

# IACT DL3 MEETING INTRODUCTION

---

*Christoph Deil, MPIK  
March 20, 2017*

# WELCOME!

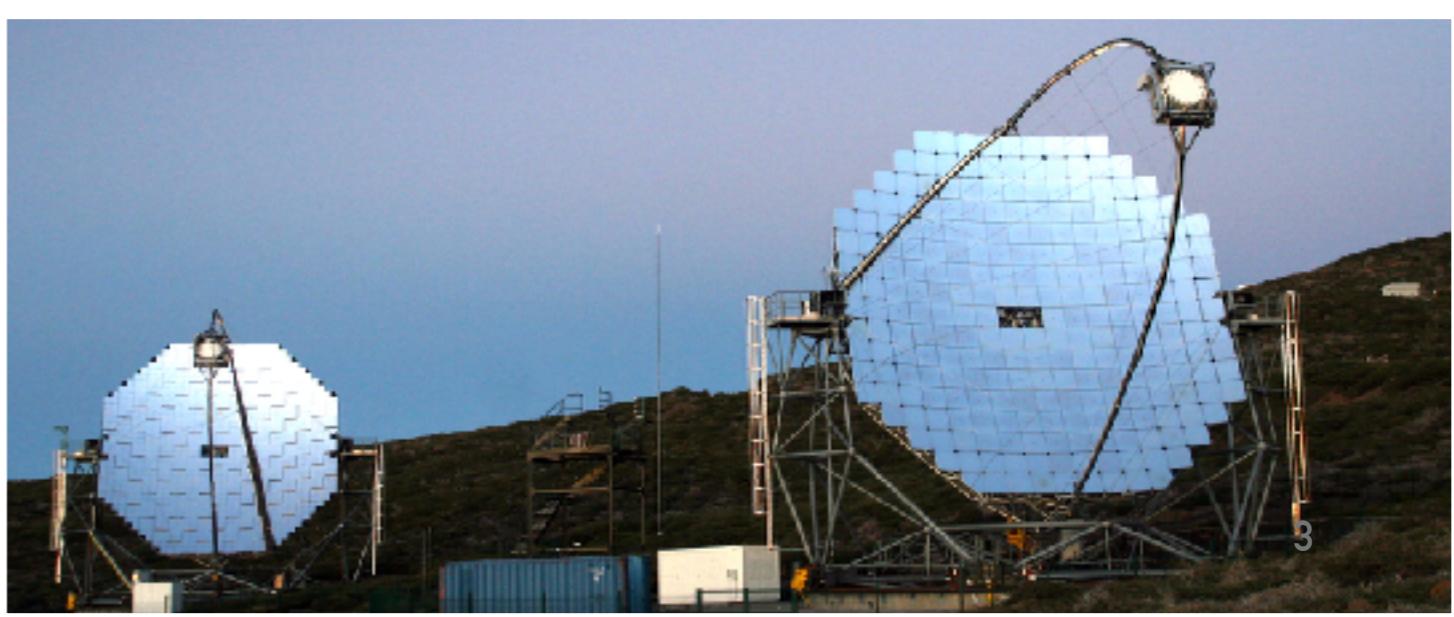
---

- Infos & slides: [https://github.com/open-gamma-ray-astro/  
2017-03\\_IACT\\_DL3\\_Meeting](https://github.com/open-gamma-ray-astro/2017-03_IACT_DL3_Meeting)
- Agenda
  - Today: presentation & discussion with eZuce connection  
(see list of presentations on webpage)
  - Rest of the week: free time to discuss work together
- Logistics
  - Will be in this room all week.
  - WIFI: eduroam or guest account
  - Dinner together?  
*(for me only Wednesday & Thursday possible)*

# DL3 CONTEXT

---

- Before & now: IACTs are experiments with private data and software, limited collaboration or joint analyses or re-use.



# DL3 CONTEXT

---

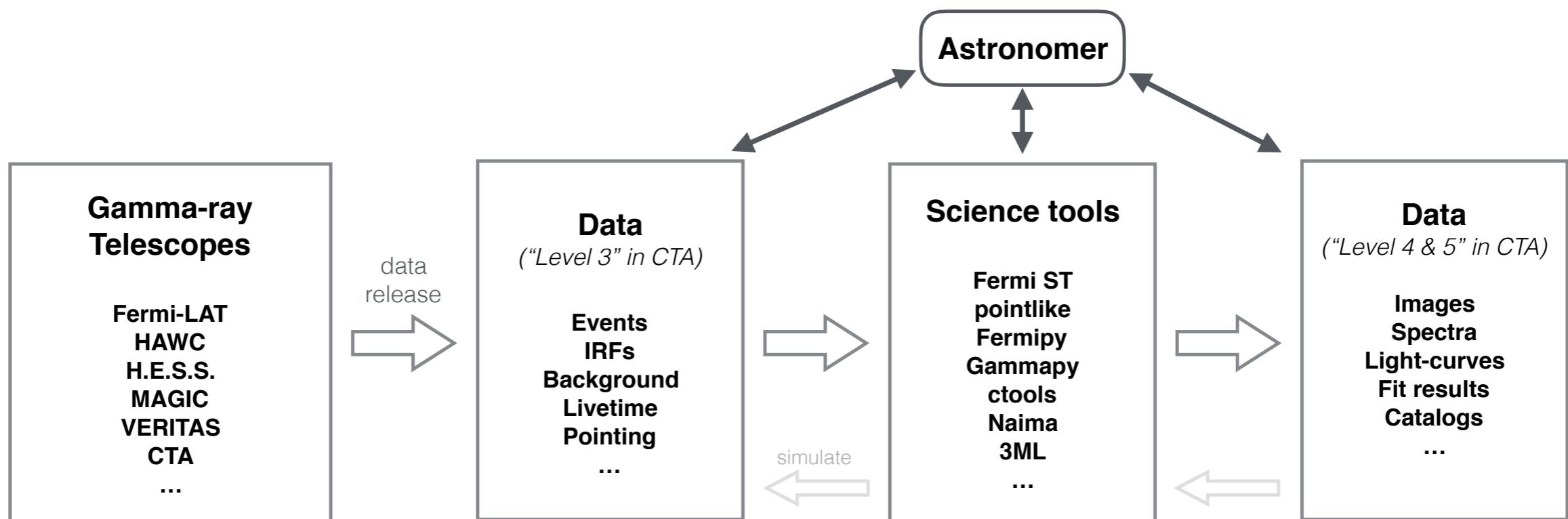
- Soon: CTA is coming as open observatory with open high-level data and analysis software
- Fermi-LAT and other space telescopes already open data. Some re-use of formats and code across missions.



# WHAT IS DL3?

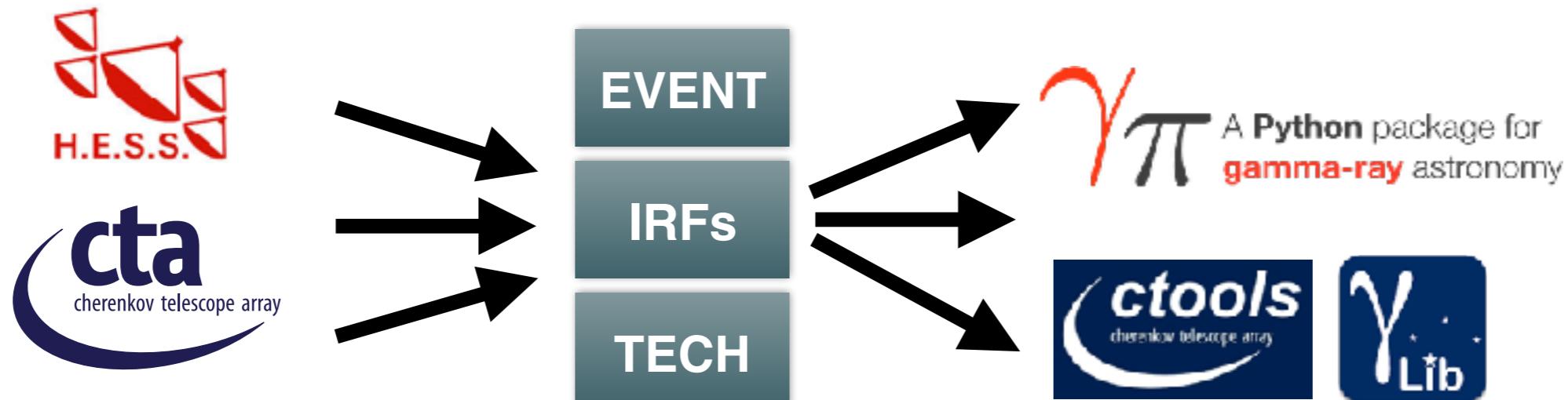
---

- DL3 is
  - an interface between pipeline & science tools
  - the files given to astronomers for science analyses

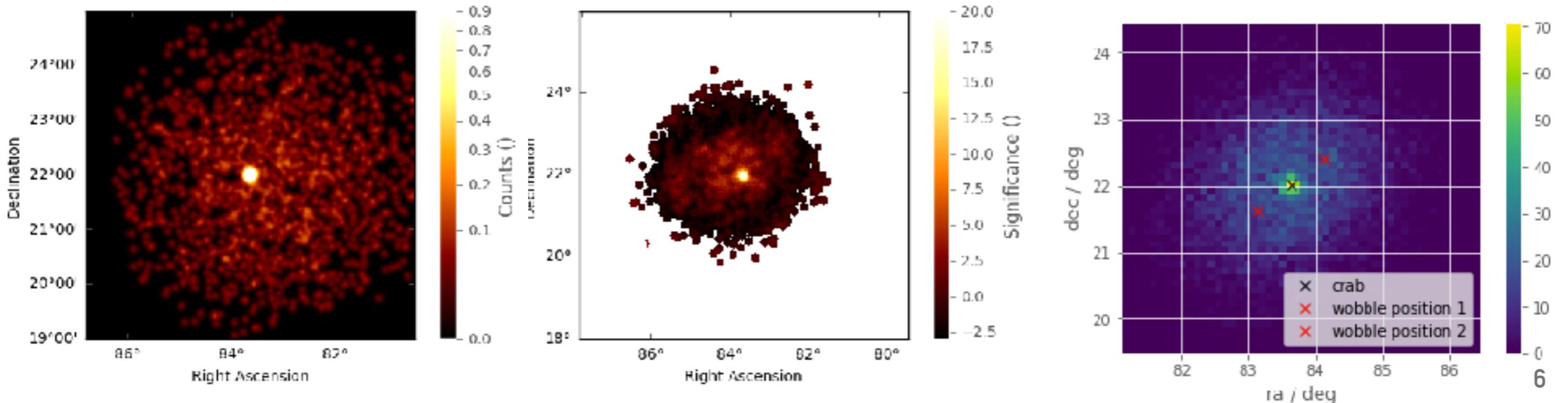


# A COMMON DL3 DATA MODEL & FORMAT CAN BE POWERFUL

Common DL3 can lead to collaboration & common science tools for different telescopes



Example: Crab nebula sky image with different IACT instruments with Gammapy



# A COMMON DL3 DATA MODEL AND FORMATS IS HARD TO ACHIEVE

---

- In X-ray astronomy there was the OGIP group that did this work to develop FITS recommendations and those were used by many missions. This was a deliberate and joint effort.
- To me it's unclear if we can achieve the same for gamma-ray astronomy now and in the coming years.
- If yes, what can we do to nourish this DL3 spec effort?
- If no, what's the real issue?
  - Telescopes / data (even from IACTs) are too different (slewing telescopes like Fermi-LAT and HAWC vs IACTs)
  - Timescales / manpower / interest to work on DL3 are different for different telescopes and science tools

- open-gamma-ray-astro effort started after discussions at the [Python for gamma-ray astronomy workshop](#) (MPIK, Heidelberg, November 2015)
- Mailing list set up by Jeremy Perkins (Fermi-LAT) [lists.nasa.gov/mailman/listinfo/open-gamma-ray-astro](https://lists.nasa.gov/mailman/listinfo/open-gamma-ray-astro)
  - *"This group is organized for the discussion of software and data formats for the gamma-ray astronomy community. If you are interested in open and common data and software formats for space- and ground-based instruments you are encouraged to join."*
  - 80 people, many major gamma-ray telescopes represented
  - Open data specifications ([Github repo](#), [docs](#))
  - F2F May 2016 in Meudon on DL3 with 17 people ([Github repo](#))
  - Gamma 2016 poster & proceeding ([arXiv:1610.01884](#))

Let's look at these links in the web browser together now,  
to give an overview of the current status.

# ACTIVITY / PROGRESS ON DL3

---

- Overall the DL3 effort is slow, very little progress since  $\sim 1$  year, much fewer people at this f2f meeting compared to last year.
- For current telescopes:
  - Little activity in HESS, MAGIC, VERITAS, FACT  
No effort in HAWC yet as far as I know to export to FITS
  - Status: not a single paper based on FITS events yet.  
Exporting / archiving to FITS DL3 not considered a priority!?
- For CTA:
  - Started too early?  
*observation modes and many other things unclear, ctapipe not producing responses yet; CTA needs prototyping phase of a few years to develop something good? Contrasting point of view: ideally data model and formats are fixed before data taking start, but maybe not possible for CTA?*
  - Open, joint effort with existing IACTs not the right approach?  
*DL3 is an interface, so for any addition and change there's several people that have to agree and implement support for the addition or change. Unclear who can make decisions on DL3 for CTA.*
  - Status: so far CTA is producing ROOT responses, see  
<https://www.cta-observatory.org/science/cta-performance/>

# GOALS FOR THIS WEEK

---

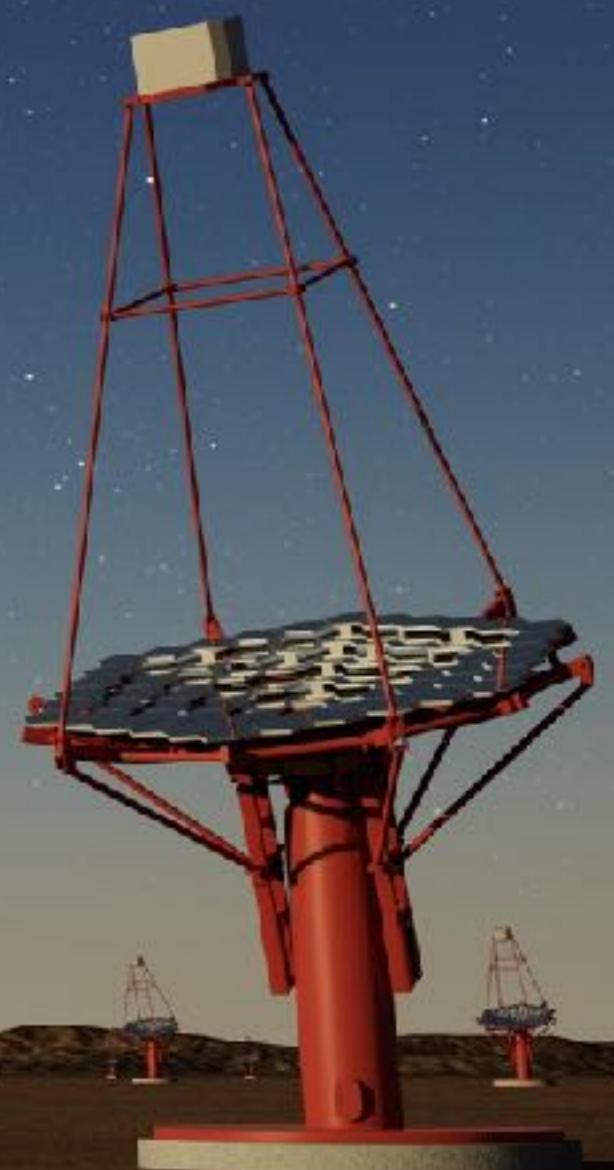
- What should we focus on and try to achieve this week?
- Small, short-term:
  - Check what we have in HESS, MAGIC & CTA and make sure that all current use cases are supported and files are compatible?
  - Review existing spec and open issues (cleanup and finalise a first version)?
  - Add support for point-like response in the formats
  - Need better support for safe energy threshold or OK now?
  - Work on HESS or MAGIC or CTA scripts to convert to DL3?
  - Finalise HESS test data release checks and document?
- Large, long-term:
  - EVENT-IRF link, event types, IRF factorisation, IRF errors, ...
  - Does it make sense to work on such big extensions or a full DL3 data model and format for CTA now? Or should we leave this for later when there's more interest / need from the CTA side?

# **BACKUP SLIDES**

**(COPY OF SLIDES FROM BOLOGNA CTA MEETING)**

# Data level 3

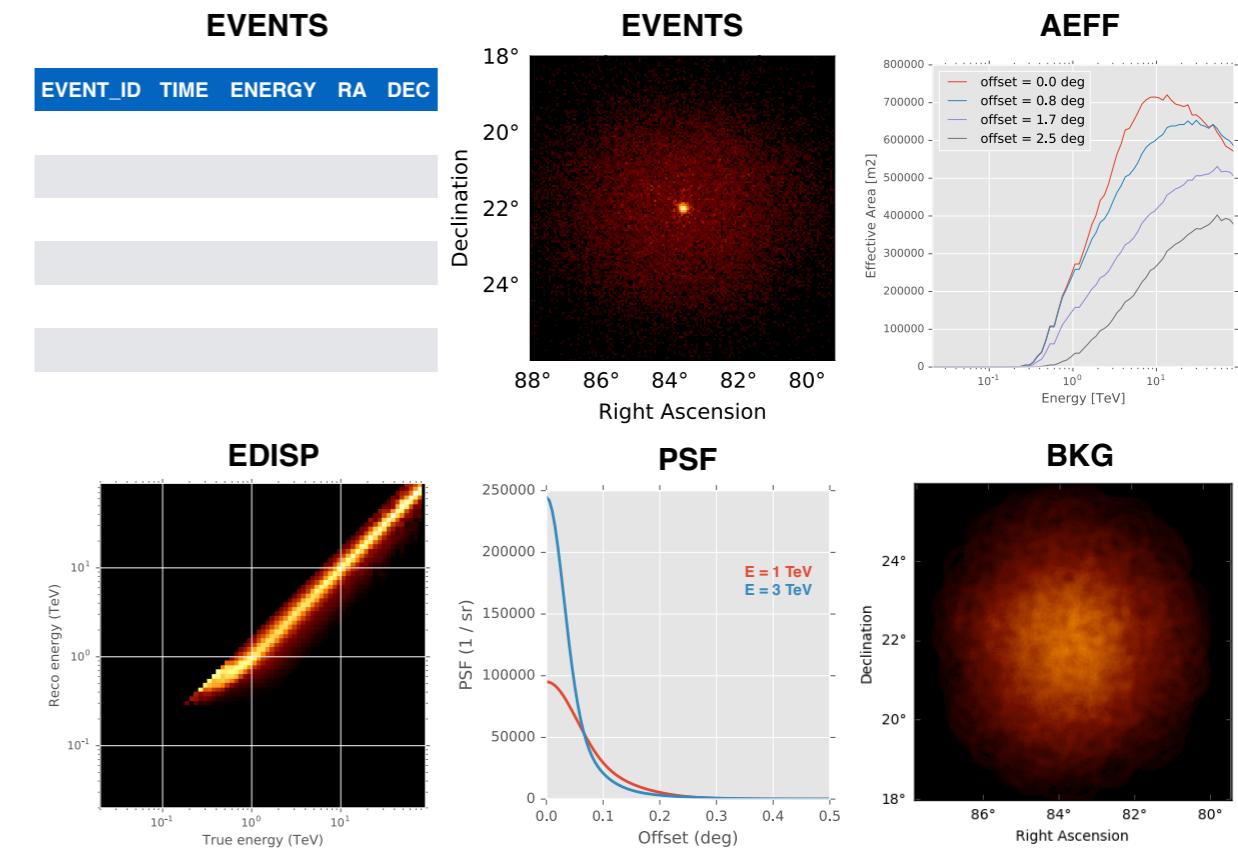
What will be released to “end users” by CTA  
Formats for events, IRFs, ...



# What we have now ...



- FITS BINTABLE formats for
  - events, gti
  - aeff\_2d
  - edisp\_2d
  - psf\_table, psf\_gtpsf  
psf\_3gauss, psf\_king
  - bkg\_2d, bkg\_3d
  - pointing (“TECH”, not used yet)
- IRFs use the multi-dim array format that stores the data and axis binning in a FITS BINTABLE
- Currently IRFs only have axes ENERGY, OFFSET, X, Y.  
not ALTITUDE, AZIMUTH, RA, DEC, ARRAY\_CONFIG, TIME, ...

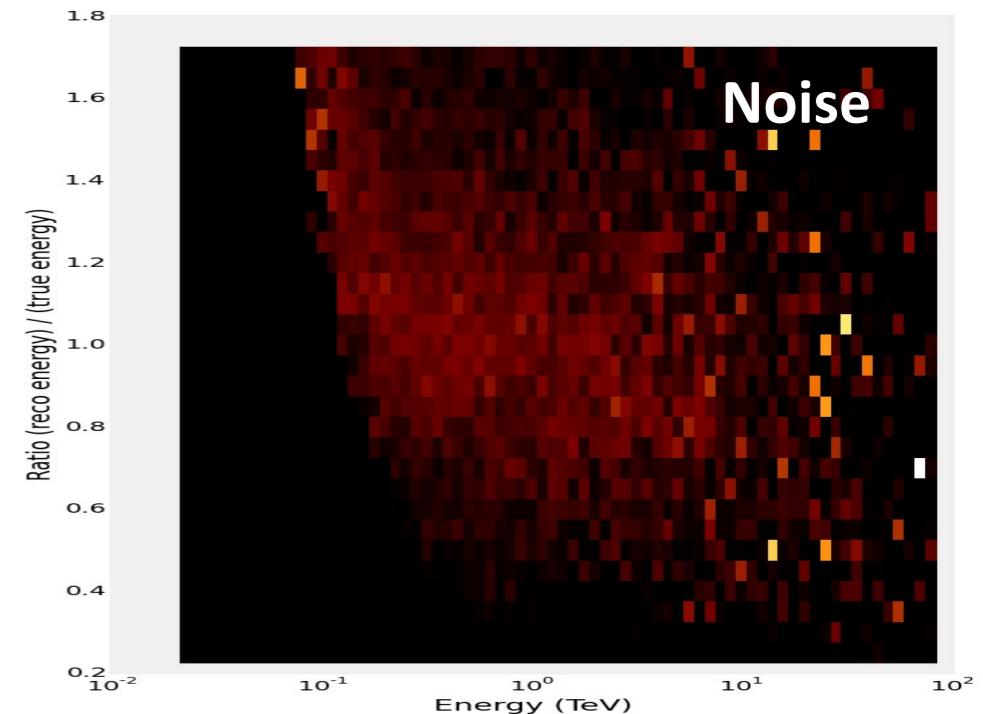


# Instrument response functions

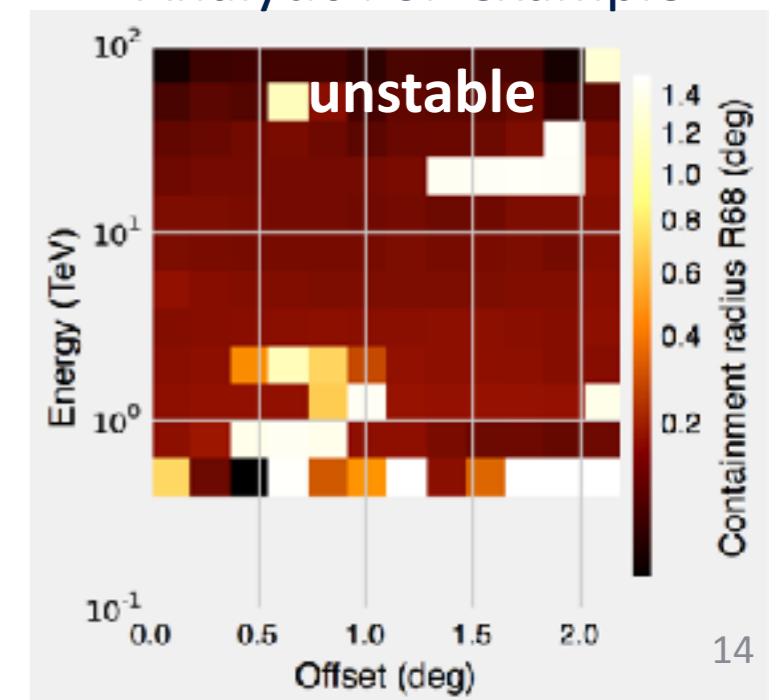


- EVENTS is the easy part
- IRFs and TECH are the hard part
- For IRFs, we currently mostly use n-dim histograms, and two analytic models (psf\_3gauss, psf\_king)
- It's not clear how to best produce and store IRFs for CTA, and how large IRFs will get compared to EVENTS
- Do we need analytic IRFs in DL3?  
*Or can we just use histogram IRFs to have simpler DL3 formats and science tools?*  
*(Can sample any analytic shape to a histogram)*
- How to communicate “IRF safe parameter range” and “IRF error” information in DL3?

Histogram EDISP example



Analytic PSF example

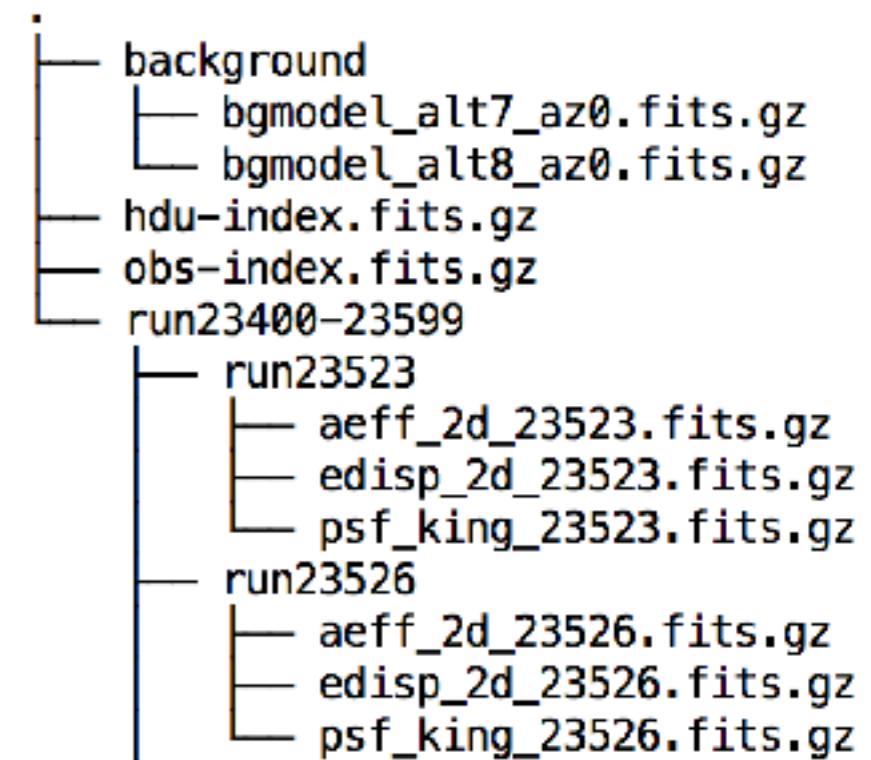
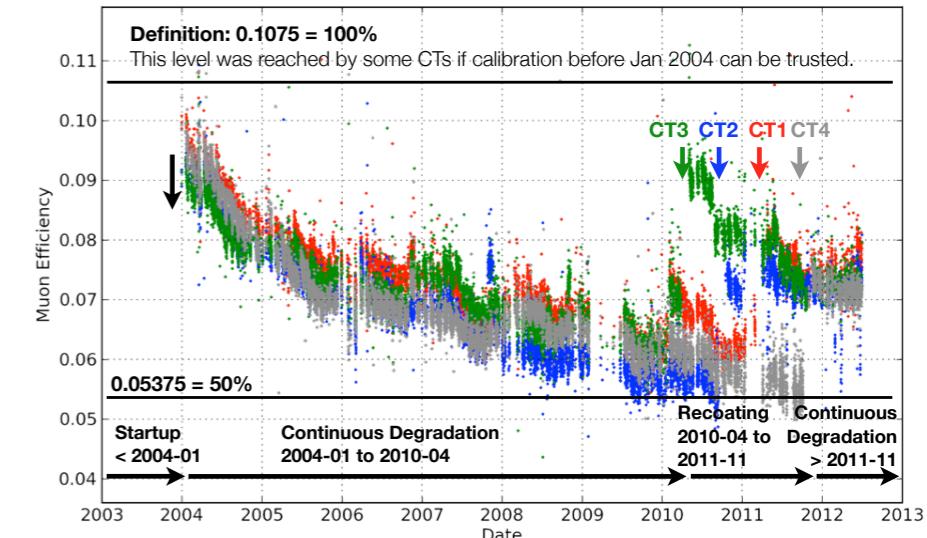


# Per-observation datasets



- Fermi-LAT ships IRFs valid for all data with ST. Instrument is stable, IRFs are small.
- **IRF DL3 database too large for IACTs.**  
Instrument not as stable. Even for HESS, very large number of IRFs and per-obs IRFs being developed now. For CTA, high IRF precision requirements and number of array configs.
- **One EVENT and set of IRF HDUs per “observation”, during which the IRFs are assumed to be stable**
  - H.E.S.S. – “observation” = 28 min
  - VERITAS – “observation” = 5 min
  - CTA – ???
- Exception: background models at the moment not per-observation

Muon Efficiency 2004 – 2012  
CT1, CT2, CT3, CT4 (HD calibration, run-by-run)



# Event – IRF linking



- How to link EVENT and IRFs data is an open question
  - Use OBS\_ID or TIME or time interval?
  - Point from EVENT header to other HDUs?
  - Require grouping in single FITS file and certain HDU names?
  - Use index files declaring the links?
- **An open question, no agreed solution!**
- Currently spec defines HDU and observation index tables, used by H.E.S.S. and supported by Gammapy & ctools
- ctools also uses an observation list XML format, not proposed as part of DL3 data releases (yet?)
- Also – we need different IRFs for each event class and type

Column Name	Description
OBS_ID	Observation ID (a.k.a. run number)
HDU_TYPE	HDU type (see below)
HDU_CLASS	HDU class (see below)
FILE_DIR	Directory of file (rel. to this file)
FILE_NAME	Name of file
HDU_NAME	Name of HDU in file

```
<observation_list title="observation library">
  <observation name="Crab" id="00001" instrument="CTA">
    <parameter name="EventList" file="events.fits"/>
    <parameter name="EffectiveArea" file="$CALDB/data/ct"
    <parameter name="PointSpreadFunction" file="$CALDB/data/ct"
    <parameter name="EnergyDispersion" file="$CALDB/data/ct"
    <parameter name="Background" file="$CALDB/data/ct"
  </observation>
</observation_list>
```

# Event classes



- “Event classes” for different reconstructions and “cuts”, usually optimised for different science use cases
- Fermi-LAT has events of different class in one EVENT list. Science tools / users can select subsets using a bitfield.
- **For current IACTs and CTA, the idea is to have completely separate sets of files for each event class.**
- *In H.E.S.S. we have a JSON file on the data server that declares which data is available. But if we keep separate files for event classes, we don't really need a format spec. Data providers document what's available and users decide what to download.*

Fermi-LAT event classes

Class Name	Class Bit	Fiducial Selection	Description
TRANSIENT020	4	CalTkr	2x P6 IGRB
TRANSIENT010	6	CalTkr	1x P6 IGRB
SOURCE	7	CalTkr	Point-Source Analysis
CLEAN	8	CalTkr	Diffuse Analysis

H.E.S.S. event classes (Crab 2006 paper)

Configuration	MRSL Min.	MRSL Max.	MRSW Min.	MRSW Max.	$\theta_{\text{cut}}^2$ Max. (degrees <sup>2</sup> )	Image Amp Min.
Standard	-2.0	2.0	-2.0	0.9	0.0125	80
	-2.0	2.0	-2.0	0.7	0.01	200
	-2.0	2.0	-2.0	1.2	0.04	40
	-2.0	2.0	-2.0	0.9	0.16	80

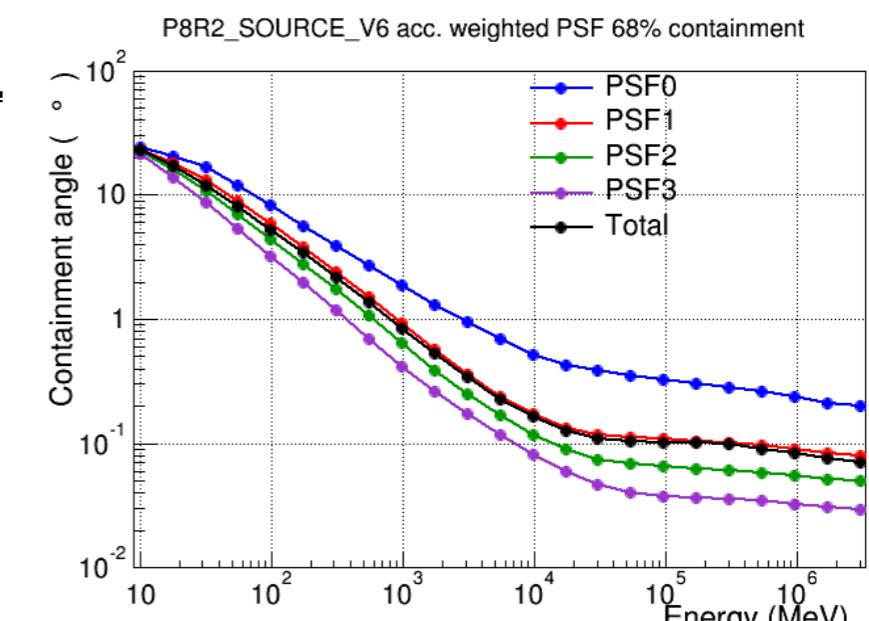
# Event types



- “Event types” partition events within a class according to certain characteristics such as reconstruction quality.  
Many IACT people use the term “event class” for “event type”. Very confusing, I suggest we adopt Fermi-LAT terminology.
- Fermi-LAT has events of different type in one EVENT list. Science tools / users can select subsets using a bitfield.
- There are a few efforts to use multiple event types (e.g. joint analysis of mono and stereo events in H.E.S.S.), but not much experience yet.
- Performance improvements via event types will be a big topic for CTA.
- **No support for IACT event types in our specs yet.**  
**Can use a hack: different OBS\_ID for each event type.**

## Fermi-LAT event types

Subclass Name	Event Type Bit	evtype argument	Description
FRONT	0	1	Front-converting
BACK	1	2	Back-converting
PSF0	2	4	0-25 % Quantile of WP8CTPSFCore
PSF1	3	8	25-50 % Quantile of WP8CTPSFCore
PSF2	4	16	50-75 % Quantile of WP8CTPSFCore
PSF3	5	32	75-100 % Quantile of WP8CTPSFCore



# Status DL3 formats for CTA



- The current DL3 spec is used and works pretty well for existing IACTs (H.E.S.S., VERITAS, MAGIC, ...), needs work for CTA
- There's many missing features and open questions:
  - How to encode info about event classes and types?
  - How to link events to IRFs?
  - How to achieve good IRF precision while keeping IRF DL3 size and MC IRF production costs at a reasonable level?
  - See presentations and minutes from the [May 2016 Meudon meeting](#) or the discussion in issues for the [gamma-astro-data-formats repo](#).
- Long way to develop the CTA DL3 data model and formats:
  - No real CTA data, pipeline not producing DL3 data and IRFs yet
  - No specs yet for what CTA is producing now:  
“IRF databases” for sets of parameters like zenith angle, array config, ...
  - Many open decisions concerning CTA observation modes, array configurations, event classes and types, IRF production, ...