

Open high-level data formats and software for gamma-ray astronomy

Christoph Deil^{1,a)}, Catherine Boisson^{3,b)}, Jeremy Perkins¹², Johannes King¹, Peter Eger¹, Michael Mayer⁶, Matthew Wood¹³, Victor Zabalza¹⁵, Jürgen Knödlseider¹¹, Tarek Hassan¹⁰, Lars Mohrmann⁵, Alexander Ziegler⁵, Bruno Khelifi⁴, Daniela Dorner⁵, Gernot Maier⁷, Giovanna Pedalletti⁷, Jaime Rosado¹⁰, José Luis Contreras¹⁰, Julien Lefaucheur³, Kai Brügge⁵, Mathieu Servillat³, Régis Terrier⁴, Roland Walter⁸ and Saverio Lombardi¹⁴

^{a)}Corresponding author: Christoph.Deil@mpi-hd.mpg.de

^{b)}Corresponding author: catherine.boisson@obspm.fr

¹*MPIK, Heidelberg, Germany*

²*NASA/GSFC, USA*

³*LUTH, Observatoire de Paris, Meudon, France*

⁴*APC, University of Paris 7, France*

⁵*FAU, Erlangen, Germany*

⁶*Humboldt University, Berlin, Germany*

⁷*DESY, Zeuthen, Germany*

⁸*Observatoire de Genève, 51 chemin des Maillettes, 1290 Sauverny, Switzerland*

⁹*Universidad Complutense de Madrid*

¹⁰*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona) Spain*

¹¹*IRAP, Toulouse, France*

¹²*NASA/GSFC*

¹³*SLAC National Accelerator Laboratory*

¹⁴*INAF, Osservatorio Astronomico di Roma, via Frascati 33, 00040 Monte Porzio Catone (Roma), Italy*

¹⁵*University of Leicester, UK*

Abstract. In gamma-ray astronomy, a variety of data formats and proprietary software exist, often developed for one specific mission or experiment. Especially for ground-based imaging atmospheric Cherenkov telescopes (IACTs), data and software have been so far mostly private to the collaborations operating the telescopes. However, there is a general movement in science towards open data and software and the next big IACT array, the Cherenkov Telescope Array (CTA), will be operated as an open observatory.

We have created a Github organisation at <https://github.com/open-gamma-ray-astro> where we are developing high-level data format specifications. A public mailing list was set up at <https://lists.nasa.gov/mailman/listinfo/open-gamma-ray-astro> and a first face-to-face meeting on the IACT high-level data model and formats took place in April 2016 in Meudon (France). The hope is that this open multi-mission effort will help to accelerate the development of open data formats and open-source software for gamma-ray astronomy, leading to synergies in the development of analysis codes and eventually better scientific results (reproducible, multi-mission). This poster will summarize what we have done so far, and has the goal to solicit comments and future contributions from the gamma-ray astronomy community.

TODO

We should probably mention these?

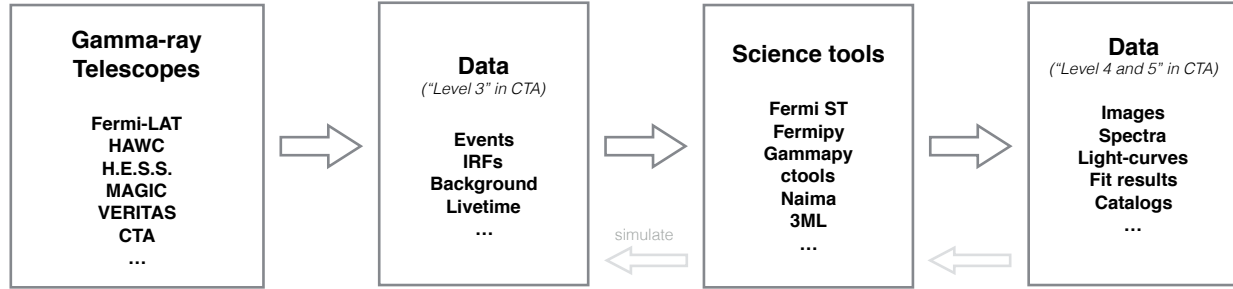


FIGURE 1. The purpose of the `gamma-astro-data-formats` effort: there's many gamma-ray data producers and consumers.

- Mention how we work together besides Github: monthly telcons + planned bi-yearly f2f meetings?
- Is there anything we can cite for CTA data challenge 1 or data challenge 2 plans?
- Mention website and some info on how CTA is releasing IRF files at the moment (prod2)?
- The Instrument Response Function Format for the Cherenkov Telescope Array [1] - has now merged with this effort.
- Previous work to define a gamma-ray lightcurve format: [2]

Introduction

Science is aiming towards open and reproducible research. In gamma-ray astronomy, especially for ground-based imaging atmospheric Cherenkov telescopes (IACT), data and software have so far been mostly private to the collaboration operating the telescopes. The Cherenkov Telescope Array (CTA), the next generation of IACT, will be operated as an open observatory, meaning that data and analysis softwares will be public at the end-user, science tools level. Momentum was thus given to the development of open data format and open-source software for gamma-ray astronomy, leading to synergies between experiments, ground based and space missions.

This is illustrated in Figure 1: there are many gamma-ray data producers and science tools to work with gamma-ray data, so agreeing on common data models and formats that cover most common analysis cases makes things simpler for data producers, tool developers and users.

Open source analysis tools for VHE gamma-ray astronomy have emerged. They all meet on common ground of using FITS files for data transfer. The current IACTs (H.E.S.S., MAGIC and VERITAS) all use their own software based on the ROOT library, which makes it impossible to analyze the respective data with anything else than the corresponding software. Current open source analysis tools provide alternative analysis techniques compared to the present standard in VHE astronomy. It is assumed that these techniques (e.g. 3D likelihood analysis such as already implemented for Fermi-LAT) improve the sensitivity of IACTs by roughly 20%. The data model for CTA is currently being developed but still work in progress. A main point of defining the high level data format for CTA is to understand the instrument including its systematics and IRF dependencies. In addition the needs from a user perspective have to be taken into account to create a solution as simple as possible. Having agreed on a common data format for files and on a way how to store and access those files (folder structure), makes mid-level (event energies, positions) and high-level (source position, morphology, spectrum) checks between the different chains, algorithms and open-source tools possible. This will also ease interoperability with other codes (e.g. to check results, combine results in one plot, ...). Currently two open-source science tools packages are being designed for current IACT and CTA data analysis, Gammapy ([3]) and Ctools ([4]). Gammapy is an in-development Astropy-affiliated package. Ctools is based on the GammaLib analysis framework, which is mainly written in C++.

We have created a Github organization where we are developing high-level data format specifications, in accordance with astronomy standards.

How to contribute

The specifications of a given data level format defines the names and semantics of data and header fields. Such specifications are made easy to understand. Specifications of this format are currently written that can form the basis

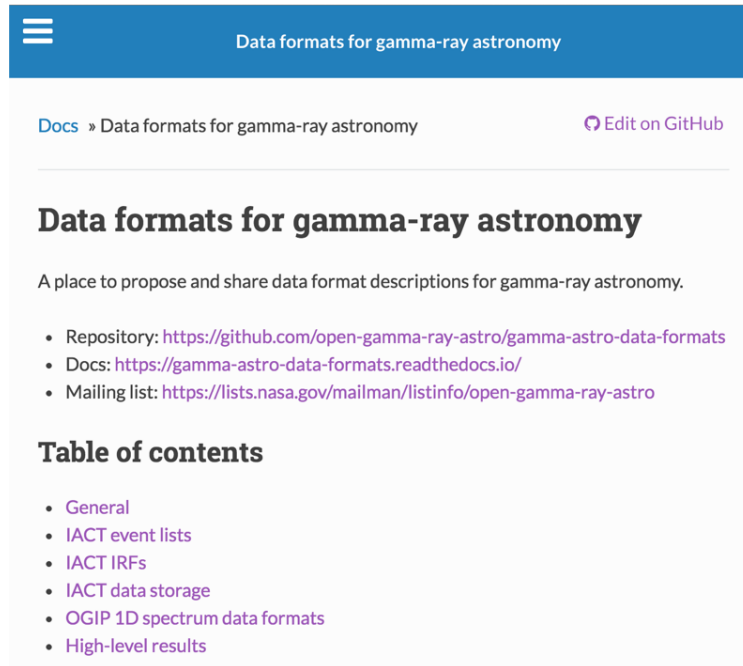


FIGURE 2. The data format specifications are available at <http://gamma-astro-data-formats.readthedocs.io/> in HTML and PDF format.

for prototyping for data producers (mainly existing IACTs and simulated CTA data) and consumers (mainly science tool codes). We include example files and some explanations, in addition to the detailed specifications for a given format.

The scope is high-level data, starting with event lists and instrument response functions (IRFs), what is called "data level 3" (DL3) in CTA. The first stable release (archived on Zenodo with a DOI) is coming soon.

If you want to contribute, it is simple:

- Use the existing format and give feedbacks. Propose additions and changes.
- Join the mailing list (see next section). Send an e-mail with an idea or proposal.
- Create a Github account. File an issue with a correction or make a pull request proposing additions.

No formal approval process is in place yet as this is a very recent effort.

Resources

- Mailing list for announcements and high level discussions (75 members, including people from all major gamma-ray collaborations :
<https://lists.nasa.gov/mailman/listinfo/open-gamma-ray-astro>
- Github issues and pull requests are used for detailed discussions:
<https://github.com/open-gamma-ray-astro/gamma-astro-data-formats>
- Data format specifications in HTML and PDF format, including example files:
<https://gamma-astro-data-formats.readthedocs.io/>

The specs are written in a markup format called "restructured text" (RST), which gets transformed by Sphinx to HTML or PDF, the latest rendered version is always available online (see Figure 2).

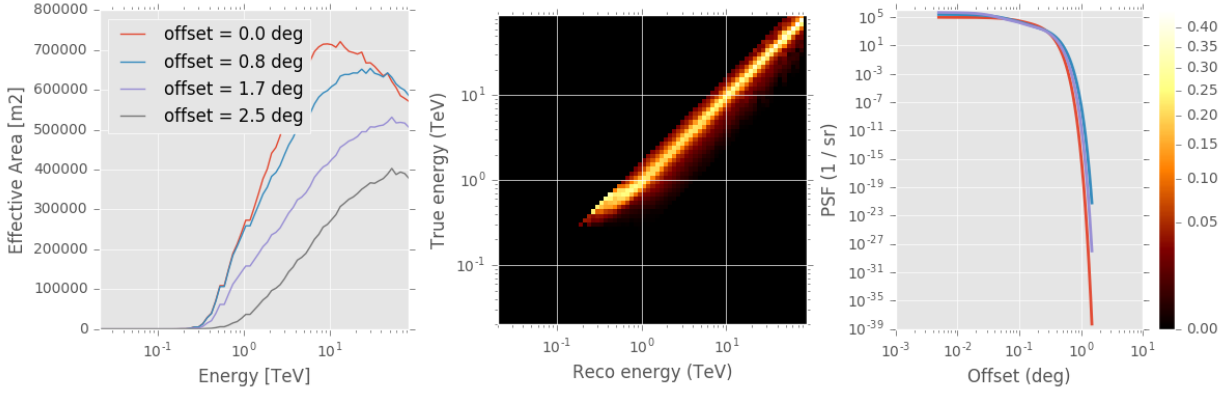


FIGURE 3. Low-level example: IACT DL3

Format specifications content

So far the focus has been mainly on the definition of the IACT DL3 formats. A "general" section defines common things like details about time scales or coordinate systems. Some work on the definition of higher-level formats for spectra and lightcurves is on going. HEALPIX (Hierarchical Equal Area isoLatitude Pixelization of a sphere) projection for images and description of data cubes are also considered.

Next sections will describe the effort toward the event list format definition as well as an example for spectra.

IACT DL3

DL3 format should describe all data released to the end users for analysis with science tools as , EVENT, IRF and TECH (technical data not directly attached to the object observed). Figure 3 illustrates some of the DL3 data content. Prototyping by existing IACTs (H.E.S.S., MAGIC, VERITAS and science tools (Gammapy, Ctools) is a major activity.

Many points are still under discussion/consideration:

- Observations modes, time intervals
- How to link EVENT and IRF
- Pointing and live time information
- IRF axis and validity ranges
- FoV coordinates

Likelihood SED and light curves

A format is developed to store spectral analysis results for exchange and publication, not only "flux points" and "upper limits", but also the full likelihood profiles (see example in Figure 4).

It was first developed in Fermipy¹ and used for Fermi-LAT (ref). It is now being adopted by ground-based high energy astronomy. Likelihood profiles are important information when those data are to be used in global spectral energy distribution (SED) of a source.

Description of light curves is another topic open for discussion.

Conclusions

TODO

¹<https://github.com/fermipy/fermipy>

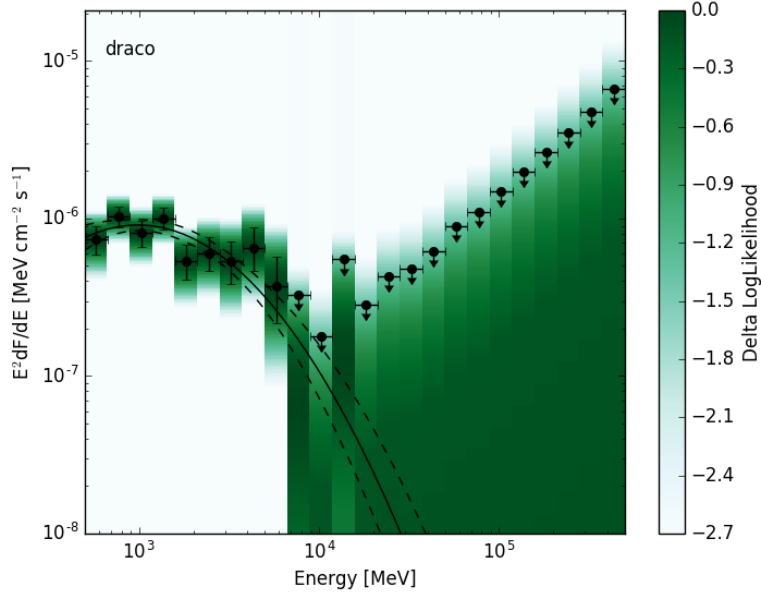


FIGURE 4. High-level example: SED likelihood profiles

Acknowledgements

We would like to thank everyone that has contributed to or supported this effort, be it directly via contributions to the format specification, or indirectly via feedback or adopting the existing formats and spending the effort to transform their existing data to the common formats defined here.

We would also like to thank the following services make this form of collaboration possible: `GitHub` for collaboration, `Sphinx` as documentation system and `Read the docs` for building and hosting the HTML and PDF version of the spec.

REFERENCES

- [1] J. E. Ward, J. Rico, T. Hassan, and f. t. CTA Consortium, ArXiv e-prints August (2015), arXiv:1508.07437 [astro-ph.IM] .
- [2] M. Tluczykont, E. Bernardini, K. Satalecka, R. Clavero, M. Shayduk, and O. Kalekin, AAP **524**, p. A48December (2010), arXiv:1010.5659 [astro-ph.HE] .
- [3] A. Donath, C. Deil, M. P. Arribas, J. King, E. Owen, R. Terrier, I. Reichardt, J. Harris, R. Bühler, and S. Klepser, ArXiv e-prints September (2015), arXiv:1509.07408 [astro-ph.IM] .
- [4] J. Knödseder, M. Mayer, C. Deil, J.-B. Cayrou, E. Owen, N. Kelley-Hoskins, C.-C. Lu, R. Buehler, F. Forest, T. Louge, H. Siejkowski, K. Kosack, L. Gerard, A. Schulz, P. Martin, D. Sanchez, S. Ohm, T. Hassan, and S. Brau-Nogué, AAP **593**, p. A1August (2016), arXiv:1606.00393 [astro-ph.IM] .