

Toward a Public MAGIC Gamma-Ray Telescope Legacy Data Portal

M. Doro^a, C. Nigro^e, E. Prandini^a, A. Tramacere^b, M. Delfino^{c,d}, J. Delgado^{c,d}, E. do Souto^c, L. Jouvin^c, J. Rico^c for the MAGIC Collaboration*

^a *University of Padova and INFN Padova, I-35131 Padova (Italy)*

^b *Department of Astronomy, University of Geneva, Chemin d'Ecogia 16 - 1290 - Versoix - Switzerland*

^e *Deutsches Elektronen-Synchrotron (DESY), D-15738 Zeuthen, Germany*

^c *Institut de Fisica d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology (BIST), E-08193 Bellaterra (Barcelona), 955 Spain*

^d *also at Port d'Informació Científica (PIC) E-08193 Bellaterra (Barcelona) Spain*

The MAGIC telescopes are one of the three major IACTs (Imaging Atmospheric Cherenkov Telescopes) for observation of gamma rays in the TeV regime currently operative. MAGIC functions since 2003, and has published data from more than 60 sources, mostly blazars. MAGIC already provides astronomical `.fits` files with basic final scientific products such as spectral energy distributions, light curves and skymaps from published results. In future, the format of the files can be complemented with further relevant information to the community: a) by including the full multi-wavelength dataset enclosed in a publication, b) providing data in alternative easy-to-use formats such as ASCII or XML, which are accessible with other commonly used packages such as `gammapy`. Finally, besides high level products, activities have started to provide photon event lists and instrument response functions in a format such that scientists within and outside the community are allowed to perform higher level analysis. A second aim is to provide a full legacy of MAGIC data. This contribution will illustrate the achievements and plans of this activity.

*36th International Cosmic Ray Conference -ICRC2019-
July 24th - August 1st, 2019
Madison, WI, U.S.A.*

*<https://magic.mpp.mpg.de/>. For collaboration list see PoS(ICRC2019)1177

1. The MAGIC Dataset



Figure 1: The MAGIC stereo system composed of two 17 m diameter dishes operating simultaneously. Telescopes are located in the Roque de los Muchachos Observatory in the Northern Hemisphere (28.7°N , 17.9°W). Courtesy of D. Lopes (IAC).

The Major Atmospheric Gamma-ray Imaging Cherenkov (MAGIC) telescopes are a pair of 17 m-diameter telescopes operating in stereo mode and sensitive to cosmic gamma rays with energies between $0.5 \div 50$ TeV [1] (see Fig. 1). The first MAGIC telescope operated in standalone mode from 2003 to 2009, when the construction of a second telescope marked the beginning of stereoscopic observations. The imaging technique is based on the detection of Cherenkov light from the charged component of the extended atmospheric showers of particles generated in the high atmosphere (10 – 20 km a.s.l.) by cosmic gamma rays impinging the Earth (as well as any charged cosmic ray). The Cherenkov light propagates as a narrow (few ns wide) front of photons travelling towards the ground that generate an ellipsoidal image on the MAGIC cameras, as shown in Fig. 2.

The Cherenkov light developed by the shower results in a compact cluster of triggered photo-multipliers in the pixelised cameras of both telescopes. If a coincidence of such signals is produced in a few \sim ns window the signal of all the pixels of both cameras is read out and stored. Such information constitutes an *event* at the detector level. To extract the physical observables (incoming direction, time, and energy) of the gamma ray producing the event, several algorithms are applied. After subtracting the uniform illumination due to the night sky background (*image cleaning*) the information contained in the aforementioned cluster of triggered pixels is parametrised as an ellipsoidal image [2], the images of the two cameras are then combined (*stereo reconstruction*) for the assignment of direction and energy, performed with a Random Forest (RF) algorithm and a Monte Carlo look-up table, respectively. The images produced by CR air showers are rejected through a RF classification algorithm [3] that utilizes Hillas parameters and stereoscopic information returning a single variable “hadronness” (close to 0 for gamma-like events). Each step allows for a significant reduction of the data volume. Roughly, of ~ 2 TB of raw data per night, ~ 8 GB

are used for the reduced images (per night, ~ 100 MB per observational run), and just ~ 100 kB of DL3 data (event lists plus IRFs) per night.

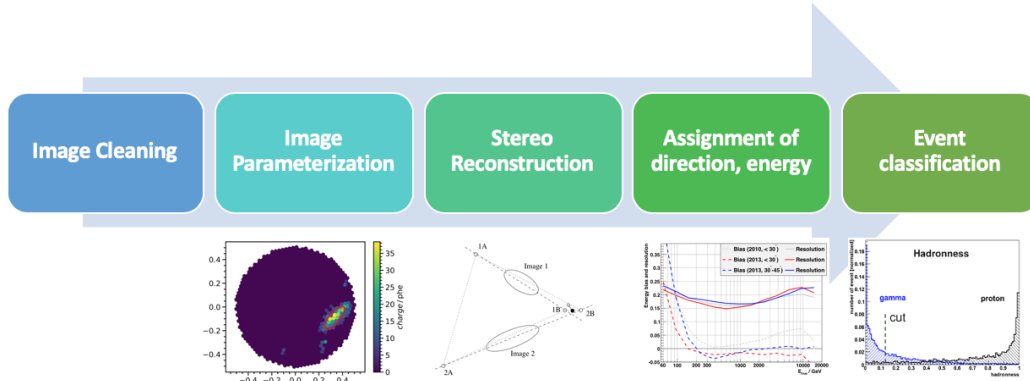


Figure 2: Basic reconstruction steps for raw MAGIC events.

From the above steps one has a list of events characterised by a time of arrival, a direction in the sky and an energy. Several selection cuts are further applied to the data, in energy bins, in order to properly select the signal and background control region.

However, in order to reconstruct a photon flux from these events, one needs to know the response of the system: the *Instrument Response Function (IRF)*, that contains the effective area, migration matrices between instrument-derived estimated and Monte Carlo matched coordinates and true energies. IRF components are generated from Monte Carlo simulated events, subject to the same analysis criteria/cuts optimized for the particular observing condition (zenith and azimuth of the observation, the sky quality and brightness, the mirror reflectivity etc...). The event list accompanied by IRFs is currently dubbed Data Level 3 (DL3) in the IACT community. Ideally, depending on the IRF on the observing conditions that continuously change with time, its components should also change for each detected event. As the change is minimal it is usual practice to generate IRFs for the entire observational run.

It is then clear that besides an event list (time energy direction) ancillary information must be annexed to enable scientists from within and outside the community to reproduce scientific results with MAGIC data. As an intermediate step to releasing the full event list, an activity that is ongoing within MAGIC [4], DL3 MAGIC data were produced for the project in Ref. [4]. MAGIC can share high level products from publications, these comprise e.g. light curves, spectral energy distributions, fits and models, and so on, which can be used for example to build multi-wavelengths data collection. The sharing of these products is extremely simplified with respect to the event list, and it is the main subject of this contribution. However, the possibilities to share the information contained in a publication are also a matter of high debate in our community regarding the technical implementation (format of files, portal server creation, ...) and specific information therein. While DL3 data have been already produced and shared for projects as the joint-Crab, these proceedings are concerned with sharing the higher level products, also called Data Level 4 (DL4).

Section 2 reports the current status of data sharing activities; Section 3 reports the future plans and Section 4 closes this report.

2. Status of data sharing

The MAGIC Collaboration currently provides a recollection of shared data accessible from the starting page <https://magic.mpp.mpg.de/index.php?id=139>. Two products are linked:

1. High-level FITS files repository: <http://vobs.magic.pic.es/fits/>
2. Low-level open data repository: <http://opendata.magic.pic.es>

The High-level FITS repository contains a single `.fits` file for each publication¹, which again contains the high-level products mentioned above as well as skymaps, detection plots and various other pieces of information. The low-level open data contains various lower level events from only a fraction of selected publications.

All of the above represent only an initial step for MAGIC data dissemination, which is going to significantly improve in the near future in several directions:

Disseminate event list (DL3). Because of a relative large number of background irreducible events, MAGIC shares candidates event list comprising energy, direction and arrival time. Such list is sometimes dubbed Data Level 3 (DL3) format.

Disseminate High Level products (DL4) MAGIC produces spectral energy distribution and light curves which are often used in a much larger multi-wavelength framework. Currently `.fits` files present MAGIC data alone. This data will be eventually augmented by adding multi-wavelength datasets published in MAGIC papers as well as the possibility of performing queries via API, for example using a [jupyter notebook](#), as in the case of the [astroquery](#) [7] package. We aim at improving this situation by providing a novel format of high-level data. This is the subject of the next section.

3. A novel high-level product file

One of the main focus of our project is the capability to provide high-level products together with all the metadata regarding the multi-wavelength campaigns, that usually are presented in MAGIC papers. In general, multi-wavelength data can be retrieved via front-end-based web queries, directly accessing the web service, such as those hosted by the Italian Space Agency (ASI) Science Data Center (SSDC) [8]; the European Southern Observatory (ESO) Archive Science Portal [10]; The Online Data Analysis (ODA) [5] hosted at the University of Geneva; or the Centre de Données astronomiques de Strasbourg [11]. Another approach, more suitable for the automated processing and for processing of large datasets, is the access to the data through API, as in the case of the ODA-API package [6] (see for example this [tutorial](#)), or the [astroquery](#) package. In this case the query can be performed using few command lines as in the following example

```
from astropy.coordinates import SkyCoord
from astroquery.sdss import SDSS
```

¹The content of the `.fits` is specified in a note accessible at the following link: <http://vobs.magic.pic.es/fits/mfits/tdas/tdas-fits.pdf>

```
pos = SkyCoord('0h8m05.63s_+14d50m23.3s', frame='icrs')
xid = SDSS.query_region(pos, spectro=True)
print(xid)
```

There are also some TeV-specific web portal for catalogs of sources such as the TeVCAT [13], the TeGeV catalogues [14]. SSDC data for example are exported in ASCII format in the following format:

```
# RA = 03 19 48.10(49.950417)
# Dec = +41 30 43.00(41.511944)
# Redshift = 0.0
# NH = 1.36E21
# BAT60AGN (15 - 55 keV) (id = 54)
# LogFreq (Hz) Bin Nufnu (erg cm^-2 s^-1) Nufnu_error TStart TStop
18.841672 0.000000 -10.337110 0.011196 53430.0 55256.0
# BAT54MCAT (15-50keV) (id = 53)
# LogFreq (Hz) Bin Nufnu (erg cm^-2 s^-1) Nufnu_error TStart TStop
18.820989 0.000000 -10.437238 0.022993 53340.5 54525.5
# 2MASS (id = 7)
# LogFreq (Hz) Bin Nufnu (erg cm^-2 s^-1) Nufnu_error TStart TStop
14.385428 0.000000 -10.443387 0.030683 0.0 0.0
14.256477 0.000000 -10.477122 0.032840 0.0 0.0
14.143015 0.000000 -10.521102 0.025126 0.0 0.0
```

Since MAGIC data need some assumptions and ancillary information, the final high-level products cannot be distributed as a single object, as in the case of the SDSS example above, where the query is returning a single `astropy Table`. In order to explain the details, we take as an example one specific MAGIC publication AA617(2018)A91 [16], related to the study of the flaring activity of the active galaxy NGC 1275 in 2016-2017.

In Fig. 3 we report four of five figures of AA617(2018)A91: Fig. 3A a MWL LC, Fig. 3B a zoom on this LC, Fig. 3C several MAGIC SEDs for different states, Fig. 3D a MWL SED. The fifth figure in the paper investigates the correlation between wavelengths and it is not of interest here. Of the above four, clearly the second is a zoom of the first and not of interest. But the remaining three are of interest. Fig. 3A reports data of MAGIC, *Fermi*/LAT and KVA, Fig. 3D of MAGIC and *Fermi*/LAT. In the following paragraphs, we try to investigate the formats of the above mentioned information that the MAGIC Collaboration may distribute to the community.

The format in which high-level data should be distributed is still a matter of debate within the TeV community. There is a forum for discussion in `github` [15], in which specific tag and fields are discussed, however, a consensus is not yet obtained. Our proposal is to store the metadata regarding each paper into a `.yaml` file, and the spectral/temporal information into `astropy tables`, using the extended `ecsv` format. Both these formats, are largely used in the astronomical community, and easily convertible (eg. `.yaml` to `.json`, and `.ecsv` to `.fits`) using well established libraries.

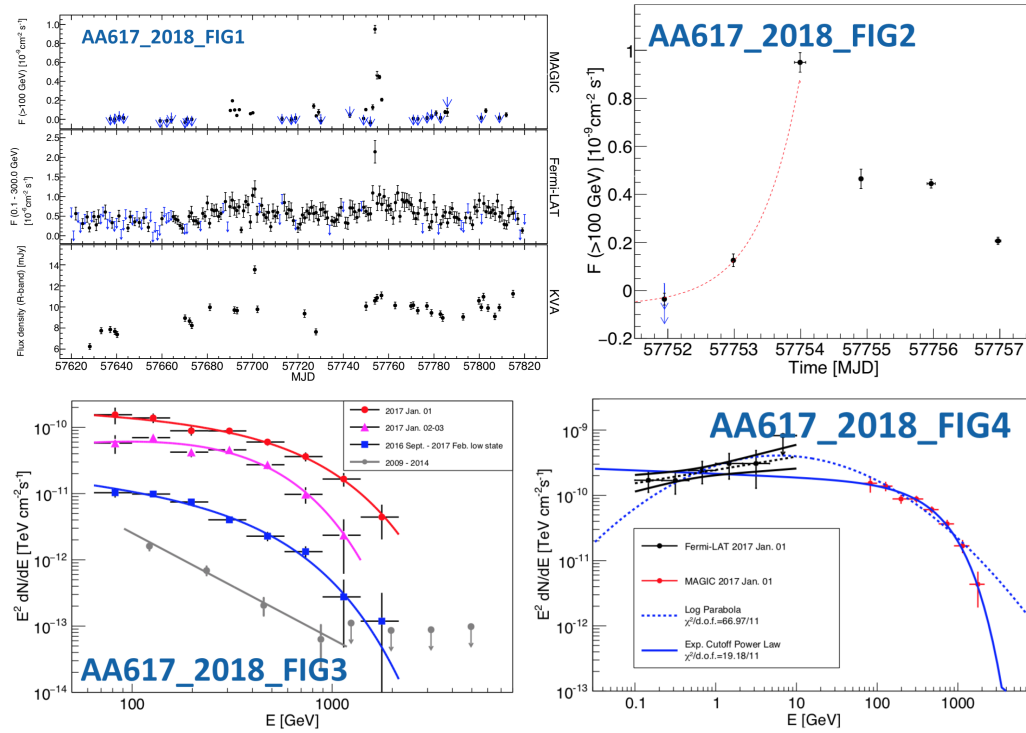


Figure 3: Figures of Ref.[16].

As an example, for AA617(2018)A91 we propose to have a .yaml file in the following format:

```

Filename:: magic_18f.yaml
File_info:
  Fdate = 20190315                                # File creation date
  Fvers = 1                                         # File version
  Fgen = Michele Doro, michele.doro@unipd.it       # Owner
  Fmail = magic_sapo@mpp.mpg.de                   # Contact
  Flink = XXXX                                     # Link
Paper_info:
  Pref: Astron.Astrophys. 617 (2018) A91
  Pdoi: https://doi.org/10.1051/0004-6361/201832895
  Parxiv: http://arxiv.org/abs/arXiv:1806.01559
  Pcoll: magic                                     # Main collaboration
  Pcauthor: XX, YY                               # Main authors of papers
  Pads: 2018A&A...617A..91M                       # ADS paper tag
  Pinspire: Ansoldi:2018sqg                        # inspire paper tag
  Targets in file Tpname: NGC1275                 # Main target name
  Taname: XX YY                                   # Other target names
File list MAGIC:                                  # List of MAGIC products
  magic_18f_lc1_fig1.ecsv
  magic_18f_lc2_fig1.ecsv
  magic_18f_sed1_fig3.ecsv

```

```

magic_18f_sed2_fig3.ecsv
magic_18f_sed3_fig3.ecsv
magic_18f_sed1_fig4.ecsv
File list MWL:                                     # List of MWL products
magic_18f_lc1_fig1_lat.ecsv
magic_18f_lc2_fig1_lat.ecsv
magic_18f_sed1_fig4_lat.ecsv
File on demands (available on request to Fmail) # Further products
magic_18f_sed1_fig3_fit.ecsv
magic_18f_sed2_fig3_fit.ecsv
magic_18f_sed3_fig3_fit.ecsv
magic_18f_sed1_fig4_fit.ecsv
magic_18f_sed1_fig4_model.ecsv
magic_18f_sed2_fig4_model.ecsv
Comments: None

```

For example `magic_18f_sed1_fig3.ecsv` would look like:

```

Filename:: magic_18f_sed1_fig3.ecsv
# %ECSV 0.9
# —
# datatype:
# - {name: e_ref, unit: TeV, datatype: float32, descr: Energy}
# - {name: e_min, unit: TeV, datatype: float32, descr: Energy bin min}
# - {name: e_max, unit: TeV, datatype: float32, descr: Energy bin max}
# - {name: dnde, unit: 1 / (cm2 s TeV), datatype: float32,
#   description: Differential photon flux at 'e_ref'}
# - {name: dnde_min, unit: 1 / (cm2 s TeV), datatype: float32,
#   description: Lower Statistical error (1 sigma) on 'dnde'}
# - {name: dnde_max, unit: 1 / (cm2 s TeV), datatype: float32,
#   description: Upper Statistical error (1 sigma) on 'dnde'}
# - {name: tstart, unit: mjd, datatype: float32, description: mjd start}
# - {name: tstop, unit: mjd, datatype: float32, description: mjd stop}
# - {name: tdurat, unit: h, datatype: float32, description: obs.time}
# schema: astropy-2.0
e_ref e_min e_max dnde dnde_min dnde_max tstart tstop tdurat
0.1567 0.0308 0.3561 1.529e-10 1.429e-10 1.729e-10 57637.1 57811.9 63
0.2484 0.0489 0.5646 4.719e-11 4.619e-11 4.819e-11 57637.1 57811.9 63
0.3937 0.0775 0.8949 1.459e-11 1.359e-11 1.659e-11 57637.1 57811.9 63
0.6239 0.1226 1.4182 4.609e-12 4.509e-12 4.809e-12 57637.1 57811.9 63
0.9888 0.1944 2.2477 1.139e-12 1.039e-12 1.239e-12 57637.1 57811.9 63
1.5670 0.3081 3.5623 3.839e-13 3.739e-13 4.139e-13 57637.1 57811.9 63

```

The above example file is significantly based on the format defined for the `gamma-cat` project [17]. This project aims at collecting all publications from current IACTs, and at being considered a reference for future projects. Our proposal provides a wider dataset that can be straightforwardly exported into the `gamma-cat` standards.

There are several possibilities to store such files. Considering we would like to provide an API, our server could be based in one of the MAGIC Collaboration institutes, and mirrored elsewhere.

The space requirement would be minor. The most critical part of the project is that of iterating all published papers and transfer all the high-level products into the proposed format. However, for novel published papers this would not constitute a problem.

4. Discussion and conclusions

We believe that providing high-level product data from MAGIC publications may be of interest to a wider astronomical community. We want to facilitate the access to these data by building an easily accessible dataset for all MAGIC published paper as well as papers to come. This paper catalogue will be released in 2019 and will be filled constantly.

Besides this effort, the MAGIC Collaboration is also working on the dissemination of the data at DL3-level, with high level data products relative to individual observations. A working group is active in MAGIC and results are expected soon.

Acknowledgement This proposal comes from discussion with several people, specially those attending the ASTERICS-OBELICS PyGamma19 meeting in Heidelberg <https://indico.cern.ch/event/783425/overview> in particular C. Boisson, C. Deil, G. Maier and R. Zanin. This project has received funding from the European Union's Horizon2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no 664931. The MAGIC Collaboration acknowledges support of institutes as in https://magic.mpp.mpg.de/acknowledgments_ICRC2019/.

References

- [1] J. Aleksić, et al. [MAGIC collaboration], *Astropart. Phys.* **72** (2016) 76
doi:10.1016/j.astropartphys.2015.02.005 [arXiv:1409.5594 [astro-ph.IM]].
- [2] Hillas, A. M. 1985, International Cosmic Ray Conference, 3,
- [3] J. Albert *et al.* [MAGIC Collaboration], *Nucl. Instrum. Meth. A* **588** (2008) 424
doi:10.1016/j.nima.2007.11.068 [arXiv:0709.3719 [astro-ph]].
- [4] C. Nigro *et al.*, *Astron. Astrophys.* **625**, A10 (2019) doi:10.1051/0004-6361/201834938
[arXiv:1903.06621 [astro-ph.HE]].
- [5] ODA: <https://www.astro.unige.ch/cdci>
- [6] ODA-API: <https://oda-api.readthedocs.io/en/latest/index.html>
- [7] astroquery: <https://astroquery.readthedocs.io/en/latest/>
- [8] SSDC: <http://www.ssdsc.asi.it>
- [9] <https://github.com/cta-observatory/ctapipe>
- [10] SSDC: ESP: archive.eso.org
- [11] CDS: <http://cds.u-strasbg.fr/>
- [12] Astropy: <https://www.astropy.org>
- [13] TeVCAT: <http://tevcat2.uchicago.edu>

- [14] TeGeV: <https://www.ssdsc.asi.it/tgevcats/>
- [15] Gamma Astro Data Format: <https://gamma-astro-data-formats.readthedocs.io/>
- [16] S. Ansoldi *et al.* [MAGIC Collaboration], *Astron. Astrophys.* **617** (2018) A91
doi:10.1051/0004-6361/201832895 [arXiv:1806.01559 [astro-ph.HE]].
- [17] Gamma-cat project: <https://github.com/gammapy/gamma-cat>.