Standardisation of Data Formats in Gamma-ray Astronomy

C. Nigro[1] and T. Hassan[2] on behalf of the *Data Formats for Gamma-ray Astronomy* forum contributors

[1]*Institut de Física d'Altes Energies (IFAE), Barcelona, Spain;*
`cosimo.nigro@ifae.es`

[2]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*

Abstract.   The operation of the future Cherenkov Telescope Array (CTA), the next generation of Imaging Atmospheric Cherenkov Telescopes (IACTs), as an Open Observatory requires data products and analysis tools to be accessible and usable by the scientific community. This requirement has impelled gamma-ray astronomers to develop open-source science tools and standardised data formats. The objective of this presentation is to provide a perspective on the current effort to define a prototypical format for high-level gamma-ray astronomical data. The initiative, started by the IACT community in 2016, has gained in the years 2018 and 2019 full recognition thanks to the first public data releases in such format by current-generation IACTs, followed by a series of papers employing it either to test the capabilities of open-source gamma-ray analysis tools or to showcase examples of reproducible, multi-instrument analysis.

1. Introduction

With the noticeable exception of *Fermi*-LAT, current instruments for gamma-ray astronomy rely on independent proprietary software and data for their scientific analyses. These technical differences not only prevent multi-instrument analyses to be performed within a common framework, but also impede any legacy data products to be disseminated without releasing the corresponding analysis software. Furthermore, the next instrument for gamma-ray astronomy at Very High Energies (VHE, $E > 100\,\text{GeV}$), CTA (CTA Consortium 2019), will be the first IACT to be operated as an Open Observatory. The requirement on its analysis tools and data to be accessible and usable to the astronomical community has led to the development of open-source science tools, `ctools` (Knödlseder et al. 2016) and `gammapy` (Deil et al. 2017a), and to the definition of a preliminary standardised data format for gamma-ray data (Deil et al. 2017b). This initiative, focus of our review, will be described in Section 2. In Section 3 we will instead describe the projects that implemented such data standard and used it either for the purpose of validating the above-mentioned science tools or to illustrate the ease of multi-instrument analysis once data standardisation has been achieved.
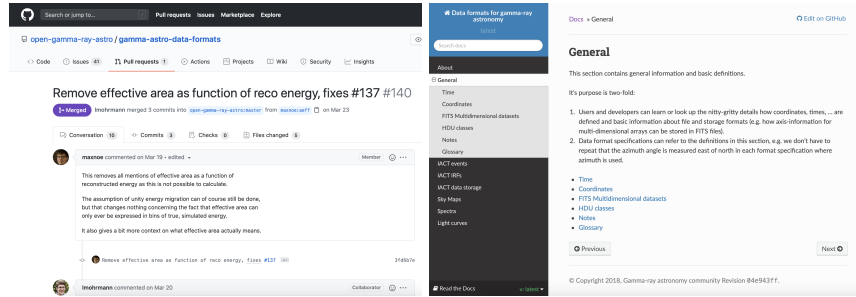
Figure 1. `github` repository (left) and online documentation (right) hosting the *Data Formats for Gamma-ray Astronomy* specifications.

2. The *Data Formats for Gamma-ray Astronomy* Forum

The discussion around a common format for high-level gamma-ray astronomical data was started in early 2016 and has been channeled in the so-called *Data Formats for Gamma-ray Astronomy* forum (Deil et al. 2017b). The data level chosen to be standardised is the so-called Data Level 3 (DL3, Contreras et al. 2015) which encapsulates lists of events classified as γ rays and the Instrument Response Function (IRF) in the data units of a `FITS` file (Wells et al. 1981). The specifications of the format, establishing the content, and the naming scheme of the `FITS` files are written in the form of a `sphinx` documentation hosted on `github`[1] and built via `readthedocs`[2], as shown in Figure 1. The open access to the format allows both gamma-ray astronomers producing the actual data and science tools developers to modify or expand it following the `github` workflow.

3. Projects using the Data Level 3 Format

Though initiated with the objective of CTA, the DL3 format can accommodate the data of current-generation IACTs. In what follows we present some projects already implementing and testing DL3 data products and open-source science tools.

3.1. The H.E.S.S. First Public Test Data Release

The H.E.S.S. collaboration was the first to make publicly available a small sample of its earliest observations (2004-2006) in DL3 format (H. E. S. S. Collaboration 2018)[3]. The release contains 30 h of observations of sources representing different science cases (Crab Nebula, PKS 2155-304, MSH 15-52 and RX J1713.7-3946) and 20 h of observations of sky patches devoid of gamma-ray sources, usable to build a background model (off-source observations). The aim of this release is to advertise the endeavour on the data standardisation and to allow for a double feedback: both on the side of the data format and on the science tools that will employ it.

---

[1] https://github.com/open-gamma-ray-astro/gamma-astro-data-formats

[2] https://gamma-astro-data-formats.readthedocs.io/en/latest

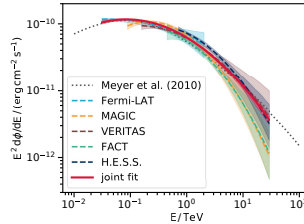[3] https://www.mpi-hd.mpg.de/hfm/HESS/pages/dl3-dr1

Figure 2.   Fit of the Crab Nebula spectrum obtained using the individual datasets in Nigro et al. (2019) and combining them in a single likelihood (red).

3.2.   Towards open and reproducible multi-instrument analysis in gamma-ray astronomy

The project, described in Nigro et al. (2019), provides the first example of fully reproducible, multi-instrument analysis of gamma-ray data. Observations of the Crab Nebula by `Fermi`-LAT and by four of the operating IACTs (MAGIC, VERITAS, FACT, and H.E.S.S.[4]) converted into DL3 format, are analysed with `gammapy`. A one-dimensional binned likelihood analysis (obtaining photon counts versus energy via aperture photometry) is performed to estimate the spectrum of the Crab Nebula using all the datasets jointly (Figure 2). A possible modification of the likelihood to account for the systematic uncertainties in the energy scales of the different instruments is also discussed. The project constitutes the first joint data release by IACTs. Full reproducibility is ensured by providing all the data, analysis scripts and tutorial notebooks publicly on `github`[5] along with a `conda` environment installing all the software dependencies. An ancillary `docker` container[6] guarantees reproducibility longer beyond life of the software used.

3.3.   Analysis of the H.E.S.S. public data release with ctools

The H.E.S.S. DL3 data release was analysed with `ctools` in Knödlseder et al. (2019). The science tool was thus far tested only with CTA simulations, and the publication represents the first test of its capabilities with actual IACT data. A recipe is presented to build a background model starting from the Off observations, the latter is used for spectral (and morphological — where possible) analysis of the 4 sources included in the H.E.S.S. DL3 data release. Results of binned and unbinned three-dimensional likelihoods (accounting at once for sources position, morphology and spectra) are compared against the ones of one-dimensional binned likelihoods. The internal consistency of the `ctools` results obtained with the different likelihood methods, and their agreement with the results from literature extracted from larger datasets of the same sources show the maturity of the analysis tool (Figure 3). The paper also illustrates the capability of `ctools` to simultaneously analyse *Fermi*-LAT and IACT data within the same likelihood framework.

---

[4]The H.E.S.S. dataset is a taken from the H.E.S.S. public data release.

[5]`https://github.com/open-gamma-ray-astro/joint-crab`

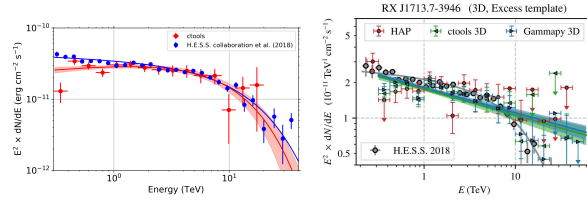[6]`https://hub.docker.com/r/gammapy/joint-crab`

Figure 3.   Validation of science tools on the extended source RX J1713.7-3946.
Left: comparison of the result of the `ctools` unbinned analysis and background
model in Knödlseder et al. (2019) against literature.   Right: comparison of
the `ctools` unbinned and `gammapy` binned analysis using the background model
in Mohrmann et al. (2019) against the H.E.S.S. proprietary software and literature.

3.4.   Validation of Open-source Science Tools and Background Model Construction in γ-ray Astronomy

Mohrmann et al. (2019) aim at validating both `ctools` and `gammapy` by comparing
their results when analysing the H.E.S.S. DL3 data release. Here another procedure
for the construction of a background model is illustrated, this time though relying on
a large private dataset of H.E.S.S. off-source observations ($\sim$ 4000 h). The results of
three-dimensional and one-dimensional analyses by the two science tools are compared
against the results obtained with one of the H.E.S.S. proprietary analysis chain on the
same datasets and against the literature. For both likelihood methods an excellent agreement is found with the results of the H.E.S.S. analysis software.

4.   Conclusion

The specifications of the *Data Formats for Gamma-ray astronomy* forum have been recently employed to produce the first public IACT data release and to validate the analysis methods of the science tools currently in development. DL3 data constitutes the
first example of multi-instrument gamma-ray data analysed with a fully-reproducible
pipeline. The VHE community has been building crucial expertise in data standardisation that will be essential in the future operation of CTA.

References

Contreras, J. L., et al. 2015, in 34th International Cosmic Ray Conference (ICRC2015), vol. 34
        of International Cosmic Ray Conference, 960. 1508.07584
CTA Consortium 2019, Science with the Cherenkov Telescope Array (World Scientific Publishing Co. Pte. Ltd.)
Deil, C., et al. 2017a, in 35th International Cosmic Ray Conference (ICRC2017), vol. 301 of
        International Cosmic Ray Conference, 766. 1709.01751
— 2017b, in 6th International Symposium on High Energy Gamma-Ray Astronomy, vol. 1792
        of American Institute of Physics Conference Series, 070006. 1610.01884
H. E. S. S. Collaboration 2018, arXiv e-prints, arXiv:1810.04516. 1810.04516
Knödlseder, J., et al. 2016, A&A, 593, A1. 1606.00393
— 2019, A&A, 632, A102. 1910.09456
Mohrmann, L., et al. 2019, A&A, 632, A72. 1910.08088
Nigro, C., et al. 2019, A&A, 625, A10. 1903.06621
Wells, D. C., et al. 1981, A&AS, 44, 363