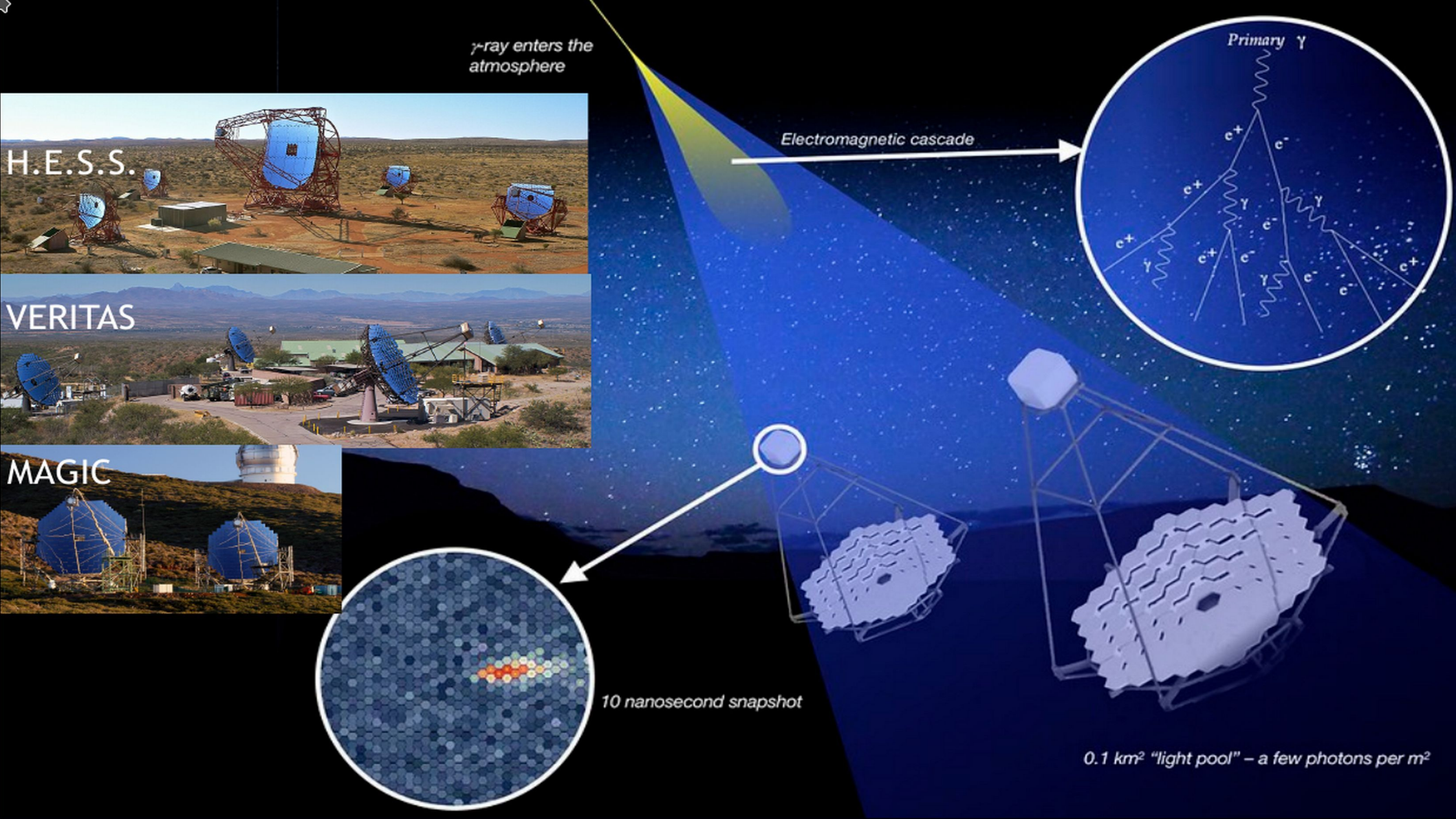
CTA - Cherenkov Telescope Array

Archiving data from a software telescope

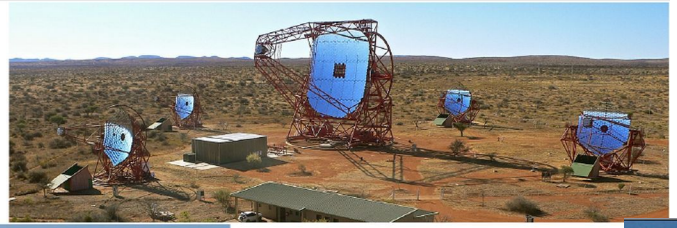
Catherine Boisson, Mathieu Servillat
Observatoire de Paris

Travail historique d'un groupe de personnes sur plusieurs instituts
(OP, CDS, LUPM, CEA)


γ-ray enters the atmosphere
H.E.S.S.
VERITAS
MAGIC
Electromagnetic cascade
Primary γ
10 nanosecond snapshot
0.1 km² "light pool" – a few photons per m²

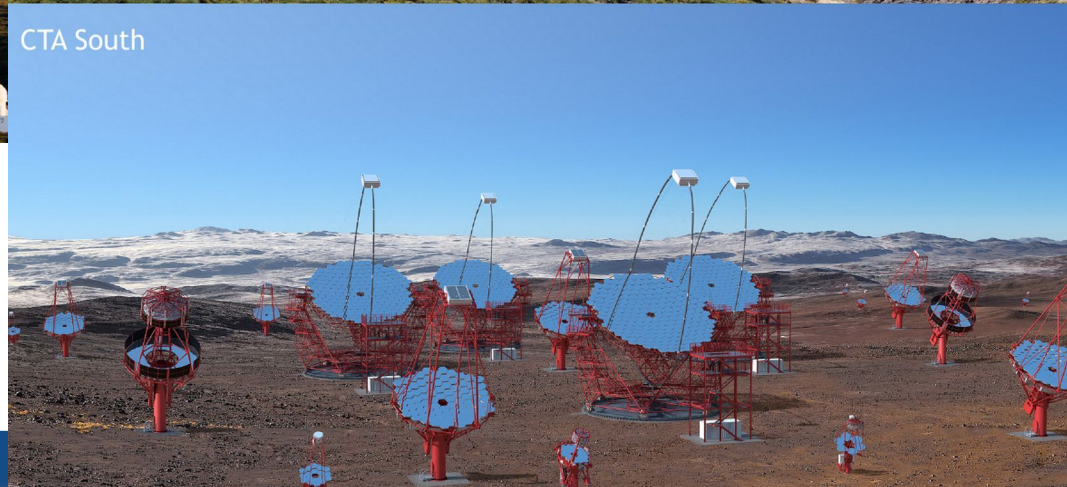
Current generation IACTs

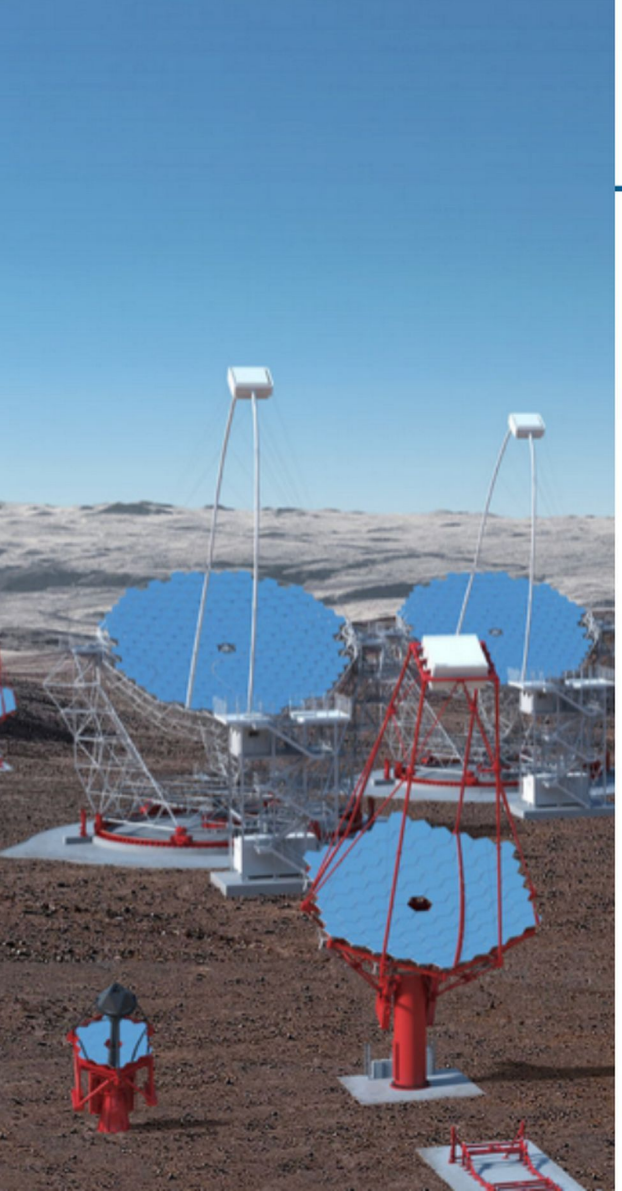
H.E.S.S.
Namibia
4 + 1 telescopes
12 m + 28 m



VERITAS
Arizona
4 telescopes
10 m

MAGIC
Canary Islands
2 telescopes
17 m





Important points for data model

Pointed instrument: **≈8° FOV**

arrays → **many telescopes, multiple sub-arrays possible**

High-Energy Astrophysics → **single photon counting**

Ground-based → **Can't point everywhere, Earth turns**

Atmosphere *part of* telescope → **impulse response varies**

- air density and aerosols
- zenith angle (atmosphere depth)
- azimuth angle (B-field)
- Night-Sky-Background light (both stars and man-made)
- ...

short duty cycle +

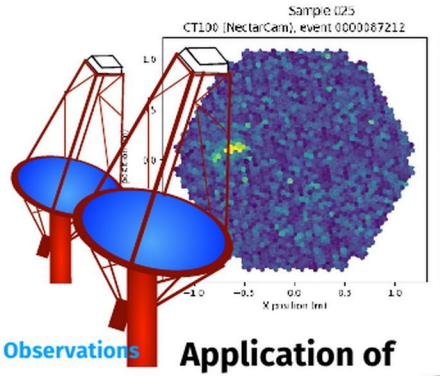
deep integration times → **many observations combined to analyze a source**

CTA Low-Level Data

Happens in CTA Observatory Data Center

Science-Ready data Products
further processed with
Science Tools (GammaPy)

DL0



Observations

Simulations

Application of Calibration + Time-Integration

Big Data

DL1

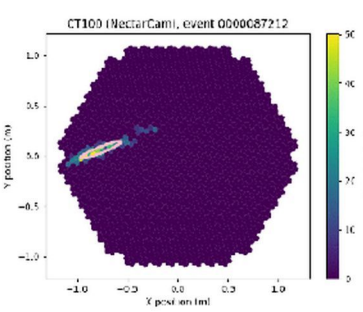
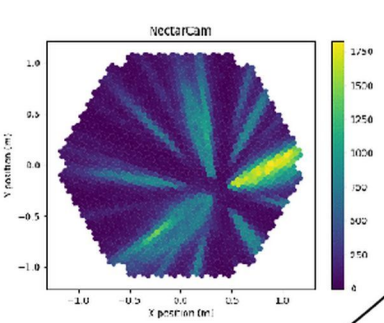
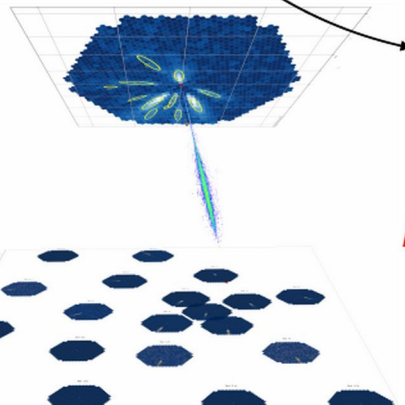


Image Processing And Feature Extraction

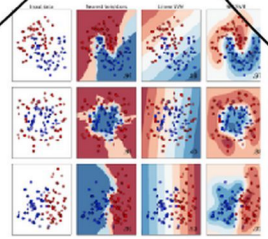
DL2

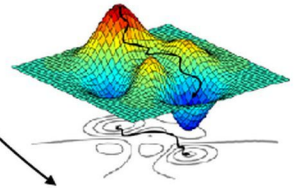


Stereo Reconstruction



Energy Reconstruction



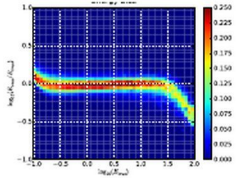


Optimization & Discrimination

Particle Classification

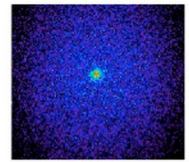
Machine Learning

DL3



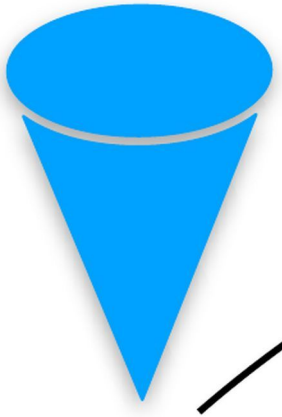
Instrumental Responses & Performance Metrics

Event Lists



Science Analysis: DL3-DL5

Happens at CTA data centers (automatic) + by users on user's laptops or e.g. ESCAPE science platform



Search for data
covering region of interest in space, time, maybe other parameters

VO tools?

DL3

Science-Ready

Retrieve Event Lists and IRFs that cover region of interest

obs_id = 84753

event_id	Time	Energy	RA	Dec
12351	55933.30	0.03	53.6	-29.2

obs_id = 99584

event_id	Time	Energy	RA	Dec
12351	55933.30	0.03	53.6	-29.2
12356	55933.31			
12378	559			
12389	559			

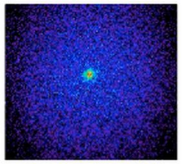
obs_id = 100202

event_id	Time	Energy	RA	Dec
12351	55933.30	0.03	53.6	-29.2
12356	55933.31	0.1	53.2	-29.3
12378	55933.36	12.0	53.8	-29.5
12389	55933.45	0.2	53.7	-28.7

DL4

Science-Binned

Use *Science Tools* to make intermediate Binned Data Products

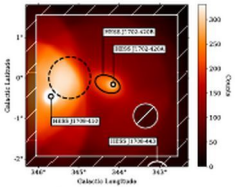


Exposure Cube
Counts Cube
Exclusion Map
...

DL5

Science-Quicklook

Use *Science Tools* to Fit models to binned data (forward-folding) and make Flux cubes



Sky model
Light-Curve
Spectra
Flux Map/Cube

DL6

Science-Observatory

Observatory or CTAC produces some legacy products

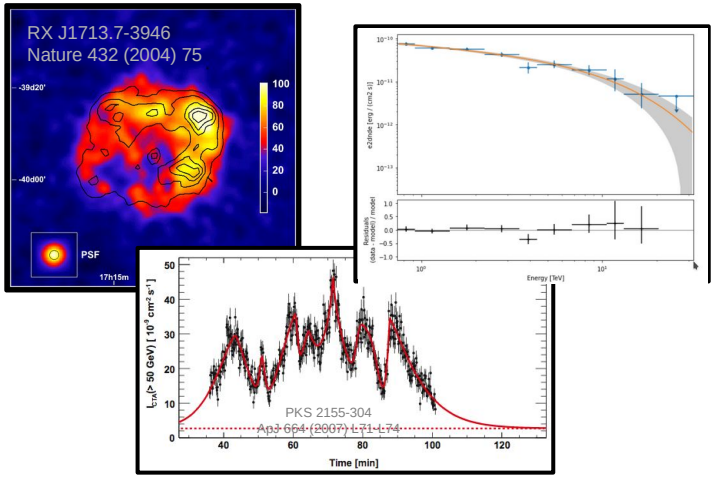
Catalogs,
Survey Products

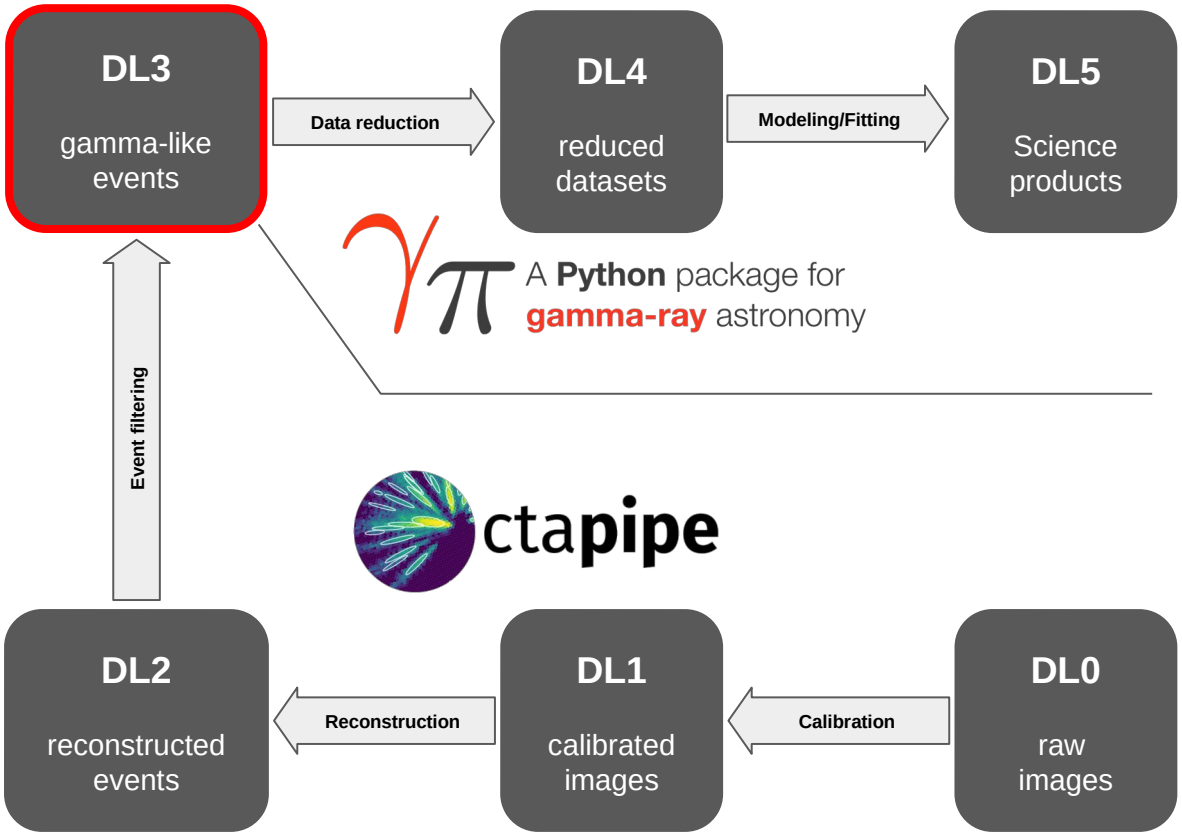
The case of imaging Cherenkov data

Data Level 3 (DL3) event lists

- TeV “candidate” photons (coordinates, time, energy)
- Low count statistics
- High background

Science products (DL5):



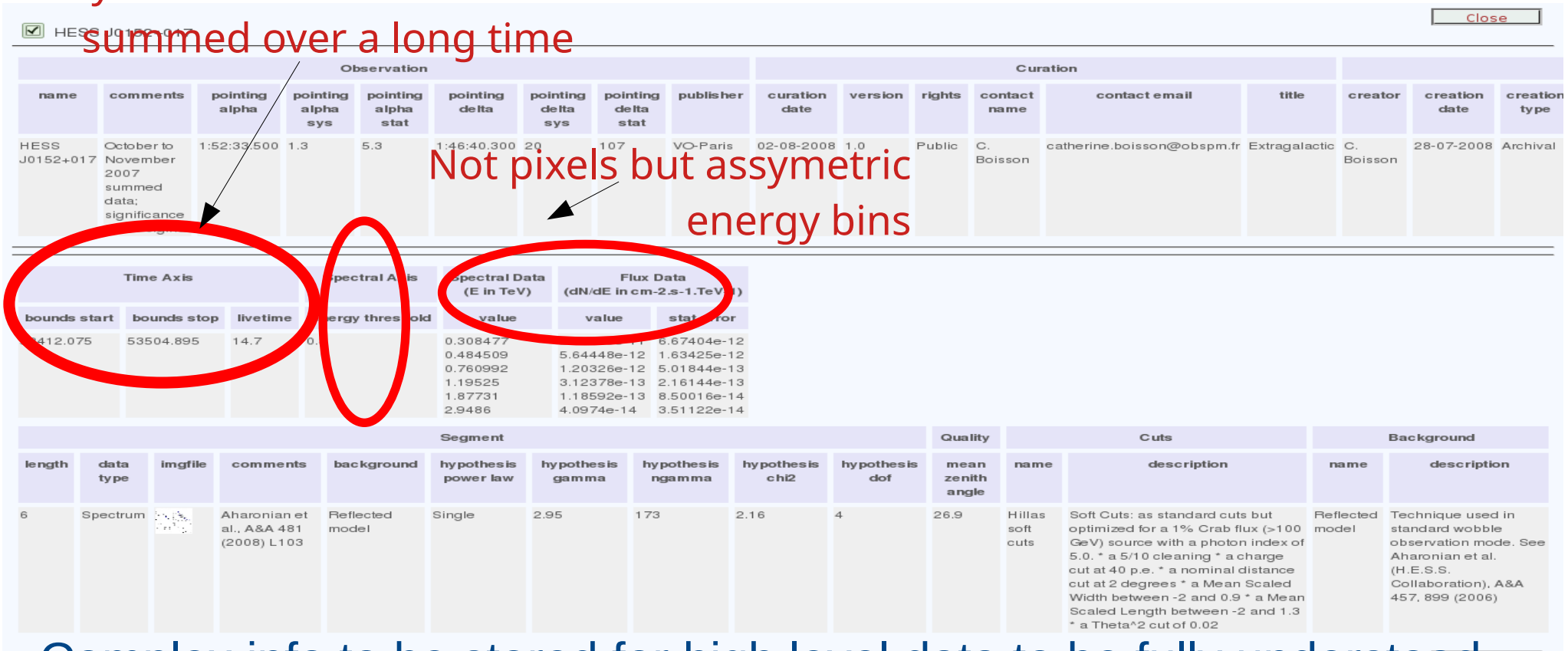


H.E.S.S. AGN - DL5 - IVOA spectral DM ?

Only a few hours of useful data

summed over a long time

Not pixels but assymetric energy bins



Close

HESS J0152-017

Observation

name	comments	pointing alpha	pointing alpha sys	pointing alpha stat	pointing delta	pointing delta sys	pointing delta stat	publisher
HESS J0152+017	October to November 2007 summed data; significance	1:52:33.500	1.3	5.3	1:46:40.300	20	107	VO-Paris

Curation

curation date	version	rights	contact name	contact email	title	creator	creation date	creation type
02-08-2008	1.0	Public	C. Boisson	catherine.boisson@obspm.fr	Extragalactic	C. Boisson	28-07-2008	Archival

Time Axis

bounds start	bounds stop	livetime
412.075	53504.895	14.7

Spectral Axis

energy threshold
0.

Spectral Data (E in TeV) / **Flux Data (dN/dE in cm-2.s-1.TeV-1)**

value	value	stat error
0.308477	1.7	6.67404e-12
0.484509	5.64448e-12	1.63425e-12
0.760992	1.20326e-12	5.01844e-13
1.19525	3.12378e-13	2.16144e-13
1.87731	1.18592e-13	8.50016e-14
2.9486	4.0974e-14	3.51122e-14

Segment

length	data type	imgfile	comments	background	hypothesis power law	hypothesis gamma	hypothesis ngamma	hypothesis chi2	hypothesis dof
6	Spectrum		Aharonian et al., A&A 481 (2008) L103	Reflected model	Single	2.95	173	2.16	4

Quality

mean zenith angle
26.9

Cuts

name	description
Hillas soft cuts	Soft Cuts: as standard cuts but optimized for a 1% Crab flux (>100 GeV) source with a photon index of 5.0. * a 5/10 cleaning * a charge cut at 40 p.e. * a nominal distance cut at 2 degrees * a Mean Scaled Width between -2 and 0.9 * a Mean Scaled Length between -2 and 1.3 * a Theta^2 cut of 0.02

Background

name	description
Reflected model	Technique used in standard wobble observation mode. See Aharonian et al. (H.E.S.S. Collaboration), A&A 457, 899 (2006)

Close

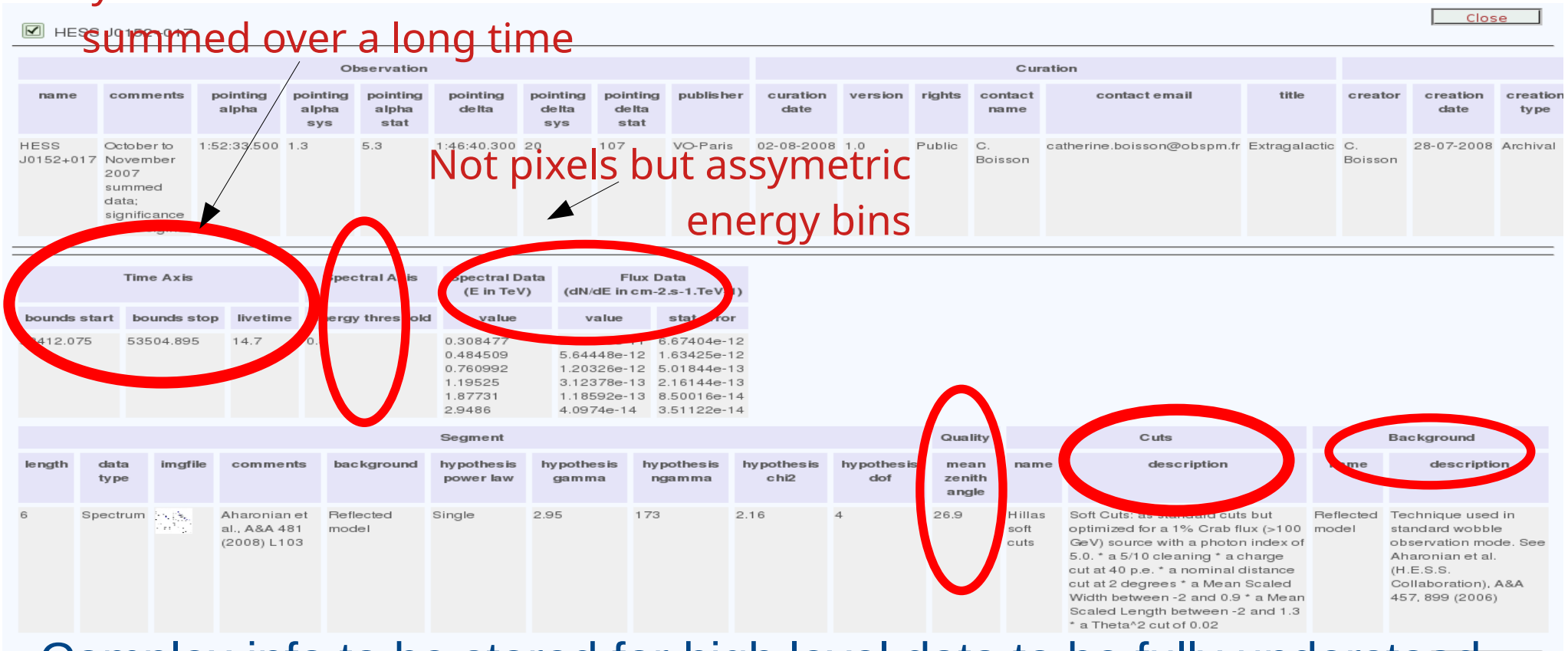
Complex info to be stored for high level data to be fully understood

H.E.S.S. AGN - DL5 - IVOA spectral DM ?

Only a few hours of useful data

summed over a long time

Not pixels but assymetric energy bins



Close

HESS J0152-017

Observation

name	comments	pointing alpha	pointing alpha sys	pointing alpha stat	pointing delta	pointing delta sys	pointing delta stat	publisher
HESS J0152+017	October to November 2007 summed data; significance	1:52:33.500	1.3	5.3	1:46:40.300	20	107	VO-Paris

Curation

curation date	version	rights	contact name	contact email	title	creator	creation date	creation type
02-08-2008	1.0	Public	C. Boisson	catherine.boisson@obspm.fr	Extragalactic	C. Boisson	28-07-2008	Archival

Time Axis

bounds start	bounds stop	livetime
1412.075	53504.895	14.7

Spectral Axis

energy threshold
0.[illegible]

Spectral Data (E in TeV) / **Flux Data (dN/dE in cm-2.s-1.TeV-1)**

value	value	stat error
0.308477	[illegible]	6.67404e-12
0.484509	5.64448e-12	1.63425e-12
0.760992	1.20326e-12	5.01844e-13
1.19525	3.12378e-13	2.16144e-13
1.87731	1.18592e-13	8.50016e-14
2.9486	4.0974e-14	3.51122e-14

Segment

length	data type	imgfile	comments	background	hypothesis power law	hypothesis gamma	hypothesis ngamma	hypothesis chi2	hypothesis dof
6	Spectrum		Aharonian et al., A&A 481 (2008) L103	Reflected model	Single	2.95	173	2.16	4

Quality

mean zenith angle
26.9

Cuts

name	description
Hillas soft cuts	Soft Cuts: as standard cuts but optimized for a 1% Crab flux (>100 GeV) source with a photon index of 5.0. * a 5/10 cleaning * a charge cut at 40 p.e. * a nominal distance cut at 2 degrees * a Mean Scaled Width between -2 and 0.9 * a Mean Scaled Length between -2 and 1.3 * a Theta^2 cut of 0.02

Background

[illegible]	description
Reflected model	Technique used in standard wobble observation mode. See Aharonian et al. (H.E.S.S. Collaboration), A&A 457, 899 (2006)

Close

Complex info to be stored for high level data to be fully understood

IVOA DMs

- No standard yet to archive high level 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 for 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
  - 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
- Not initially foreseen by IVOA
  - DL3 : event list + IRFs
- 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

What are Science-Ready Data? (DL3)

Gamma-like *Event Lists* *(≈ a photon list)*

- Particle **parameters reconstructed** from air-shower measurements
  - time of event arrival
  - reconstructed position on sky + ground
  - reconstructed energy
  - reconstructed shower h_{max} or X_{max} (optional)
  - a background class (how likely it is a gamma)
  - a reconstruction class (how well reconstructed)
- **True** particle **parameters** (if from simulations)

Instrument Monitoring Tables

Good-Time Intervals (pre-made or user)

Instrument Response Function (IRF):
& Background Model: *B*

- IRF (R) : Probability distributions that relate **Reconstructed** (instrument) to **True** (physics) parameters
  - $N_{\text{predicted}} = F_{\text{true}} \circledast \hat{\mathbf{R}} + B_{\text{predicted}}$
- **Assumptions**:
  - time-invariant per "good time interval"
  - pointing is much better than PSF (no bias)
  - we can factor ***R*** as

$$\hat{\mathbf{R}} = A_{\text{eff}}(E) \cdot \hat{E}_{\text{mig}}(E, E') \cdot \text{PSF}(p, p')$$

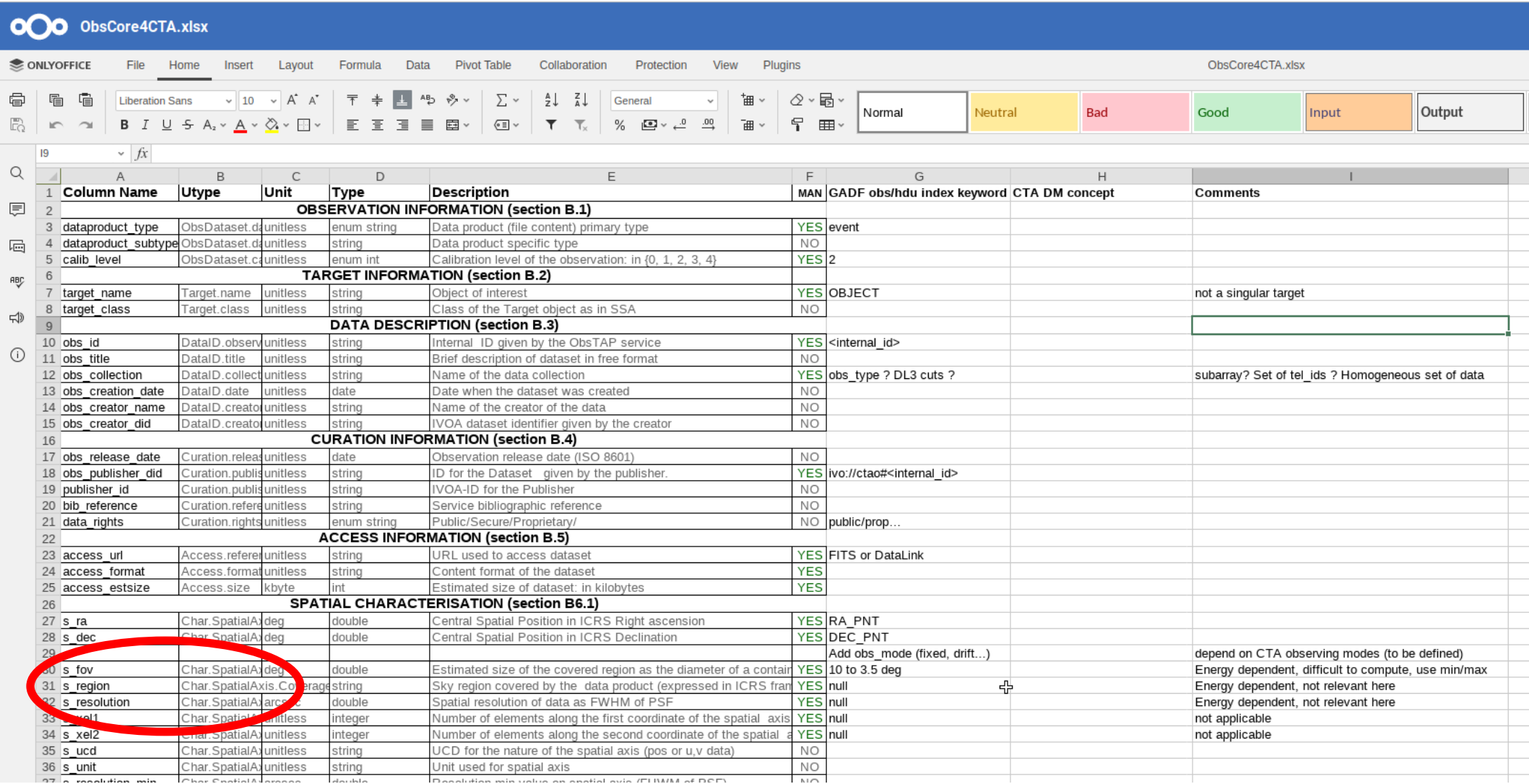
effective collection area — Energy Migration Matrix (resolution & bias) — Point-Spread Function

DL3 metadata and IVOA standard data models

Explore existing metadata descriptions to map to ObsCore metadata
Identify data descriptions uncovered by IVOA data models at the time
—> proposition for updates

- **DL3 files in FITS format**
  - EVENTS frame (time, sky position, energy)
  - Additional frames
    - TELARRAY : list of telescopes
    - GTI : Good Time Intervals (list of start and stop)
    - POINTING : time stamp with pointing position in different coordinate systems
  - Some metadata
    - obs_id
    - Telescope type list
    - Start, stop times
    - mean zenithal angle, pointing

- **IRF files**
  - From simulations
    - Aeff : effective area
    - Edisp : energy dispersion
    - Psf : point spread function after reconstruction (variation function of t, E, position across image)
  - From observations
    - bkg estimation of the hadronic background noise
- Index of observations
  - Collections are stored on disk for gammapy DataStore
  - obs-index et hdu-index tables per data subset
  - Contain characteristics for requests and data selection



	Column Name	Utype	Unit	Type	Description	MAN	GADF obs/hdu index keyword	CTA DM concept	Comments
2	**OBSERVATION INFORMATION (section B.1)**								
3	dataproduct_type	ObsDataset.da	unitless	enum string	Data product (file content) primary type	YES	event		
4	dataproduct_subtype	ObsDataset.da	unitless	string	Data product specific type	NO			
5	calib_level	ObsDataset.ca	unitless	enum int	Calibration level of the observation: in {0, 1, 2, 3, 4}	YES	2		
6	**TARGET INFORMATION (section B.2)**								
7	target_name	Target.name	unitless	string	Object of interest	YES	OBJECT		not a singular target
8	target_class	Target.class	unitless	string	Class of the Target object as in SSA	NO			
9	**DATA DESCRIPTION (section B.3)**								
10	obs_id	DataID.observ	unitless	string	Internal ID given by the ObsTAP service	YES	<internal_id>		
11	obs_title	DataID.title	unitless	string	Brief description of dataset in free format	NO			
12	obs_collection	DataID.collect	unitless	string	Name of the data collection	YES	obs_type ? DL3 cuts ?		subarray? Set of tel_ids ? Homogeneous set of data
13	obs_creation_date	DataID.date	unitless	date	Date when the dataset was created	NO			
14	obs_creator_name	DataID.creato	unitless	string	Name of the creator of the data	NO			
15	obs_creator_did	DataID.creato	unitless	string	IVOA dataset identifier given by the creator	NO			
16	**CURATION INFORMATION (section B.4)**								
17	obs_release_date	Curation.releas	unitless	date	Observation release date (ISO 8601)	NO			
18	obs_publisher_did	Curation.publis	unitless	string	ID for the Dataset given by the publisher.	YES	ivo://ctao#<internal_id>		
19	publisher_id	Curation.publis	unitless	string	IVOA-ID for the Publisher	NO			
20	bib_reference	Curation.refere	unitless	string	Service bibliographic reference	NO			
21	data_rights	Curation.rights	unitless	enum string	Public/Secure/Proprietary/	NO	public/prop…		
22	**ACCESS INFORMATION (section B.5)**								
23	access_url	Access.referer	unitless	string	URL used to access dataset	YES	FITS or DataLink		
24	access_format	Access.format	unitless	string	Content format of the dataset	YES			
25	access_estsize	Access.size	kbyte	int	Estimated size of dataset: in kilobytes	YES			
26	**SPATIAL CHARACTERISATION (section B6.1)**								
27	s_ra	Char.SpatialAx	deg	double	Central Spatial Position in ICRS Right ascension	YES	RA_PNT		
28	s_dec	Char.SpatialAx	deg	double	Central Spatial Position in ICRS Declination	YES	DEC_PNT		
29							Add obs_mode (fixed, drift…)		depend on CTA observing modes (to be defined)
30	s_fov	Char.SpatialAx	deg	double	Estimated size of the covered region as the diameter of a contain	YES	10 to 3.5 deg		Energy dependent, difficult to compute, use min/max
31	s_region	Char.SpatialAxis.Co	verage	string	Sky region covered by the data product (expressed in ICRS fram	YES	null		Energy dependent, not relevant here
32	s_resolution	Char.SpatialAx	arcsec	double	Spatial resolution of data as FWHM of PSF	YES	null		Energy dependent, not relevant here
33	s_xel1	Char.SpatialAx	unitless	integer	Number of elements along the first coordinate of the spatial axis	YES	null		not applicable
34	s_xel2	Char.SpatialAx	unitless	integer	Number of elements along the second coordinate of the spatial a	YES	null		not applicable
35	s_ucd	Char.SpatialAx	unitless	string	UCD for the nature of the spatial axis (pos or u,v data)	NO			
36	s_unit	Char.SpatialAx	unitless	string	Unit used for spatial axis	NO			

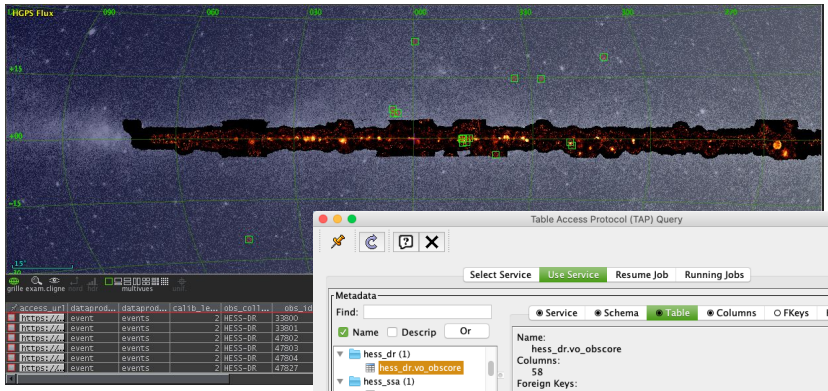
VO access to H.E.S.S. public data

Virtual Observatory (VO) standards tailored to make data findable:

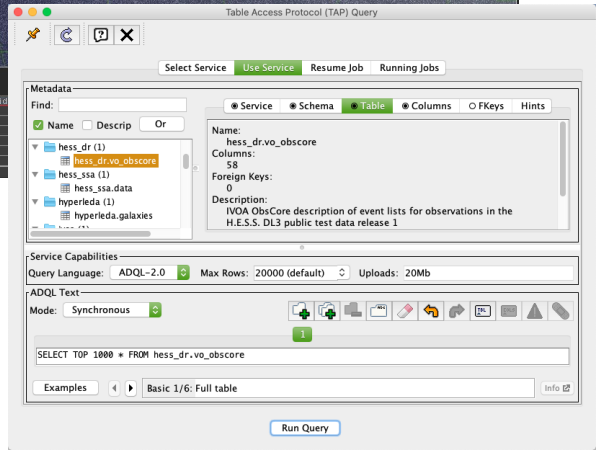
- H.E.S.S. experiment first public data release
  → https://zenodo.org/record/1421099
- **VO ObsTAP Service** implemented at the Observatoire de Paris:
  - **IVOA** Observation Data Model Core Components (**Obs**Core) [link to IVOA REC]
    → adapted to Cherenkov data
  - **IVOA** Table Access Protocole (**TAP**) [link to IV
  - Registered to the **VO Registry** via PADC (Paris Astronomical Data Centre)
- Data widely findable
  - e.g. Aladin, TOPCAT, TapHandle, PyVO…
  - + dedicated web pages

https://hess-dr.obspm.fr

TapHandle



Aladin



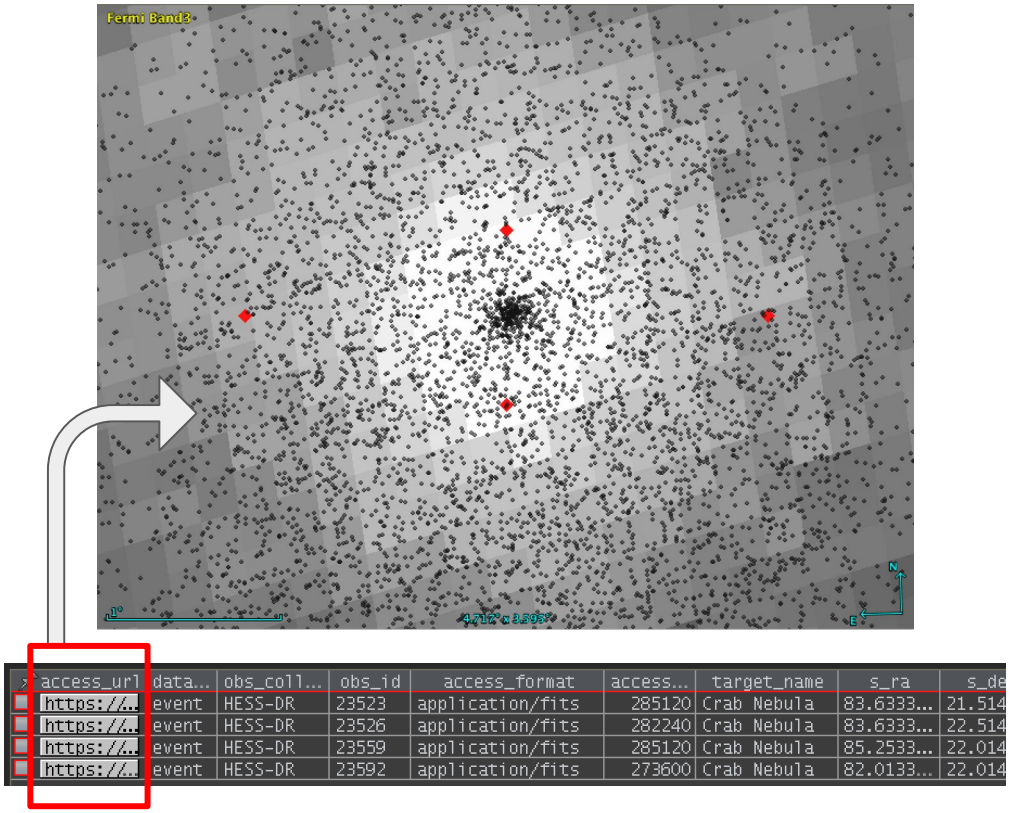
TOPCAT

Obtaining the data

- ObsCore **`access_url`**
  - Direct download link to the FITS file
  - IVOA DataLink (to be implemented):
    - Associated calibration data
    - Service that packages the requested **collection** of data + calibration for analysis
- Access rights
  - Public data: no restrictions
  - Anticipating need for permissions:
    - PI proprietary period
    - Federation authentication
    - e.g. IAM ESCAPE service

      → https://indigo-iam.github.io/escape-docs

DL3 event list directly opened in Aladin
(each black dot is an event)



Common data format and FAIR software

https://gamma-astro-data-formats.readthedocs.io
A community initiative to define common DL3 data formats for gamma-ray astronomy based on FITS

- Work and discussions in progress
- Includes formats for: event lists, effective area, energy resolution, point spread function, instrumental background
- A prototype data format for CTA
- Used for the H.E.S.S. public data release
- Partially in use by current instruments:
  Fermi-LAT, HESS, VERITAS, MAGIC, FACT, ...

[A&A 625, A10, 2019]
[A&A 632, A72, 2019]
[A&A 632, A102, 2019]

+

A **Python** package for **gamma-ray** astronomy

→ [ADASS XXX poster]
→ https://gammapy.org

- Open-source Python package
  (Astropy affiliated package)
- Core library for the Science Tools of CTA
- Used in the analysis of existing gamma-ray instruments, such as H.E.S.S., MAGIC, VERITAS, HAWC...

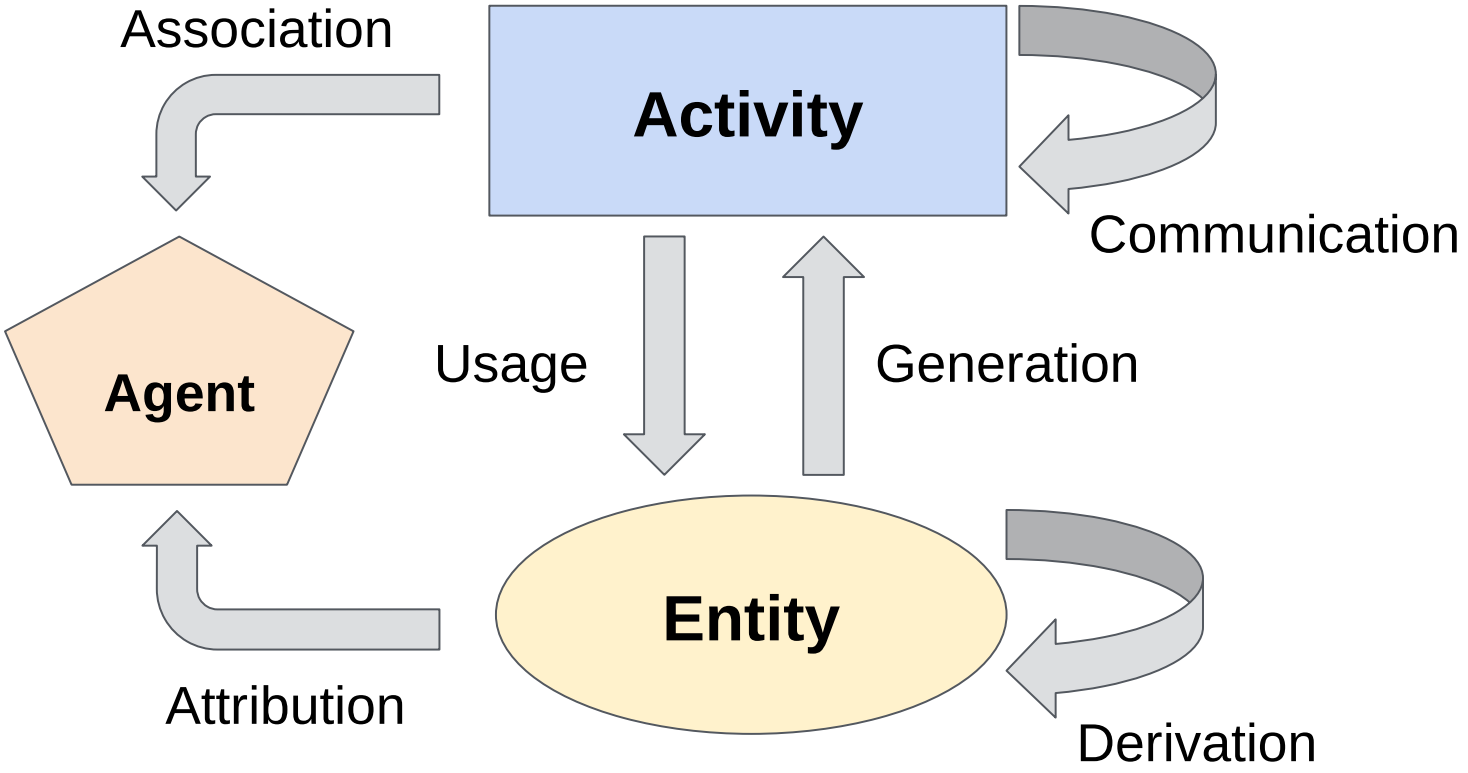
- **FAIR4RS:** FAIR Principles for Research Software
  → https://doi.org/10.15497/RDA00065

- **ESCAPE OSSR**: open-access repository to share scientific software and services
  → https://escape2020.pages.in2p3.fr/wp3/ossr-pages
  → [ADASS XXX poster]

Provenance

- Complexity in the detection process
  - need structured and detailed provenance information
  - Provenance information of data products are necessary to the user to perform a correct scientific analysis
  - capture of relevant provenance along the processing/analysis workflow
- Provenance concepts are relevant for different aspects of CTA
  - Data diffusion: the data products have to contain all the relevant context information with the assumptions made as well as a description of the methods and algorithms used during the data processing.
  - Pipeline : the CTA Observatory must ensure that data processing is traceable and reproducible.
  - Instrument Configuration : the characteristics of the instrument at a given time have to be available and traceable (hardware changes, measurements of e.g. a reflectivity curve of a mirror, …)
- To be studied :
  - chaining, concatenation provenance: DL0 -> DL5

Provenance glossary : designed around 3 poles



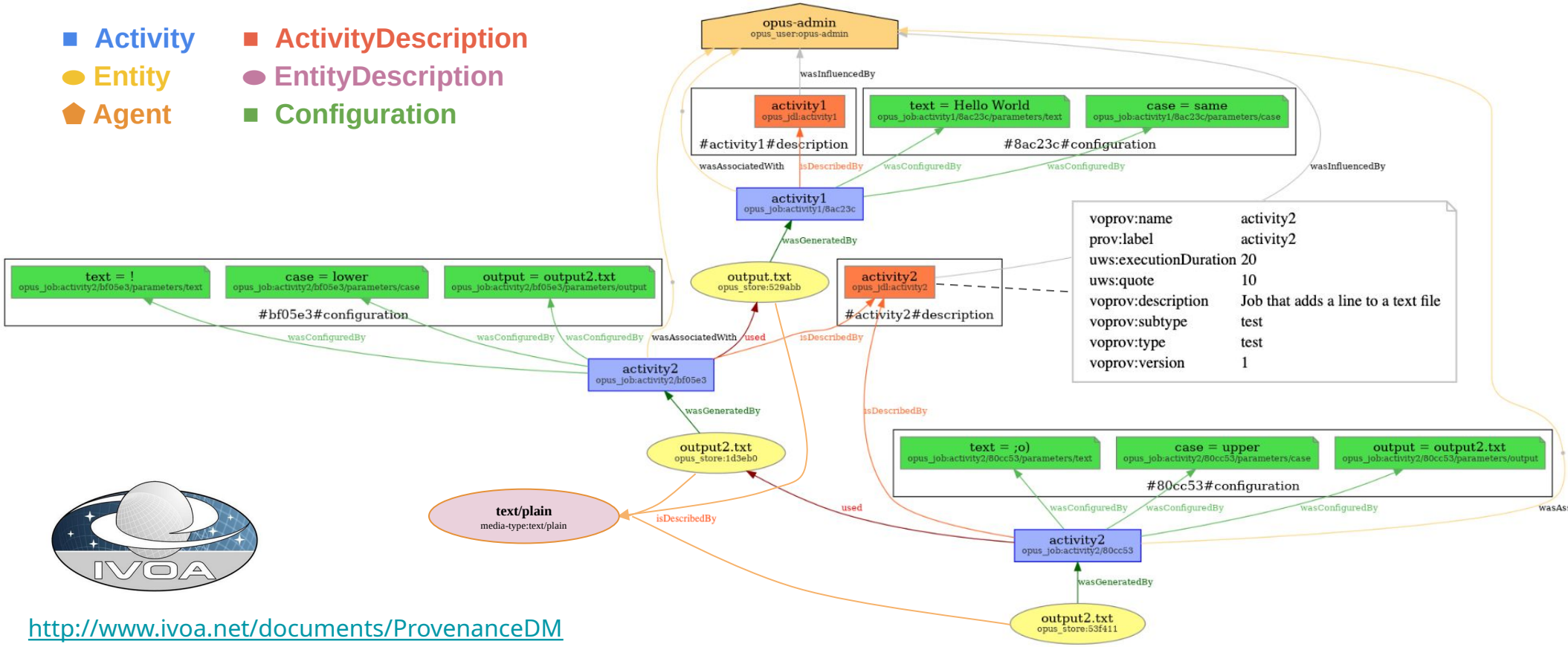
Word Wide Web Consortium

http://www.w3.org/TR/prov-overview

Prov Speaking: we want to describe the activities leading to the entities

Full IVOA Provenance graph

- Activity
- Entity
- Agent
- ActivityDescription
- EntityDescription
- Configuration



http://www.ivoa.net/documents/ProvenanceDM

Some terminology

- **full provenance**: graph/tree/chain that **traces** activities and entities up to the raw data. This information is not hosted by the entities themselves, it should be stored in a central database, or as separate files.
- **end-user/specific "provenance"**: can be embedded into an entity, keywords or data that provides project specific **key information to use/analyse** the entity (e.g. for CTA: event class/type, telescope configuration, sky conditions, reco method...)
- **last-step provenance**: embedded into an entity as a list of keywords that gives some context and info on **last activity** (general workflow, software, versions, contact…), including the list of generated and used entity ids, so that a full provenance may be reconstructed from this minimum provenance.

See ADASS XXX BoF proceedings : https://arxiv.org/abs/2101.08691
ESCAPE workshop on provenance : https://indico.in2p3.fr/event/21913/page/2641-summary

A provenance management system

Science portal for FAIR Cherenkov data

- Making data FAIR requires discussions and anticipation
  - VO compliance as a requirement
  - Capture of relevant metadata along the processing
- Towards a Science portal
  - Advanced search dedicated to Cherenkov data
  - Data preview and selection
  - Online post processing (with provenance tracing!)
  - User management and space
- Test implementations
  - TAP Distiller → https://voparis-cta-test.obspm.fr
  - OPUS → https://voparis-uws-test.obspm.fr
    → [ADASS XXX proceedings]
  - ESCAPE ESAP (data + software + resource)
    → [ADASS XXXI poster X0-010]