gamma-cat: an open data collection and source catalog for gamma-ray astronomy

Christoph Deil,[1] Arjun Voruganti,[1] Axel Donath,[1] Johannes King,[1]
Catherine Boisson,[2] Konstancja Satalecka,[3] and Matthias Wegen[3]
[1]*MPIK, Heidelberg, Germany;* Christoph.Deil@mpi-hd.mpg.de

[2]*LUTH, Observatoire de Paris, Meudon, France*

[3]*DESY, Zeuthen, Germany*

Abstract. gamma-cat (https://github.com/gammapy/gamma-cat) is an open data collection and source catalog for TeV gamma-ray astronomy. Currently data from 32 papers is available, and the catalog contains 163 sources. Data is input using the hierarchical, human- and machine-readable YAML format and the tabular ECSV text formats, processed using Python scripts into an as-uniform form as possible. The data can be browsed on the gamma-sky.net web page, analyzed using Gammapy, or fully downloaded in FITS and other formats and used in whatever way the user likes. Data is collected in a git repository on Github, providing transparency, version control as well as simple maintenance and contribution workflow. This data repository was started in August 2016, the data collection as well as the specification of the input and output formats is work in progress. Here we present the project for the first time, and discuss it's context, implementation, status, plans as well as some possible use cases for science analysis.

1. Introduction

The first cosmic TeV gamma-ray source detected from the ground was the Crab nebula in 1989. Since then, TeV astronomy has seen rapid growth. As of October 2016, 163 sources have been detected (see Figure 1). Information on the position, morphology, spectra and lightcurves of these sources have been published by HEGRA, H.E.S.S., VERITAS, MAGIC, HAWC and other telescopes in a few hundred individual papers. There is a need for a machine-readable, curated, up-to-date collection of TeV source data and a TeV catalog. Here we give an overview of other TeV gamma-ray data collections and catalogs in Section 2, then present the new project gamma-cat in Section 3 and conclude in Section 4.

2. Other TeV gamma-ray data collections and catalogs

For a given paper, the measurements on TeV gamma-ray sources are often not available in a machine-readable format. Typically the source position, morphology and spectral model are fitted to the data, and the results of that fit are given in the paper text. Sometimes, spectral points or lightcurve points are given in ASCII or FITS format on some

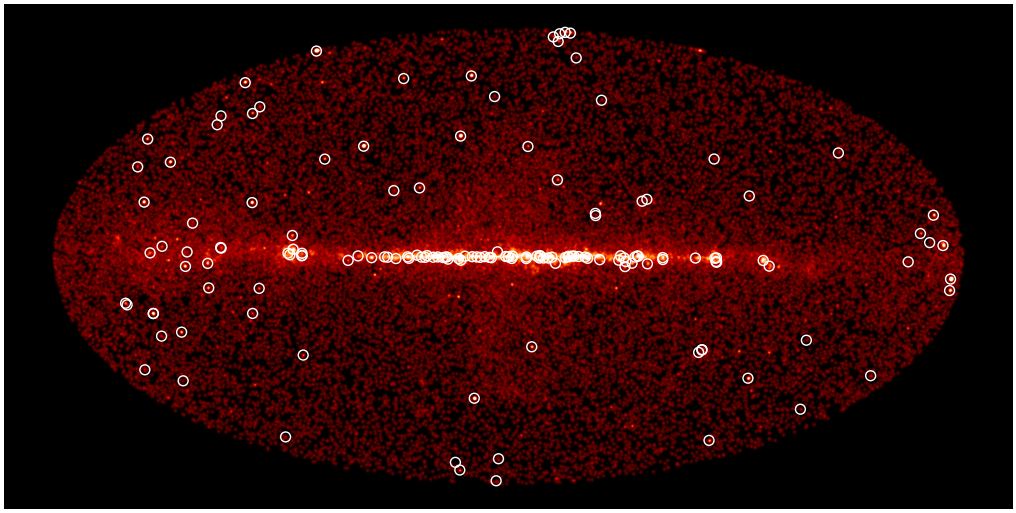

Figure 1. TeV gamma-ray sources from `gamma-cat` (white circles, 163 sources, status October 2016). The image (counts, smoothed with a Gaussian of width σ = 0.3 deg) shows the gamma-ray sky above 50 GeV using the Fermi-LAT 2FHL dataset (Ackermann et al. 2016).

webpage, or can be obtained from the corresponding author. In this section we describe prior efforts to gather available TeV gamma-ray data. A H.E.S.S. source list is available in HTML (Hypertext markup language) and CSV (character-separated values) formats at `https://www.mpi-hd.mpg.de/hfm/HESS/pages/home/sources/`, some spectra and lightcurves of blazars measured by H.E.S.S. in FITS (Flexible Image Transport system) and VOTable (Virtual Observatory table) formats at `http://hess.obspm.fr/`. Some VERITAS images, lightcurves and spectra in FITS format are available at `http://veritas.sao.arizona.edu/veritas-science/veritas-results-mainmenu-72`. MAGIC at `http://vobs.magic.pic.es/fits/`. The release of the H.E.S.S. Galactic plane survey catalog and the first all-sky HAWC catalog is expected soon. FITS and VOTABLE data for some blazars is available at `https://astro.desy.de/gamma_astronomy/magic/projects/light_curve_archive/`, which is an evolution of the data format and dataset described in Tluczykont et al. (2010). There's also 2WHSP (`http://www.asdc.asi.it/2whsp/`, Chang et al. (2016)), a very recent catalog of of HE and VHE gamma-ray blazars and blazar candidates.

An online TeV source catalog is available at `http://tevcat.uchicago.edu/` (Wakely & Horan 2008). `TeVCat` is not available for download, and the terms and conditions page explicitly forbids scraping the `TeVCat` web page to download a copy of the data. `TeVCat` does not contain spectral points or lightcurve or image data.

The `TeGeV` catalog is available at `http://www.asdc.asi.it/tgevcat/` (Carosi et al. 2015). `TeGeV` is a larger collection of data compared to `TeVCat`, it does include spectral points and lightcurves. `TeGeV` is part of a larger collection of multi-wavelength data and web-based tools to browse and analyze the data. The `TeGeV` data is available for download in CSV format, downloading all data (including spectral points and lightcurves) would require scraping the website or contacting the maintainers.

Given that these gamma-ray data collection exist, the question "why start a new one?" needs to be answered. We will first present `gamma-cat` in the next section, and then address this question in the conclusions.

3. `gamma-cat`

3.1. An open TeV data collection and catalog

`gamma-cat` (`https://github.com/gammapy/gamma-cat`) is both a *TeV data collection* and a *TeV source catalog*. The data collection consists of measurements from papers in machine-readable form. From this collection, a TeV source catalog is derived as a higher-level data product. The source catalog is somewhat subjective and doesn't contain all data from the collection. For a given source, usually each paper contains a set of measurements (e.g. source position, morphology, spectrum) and for the catalog, the "best" available parameters are chosen. Sometimes, especially for sources in the inner Galactic plane region where the source density is high, it can even happen that a later higher-resolution image reveals that what was previously thought of as one "source" really consists of multiple "sources". Extra-galactic sources usually don't have this issue of source confusion, but the most common source class, active galactic nuclei (AGN), are often variable in flux and spectral shape, implying that no single measurement can be the "best". The source identification and association status can change over time, as new potential counterparts (e.g. pulsars, supernova remnants, AGN, ...) are discovered in multi-wavelength data. For these reasons, we consider the *TeV data collection* aspect of primary importance — `gamma-cat` is a service to the astronomical community where information from hundreds of individual papers has been collected into a machine-readable form that can be easily queried and used. The `gamma-cat` catalog is a secondary data product that provides a simplified summary of the available information on TeV gamma-ray sources.

The `gamma-cat` scope is not completely decided. Currently the focus is on collecting the published measurements for source position, morphology, spectrum and lightcurves for TeV gamma-ray sources, mostly from the past decade and the H.E.S.S., VERITAS, MAGIC and HAWC telescopes. However, if we receive contributions of similar datasets in machine-readable formats that are not readily available elsewhere, we will consider adding them (but not exposing them in the gamma-ray source catalog). Examples that come to mind are: Fermi-LAT gamma-ray source measurements from individual papers in the GeV–TeV energy range, TeV gamma-ray diffuse emission measurements or models, cosmic ray electron spectra. Whether `gamma-cat` will remain solely focused on TeV gamma-ray sources, or will expand to become more of a "TeV gamma-ray very-high-level datasets" repository remains to be seen. Another choice we will have make is how to expose or merge the information from upcoming TeV source catalogs by H.E.S.S. and HAWC.

3.2. Guiding principles and implementation

The creation of `gamma-cat` was inspired by the other open astronomy catalogs at `https://astrocats.space/`. In the paper describing the *Open supernova catalog*, Guillochon et al. (2016) explain their guiding principles and describe the resulting implementation. `gamma-cat` is very similar, and in this section we briefly summarize our guiding principles and implementation.

gamma-cat data is fully open-access. Anyone can download all available data as a `zip` file, or clone the git repository to obtain the full version history of `gamma-cat` and use it however they like (except for an attribution request, which we state in our terms of use). Currently the size is very small, roughly 1 MB. We expect this to grow over time, but to stay within the limit of roughly 1 GB that comfortably fits in a single git repository. If large data files were added at some point (e.g. FITS images), extra repositories or other data stores could be used.

gamma-cat is set up for fully open collaboration. The input data collection is stored using the hierarchical, human- and machine-readable `YAML` format and the tabular `ECSV` text formats. The output data products (source catalog, spectra, lightcurves) are created using Python scripts (just run one command `python make.py all`). All scripts and Python packages we use are open-source: PyYAML (`http://pyyaml.org/`), Astropy (Astropy Collaboration 2013), Gammapy (Donath et al. 2015) and ADS (`https://github.com/andycasey/ads`). By making the data, tools and process fully open, we hope to create a community-driven project with a large scope and lifetime. Using a git repo and the simple "collection of text files as database" implementation means that the project is not tied to a given data center, website or researcher. Github pull requests provide a simple way to accept and review contributions without giving new people write access from the start.

3.3. Data formats, products, website, Python tools

The data products offered for `gamma-cat` are a source catalog as well as per-paper data. The source catalog is a flat table (available in `ECSV`, `FITS` and `VOTABLE` format) that contains one row per source with the most important information listed in a simple format. The catalog does not contain spectral points or lightcurve points or measurements from all papers for a given source. The full data available in `gamma-cat` is given as a collection of files that are generated based on collected input `YAML` and `ECSV` data, and derived information from the Python scripts. E.g. we compute the integral flux above 1 TeV according to the best-fit spectral model, even if that number wasn't given in the paper. Currently we are using `JSON` for hierarchical data and `ECSV` for tabular output data. Other formats (e.g. `YAML` or `FITS` and `VOTABLE`) could easily be generated as well, this will likely change as we receive feedback on which formats are preferred by the community. Development of the data format specifications (e.g. for spectral points or lightcurves) is work in progress and happens at `http://gamma-astro-data-formats.readthedocs.io/` (Deil et al. 2016), the formats we are using now should be considered unstable prototypes. We have started to develop a Python package, `gammapy.catalog` (`http://docs.gammapy.org/en/latest/catalog/`, Donath et al. (2015)), that allows easy querying of `gamma-cat` and working with the data, e.g. plotting spectra or lightcurves or analyzing them, or generating a TeV sky model from the catalog.

4. Conclusions

`gamma-cat` is a TeV gamma-ray astronomy data collection and source catalog. Compared to previous similar projects, `gamma-cat` stands out be being fully open-access and set up in a way that makes maintenance and community contributions easy, following the lead of other open astronomy catalogs from `https://astrocats.space/`. Merging of some previous TeV data collections (e.g. the light curve archive as DESY)

as well as communication about possible collaboration is ongoing. `gamma-cat` is being developed in close collaboration with ongoing efforts to develop gamma-ray data format standards (`gamm-astro-data-formats`), open-source science tools (`Gammapy`) as well as a website dedicated to the gamma-ray sky (`gamma-sky.net`). `gamma-cat` lets you quickly access all published information for all TeV sources, ready for data exploration or analysis. We hope that this will be a useful resource for daily work as well as the basis for novel studies involving the extensive set of available archival gamma-ray data. Concretely we propose `gamma-cat` to be used as one of the inputs to build the sky model for the planned CTA data challenge. Feedback and contributions to `gamma-cat` are highly welcome!

Acknowledgments. We thank Fabrizio Lucarelli and Gernot Maier for comments about `gamma-cat`. This research has made use of NASA's Astrophysics Data System Bibliographic Services, the SIMBAD database, operated at CDS, Strasbourg, France, the TeVCat online source catalog and the TeGeV catalog at ASDC, as well as the following astronomy Python packages: Astropy, Gammapy, ads.

References

Ackermann, M., et al. 2016, ApJS, 222, 5. 1508.04449
Astropy Collaboration 2013, A&A, 558, A33
Carosi, A., Lucarelli, F., Antonelli, L. A., & Giommi, P. 2015, ArXiv e-prints. 1510.08681
Chang, Y.-L., Arsioli, B., Giommi, P., & Padovani, P. 2016, ArXiv e-prints. 1609.05808
Deil, C., et al. 2016, ArXiv e-prints. 1610.01884
Donath, A., et al. 2015, ArXiv e-prints. 1509.07408
Guillochon, J., Parrent, J., & Margutti, R. 2016, ArXiv e-prints. 1605.01054
Tluczykont, M., et al. 2010, A&A, 524, A48. 1010.5659
Wakely, S. P., & Horan, D. 2008, in ICRC, vol. 3 of ICRC, 1341