



FAIR high level data for Cherenkov astronomy

Mathieu Servillat, Catherine Boisson, Matthias Fuessling, Bruno Khelifi

► To cite this version:

Mathieu Servillat, Catherine Boisson, Matthias Fuessling, Bruno Khelifi. FAIR high level data for Cherenkov astronomy. ADASS XXXI, Oct 2021, Cape Town, South Africa. obspm-03516688

HAL Id: obspm-03516688

<https://hal-obspm.ccsd.cnrs.fr/obspm-03516688>

Submitted on 7 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FAIR high level data for Cherenkov astronomy

Mathieu Servillat,¹ Catherine Boisson,¹ Matthias Fuessling,² and Bruno Khelifi³

¹*LUTH, Observatoire de Paris, Université PSL, Université de Paris, CNRS, F-92190 Meudon, France; mathieu.servillat@obspm.fr*

²*CTAO gGmbH, 69117, Heidelberg, Germany*

³*APC, IN2P3/CNRS - Université de Paris, 75013 Paris, France*

Abstract. We highlight here several solutions developed to make high-level Cherenkov data FAIR: Findable, Accessible, Interoperable and Reusable. The first three FAIR principles may be ensured by properly indexing the data and using community standards, protocols and services, for example provided by the International Virtual Observatory Alliance (IVOA). However, the reusability principle is particularly subtle as the question of trust is raised. Provenance information, that describes the data origin and all transformations performed, is essential to ensure this trust, and it should come with the proper granularity and level of details.

We developed a prototype platform to make the first H.E.S.S. public test data findable and accessible through the Virtual Observatory (VO). The exposed high-level data follows the gamma-ray astronomy data format (GADF) proposed as a community standard to ensure wider interoperability. We also designed a provenance management system in connection with the development of pipelines and analysis tools for CTA (ctape and gammapy), in order to collect rich and detailed provenance information, as recommended by the FAIR reusability principle. The prototype platform thus implements the main functionalities of a science gateway, including data search and access, online processing, and traceability of the various actions performed by a user.

1. Introduction and context

An effort is on-going to make science data FAIR: Findable, Accessible, Interoperable and Reusable (Wilkinson et al. 2016). In the astronomy domain, the International Virtual Observatory Alliance (IVOA, <https://www.ivoa.net>) provides standards since 2002 with similar objectives. We thus observe a connection between the FAIR principles and IVOA standards, demonstrating that, with these standards implemented, data management systems are close to compliance, with only a few extra components required to make astronomical data FAIR (O'Toole 2021).

Though the first three principles can be fulfilled with technical solutions, the reusability principle is more subtle because the question of trust is raised. It is stated that "data and metadata should be richly described with a plurality of accurate and relevant attributes". The wording is more qualitative than quantitative, indicating that the richness, the accuracy and relevance are to be adapted to each project. Detailed provenance of the provided data is indicated as essential to ensure this trust. An

IVOA recommendation to model astronomy provenance metadata has been released in 2020 as to set up the basics to structure provenance information (Servillat et al. 2020). The provenance should then be provided at the appropriate granularity and level of detail for each project. In order to make provenance of astronomy data more practical, we organized discussions on common approaches across all wavebands with current and future large observatories within the ESCAPE European project (<https://www.projectescape.eu>), the IVOA and at ADASS (Servillat et al. 2021b).

Cherenkov astronomy provides an interesting use case, as complex data is acquired and processed based on the detection of Cherenkov light generated in the atmosphere by air showers induced by energetic cosmic particles. Unlike current experiments such as H.E.S.S., MAGIC or VERITAS, the next generation Cherenkov Telescope Array (CTA) will be an open observatory providing a public access to its high level science data products, with the requirement to provide FAIR data. The complex treatment of raw Cherenkov data thus implies to attach a detailed description of the high level data and their provenance, so as to explain the different assumptions made during the data processing, and building users' trust.

We developed a prototype platform to expose the first H.E.S.S. public test data release through the Virtual Observatory, making it findable and accessible (see §2). This high-level data follows a proposed standard to ensure wider interoperability (see §3). Along with this standardization, the development of dedicated Science Tools makes the data even more accessible and interoperable to end users (see §4). We finally designed a provenance management system in connection with the development of pipelines and analysis tools for CTA (ctapipe and gammapy), in order to collect rich and detailed provenance information (see §5). The prototype platform thus implements the main functionalities of a science gateway, including data search and access, online processing, and traceability of the various actions performed by a user.

2. VO access to H.E.S.S. public test data

A collection of data has been publicly released¹ by the H.E.S.S. collaboration in 2018. The released data contains lists of candidate gamma-ray photons detected by H.E.S.S. during runs of observations on specific targets, along with instrumental response functions. This high-level data (defined as data level 3, DL3), has already been specifically processed from raw data and serves as a base for further analysis by the end users.

We created a VO access to this H.E.S.S. DL3 public data², for which we adapted the IVOA observation core metadata (ObsCore, Louys et al. 2017) in order to make Cherenkov data findable and accessible. The service can be queried using the IVOA Table Access Protocole (TAP) to further explore the available data. The resulting IVOA ObsTAP service is declared to the VO registry, so that it can be found by VO tools (such as Aladin or Topcat).

¹<https://www.mpi-hd.mpg.de/hfm/HESS/pages/dl3-dr1>

²<https://hess-dr.obspm.fr>

3. A common format for interoperability

In gamma-ray astronomy, a variety of data formats and proprietary software have been traditionally used, often developed for one specific mission or experiment. For ground-based imaging atmospheric Cherenkov telescopes (IACTs), data and software are mostly private to the collaborations operating the telescopes. To foster interoperability, the community is developing high-level data format specifications through the project Gamma-ray Astronomy Data Format³ (GADF, Deil et al. 2017a). This effort is necessary to ensure a wider use of the data, but it also helps to build confidence in the data.

4. The open-source science tool gammapy

Gammapy⁴ is an open-source Python package for gamma-ray astronomy built on Numpy, Scipy and Astropy (Deil et al. 2017b; Nigro et al. 2019). It is used as core library for the Science Analysis tools of the Cherenkov Telescope Array, is the recommended tool by the H.E.S.S. collaboration to be used for Science publications, and is already widely used in the analysis of existing gamma-ray instruments, such as MAGIC, VERITAS and HAWC. The development of gammapy is carried out taking into account most of the FAIR4RS principles (Chue Hong et al. 2021), in order to improve the software productivity, quality, reproducibility, and sustainability. This effort again promotes interoperability and builds confidence in data analysis.

5. Providing provenance to foster reusability

Provenance information has been discussed early within the IVOA⁵. Because of the complexity of Cherenkov data, CTA served as a major use case for the development of the IVOA Provenance data model (Servillat et al. 2019). This served to further develop a provenance management system adapted to astronomical projects needs (Servillat et al. 2021a).

Additional use cases have been discussed within the IVOA and the European ESCAPE project (Servillat et al. 2021b). We observed that some projects already have data collections generated and archived from which the provenance has to be extracted (provenance "on top"). Other projects, such as the CTA project, are building complex pipelines that can be used to test the automatic capture of the provenance information (capture "inside"). Different tools and prototypes have been developed and tested to capture, store, access and visualize the provenance information, which participate to the shaping of a full provenance management system able to handle detailed provenance information.

Provenance graphs can become extremely complex. For simpler use, the concept of a last-step flat provenance has recently been proposed: it is information on the last activity (execution and software description) and the context (workflow, instrument), that can be embedded into an entity as a list of keywords. It is then important to include

³<https://gamma-astro-data-formats.readthedocs.io>

⁴<https://gammapy.org>

⁵<https://wiki.ivoa.net/twiki/bin/view/IVOA/ProvenanceDataModelLegacy>

the list of identifiers of generated and used entities, so that a full provenance may be reconstructed from this last-step provenance.

Acknowledgments. We acknowledge support from the ESCAPE project funded by the EU Horizon 2020 research and innovation program (Grant Agreement n.824064). Additional funding was provided by the INSU (Action Spécifique Observatoire Virtuel, ASOV), the Action Fédératrice CTA at the Observatoire de Paris and the Paris Astronomical Data Centre (PADC).

References

- Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A.-L., Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., Honeyman, T., & et al. 2021, FAIR Principles for Research Software (FAIR4RS Principles), Research Data Alliance. URL <https://doi.org/10.15497/RDA00065>
- Deil, C., Boisson, C., Kosack, K., Perkins, J., King, J., Eger, P., Mayer, M., Wood, M., Zabalza, V., Knödseder, J., Hassan, T., Mohrmann, L., Ziegler, A., Khelifi, B., Dorner, D., Maier, G., Pedaletti, G., Rosado, J., Contreras, J. L., Lefaucheur, J., Brügge, K., Servillat, M., Terrier, R., Walter, R., & Lombardi, S. 2017a, in 6th International Symposium on High Energy Gamma-Ray Astronomy, vol. 1792 of American Institute of Physics Conference Series, 070006. [arXiv:1610.01884](#)
- Deil, C., Zanin, R., Lefaucheur, J., Boisson, C., Khelifi, B., Terrier, R., Wood, M., Mohrmann, L., Chakraborty, N., Watson, J., Lopez-Coto, R., Klepser, S., Cerruti, M., Lenain, J. P., Acero, F., Djannati-Ataï, A., Pita, S., Bosnjak, Z., Trichard, C., Vuillaume, T., Donath, A., Consortium, C., King, J., Jouvin, L., Owen, E., Sipocz, B., Lennarz, D., Voruganti, A., Spir-Jacob, M., Ruiz, J. E., & Arribas, M. P. 2017b, in 35th International Cosmic Ray Conference (ICRC2017), vol. 301 of International Cosmic Ray Conference, 766. [arXiv:1709.01751](#)
- Louys, M., Tody, D., Dowler, P., Durand, D., Michel, L., Bonnarel, F., Micol, A., & IVOA DataModel Working Group 2017, Observation Data Model Core Components, its Implementation in the Table Access Protocol Version 1.1, IVOA Recommendation 09 May 2017
- Nigro, C., Deil, C., Zanin, R., Hassan, T., King, J., Ruiz, J. E., Saha, L., Terrier, R., Brügge, K., Nöthe, M., Bird, R., Lin, T. T. Y., Aleksić, J., Boisson, C., Contreras, J. L., Donath, A., Jouvin, L., Kelley-Hoskins, N., Khelifi, B., Kosack, K., Rico, J., & Sinha, A. 2019, *A&A*, 625, A10. [arXiv:1903.06621](#)
- O’Toole, S. 2021, in ADASS XXXI, edited by O. Smirnov, R. Taylor, C. Ödman, & B. Hugo (San Francisco: ASP), vol. TBD of ASP Conf. Ser., 999 TBD
- Servillat, M., Boisson, C., Lefaucheur, J., Bregéon, J., Sanguillon, M., & Contreras, J.-L. 2019, in Astronomical Data Analysis Software and Systems XXVI, edited by M. Molinaro, K. Shortridge, & F. Pasian, vol. 521 of Astronomical Society of the Pacific Conference Series, 469. [arXiv:1706.06512](#)
- Servillat, M., Bonnarel, F., Boisson, C., Louys, M., Ruiz, J. E., & Sanguillon, M. 2021a, in Provenance and Annotation of Data and Processes, edited by B. Glavic, V. Braganholo, & D. Koop (Cham: Springer International Publishing), 244
- Servillat, M., Bonnarel, F., Louys, M., & Sanguillon, M. 2021b, in ADASS XXX, edited by J.-E. Ruiz, & F. Pierfederici, vol. TBD of ASP Conf. Ser., TBD. [arXiv:2101.08691](#)
- Servillat, M., Riebe, K., Boisson, C., Bonnarel, F., Galkin, A., Louys, M., Nullmeier, M., Renault-Tinacci, N., Sanguillon, M., & Streicher, O. 2020, IVOA Provenance Data Model Version 1.0, IVOA Recommendation 11 April 2020
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. 2016, *Scientific Data*, 3