Virtual Observatory and High Energy Astrophysics

Version 0.1

IVOA Note 2023-11-10

Working Group
DM
This version
https://www.ivoa.net/documents/VOHE/20231110
Latest version
https://www.ivoa.net/documents/VOHE
Previous versions
This is the first public release
Author(s)
HE club
Editor(s)
HE club

Abstract

Virtual Observatory and High Energy Astrophysics

Status of this document

This is an IVOA Note expressing suggestions from and opinions of the authors. It is intended to share best practices, possible approaches, or other perspectives on interoperability with the Virtual Observatory. It should not be referenced or otherwise interpreted as a standard specification.

A list of current IVOA Recommendations and other technical documents can be found at https://www.ivoa.net/documents/.

Contents

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).

Conformance-related definitions

The words "MUST", "SHALL", "SHOULD", "MAY", "RECOMMENDED", and "OPTIONAL" (in upper or lower case) used in this document are to be interpreted as described in IETF standard RFC2119 (Bradner, 1997).

The *Virtual Observatory (VO)* is a general term for a collection of federated resources that can be used to conduct astronomical research, education, and outreach. The International Virtual Observatory Alliance (IVOA) is a global collaboration of separately funded projects to develop standards and infrastructure that enable VO applications.

1 Introduction

High Energy (HE) astronomy typically includes X-ray astronomy, gamma-ray astronomy, neutrino astronomy, studies of cosmic rays, and more recently gravitational wave astronomy. This domain is now sufficiently developed to provide high level data such as catalogs, images, including full-sky surveys for some missions, and sources properties in the shape of spectra and time series. Such high level, HE observations have been included in the VO, via data access endpoints provided by observatories or by agencies and indexed in the VO Registry.

However, after browsing this data, users may want to download lower level data and reapply data reduction steps relevant to their Science objectives. A common scenario is to download HE "event" lists, i.e. lists of detected events on a HE detector, that are expected to be detection of particles (e.g. a HE photon), and the corresponding calibration files, including instrument response functions. The findability and accessibility of these data via the VO is the focus of this note.

We first report typical use cases for data access and analysis of data from current HE observatories. From those use cases, we note that some existing IVOA Recommendations are of interest to the domain. They should be further explored by HE observatories. We then discuss how standards could evolve to better integrate specific aspects of HE data, and if new standards should be developed.

2 Use Cases

Several observatories distribute high-energy data, with their dedicated solutions for data access and analysis.

2.1 Gamma-ray programs

2.1.1 CTAO

The Cherenkov Telescope Array Observatory (CTAO) is the next generation ground-based instrument for gamma-ray astronomy at Very-High Energies (VHE). With 64 telescopes located in the northern and southern hemispheres, the CTAO will be the first open ground-based gamma-ray observatory and the world's largest and most sensitive instrument to study high-energy phenomena in the Universe. Building on the technology of current generation ground-based gamma-ray detectors (H.E.S.S., MAGIC and VERITAS), the CTAO will be between five and 10 times more sensitive and have unprecedented accuracy in its detection of high-energy gamma rays.

CTAO will distribute data as an open observatory, for the first time in this domain, with calls for proposals and publicly released data after a proprietary period. CTAO will ensure that the data provided will be FAIR: Findable, Accessible, Interoperable and Reusable, by following the FAIR Principles for data management (Wilkinson and Dumontier et al., 2016). In particular, because of the complex data processing and reconstruction step, the provision of provenance metadata for CTAO data has been a driver for the development of a provenance standard in Astronomy.

CTAO will also ensure VO compatibility of the distributed data and access systems. CTAO participated to the ESCAPE European Project, and is now part of the ESCAPE Open Collaboration to face common challenges for Research Infrastructure in the context of cloud computing, including data analysis and distribution.

A focus of CTAO is to distribute in this context their Data Level 3 (DL3) datasets, that correspond to lists of Cherenkov events detected by the telescopes along with the proper IRFs.

2.1.2 H.E.S.S

H.E.S.S. is a system of Imaging Atmospheric Cherenkov Telescopes located in Namibia that investigates cosmic gamma rays in the energy range from 10s of GeV to 10s of TeV. It is constituted of four telescopes officially inaugurated in 2004, and a much larger fifth telescope operational since 2012, extending the energy coverage towards lower energies and further improving sensitivity.

The H.E.S.S. collaboration operates the telescopes as a private experiment and published mainly high level data, i.e. images, time series and spectra in scientific publications after dedicated analyses.

In September 2018, the H.E.S.S. Collaboration has, for the first time and unique time, released a small subset of its archival data in Flexible Image Transport System (FITS) format, an open file format widely used in astronomy. The release consists of event-lists and instrument response functions for observations of various well-known gamma-ray sources (H.E.S.S. Collaboration, 2018).

This test data collection has been registered in the VO via a TAP service hosted at the Observatoire de Paris, with a tentative ObsCore description of each dataset. We hope that in the future, the H.E.S.S. legacy archive will be published in a similar way and accessible through the VO.

2.2 X-ray programs

2.2.1 Chandra

The Chandra X-ray Observatory, luanched in 1999, is part of NASA's fleet of "Great Observatories". Chandra allows scientists from around the world to obtain X-ray images of exotic environments to help understand the structure and evolution of the universe.

To be completed: CXC; catalog, data... and VO access.

2.2.2 XMM-Newton

The European Space Agency's (ESA) X-ray Multi-Mirror Mission (XMM-Newton) was launched in 1999. XMM-Newton is ESA's second cornerstone of the Horizon 2000 Science Programme. It carries 3 high throughput X-ray telescopes with an unprecedented effective area, and an optical monitor, dedicated to the study of celestial X-ray sources.

To be completed: XMM catalogs, data... and VO access.

2.2.3 SVOM

The SVOM mission (Space-based multi-band astronomical Variable Objects Monitor) is a Franco-Chinese mission dedicated to the study of the most distant explosions of stars, the gamma-ray bursts. It is to be launched end 2023.

To be completed

2.3 KM3Net and neutrino detection

To be completed

3 Common practices in the High Energy community

3.1 Event-counting

Observations of the Universe at high energies are based on techniques that are radically different compared to the optical, or radio domain. HE observatories are generally designed to detect particles, e.g. individual photons, cosmic-rays, or neutrinos, with the ability to estimate several characteristics of those particles. This technique is commonly named **event-counting**, where an event has some probability of being due to the interaction of an astronomical particle with the detectors.

The data corresponding to an **event** is first an instrumental signal, which is then calibrated and processed to estimate event characteristics such as a time of arrival, coordinates on the sky, and the energy associated to the event. Several other intermediate and qualifying characteristics can be associated to a detected event.

When observing during an interval of time, the data collected is a list of the detected events, commonly named an **event-list** in the HE domain.

3.2 Instrument Response Functions

Though an event-list can express calibrated physical values, this data still have to be corrected for the response of the instruments used. Several Instrument Response Functions (IRFs) thus have to be used to enable a scientific analysis of an event-list. The IRFs are applied to convert the events that were detected into an estimation of the real flux of particles arriving at the instrument and morphology of the source.

3.3 Event selection

When processing an event-list, it is also important to perform an optimal selection of the events that are more likely to be due to the incident particles expected. This selection may depend on the source targeted or on the science objectives.

3.4 Assumptions and probabilistic approach

In order to produce advanced data products like light curves or spectra, assumptions about the kind of particles, noise, source type and its expected energy distribution must be introduced. This is one of the main driver for enabling a full and well described access to event-list data, as scientific analyses generally start from this data level.

3.5 Multi-wavelength and multi-messenger science

Though there are scientific results based on HE data only, the multi-wavelength and multi-messenger approach is particularly developped in the HE domain. An astrophysical source of HE radiations is indeed generally radiating energy in several domains across the electromagnetic spectrum and may be a strong source of other particles. It is not rare to observe a HE source in radio and to look for counterparts in the infrared, optical or UV domain. Spectroscopy is also widely used to identify HE sources.

The HE domain is thus confronted to different kinds of data types and data archives, which leads to interesting use cases for the development of the VO.

3.6 Granularity of data products

The efficient granularity for distributing HE data products seems to be the full combination of data and IRFs, although some of the IRFs may also be recomputed by a service or script after parameters selection.

In order to allow for multi-wavelength data discovery of HE data products and compare observations across different regimes, it seems appropriate to distribute the metadata in the VO ecosystem together with an access link to the data file in community format for finer analysis.

The event-list dataset is generally stored as a table, with one row for a candidate detection (event) and several columns for the estimated physical parameters, at least time, sky position, energy, and extra parameters not standardized across projects: errors, flags, etc.

The list of columns present in the event-list is for example described in the data format in use in the HE domain, such as OGIP or GADF as introduced below. The data formats in use generally describe the event-list data together with the IRFs and other relevant information, such as: Stable or Good Time Interval, Effective Area, Energy Dispersion, Point Spread Function, Background,...

3.7 OGIP

The HEASARC FITS Working Group, also known as the OGIP (Office of Guest Investigator Programs) FITS Working Group, has promoted multi-mission standards for the format of FITS data files in high-energy astrophysics. Those recommendations[1] include standards on keyword usage in

[1] https://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html

metadata, on storage of time information, and representation of response function.

To be completed

3.8 GADF and VODF

The data formats for gamma-ray astronomy[2] (GADF) is a community-driven initiative for the definition of a common and open high-level data format for gamma-ray instruments (Nigro and Hassan et al., 2021). GADF is based on the OGIP standards and is specialised for Very High Energy data.

The Very-high-energy Open Data Format[3] (VODF), is an open data model and format for Very-High-Energy (VHE) gamma-ray and neutrino astronomy. Its goal is to provide a standard set of file formats and standards for data starting at the reconstructed event level as well as higher-level products such as N-dimensional binned data cubes (including sky images, light curves, and spectra) and source catalogues. With these standards, common science tools can be used to analyze data from multiple high-energy instruments. VODF aims to follow as much as possible the IVOA standards.

To be completed

3.9 Multiple Imaging Atmospheric Cherenkov Telescopes extraction example

In order to exploit high energy data across a large interval of energy values, and from various IACTs, there is a need to harmonise metadata description. Datasets can then be mixed together to create a fused event-list dataset, to expand the analysis along the spectral energy axis and study the spectral behaviour of an astronomical object.

This was proposed in (Nigro and Deil et al., 2019) by a group of HE astronomers of various HE facilities. In this work, the authors implemented a prototypical data format (GADF) for a small set of MAGIC, VERITAS, FACT, and H.E.S.S. Crab nebula observations, and they analyzed them with the open-source gammapy software package (). By combining data from Fermi-LAT, and from four of the currently operating imaging atmospheric Cherenkov telescopes, they produced a joint maximum likelihood fit of the Crab nebula spectrum.

Such a work has been more recently extended with the HAWC data (Albert and Alfaro et al., 2022).

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

[3] https://vodf.readthedocs.io/

4 IVOA Recommendations of interest for HE

4.1 ObsCore and TAP

Event-list datasets can be described in ObsCore using a dataproduct_type set to "event". However, this is not widely used in current services, and we observe only a few services with event-list datasets declared in the VO Registry, and mainly the H.E.S.S. public data release (see 2.1.2).

As services based on the Table Access Protocol (Dowler and Rixon et al., 2019) and ObsCore are well developed within the VO, it would a straightforward option to discover HE event-list datasets, as well as multi-wavelength and multi-messenger associated data.

4.2 HiPS

Several HE observatories are well suited for sky survey, and the Hierarchical Progressive Survey (HiPS) standard is well suited for sky survey exploration. We note that the Fermi facility provides a useful sky survey in the GeV domain.

4.3 MOCs

Cross-correlation of data with other observations is an important use case in the HE domain. Using the Multi-Order Coverage map (MOC) standard, such operations become more efficient. Distribution of MOCs associated to HE data should thus be encouraged.

5 Discussion on the possible evolution of Standards

5.1 Definition of a HE event in the VO

The IVOA standards incude the concept of event-list, for example in ObsCore v1.1 (Louys and Tody et al., 2017), where event is a dataproduct_type with the following definition:

> **event**: an event-counting (e.g. X-ray or other high energy) dataset of some sort. Typically this is instrumental data, i.e., "event data". An event dataset is often a complex object containing multiple files or other substructures. An event dataset may contain data with spatial, spectral, and time information for each measured event, although the spectral resolution (energy) is sometimes limited. Event data may be used to produce higher level data products such as images or spectra.

More recently, a new definition was proposed in the product-type vocabulary[4] (draft):

> **event-list**: a collection of observed events, such as incoming high-energy particles. A row in an event list is typically characterised by a spatial position, a time and an energy.

Such a definition remains vague and general, and could be more specific, including a definition for a HE event, and the event-list data type.

5.2 ObsCore metadata description of an event-list

5.2.1 Mandatory fields

ObsCore (Louys and Tody et al., 2017) can provide a metadata profile for a data product of type event-list and a qualified access to the distributed file using the Access class from ObsCore (URL, format, file size).

In the ObsCore representation, the event-list data product is described in terms of curation, coverage and access. However, several properties are simply set to NULL following the recommendation: Resolutions, Polarization States, Observable Axis Description, Axes lengths (set to -1)...

We also note that some properties are energy dependent, such as the Spatial Coverage, Spatial Extent.

- dataproduct_subtype = DL3, maybe specific data format (VODF)
- calib_level = between 2 and 3
- obs_collection could contain many details : obs_type (calib, science), obs_mode (subarray configuration), pointing_mode, tracking_mode, event_type, event_cuts, analysis_type. . .
- s_ra, s_dec = telescope pointing coordinates
- target_name : several targets may be in the field of view
- s_fov, s_region, s_resolution, em_resolution. . . all those values are energy dependent, one should specifiy that the value is at a given energy, or within a range of values.
- em_min, em_max : add fields expressed in energy (e.g. TeV)
- t_exptime : ontime, livetime, stable time intervals. . . maybe a T-MOC would help
- facility_name, instrument_name : minimalist, would be e.g. CTAO and a subarray.

[4] https://www.ivoa.net/rdf/product-type

5.2.2 Additional fields

The searching criteria in terms of time coverage requires the list of stable/-good time intervals to pick appropriate datasets. t_min, t_max is the global time span but t_gti could contain the list of GTI as a T_MOC description following the Multi-Order-Coverage (MOC) IVOA standard (Fernique and Nebot et al., 2022). This element could then be compared across data collections to make the data set selection via simple intersection or union operations in T_MOC representation. On the data provider's side, the T-MOC element can be computed from the Stable/Good Time Interval table in OGIP or GADF to produce the ObsCore t_gti field.

A data release

5.2.3 Access and Description of IRF files

Each IRF file can have an Access object from ObsCore DM to describe a link to the IRF part of the data file. This can be reflected in an extension of ObsTAP TAP_SCHEMA.

In the TAP service we can add an IRF Table, with the following columns:

- event-list datapublisher_id
- irf_type, category of response: EffectiveArea, PSF, etc.
- irf_description, one line explanation for the role of the file
- Access.url, URL to point to the IRF
- Access.format, format of IRF
- Access.size, size of IRF file

5.3 Event-list Context Data Model

The event-list concept may include, or may be surrounded by other connected concepts. Indeed, an event-list dataset alone cannot be scientifically analysed without the knowledge of some contextual data and metadata, starting with the good/stable time intervals, and the corresponding IRFs.

The aim of the Event-list Context Data Model is to name and indicate the relations between the event-list and its contextual information. It is presented in Figure 1.

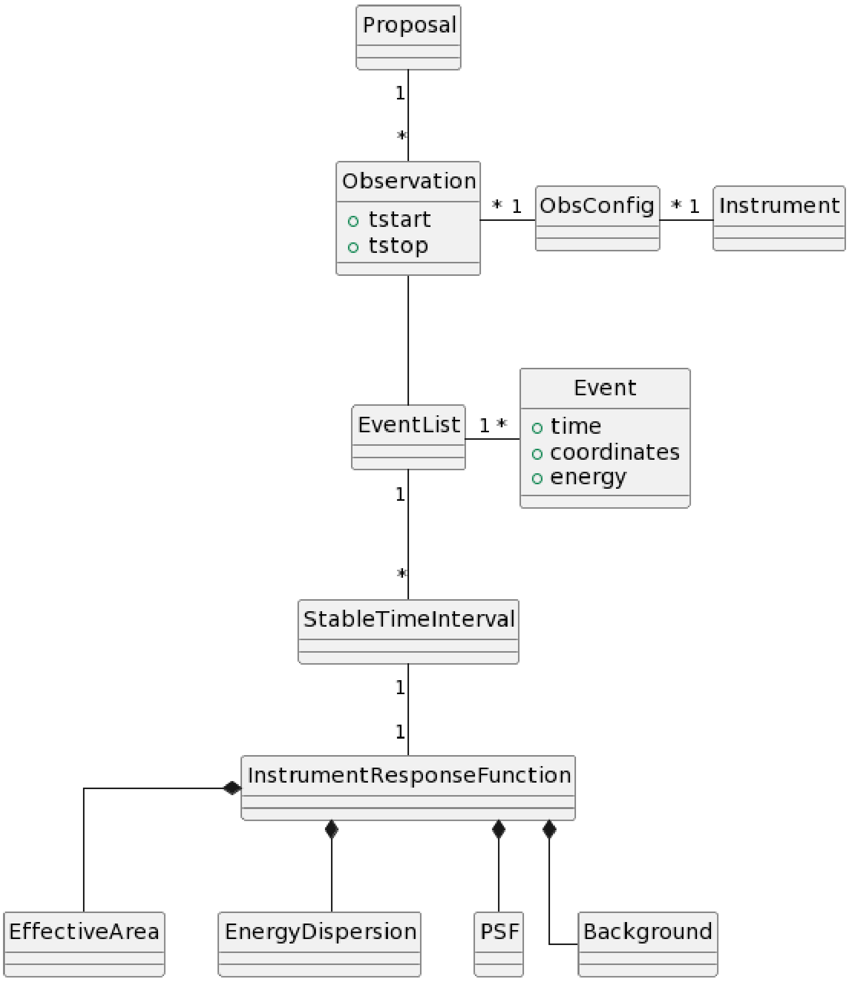

Figure 1: event-list Context Data Model

5.4 Use of Datalink for HE products

There are two options to provide an access to a full event-list package.

First, the event-list dataset itself can contain all the relevant information, e.g. several frames in FITS file, one corresponding to the event-list itself, and the other providing good/stable time intervals, or any IRF file. This is what was done in the current GADF data format (see 3.8).

The second option is to provide links to the relevant information from the base event-list dataset. This could be done using Datalink and a list of link to each contextual information such as the IRFs. The Event-list Context Data Model (see 5.3) would provide the concepts and vocabulary to characterise the IRFs and other information relevant to the analysis of an event-list.

In the first option, The content of the event-list package should be properly defined in its description: what information is included and where is it in the dataset structure? HEre again, the Event-list Context Data Model (see 5.3) would be useful.

5.5 Event-list and the Cube Data Model

To be completed

References

Albert, A., Alfaro, R., Arteaga-Velázquez, J. C., Ayala Solares, H. A., Babu, R., Belmont-Moreno, E., Brisbois, C., Caballero-Mora, K. S., Capistrán, T., Carramiñana, A., Casanova, S., Chaparro-Amaro, O., Cotti, U., Cotzomi, J., Coutiño de León, S., De la Fuente, E., Diaz Hernandez, R., DuVernois, M. A., Durocher, M., Espinoza, C., Fan, K. L., Fernández Alonso, M., Fraija, N., García-González, J. A., Goksu, H., González, M. M., Goodman, J. A., Harding, J. P., Hinton, J., Huang, D., Hueyotl-Zahuantitla, F., Hüntemeyer, P., Jardin-Blicq, A., Joshi, V., Linnemann, J. T., Longinotti, A. L., Luis-Raya, G., Malone, K., Marandon, V., Martinez, O., Martínez-Castro, J., Matthews, J. A., Miranda-Romagnoli, P., Morales-Soto, J. A., Moreno, E., Mostafá, M., Nayerhoda, A., Nellen, L., Nisa, M. U., Noriega-Papaqui, R., Olivera-Nieto, L., Pérez-Pérez, E. G., Rho, C. D., Rosa-González, D., Ruiz-Velasco, E., Salazar-Gallegos, D., Salesa Greus, F., Sandoval, A., Schoorlemmer, H., Serna-Franco, J., Smith, A. J., Son, Y., Springer, R. W., Tollefson, K., Torres, I., Torres-Escobedo, R., Turner, R., Ureña-Mena, F., Villaseñor, L., Wang, X., Watson, I. J., Willox, E., Zhou, H., de León, C., Zepeda, A., HAWC Collaboration, Donath, A. and Funk, S. (2022), 'Validation of standardized data formats and tools for ground-level particle-based gamma-ray observatories', A&A **667**, A36, arXiv:2203.05937.
http://doi.org/10.1051/0004-6361/202243527

Bradner, S. (1997), 'Key words for use in RFCs to indicate requirement levels', RFC 2119.
http://www.ietf.org/rfc/rfc2119.txt

Dowler, P., Rixon, G., Tody, D. and Demleitner, M. (2019), 'Table Access Protocol Version 1.1', IVOA Recommendation 27 September 2019.
http://doi.org/10.5479/ADS/bib/2019ivoa.spec.0927D

Fernique, P., Nebot, A., Durand, D., Baumann, M., Boch, T., Greco, G., Donaldson, T., Pineau, F.-X., Taylor, M., O'Mullane, W., Reinecke, M. and Derrière, S. (2022), 'MOC: Multi-Order Coverage map Version 2.0',

IVOA Recommendation 27 July 2022.
https://ui.adsabs.harvard.edu/abs/2022ivoa.spec.0727F

H.E.S.S. Collaboration (2018), 'H.e.s.s. first public test data release'.
https://zenodo.org/record/1421098

Louys, M., Tody, D., Dowler, P., Durand, D., Michel, L., Bonnarel, F., Micol, A. and 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.
http://doi.org/10.5479/ADS/bib/2017ivoa.spec.0509L

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. and Sinha, A. (2019), 'Towards open and reproducible multi-instrument analysis in gamma-ray astronomy', A&A **625**, A10, arXiv:1903.06621.
http://doi.org/10.1051/0004-6361/201834938

Nigro, C., Hassan, T. and Olivera-Nieto, L. (2021), 'Evolution of data formats in very-high-energy gamma-ray astronomy', *Universe* **7**(10), 374.
http://dx.doi.org/10.3390/universe7100374

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., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. (2016), 'The fair guiding principles for scientific data management and stewardship', *Scientific Data* **3**(1), 160018.
https://doi.org/10.1038/sdata.2016.18

A Changes from Previous Versions

No previous versions yet.