International
Virtual
Observatory
Alliance

Virtual Observatory and High Energy Astrophysics

Version 0.4

IVOA Note 2024-05-16

Working Group
DM
This version
https://www.ivoa.net/documents/VOHE-Note/20240516
Latest version
https://www.ivoa.net/documents/VOHE-Note
Previous versions
This is the first public release
Author(s)
HE club
Editor(s)
Mathieu Servillat

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, and studies of cosmic rays. 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 (IRFs). 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.

1.1 Objectives of the document

The main objective of the document is to analyse how HE data can be better integrated to the VO.

We first identify and expose the specificities of HE data from several HE observatories. Then we intend to illustrate how HE data is or can be handled using current IVOA standards. Finally, we explore several topics that could lead to HE specific recommandations.

A related objective is to provide a context and a list of topics to be further discussed within the IVOA by a dedicated HE Interest Group.

1.2 Scope of the document

This document mainly focuses on HE data discovery through the VO, with the identification of common use cases in the HE Astrophysics domain, which provides an insight of the specific metadata to be expose through the VO for HE data.

To some extend, all current existing IVOA recommandation could be discussed in this document in the HE context.

2 High Energy observatories and experiments

There are various observatories, either from ground- or space-based, that distribute high-energy data with different level of involvement in the VO. We list here the observatories currently represented in the VO HE group. There are also other observatories that are connected to the VO in some way, and may join the group discussions at IVOA.

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 tens of 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. CTAO is planning an internal and a public Science Data Challenge, which represent opportunities to build "VO inside" solutions.

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 100s 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 Cherenkov event-lists and IRFs 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

Part of NASA's fleet of "Great Observatories", the Chandra X-ray Observatory (CXO) was launched in 1999 to observe the soft X-ray universe in the 0.1 to 10 keV energy band. Chandra is a guest observer, pointed-observation mission and obtains roughly 800 observations per year using the Advanced CCD Imaging Spectrometer (ACIS) and High Resolution Camera (HRC) instruments. Chandra provides high angular resolution with a sub-arcsecond on-axis point spread function (PSF), a field of view up to several hundred square arcminutes, and a low instrumental background. The Chandra PSF varies with X-ray energy and significantly with off-axis angle, increasing to R50 $\sim$25 arcsec at the edge of the field of view. A pair of transmission gratings can be inserted into the X-ray beam to provide dispersed spectra with E/DeltaE $\sim$1000 for bright sources. The Chandra spacecraft normally dithers in a Lissajous pattern on the sky while taking data, and this motion must be removed from the time-resolved X-ray event lists when constructing X-ray images using the motion of optical guide stars tracked by the Aspect camera.

The Chandra X-ray Center (CXC) processes the spacecraft data through a set of Standard Data Processing Level 0 through Level 2 pipelines. These pipelines perform numerous steps including decommutating the telemetry data, applying instrument calibrations (e.g., detector geometric, time- dependent gain, and CCD charge transfer efficiency [CTI] corrections, bad and hot pixel flagging), computing and applying the time-resolved Aspect solution to de-dither the motion of the telescope, identifying good time intervals (GTIs), and finally filtering out bad times and X-ray events with bad status. All data products are archived in the Chandra Data Archive (CDA) in FITS format following HEASARC OGIP standards. The CDA manages the proprietary data period (currently 6 months, after which the data become public) and provides dedicated interactive and IVOA-compliant interfaces to locate and download datasets.

The CXC also provides the Chandra Source Catalog, which in the latest release (2.1) includes data for ∼407K unique X-ray sources on the sky and more than 2.1 million individual detections and photometric upper limits. For each X-ray source abd detection, the catalog provides a detailed set of more than 100 tabulated positional, spatial, photometric, spectral, and temporal properties. An extensive selection of individual observation, stacked-observation, detection region, and master source FITS data products (e.g., RMFs, ARFs, PSFs, spectra, light curves, aperture photometry MPDFs) are also provided that are directly usable for further detailed scientific analysis.

Finally, the CXC distributes the CIAO data analysis package to allow users to recalibrate and analyze their data. A key aspect of CIAO is to provide users the ability to create instrument responses (RMFs, ARFs, PSFs, instrument and exposure maps, etc.) for their observations using their choice of spectral models and weightings. The Sherpa modeling and fitting package supports N-dimensional model fitting and optimization in Python, and supports advanced Bayesian Markov chain Monte Carlo analyses.

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 Program. 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 in 2024.

To be completed

2.3 KM3Net and neutrino detection

The KM3NeT neutrino detectors are an array of water-based Cherenkov detectors currently under construction in the deep Mediterranean Sea. With its two sites off the French and Italian coasts the KM3NeT collaboration aims at single particle neutrino detection for neutrino physics with the more densely instrumented ORCA detector in the GeV to TeV range, and high-energy astrophysics with the ARCA detector in the TeV range and above.

Using Earth as a shield from atmospheric particle interference by searching for upgoing particle tracks in the detectors, the measurement of astrophysical neutrinos can be performed almost continuously for a wide field of view that covers the full visible sky. For these particle events, extensive Monte Carlo simulations are performed to evaluate the statistical significance towards the various theoretical assumptions for galactic or cosmic neutrino signals.

During the construction phase, the KM3NeT collaboration develops its interfaces for open science and builds on the data gathered by its predecessor ANTARES, from which neutrino event lists have already been published on the KM3NeT VO server as TAP service. However, reproducibility of the searches for point-like sources require information derived from simulations like background estimate, point spread function and detector acceptance which require linking to the actual event list and interpretation for a given observation, usually as neutrino flux limits for non-significant detection attributable to background rather than an observation.

With multiple detectors targeting high-energy neutrinos like IceCube, ANTARES, KM3NeT, Baikal and future projects, the chance to detect a significant amount of cosmic and galactic neutrinos increases, requiring an integrated approach to link event lists with instrument responses and to correctly interpret observation time and flux expectations.

3 Common practices in the High Energy community

3.1 Data flow specificities

3.1.1 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 generally 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 proxy 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, named an **event list** in the HE domain, and event-list in this document.

3.1.2 Data levels

After the detection of events, data processing steps are applied to generate data products. We may distinguish at least 3 main data levels. However, those data levels can vary significantly from facility to facility, and may not map directly to separate ObsCore calib_levels.

For example, in the VHE Cherenkov astronomy domain (e.g. CTA), those data levels are labelled DL3[1] to DL5. In the X-ray domain, this generally correspond to L1, L2, L3.

1 e.g. L1/DL3: an event-list is first a list of events with calibrated temporal and spatial characteristics, e.g. sky coordinates for a given epoch, time with a reference and a proxy of energy

2 e.g. L2/DL4: the event-list can then be binned or filtered to prepare the generation of science images, spectra or light-curve, and corresponding instrument response correction are associated or calculated but not yet applied (exposure maps, sensitivity maps...)

3 e.g. L3/DL5: Calibrated maps, or spectral energy distributions for a source, or light-curves in physical units

For observations that use transmission gratings (e.g. for chandra or XMM-Newton), grating data products are created in an intermediate L1.5.

An additional data level corresponds to catalogs, e.g. a source catalog pointing to several data products (e.g. collection of L3 products), each one corresponding to a source.

3.1.3 Background signal

Observations in HE may contain a high background component, that may be due to instrument noises, or to unresolved astrophysical sources, emission from extended regions or other terrestrial sources producing particles similar to the signal. The characterization and estimation of this background may

---

[1] events being reconstructed, lower level data is specific this domain (DL0-DL2).

be particularly important to then apply corrections during the analysis of a source signal.

In the VHE domain with the IACT, WCD and neutrino techniques, the background is created by cosmic-ray induced events. The case of unresolved astrophysical sources, emission from extended regions are treated as a model of a gamma-ray or neutrino emission.

3.1.4 Time intervals

Depending on the stability of the instruments and observing conditions, a HE observation can be decomposed into several intervals of time that will be further analysed. For example, Stable Time Intervals (STI) are defined in Cherenkov astronomy to characterize the instrument response over a stable period of time. In the X-ray domain, Good Time Intervals (GTI) are computed, e.g. to reject intervals of time contaminated by solar flares. In contrast, for neutrino physics, relevant observation periods can cover up to several years due to the low statistics of the expected signal and a continuous observational coverage of the full field of view.

3.1.5 Instrument Response Functions

Though an event-list can contain calibrated physical values, this data still have to be corrected for the response of the instruments used. Several IRFs thus have to be applied 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.

3.1.6 Granularity of data products

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 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, e.g. for X-ray data, so further files allowing for this reprocessing could also be considered to be part of a package.

The event-list dataset is generally stored as a table, with one row per candidate detection (event) and several columns for the estimated physical parameters, e.g. arrival time, position (on detector or in the sky), energy or pulse height, and different extra indicators 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,...

3.2 Work flow specificities

3.2.1 Event selection

When processing an event-list, it is 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. The selection can be performed on the event characteristics, e.g. time, energy or more specific indicators (patterns, shape...)

3.2.2 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 at this data level.

3.3 Data formats

3.3.1 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[2] include standards on keyword usage in metadata, on storage of time information, and representation of response function.

To be completed

3.3.2 GADF and VODF

The data formats for gamma-ray astronomy[3] (GADF) is a community-driven initiative for the definition of a common and open high-level data format for gamma-ray instruments (Deil and Boisson et al., 2017; Nigro and Hassan

---

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

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

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[4] (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: Bruno

3.4 Tools for data extraction and visualisation

HE data is particularly complex and diverse at lower levels. It is common to find specific tools to process the data for a given facility, e.g. CIAO for Chandra, SAS fro XMM-Newton, of Gammapy for gamma-ray data, with a particular focus on Cherenkov data as foreseen for CTA.

Those tools can generally handle data from several other obsevatories, that have some level of commonalities.

Several other HE software are build to handle the existing data format standards, hence enabling mutli-instrument studies, e.g. XSpec, Sherpa, or Gammapy.

To be completed (e.g. ???)

4 Use Cases

4.1 UC1: re-analyse event-list data for a source in a catalog

After the selection of a source of interest, or a group of sources, one may access different HE data products such as images, spectra and light-curves, and then want to download the corresponding event-lists and calibrations to further analyse the data.

One of the characteristics of the HE data is that, contrary to what is usually done in optics for example, their optimal use requires providing users with a view of the processing that generated the data. This implies providing ancillary data, products with different calibration levels, and possibly linking together products issued by the same processing.

---

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

4.2 UC2: observation preparation

When planning for a new HE observation, one needs to search for any existing event-list data already available in the targeted sky regions, and assess if this data is sufficient to fulfill the science objectives.

To be completed (e.g. Bruno)

4.3 UC3: transient or variable sources

To be completed (e.g. Ada)

4.4 UC4: 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 developed 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.

To be completed (e.g. Bruno)

4.5 Examples of multi-wavelength analysis

4.5.1 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), and included neutrino data in a common CTA and KM3NeT source search (T. Unbehaun, 2023).

5 IVOA standards of interest for HE

5.1 IVOA Recommendations

5.1.1 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 be a straightforward option to discover HE event-list datasets, as well as multi-wavelength and multi-messenger associated data.

Here is the evaluation of the ObsCore metadata for distributing high energy data set, some features being re-usable as such, and some other features requested for addition or re-interpretation.

5.1.2 DataLink

DataLink specification (Bonnarel and Dowler et al., 2023) defines a {links} endpoint providing the possibility to link several access items to each row of the main response table. These links are described and stored in a second table. In the case of an ObsCore response each dataset can be linked this way (via the via the access_url FIELD content) to previews, documentation pages, calibration data as well as to the dataset itself. Some dynamical links to web services may also be provided. In that case the service input parameters are described with the help of a "service descriptor" feature as described in the same DataLink specification.

5.1.3 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.

5.1.4 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 and especially ST-MOCs (space + time coverage) that make easier the study of transient phenomena.

5.1.5 MIVOT

Model Instances in VOTables (MIVOT) defines a syntax to map VOTable data to any model serialized in VO-DML. The annotation operates as a bridge between the data and the model. It associates the column/param metadata from the VOTable to the data model elements (class, attributes, types, etc.) [...]. The data model elements are grouped in an independent annotation block complying with the MIVOT XML syntax. This annotation block is added as an extra resource element at the top of the VOTable result resource. The MIVOT syntax allows to describe a data structure as a hierarchy of classes. It is also able to represent relations and composition between them. It can also build up data model objects by aggregating instances from different tables of the VOTable. In the case of HE data, this annotation pattern, used together with the MANGO model, will help to make machine-readable quantities that are currently not considered in the VO, such as the hardness ratio, the energy bands, the flags associated with measurements or extended sources.

To be completed

5.1.6 Provenance

To be completed (e.g. Mathieu)

5.2 Data Models in working drafts

The HE domain and practices could serve as use cases for the developments of data models, such as Dataset DM, Cube DM or MANGO DM.

6 Topics for discussions in an Interest Group

6.1 Definition of a HE event in the VO

6.1.1 Current definition 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[5] (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.

6.1.2 Proposed definition to be discussed

A first point to be discuss would be to converge on a proper definition of HE specific data products:

- Propose definitions for a product-type **event-list**: A collection of observed events, such as incoming high-energy particles, where an event is generally characterised by a spatial position, a time and a spectral value (e.g. an energy, a channel, a pulse height).

- Propose definitions for a product-type **event-bundle**: An event-bundle dataset is a complex object containing an event-list and multiple files or other substructures that are products necessary to analyze the event-list. Data in an event-bundle may thus be used to produce higher level data products such as images or spectra.

An ObsCore erratum could then propose to change event for event-list and event-bundle.

The precise content of an event-bundle remains to be better defined, and may vary significantly from a facility to another.

For example, Chandra primary products distributed via the Chandra Data Archive include around half a dozen different types of products necessary to analyze Chandra data (for example, L2 event-list, PHA spectrum, Aspect solution, bad pixel map, spacecraft ephemeris, V&V Report). It is also possible to retrieve secondary products, containing more products that are needed to recalibrate the data with updated calibrations.

---

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

6.2 ObsCore metadata description of an event-list

6.2.1 Usage of the mandatory terms in ObsCore

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, PSF.

TODO: show a table with all reused terms , and provide an example

- dataproduct_subtype = DL3, maybe specific data format (VODF)
- calib_level = between 1 and 2
- obs_collection could contain many details : obs_type (calib, science), obs_mode (subarray configuration), pointing_mode, tracking_type, event_type, event_cuts, analysis_type. . .
- s_ra, s_dec = maybe 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. eV, keV or 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.

6.2.2 Metadata re-interpretation for the VOHE context

observation_id In the current definition of ObsCore, the data product collects data from one or several observations. The same happens in HE context.

access_ref, access_format The initial role of this metadata was to hold the access_url allowing data access. Depending on the packaging of the event bundle in one compact format (OGIP, GADF, tar ball, ...) or as different files available independently in various urls, a datalink pointer can be used for accessing the various parts of IRFs, background maps, etc. Then in such a case the value for access_format should be "application/x-votable+xml;content=datalink". The format itself of the data file is then given by the datalink parameter "content-type". See next section 6.4.

o_ucd For the even-list table, we can consider all measures stored in columns values have been observed . The nature of items along time, position and energy axis are identifed in Obscore with ucd as 'time', 'pos.eq.*', 'em.*' and counted as t_xel, s_xel1, s_xel2, em_xel which correspond to the number of rows/events candidates observed.

The signal observed is the result of event counting and would be PHA (Pulse height amplitude at detector level) or a number of counts for photons or particles, or a flux, etc.., depending on the data calibration level considered. ObsCore uses o_ucd to characterise the nature of the measure. various UCDs are used for that: o_ucd=phys.count, phot.count, phot.flux, etc. there is currently no UCD defined for a raw measure like PulseHeightAmplitude, but if needed this can be requested for addition in the UCDList vocabulary. See VEP-UCD-15_pulseheight.txt proposed at 'https://voparis-gitlab.obspm.fr/vespa/ivoa-standards/semantics/vep-ucd/-/blob/master/'.

Note that these parameters vary between the dataset of calib_level of 1 (Raw) to the a more advanced data products (calib_level 2 or 3), which are filtered and rebinned from the original raw event-list.

6.2.3 Metadata addition required

ev_number The event list contains a number of rows, representing detections candidates, that have no metadata keyword yet in Obscore. We propose 'ev_number' to record this. In fact the t_xel, s_xel1 and s_xel2, em_xel elements do not apply for an event list in raw count as it has not been binned yet.

Adding MIME-type to access_format table As seen in section 3.3 current HE experiments and observatories use their community defined data format for data dissemination. They encapsulate the event-list table together with ancillary data dedicated to calibration and observing configurations and parameters. Even if the encapsulation is not standardized between the various projects, it is useful for a client application to rely on the access_format property in order to send it to an appropriate visualizing tool.

Therefore these can be included in the MIME-type table of ObsCore section 4.7. suggestion for new terms like :

- application/x-fits-ogip ...
- application/x-gadf ...
- application/x-vodf ...

to be completed with proper definition

energy_min, energy_max It is not user-friendly for the user to select dataset according to an energy range when the spectral axis is expressed in wavelength and meters. The units and quantities are not familiar to this community. Moreover the numerical representation of the spectral range in em_min leads to quantities with many figures and a power as -18 not easily comparable with the current usage.

cf. example HESS data shown in Aladin

t_gti The searching criteria in terms of time coverage require 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.

6.2.4 Access and Description of IRFs

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 could 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

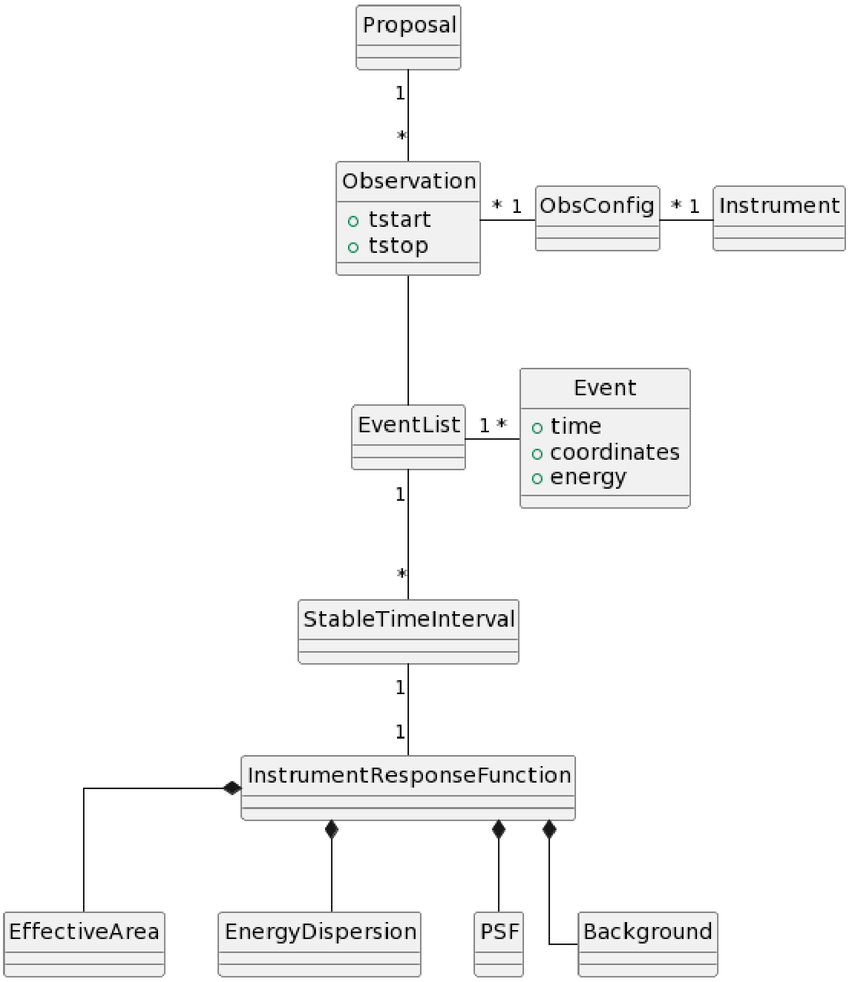

Figure 1: event-list Context Data Model. Notes: STIs and GTIs are slightly different concepts, and multiplicities should be adapted, energy is to specific for an event (intensity?), more products may be attached to a STI/GTI or to IRF.

- Access.format, format of IRF
- Access.size, size of IRF file

6.3 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.

6.4 Use of Datalink for HE products

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

In the first option, the "event-bundle" dataset (6.1) exposed in the discovery service contains all the relevant information, e.g. several frames in the FITS file, one corresponding to the event-list itself, and the others providing good/stable time intervals, or any IRF file. This is what was done in the current GADF data format (see 3.3.2). In this 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? Obviously the Event-list Context Data Model (see 6.3) would be useful to provide that.

In the second option we provide links to the relevant information from the base "event-list" (6.1) exposed in the discovery service. This could be done using Datalink and a list of links to each contextual information such as the IRFs. The Event-list Context Data Model (see 6.3) would provide the concepts and vocabulary to characterise the IRFs and other information relevant to the analysis of an event-list. These specific concepts and terms describing the various flavors of IRFs and GTI will be given in the semantics and content_qualifier FIELDS of the DataLink response to qualify the links. The different links can point to different dereferencable URLs or alternbatively to different fragments of the same drefereencable URL as stated by the DataLink specification.

To be completed: show an example ?

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. doi:10.1051/0004-6361/202243527, https://ui.adsabs.harvard.edu/abs/2022A&A...667A..36A.

Bonnarel, F., Dowler, P., Michel, L., Demleitner, M. and Taylor, M. (2023), 'IVOA DataLink Version 1.1', IVOA Recommendation 15 December 2023. https://ui.adsabs.harvard.edu/abs/2023ivoa.spec.1215B.

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

Deil, C., Boisson, C., Kosack, K., Perkins, J., King, J., Eger, P., Mayer, M., Wood, M., Zabalza, V., Knödlseder, 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. and Lombardi, S. (2017), 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*, AIP, p. 070006. doi:10.1063/1.4969003, https://ui.adsabs.harvard.edu/abs/2017AIPC.1792g0006D.

Dowler, P., Rixon, G., Tody, D. and Demleitner, M. (2019), 'Table Access Protocol Version 1.1', IVOA Recommendation 27 September 2019. doi:10.5479/ADS/bib/2019ivoa.spec.0927D, https://ui.adsabs.harvard.edu/abs/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'. doi:10.5281/ZENODO.1421098, 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. doi:10.5479/ADS/bib/2017ivoa.spec.0509L, https://ui.adsabs.harvard.edu/abs/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. doi:10.1051/0004-6361/201834938, https://ui.adsabs.harvard.edu/abs/2019A&A...625A..10N.

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

T. Unbehaun, L. Mohrmann, e. a. (2023), 'Prospects for combined analyses of hadronic emission from γ-ray sources in the milky way with cta and km3net', arXiv:2309.03007.

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. doi:10.1038/sdata.2016.18.

A Changes from Previous Versions

No previous versions yet.