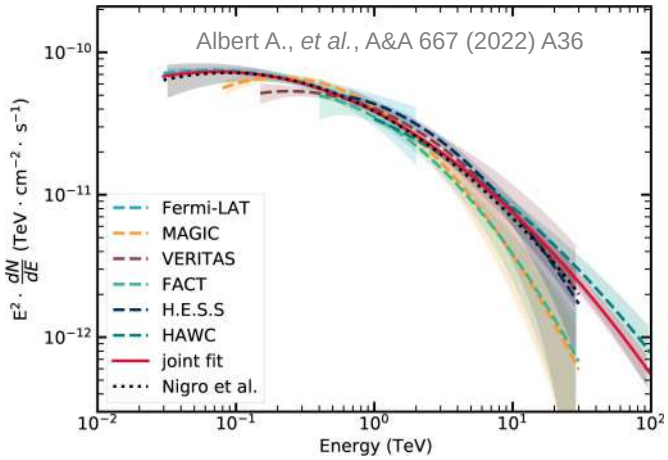
VHE γ-rays: GADF

Proved to be well suited for VHE needs and can serve as standard

As demonstrated by the joint Crab analysis with `Gammapy`:

Or with astrophysical papers

HGPS, Abdalla H., *et al.*, A&A 612 (2018) A1



But some drawbacks appear

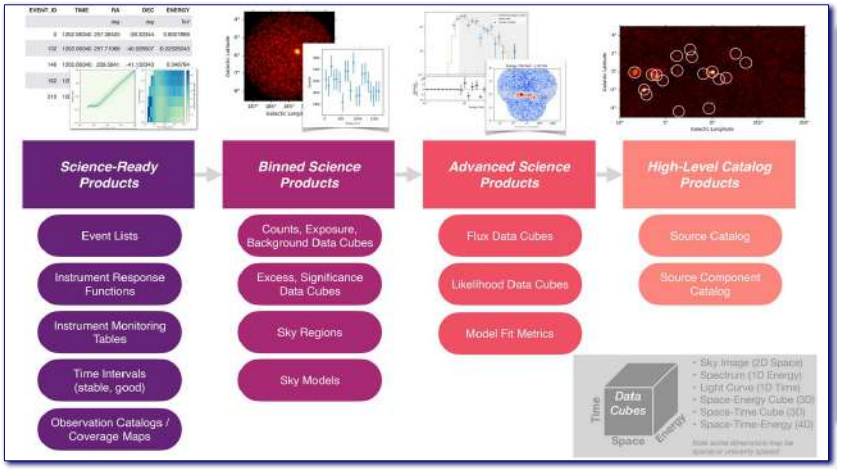
- GADF had no clear organizational structure
- No clear resolution of contentious issues
- No clear roadmap
- Not FAIR enough, no clear standards for DL4+

→ Hand-over to VODF

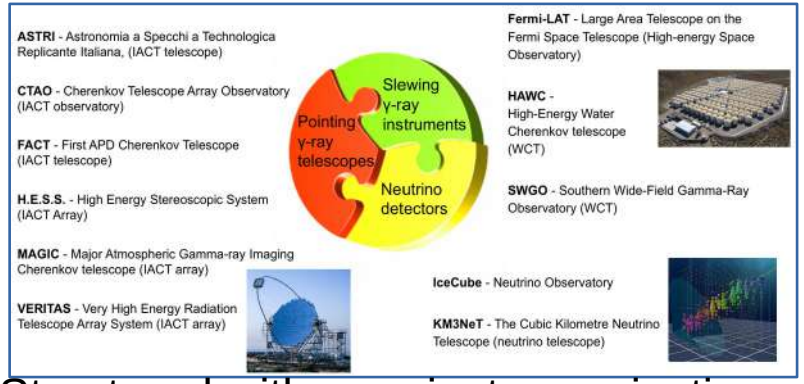
VHE data: VODF

Open Initiative ‘Very-high-energy Open Data Format’ (link)

- Aims to format VHE data (gamma and neutrino)
- Officially supported by 11 experiments
- Structured with a project organization



Khélifi, B., *et al.*, Proc. of 38th ICRC (2023)



Coordination Committee, Conveners: R. Zanin, B. Khélifi
Lead Editors: K. Kosack, L. Olivera-Nieto, J. Schnabel

```
/*****************************************************************************
/* FITS FILE:
/*    VODF Level-1 Event Data
/*
/* EXTENSIONS SUMMARY:
/*  IDX  NAME              VER CLASS               TYPE
/*  ---  ----------------- --- ------------------- ----------------
/*    0. EVENTS              0 OGIP.EVENTS         [TableExtension]
/*    1. SOI                 0                     [TableExtension]
/*****************************************************************************

/ ###########################################################################
/ HDU: EVENTS
/ DESCRIPTION:
/    VODF Level 1 Event List
/ ###########################################################################
XTENSION = BINTABLE
EXTNAME  = EVENTS
EXTVER   = 0
```

© Kosack, K. and Khélifi, B.

Poisson Log-Likelihood

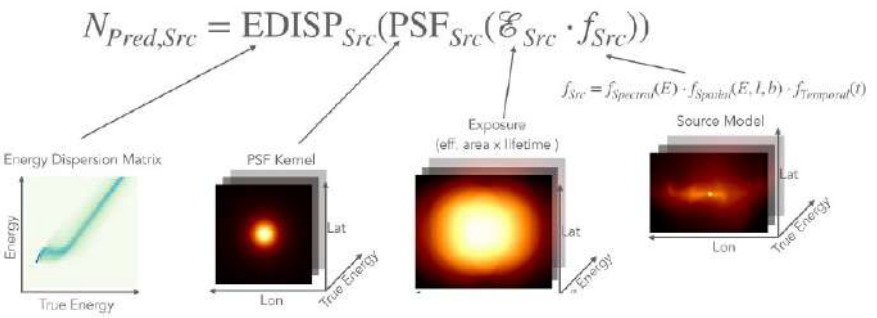
Common algorithms for the libraries: Poisson Log-Likelihood

"Cash statistics": summed over all "bins"

$$\mathscr{C} = 2\sum_i N^i_{Pred} - N^i_{Obs} \cdot \log N^i_{Pred}$$

i: spectral channels or 3D voxels

$$N_{Pred} = N_{Bkg} + \sum_{Src} N_{Pred,Src}$$



- Bins in the spectral, spatial, temporal domain

→ Need of a "global" background model template with "correction parameters"

→ Need of the "signal" IRFs and source models

Most of the time,
Identical factorization of the IRFs
for X-rays → UHE & neutrino exp.

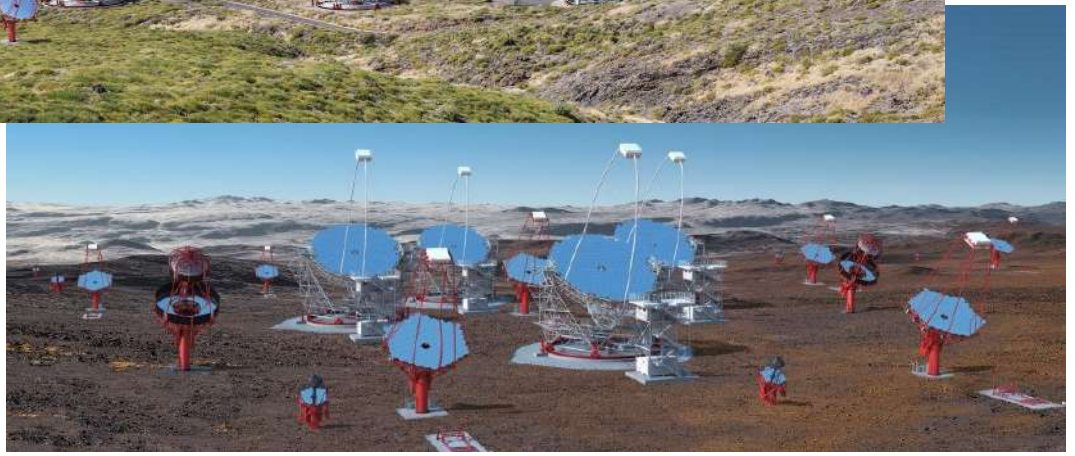
Cherenkov Telescope Array

Gamma-ray observatory

- First Open VHE observatory
- O(10)GeV → 200TeV
- PSF: 0.3° → 0.02°
- Observations with FoV of 3°-10° during dark mights

Full sky observations

Under construction



© CTAO

Array of the Northern site (La Palma)

- 9 MSTs (88 m²), 4 LSTs (400 m²)

Array of the Southern site (Paranal)

- 2/3 LSTs (400 m²), 14 MSTs (88 m²), 42 SSTs (5 m²)

Southern Wide-field Gamma-ray Observatory

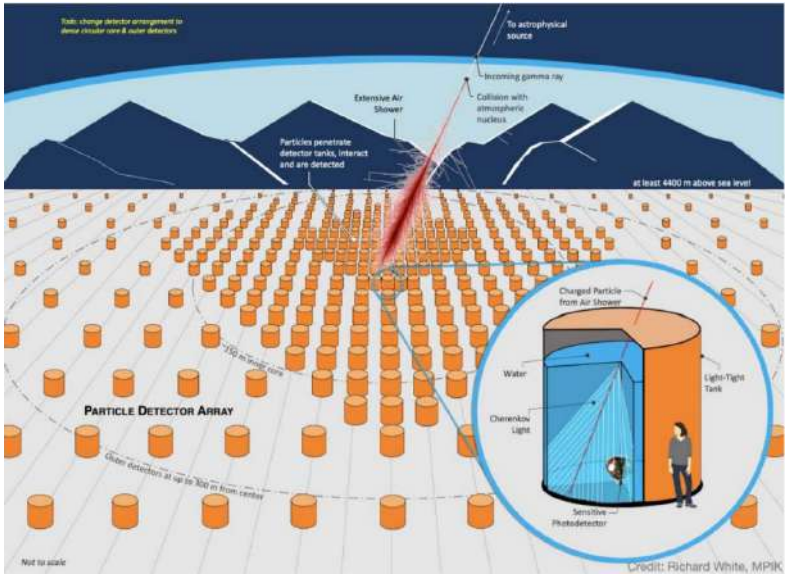
Project of new UHE observatory

- O(100)GeV → PeV
- PSF: 1°-2° →0.04°
- FoV of 60° during day and nights

Under Design Study

M1	R&D Phase Plan Established
M2	Science Benchmarks Defined
M3	Reference Configuration & Options Defined
M4	Site Shortlist Complete
M5	Candidate Configurations Defined
M6	Performance of Candidate Configurations Evaluated
M7	Preferred Site Identified
M8	Design Finalised
M9	Conceptual Design Report Complete

Water Cherenkov Detector
Ex: HAWC, LHAASO



Selected site (Aug. 24): Atacama Astronomical Park, Chili

The project Gammapy

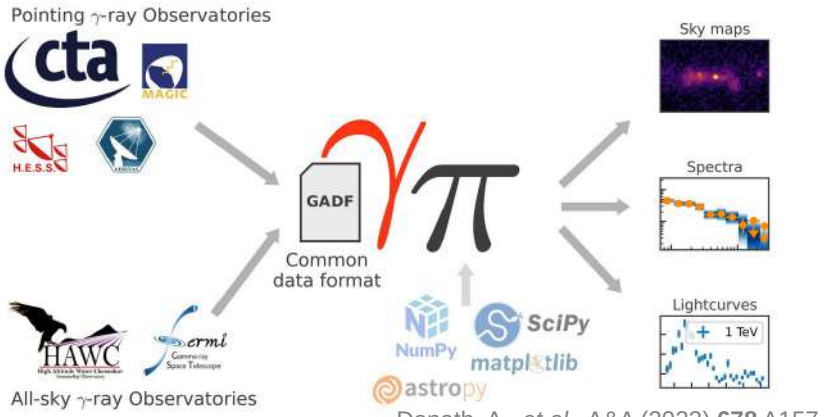
A Python package for gamma-ray astronomy

Open Python analysis library

- Uses data written in the GADF format
- Inserted into the Python ecosystem
- Respecting the FAIR4RS principles
- Making multi-instrument joint analyses

Open Research Software

- Open contributions within an open organisation with an open governance
- Reference library for the VHE gamma-ray astronomy
  - Selected as core library of the open Science Analysis Tool of CTAO
- Well recognized : jury's prize of the CoSO (2022)



Donath, A., *et al.*, A&A (2023) **678** A157