PROCEEDINGS OF SCIENCE

Using the virtual observatory to study the gamma-ray sky

P. Kornecki,[1,*] **M. Servillat,**[1] **C. Boisson**[1] **and M. Fuessling**[2]

[1]*Observatoire de Paris-Meudon, 5 place Jules Janssen, Paris, France*
[2]*CTAO gGmbH*
E-mail: paula.kornecki@obspm.fr, mathieu.servillat@obspm.fr, catherine.boisson@obspm.fr, matthias.fuessling@cta-observatory.org

The growth of astrophysical data from space-borne missions, ground-based telescopes, and simulations from theoretical models, which have different services and data stores, has been exponential. The latter led to the necessity of unified ways of describing and accessing the data. Specific standards have to be defined and maintained to achieve this goal. The International Virtual Observatory Alliance (IVOA) regulates these standards which also must follow the FAIR principles for data management (Findable, Accessible, Interoperable, Reusable). The Virtual Observatory (VO) provides access to tools and utilities to work with data at different wavelengths. The exploration of gamma-ray data analysis in the FAIR context is still under development.

In this contribution, we illustrate on how the VO can facilitate scientific work in the gamma-ray astronomy domain nowadays. In the context of the FAIR principles, we present a system that performs a full analysis of H.E.S.S. public data on-the-fly using Gammapy on an interoperable job management platform. To comply with those principles, we must provide the detailed provenance of the analysis results. The coarse provenance is automatically stored on the platform, but we also improved the data reusability by capturing detailed provenance information directly in the Gammapy analysis script.

7th Heidelberg International Symposium on High-Energy Gamma-Ray Astronomy (Gamma2022)
4-8 July 2022
Barcelona, Spain

*Speaker

PoS(Gamma2022)227

1. Introduction

The management of data and data-driven research in astronomy and particle physics has become very important due to the enormous amount of astronomical data of different types to handle. The Virtual Observatory (VO) emerged with the purpose of collecting this variety of astronomical data archives and software tools, and to make them interoperable. For this, the International VO Alliance (IVOA) defined specific standard which are also maintained and regulated following the needs of the scientific community.

ESCAPE[1] is a European H2020 project which focuses on these big challenges. One of the objectives of ESCAPE is to make the data of several multi-messenger facilities openly accessible. In particular, this includes developing and evaluating VO standards to make the scientific data from very-high-energy facilities, in order to follow the FAIR principles (Findable, Accessible, Interoperable, Reusable) [1]. Nowadays, the VO is a framework for astronomical datasets, tools, and services to work seamlessly, connecting the scientist with the data and service providers. In particular, the VO provides tools and utilities to work with gamma-ray data, but exploring new solutions to analyse it in the context of the FAIR principles for data management is still necessary.

In this proceeding, we discuss and analyse the state of the art of very-high-energy gamma-ray data in the context of the VO. In section 2, we discuss the role of standards when exploring gamma-ray data in the VO framework. In section 3, we show an example of the use of the VO to manage gamma-ray data. In section 4, we show an on-the-fly analysis of H.E.S.S data using Gammapy [2] and the OPUS (Observatoire de Paris Universal Worker System) job management platform [3]. We also go deeper in the difficult task of tracing the detailed provenance. We end up in section 5 with a discussion and a description of future work.

2. The uses of VO standards with gamma-ray data

In order to make gamma-ray data products easily findable in the VO framework, we associated specific metadata to the standardised fields of the IVOA ObsCore (Observation Core components) data model [4]. This provides standard description records for each gamma-ray data product. This metadat also includes a link and description on how to access the data. Those metadata records are then ingested in an IVOA TAP (Table Access Protocol) [5] server, so that data can be searched via generic tools.

Tools that are VO compatible (like TOPCAT [6] and Aladin [7]) can thus be used to browse, search, access, and even manipulate these data (for example, *Fermi* and public H.E.S.S data can be found via the VO). Furthermore, we can easily transfer data from one VO tool to another, ensuring a large interoperability within the VO framework. Data can also be analysed with Gammapy by defining a dedicated job on the OPUS platform[2], as this platform is VO compatible, as it follows the IVOA UWS (Universal Worker System pattern) [8].

To foster their reusability, we need to provide the detailed provenance of the gamma-ray data products. By provenance, we mean the information involved in producing a piece of data, as a succession of various steps (activities) with intermediate data products (entities). This provenance

[1]https://projectescape.eu/

[2]https://voparis-uws-test.obspm.fr/client/

description is essential to assess its quality and trustworthiness of the data product. The IVOA Provenance data model [9] describes how provenance information can be modelled, stored and exchanged. VOPROV [10] is a python library that implements this IVOA Provenance Data Model. It allows to, describe and configure classes and relations between agents (e.g. users, servers), activities (e.g. dataset selection, background subtraction) and entities (e.g. input data, images, spectrum). It also permits the visualisation of the data provenance in several formats.

3. How to quickly look at gamma-ray data using VO tools

Before delving into the study of one or more gamma-ray sources, it is highly recommended to take a first look at the available observational data on these sources, study their position or spatial distribution, and check for the existence of observations at other wavelengths. This would give a broader view of the sources environment. One can already use the VO to get a first look of gamma-ray data.

Here is an example with H.E.S.S and *Fermi* data, described step by step:

1. We start with a search on the High Energy Spectroscopic System first data release ((H.E.S.S. DL3 DR1)[3]. One can use TOPCAT to search gamma-ray catalogues. We here use the "TAP Query button" in TOPCAT, and run the query "`SELECT * FROM hess_dr.vo_obscore_2`" on the H.E.S.S. DL3 DR1 service.

2. We then send the catalogue to ALADIN Desktop using an interoperable protocole (IVOA SAMP, for Simple Application Messaging Protocol), in order to visualise the source distribution in Galactic coordinates (chose the Gal Frame).

3. With ALADIN, we also load a *Fermi* all-sky map (*Fermi* colours HEALPix survey in Collections/Images/Gamma-ray),

4. We finally overlay the H.E.S.S Galactic plane survey (HGPS integral flux above 1 TeV in Collections/Images/Gamma-ray/HESS).

We show the resulting image in Fig. 1 that gives a global view of the localisation of the H.E.S.S. DL3 DR1 datasets.

Another helpful task is to search for spatial correlations between different catalogues of high energy sources. For example, the H.E.S.S source catalogue can be correlated with the 4th *Fermi* catalogue (4FGL DR3). To do this, we load in TOPCAT the 4FGL catalogue through a TAP server and then choose the cross match tool option.

4. Online analysis with Gammapy and OPUS based on standards

We developed a job to analyse Cherenkov gamma-ray data using Gammapy. This job can be run asynchronously on our work cluster with the OPUS job manager. Gammapy is a community-developed, open-source Python package now widely used. It is also the official Science Tools library

[3]https://hess-dr.obspm.fr

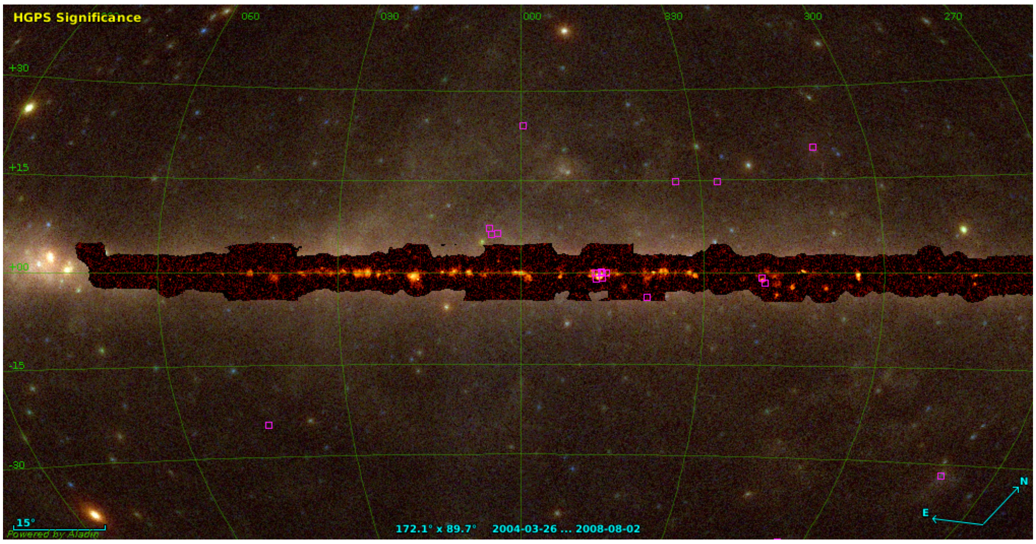

Figure 1: Aitoff projection including the *Fermi* 3 color HEALPix survey at the background, and the HGPS (H.E.S.S Galactic plane survey) gamma-ray flux above 1 TeV for regions where the sensitivity is better than 10% Crab. The pink squares are the observations of the H.E.S.S. DL3 DR1.

for the Cherenkov Telescope Array Observatory. OPUS is a job control system that facilitates access to analysis and simulation processes. It is based on two IVOA standards: the Universal Worker System pattern (UWS) to make the tool interoperable, and the Provenance Data Model to capture and expose the provenance information of jobs and results for reusability. As the tool is interoperable, jobs can be run either through a web interface or a script.

Data from the H.E.S.S. DL3 DR1 may thus be analysed with Gammapy by running it on the OPUS platform [3]. The job `point_source` in OPUS performs a complete spectral analysis of a generic point source covered by the H.E.S.S. DL3 DR1, simply by giving its sky coordinates as input parameters (*ra* and *dec*).

We show in Fig. 2 two of the `point_source` job results (a counts map and a spectrum) when supplying the source PKS 2155-304.

OPUS directly records the agents involved (opus-admin that defined the job and the current user that executed the job), the activity (an instance of the job `point_source`), the result entities (stacked_CountsMap.png, spectra.png and a SpectralPoints_table.tex), and input parameters (*ra* and *dec*). OPUS thus tracked the **coarse provenance** of `point_source`, however, it cannot capture what happened “inside" the job.

In order to capture this **detailed provenance** of the job results, we use the VOPROV routines from within the job script and store the internal provenance as a job result. This allowed us to track in particular the observations selected as input entities in the initial step of the script, and used during the following analysis steps.

OPUS then “merged" the coarse with the internal provenance to provide a more detailed provenance. This **merged provenance** is thus completed by essential information not available elsewhere, as shown in Fig. 3.

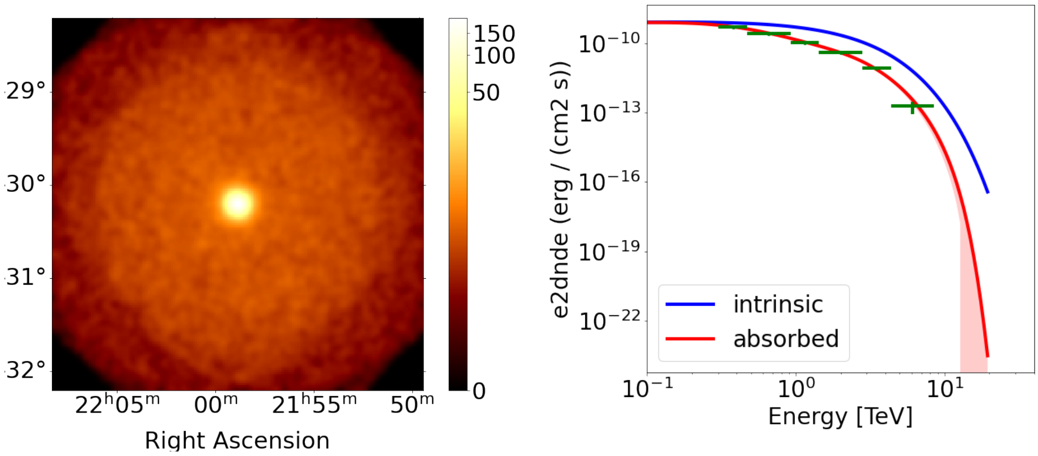

Figure 2: PKS 2155-304 data products. *Left panel*: Counts map with energies between 0.1 a 20 TeV. *Right panel*: Source spectrum. Blue line: intrinsic spectrum assuming a power-low with an exponential cutoff model; Red line: same model with the EBL absorption.

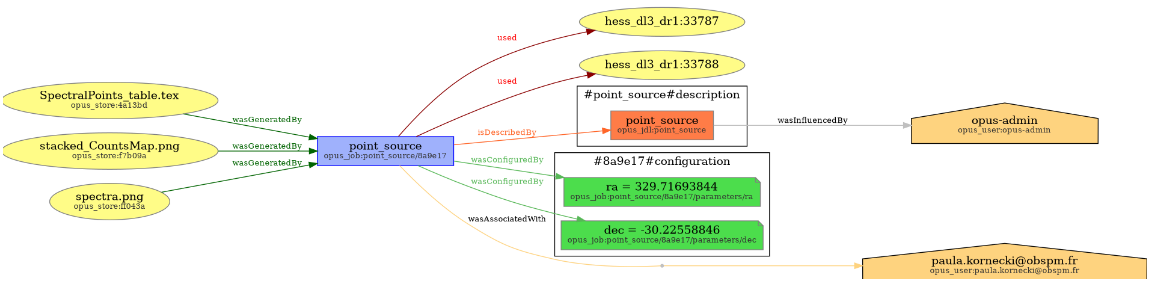

Figure 3: Merged graph of the coarse and detailed provenance of the data products, exposing in particular the central role of the `hess_dl3_dr1` data entities.

5. Discussions and next steps

The upcoming Cherenkov Telescope Array (CTA) will be a public observatory, and the data produced will thus have to be accessible through the VO, and follow the FAIR Principles. The IVOA standardisation provided so far allows the scientists to quickly discover and access gamma-ray data through the VO. However, gamma-ray data still needs to be further adapted to IVOA standards, and IVOA standards needs to evolve to cover the specificity of gamma-ray data, in particular the description of event lists.

We now would like to provide a more complete analysis job, based on a set of parameters that will enable various behaviours of the analysis job. Such a set of parameters will be defined by looking at different works that report analysis of H.E.S.S. sources, so that each analysis use case can be performed with one or several OPUS jobs. We would thus provide an advanced, on-the-fly analysis tool that takes advantage of the VO and provides FAIR data results, including their detailed provenance.

Provenance capture can be a complex task, with the need to merge different provenance information during some post-processing. The next step would be to increase the granularity of the internal provenance to trace each sub-activity involved in the generation of each final product. The necessary detail of provenance tracing depends on the context and is still under debate. One of the open questions is, what is the optimal amount of provenance to guarantee data reusability? The study of provenance is highly complex but worthwhile, as it is a crucial tool to prove the reliability to scientific works.

Acknowledgements: We acknowledge support from the ESCAPE project (European Science Cluster of Astronomy and Particle physics ESFRI – European Strategy Forum on Research Infrastructures) funded by the EU Horizon 2020 research and innovation program (Grant Agreement n.824064). We made use of Gammapy, a community-developed core Python package for TeV gamma-ray astronomy.

References

[1] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak et al., *The fair guiding principles for scientific data management and stewardship*, *Scientific Data* **3** (2016) .

[2] C. Deil, C. Boisson, K. Kosack, J. Perkins, J. King, P. Eger et al., *Open high-level data formats and software for gamma-ray astronomy*, in *6th International Symposium on High Energy Gamma-Ray Astronomy*, vol. 1792 of *American Institute of Physics Conference Series*, p. 070006, Jan., 2017, DOI [arXiv:1610.01884].

[3] M. Servillat, S. Aicardi, B. Cecconi and M. Mancini, *OPUS: an Interoperable Job Control System Based on VO Standards*, in *Astronomical Society of the Pacific Conference Series*, J.E. Ruiz, F. Pierfedereci and P. Teuben, eds., vol. 532 of *Astronomical Society of the Pacific Conference Series*, p. 451, July, 2022 [2101.08683].

[4] M. Louys, D. Tody, P. Dowler, D. Durand, L. Michel, F. Bonnarel et al., "Observation Data Model Core Components, its Implementation in the Table Access Protocol Version 1.1." IVOA Recommendation 09 May 2017, May, 2017. 10.5479/ADS/bib/2017ivoa.spec.0509L.

[5] P. Dowler, G. Rixon, D. Tody and M. Demleitner, "Table Access Protocol Version 1.1." IVOA Recommendation 27 September 2019, Sept., 2019. 10.5479/ADS/bib/2019ivoa.spec.0927D.

[6] M.B. Taylor, *TOPCAT & STIL: Starlink Table/VOTable Processing Software*, in *Astronomical Data Analysis Software and Systems XIV*, P. Shopbell, M. Britton and R. Ebert, eds., vol. 347 of *Astronomical Society of the Pacific Conference Series*, p. 29, Dec., 2005.

[7] F. Bonnarel, P. Fernique, O. Bienaymé, D. Egret, F. Genova, M. Louys et al., *The ALADIN interactive sky atlas. A reference tool for identification of astronomical sources*, *A&AS* **143** (2000) 33.

[8] P.A. Harrison and G. Rixon, "Universal Worker Service Pattern Version 1.1." IVOA Recommendation 24 October 2016, Oct., 2016. 10.5479/ADS/bib/2016ivoa.spec.1024H.

[9] M. Servillat, K. Riebe, C. Boisson, F. Bonnarel, A. Galkin, M. Louys et al., "IVOA Provenance Data Model Version 1.0." IVOA Recommendation 11 April 2020, Apr., 2020. 10.5479/ADS/bib/2020ivoa.spec.0411S.

[10] M. Sanguillon, F. Bonnarel, M. Louys, M. Nullmeier, K. Riebe and M. Servillat, *Provenance Tools for Astronomy*, in *Astronomical Data Analysis Software and Systems XXVII*, P. Ballester, J. Ibsen, M. Solar and K. Shortridge, eds., vol. 522 of *Astronomical Society of the Pacific Conference Series*, p. 545, Apr., 2020 [1812.00878].