



*International
Virtual
Observatory
Alliance*

IVOA ObsCore Extension and Discovery of High Energy Astrophysics Data

Version 1.0

Proposed Endorsed Note 2025-06-26

Working Group

High Energy Interest Group

This version

<https://www.ivoa.net/documents/HighEnergyObsCoreExt/20250626>

Latest version

<https://www.ivoa.net/documents/HighEnergyObsCoreExt>

Previous versions

This is the first public release

Author(s)

I. Evans, M. Servillat, B. Khélifi, J. Evans, M. Louys, M. Kettenis,
F. Bonnarel, L. Michel, B. Boisson, M. Cresitello-Dittmar, O. Ates,
K. Kosack, and the IVOA High Energy Interest Group

Editor(s)

The IVOA High Energy Interest Group

Abstract

This document describes a proposed extension to the ObsCore specification for data description, discovery and selection of **High Energy Astrophysics (HEA)** data. It includes proposed updates to the data product vocabulary, UCDs, and MIME-types to support discovery of **HEA** data.

Status of this document

This is an IVOA Proposed Endorsed Note for review by IVOA members and other interested parties. It is appropriate to reference this document only as a Proposed Endorsed Note that is under review and may change before it is endorsed or may not be endorsed.

A list of current IVOA Recommendations and other technical documents can be found in the IVOA document repository¹.

Contents

1	Introduction	5
2	High Energy Astrophysics Data	5
3	ObsCore Attribute Definitions for High Energy Astrophysics Data	7
3.1	<i>datapduct_type</i>	7
3.2	<i>datapduct_subtype</i>	10
3.3	<i>calib_level</i>	10
3.4	<i>access_url</i>	11
3.5	<i>access_format</i>	11
3.6	<i>s_ra/s_dec</i>	11
3.7	<i>s_calib_status</i>	12
3.8	<i>t_calib_status</i>	12
3.9	<i>em_calib_status</i>	13
3.10	<i>o_ucd</i>	13
3.11	<i>proposal_id</i>	14
4	Extensions to ObsCore Specific to High Energy Astrophysics Data	14
4.1	<i>ev_xel</i>	14
4.2	<i>s_ref_energy/em_ref_energy/s_ref_oaa/em_ref_oaa</i> . . .	14
4.3	<i>t_intervals</i>	15
4.4	<i>energy_min/energy_max</i>	15
4.5	<i>obs_mode</i>	16
4.6	<i>tracking_mode</i>	16
4.7	<i>analysis_mode</i>	17
4.8	<i>event_type</i>	17
4.9	Additional Columns	18

¹<https://www.ivoa.net/documents/>

5	Vocabulary Enhancements	18
5.1	Evolution of the Data Product Type vocabulary	18
5.1.1	Event list	18
5.1.2	Response functions	19
5.1.3	Event bundle	19
5.1.4	Advanced data products	20
5.1.5	Summary table	22
5.1.6	Clarification of “flux” in some term definitions	24
5.2	DataLink vocabularies	25
5.3	UCD Enhancements	25
5.3.1	Pulse Height	25
5.3.2	Electromagnetic spectrum description in UCD	25
5.3.3	Event Type	25
5.3.4	Particles	26
5.3.5	Statistical UCDs	26
5.3.6	Evolution of UCD list	26
5.4	MIME-types Enhancements	27
6	Proposal of the <i>ivoa.obscore_healpix</i> attributes	27
7	Summary	29
	Glossary	30
	References	30
A	Detailed Science Use Cases for ObsCore	32
A.1	Event-List Data and Responses	32
A.1.1	Use Case — Search for event lists surrounding Sgr A*, for exemple for an X-ray morphological study	32
A.1.2	Use Case — Search for event bundles that include Cas A for X-ray spectrophotometric evolution studies	32
A.1.3	Use Case — Search for event bundles via DataLink that include Cas A for a TeV spectromorphology study	33
A.1.4	Use Case — Identify PSF response-functions for fur- ther analysis of previously downloaded data products	34
A.1.5	Use Case — Get all the Instrument Response Func- tions (IRFs) for a given CTAO observation, for simu- lation purposes	34
A.1.6	Use Case — Search for event lists that include a fully calibrated spectral axis for BL Lac for rapid X-ray spectrophotometric evaluation	35

A.1.7	Use Case — Search for spatially resolved spectropolarimetric observation of the Crab with spectral resolution $R > 100$	35
A.1.8	Use Case — Search for SWGO event lists and their IRFs for the event type ‘very-good’ in the region of Cygnus loop for a TeV spectromorphology study	36
A.1.9	Use Case — Search for event lists and their IRFs of CTAO South observations at energies above 10 TeV for blind search of PeVatrons from a data release using DataLink	37
A.2	Very-High-Level Data Products	38
A.2.1	Use Case — Search for Chandra Source Catalog position error MCMC draws for X-ray detections in the vicinity of Gaia DR3 486718823701242368	38
A.2.2	Use Case — Search for flux maps for CTAO-North observations between two observations ID	39
A.2.3	Use Case — Search for M31 source light curves and aperture photometry probability density functions that intersect a specific time interval	39
A.2.4	Use Case — Search for the CTAO flux light curves of PKS 2155-304 in 2030	40
B	Changes from Previous Versions	41
C	Contributions to the Note	41

Acknowledgments

The authors would like to thank all the participants in the IVOA High Energy Interest Group and Data Model Working Group for their ideas, discussions, critical reviews, and contributions to this document.

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

The International Virtual Observatory Alliance (IVOA) High Energy Interest Group (HEIG) was formed in the Fall of 2024, and developed an IVOA Note (Servillat and Boisson et al., 2024) that explores the connections between the VO and HEA. Here, the HEA covers experiments and observatories from the X-ray range up to the PeV range, as well as the astrophysical neutrinos above the TeV range, called here the HEA domain. The HEIG Note includes an outline of several important topics that have formed a roadmap for the group. An ObsCore (Louys and Tody et al., 2017) extension for HEA data is the first priority in order to meet the needs of HEA, and to coincide with similar work being carried out by the Radio IG, Time Domain IG, and discussions on DM standards, such that current and future HEA experiments and observatories are able to release data on the VO.

The goal is to explore elements needed to reliably discover and select HEA data through IVOA interfaces. It implies defining an extension to ObsCore with the possibility to use the DataLink mechanism and to enhance vocabularies of keywords for ObsCore and DataLink. We suggest that, if an attribute is unique to HEA data, that element should appear in an HEA ObsCore extension. Whereas, if an attribute makes sense for more than one domain and can be shared across those domains, then that element should be added to the base ObsCore model. This note proposes recommendations in both of these categories. We also discuss enhancements to the vocabulary of data products, DataLink semantics, UCDs and MIME-types to correctly represent HEA data. Topics related to the Registry are currently outlined in the proposed Radio extension document and are not discussed here.

2 High Energy Astrophysics Data

HEA data include observations obtained using photon detectors covering X-ray (from ~ 0.1 eV to ~ 100 keV) through gamma-ray (from 100 MeV up to $\gtrsim$ PeV) energies, as well as cosmic-ray and astrophysical neutrino ($\gtrsim$ TeV) detectors, or other messenger related to HEA phenomena. The domain is now sufficiently mature to provide open data that are science-ready and work with open analysis tools (*e.g.* CIAO (Fruscione and McDowell et al., 2006) or Gammapy (Donath and Terrier et al., 2023)). The science output of the HEA domain already includes high-level products such as images, cubes, spectra, and time series such as light curves and time-resolved spectra. Additional data products include fitted sky models with a spatial, spectral and/or temporal component(s), along with their confidence intervals or confidence limits, and covariance matrices. Finally, multiple HEA instruments produce source catalogs and surveys covering up to the full the sky, which include maps of photon or particle flux, exposure, sensitivity, and

aperture-photometry likelihood profiles.

Observations of the Universe at the highest energies are based on techniques that are radically different compared to the UV through radio domains. **HEA** observatories² are generally designed to detect particles (*e.g.*, individual photons, cosmic-rays or neutrinos) with the ability to estimate multiple observables for those particles. These detection techniques rely all on *event counting*³, where an event has some probability of being due to the interaction of an astronomical particle with the detectors, but also some probability from being from instrumental or background effects. The data corresponding to an event are first an instrumental signal, which is then calibrated and processed to estimate physical quantities such as a time of arrival, point-of-origin on the sky, and an energy proxy associated with the event. Several other intermediate and qualifying characteristics may be associated with a detected event, depending on the detection technique. The ensemble of events detected over a given time interval and spatial field-of-view is referred to as an *event list*.

Though **event-lists** *may* include estimators for calibrated physical values, they typically still have to be corrected for the photometric, spectral, spatial, and/or temporal responses of the telescope and detector combination to yield scientifically interpretable information. The mappings between the observables and physical measurements of the source properties are called **IRFs**⁴; using techniques like forward-folding, they enable one to fit a model (with any combination of spectral, spatial, and temporal components) of the true flux of particles from a source arriving at the instrument to the measured quantities. The **IRFs** generally evolve over time with the instrument and observation characteristics, and are usually defined for a specific time interval and decomposed into a standard set of independent components (see section 3.1.5 of (Servillat and Boisson et al., 2024)), such as the spatial point-spread function or the energy-migration matrix, where each component may be stored or computed separately. Some of the **IRFs** are probabilistic in nature⁵, and in addition may depend on the set of events selected for analysis by the end user. They are usually not invertible, so methods such as forward-folding fitting (using source models that are estimated) are needed to estimate physical properties given observables. Since both **IRFs** and **event-lists** are required to process **HEA** data, some **IVOA**

²One can cite Chandra, XMM-Newton, Fermi, H.E.S.S., MAGIC, VERITAS, HAWC, LHAASO, IceCube, Auger and soon CTAO and KM3NeT, SWGO.

³As opposed to signal integrating using a light detector.

⁴We try to avoid using the term **IRF** in a normative sense since historical usage across the broad **HEA** community (and from facility to facility) varies. In some cases, **IRF** has been used to mean specifically the X-ray product of the “ARF” and “RMF”, whereas in other cases **IRF** has been used more generally to mean any instrumental response function regardless of type.

⁵The energy matrix is a probability density function

standards must be modified in order to expose both of them via the VO.

In the following, the current ObsCore standard will be discussed in Section 3, focusing on attributes that need to be modified. Then, we propose the creation of a HEA extension of ObsCore in Section 4, as some attributes are very specific to our domain (see §6). In these two sections, the discussion focuses on the attribute definitions rather on the attribute values. In Section 5, enhancement of vocabulary is proposed for some ObsCore attributes, DataLink semantics, UCDs and MIME-types.

3 ObsCore Attribute Definitions for High Energy Astrophysics Data

The ObsCore representation of any HEA **event-list** data products is described in terms of curation, coverage, and access. However, given the HEA data specificities, several properties, including resolutions, observable axis descriptions, and polarization states would be simply set to “NULL”, and data axis lengths are set to “−1”. Therefore, for these data products and associated IRFs, the definitions of some ObsCore attributes should be adjusted so that they better represent the content of the data from the perspective of data discovery. We note that many properties, including spatial and spectral coverage and resolution can vary strongly with energy and off-axis angle. These adjustments will also typically apply to advanced, high-level data products derived from **event-list** data.

In addition, the hereafter modification proposal faces to the issue that some values of ObsCore attributes (*dataprodect_type* and *calib_level*) are defined both into the ObsCore standard document (Louys and Tody et al., 2017) and in the vocabularies documents (Demleitner and Gray et al., 2023, 2021), which might create some issues for the users. In this context, we have opted to propose in this document some modifications of both standards, even if we would have preferred that everything is uniquely defined in the IVOA Vocabulary. Some harmonization should be taken by the Data Model and Semantics working groups in order to avoid duplications. But until such work is achieved, we require modifications in ObsCore and Vocabulary.

3.1 *dataprodect_type*

The attribute *dataprodect_type* provides a scientific classification of the data product and is of primary importance for data discovery, especially when there may be many different types of data product associated with an observation⁶ (as is often the case for HEA datasets).

⁶We use the term “observation” in the broad sense, as is done in the ObsCore Recommendation. We note that in this context an “observation” may not correspond to a single pointed observation defined in the traditional sense.

The ObsCore v1.1 recommendation (Louys and Tody et al., 2017) only defines an **event** *datapduct_type* as:

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.

We propose to add the following *datapduct_type* term in both the ObsCore standard and into the IVOA Vocabulary of Data Product Types⁷ to better define an **HEA event-list** and an **event-bundle** that includes the event-list and its associated data:

event-list: a collection of observed particle-detection events, such as incoming high-energy particles. The table of event list is typically characterised by a spatial position, a time and an energy proxy.

event-bundle: compounded dataset 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 calibrated in physical units when containing **IRFs** or other data products that can be used to construct **IRFs**.

It may be worth mentioning that the term “event” caused confusion in the past, as it also is used for astrophysical events like supernova explosions (*e.g.* VOEvent), and that is not the type of event that is being described here, which are particle detection events. Using “event-list” was meant to help to resolve this ambiguity.

In addition to *datapduct_type* terms that focus on event data, we note that existing ObsCore definitions do not adequately span the breadth of advanced data products (with *calib_level* ≥ 3) that may be generated from astronomical observations by users or observatories. The computational complexity of analyzing **HEA** data robustly in the extreme Poisson regime (*e.g.*, Bayesian X-ray aperture photometry applied simultaneously to multiple overlapping detections and observations) means that data providers may choose to provide such analysis products directly to the end user. For example, the Chandra Source Catalog includes 38 types of advanced data products

⁷See <https://www.ivoa.net/rdf/product-type>.

(for a total of ~ 90 million files) and $\sim 50\%$ of these data product types are not well represented by a *dataprodect_type* value that allows for meaningful data discovery. Users will certainly want to discover these data products independently from the associated observation data (and many of these data products combine data from multiple observations). We therefore propose the following additional *dataprodect_type* (or *dataprodect_subtype*) terms for these advanced data products, and note that these terms will certainly be useful independent of waveband (*i.e.*, they can be equally applicable to UV/optical, IR, and radio datasets):

draws: a dataset that represents draws computed from a probability distribution, for example the Markov chain Monte Carlo (MCMC) draws used when computing the Bayesian marginal probability density function for a random variable. The draws can be interpreted to provide a robust estimation of the probability distribution of variable, and correlations between the draws provide information about how well the draws converge to the parent probability distribution.

pdf: a dataset that represents the probability density function of a quantity, for example the Bayesian marginal probability density function for a random variable. The probability density function provides a robust estimation of the variable and allows arbitrary confidence intervals to be computed directly from the distribution.

region: a dataset that includes an encoding of (one or more) regions of parameter space, for example a spatial region or a region of phase space covered by a dataset. The set of dimensions represented by the region can be arbitrary.

response-function: a dataset that represents a mapping from a physical quantity to an observable. For **HEA**, this may be the components of the composite **IRF** such as an Auxiliary Response File (**ARF**), Redistribution Matrix File (**RMF**), Effective Area (**AEFF**), Energy Dispersion (**EDISP**), the Background Rate (**BKGRATE**). The Point Spread Function (**PSF**) is a response function that is generally applicable across multiple wavebands. While these datasets may generally be represented as an N-dimensional data cube, designating them as **response-functions** enhances data discovery for very common types of **HEA** dataset (see the use cases in appendix A).

The **measurements** data product type is quite useful for many different types of advanced data products (that may be derived from multiple observations) but users of those products often may not be interested the progenitor

datasets, especially if many advanced data products are extracted from a single or a few progenitors (*e.g.*, measurements associated with sources detected in a single observation field). We propose to delete the caveat associated with **dataprod-uct_type** = “measurements” in the ObsCore IVOA Recommendation (§4.1.1) that requires the derived data products be exposed “together with the progenitor observation dataset”.

Note that these terms will be repeated in the section 5, as mentioned in the introduction of this sub-section.

3.2 *dataprod-uct_subtype*

The optional attribute *dataprod-uct_subtype* may be used by the data provider to specify additional information about the nature of the data product. For some datasets this attribute may specify the data level (*e.g.*, DL3–6, see [Servillat and Boisson et al. 2024](#), §3.1.2), or may be combined with *dataprod-uct_type* to more precisely define the content of the dataset (*e.g.*, *dataprod-uct_type* = **image** + *dataprod-uct_subtype* = **exposuremap**). A vocabulary of such data product (sub-)types is proposed in section 5 to support discovery of advanced data products.

3.3 *calib_level*

ObsCore defines calibration **Level 1** as “Instrumental data in a standard format (FITS, VOTable, SDFITS, ASDM, etc.) which could be manipulated with standard astronomical packages.” and **Level 2** as “Calibrated, science ready data with the instrument signature removed.”

However, some **HEA event-lists** include spatial and time axes that are calibrated physical quantities, but the spectral axis is instrumental and requires application of the IRFs to remove this signature. In X-ray, this is typically done because the **response-functions** can depend on the choice of region (spatial/time) from which the events are extracted (especially for telescope/detector combinations where the telescope position dithers on the sky during the exposure), which depends on the specific science case and therefore cannot be determined *a priori*. Such **event-lists** fall “between” *calib_level* 1 and 2.

On the other hand, other **event-lists** may not have any calibrated axes or may have all axes calibrated, and it is important to be able to differentiate between these for data discovery. While the value for *calib_level* for any data product is left for the data provider to determine, we suggest that individual data providers set *calib_level* = 1 if an **event-list** is considered to be “uncalibrated” according to normal usage for their data products and set *calib_level* = 2 if an **event-list** is considered to be “calibrated” according to normal usage for their data products.

Also, we propose that the calibration status of the spatial/spectral/time data axes be identified using the appropriate axis ObsCore *calib_status* keyword (*s_calib_status* for the spatial axes, *em_calib_status* for the spectral axis, and *t_calib_status* for the time axis).

3.4 *access_url*

Given the complexity and number of HE data products, the *access_url* may point either directly to a file (*e.g.* to the **event-list** or an **event-bundle**), or to a DataLink service that will provide links to the data and to associated data (*e.g.* **response-functions**).

If DataLink is provided, it should be indicated that the URL points to a Datalink service via the *access_format* = application/x-votable+xml;content=datalink.

If the *access_url* points to a bundle, the detailed content of the bundle is not exposed; therefore using a DataLink service has advantages.

3.5 *access_format*

The *access_format* attribute specifies the format of the data product when downloaded as a file from the *access_url*. The analysis of **HEA** data often requires use of multiple, related data products, for example an **event-list** combined with associated **IRFs** or ancillary files that can be employed by the user to create **IRFs**. These associated products are often bundled together with the **event-list** and we proposed in § 3.1 to assign such bundles *datapduct_type* = **event-bundle**. While these bundles are typically not standardized across different projects, knowledge of the bundle content is useful for client applications to properly handle the bundles (for example to send the data to an appropriate visualization tool). This is readily achieved by encoding an appropriate MIME-type using the *access_format* attribute. In Section 5.4 we propose additional MIME-types for some common **event-bundles**.

3.6 *s_ra/s_dec*

We propose that the attributes *s_ra/s_dec* be redefined to be the ICRS right ascension and ICRS declination of “a reference position (typically the center)” of an observation on the sky, rather than the ICRS right ascension and ICRS declination of “the center” of the observation. The center of an observation often is not useful for advanced data products that may be extracted from a cut-out from the progenitor observation, and many facilities allow an instrument to be displaced from the optical axis of the telescope, which means that the definition of “the center” of an observation may be unclear (especially when the tracking is not fixed in the ICRS system).

For non-pointing instruments (which may include all-sky instruments such as KM3NeT or HAWC), these fields are poorly defined (as is the case, generally for observations that are drift scans). For the time duration of the observation, one can compute an effective center position of the exposure skymap and the maximum radius of the covered area (*i.e.*, for an all-sky instrument this would be 2π Sr solid angle in Alt/Az, which can be converted into a rotated area in RA/Dec). However, the utility of such a characterization depends on both the duration of the observation and the use case.

3.7 *s_calib_status*

We propose that *s_calib_status* encode the calibration status of an **event-list** dataset’s spatial axes. Where multiple spatial axes are included in a dataset (*e.g.*, physical detector pixel coordinates, virtual detector coordinates corrected for distortions, world coordinates), then we recommend that the data provider use the coordinate system that is most likely to be preferred by the end user (typically the most fully calibrated spatial axes) to define *s_calib_status*.

Under the (reasonable) assumption that an end-user searching for **event-bundle** datasets is typically querying based on the properties of the primary **event-list**, we suggest that those values also be used for the **event-bundle**. However, the data provider should ultimately decide which value best describes their **event-bundle** dataset.

For dataset types that do not encode sky coordinates, we suggest setting this value to “NULL”.

3.8 *t_calib_status*

We propose that *t_calib_status* encode the calibration status of an **event-list** dataset’s time axis. Where multiple time axes are included in a dataset (*e.g.*, instrument counter, absolute time), then we recommend that the data provider use the coordinate system that is most likely to be preferred by the end user (typically the most fully calibrated time axis) to define *t_calib_status*.

Under the (reasonable) assumption that an end-user searching for **event-bundle** datasets is typically querying based on the properties of the primary **event-list**, we suggest that those values also be used for the **event-bundle**. However, the data provider should ultimately decide which value best describes their **event-bundle** dataset.

For dataset types that do not not encode time coordinates, we suggest setting this value to “NULL”.

3.9 *em_calib_status*

We propose that *em_calib_status* encode the calibration status of an **event-list** dataset’s spectral axis. Where multiple spectral axes are included in a dataset (*e.g.*, PHA, PI, energy), then we recommend that the data provider use the coordinate system that is most likely to be preferred by the end user (typically the most fully calibrated spectral axis) to define *t_calib_status*.

Under the (reasonable) assumption that an end-user searching for **event-bundle** datasets is typically querying based on the properties of the primary **event-list**, we suggest that those values also be used for the **event-bundle**. However, the data provider should ultimately decide which value best describes their **event-bundle** dataset.

For dataset types that do not encode spectral coordinates, we suggest setting this value to “NULL”.

3.10 *o_ucd*

For an **event-list**, we can consider that all measures stored in column values are observables. This is *the* fundamental difference between **HEA event-lists** and typical pixelated datasets. The current ObsCore Recommendation suggests that *o_ucd* be set to “NULL” for event lists. However this significantly hampers data discovery for **HEA** datasets. Since the data content of **event-lists** may vary significantly from facility to facility, meaningful discovery of **HEA** datasets *requires* the user be able to query the UCDs of the set of observables included in an **event-list**.

A natural way of doing this that is consistent with current usage would be to extend *o_ucd* to allow specification of *multiple* observables for **event-lists** (and **event-bundles**), for example, *o_ucd = pos.eq,time,phys.pulseHeight*.

Note that real **event-lists** may include an extensive set of columns (*e.g.*, a *Chandra* ACIS Level 1 **event-list** includes ~ 20 columns, depending on observing mode) and several columns may represent similar (but not identical) observables (*e.g.*, event position in detector pixel coordinates, projected onto the focal surface, corrected for geometric distortions, corrected for spacecraft dither motion, mapped to world coordinates). Currently defined UCDs are not sufficiently fine-grained to be able to differentiate between these various cases. But that is likely not necessary, since for data discovery purposes the user is typically interested in the “most calibrated” properties in each of the spatial/spectral/time(/polarization) axes (*e.g.*, world coordinates in the above example).

In the example *o_ucd* above, the example UCD *phys.pulseHeight* is used to represent the detector Pulse Height Amplitude (PHA). There is currently no UCD defined for a raw measure like PHA, but we propose the addition of *phys.pulseHeight* to the UCDList vocabulary, together with other UCDs that are relevant for **HEA** data, in Section 5.3.

Advanced data products may similarly record multiple observables that can only be differentiated through their UCDs. For example, a *Chandra* Source Catalog **pdf** dataset for a detection may include multiple marginalized probability density functions computed using a Bayesian X-ray aperture photometry algorithm in units of net counts, net count rates, photon fluxes, and energy fluxes in multiple apertures. The observables recorded in the different MPDFs may be distinguished by their UCDs which then become relevant for data discovery when a user is searching for specific aperture photometry datasets.

Finally, we note that extending *o_ucd* to allow specification of multiple observables would require similar adjustments to the other observable axis attributes *o_unit*, *o_calib_status*, and *o_stat_err*.

3.11 *proposal_id*

To support advanced data products that may be constructed using data from multiple progenitor observations, we propose to modify the ObsCore Recommendation for *proposal_id* to allow multiple values, similar to *facility_name* and *instrument_name*.

4 Extensions to ObsCore Specific to High Energy Astrophysics Data

4.1 *ev_xel*

The lengths of each data axis (spatial, spectral, time, polarization) captured in attributes *s_xel1*, *s_xel2*, *em_xel*, *t_xel*, *pol_xel* do not apply non-pixelated data including **event-lists**, and ObsCore recommends that these attributes be set to -1 . However, the dimensionality of an event list is an important property for data discovery, as the number of events often scales with signal-to-noise (and also data volume scales with number of events). We propose to add a new, optional attribute *ev_xel* that records the number of events in an **event-list** (effectively, the length of the “events” axis in the **event-lists** table).

4.2 *s_ref_energy/em_ref_energy/s_ref_oaa/em_ref_oaa*

For **HEA** datasets that typically span decades of energy, spatial resolution and sky coverage, and spectral resolution, can be strongly dependent on particle energy. The ObsCore Recommendation suggests that in such circumstances a *characteristic* value be specified for the spatial and spectral characterization attributes (*e.g.*, *s_fov*, *s_region*, *s_resolution*, *em_res_power*, *em_resolution*). We propose adding optional attributes (*s_ref_energy* for

spatial characterization attributes and *em_ref_energy* for spectral characterization attributes) that define the energy (in units of eV) at which these characteristic values are specified.

For some HEA datasets, these attributes vary strongly with position in the field of view, typically as a function of off-axis angle (*i.e.*, the angular separation of the target or source from the telescope optical axis). We similarly propose adding optional attributes (*s_ref_oaa* for spatial characterization attributes and *em_ref_oaa* for spectral characterization attributes) that define the off-axis angle (in units of degrees) at which these characteristic values are specified.

4.3 *t_intervals*

The global time bounds described by *t_min/t_max* in general are not sufficiently flexible when representing HEA datasets and advanced data products from any waveband. The former are typically composed of many Stable Time Intervals (STIs)/Good Time Intervals (GTIs), where data are only valid during the stable or good intervals, while advanced data products may be constructed from multiple progenitor observations that can span decades from the start time of the first observations to the stop time of the last observation (albeit very sparsely). For both cases, data queries using only *t_min/t_max* may not be adequate to determine whether useful scientific data coincide with a transient cosmic phenomenon. In such cases, a more detailed knowledge of the observation time coverage is necessary. We propose to add a new optional attribute *t_intervals* that would contain the list of observation intervals or STIs/GTIs as a TMOC 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 TMOC representation.

We recognize that performing such queries will require enhancements to ADQL, but this capability is sufficiently important for some HEA data discovery scenarios that we have chosen to add *t_intervals*, in anticipation that ADQL will eventually provide this functionality.

4.4 *energy_min/energy_max*

The existing attributes *em_min* and *em_max* that define the coverage of the spectral axis (defined as wavelength expressed in units of m) are not user friendly for HEA where datasets are generally selected according to an energy range (*i.e.*, inverse wavelength) in units of eV (or scaled units of eV, for example keV, MeV, GeV, TeV, PeV). Unlike the radio domain where $\lambda = c/\nu$, where c is an almost universally remembered physical constant, the conversion $\lambda = hc/E$ is not simple for the user to express. As the

spectral range covered by HEA data is many decades larger than for other wavebands, the accurate numerical representations of typical HEA spectral ranges as em_min/em_max requires quantities with many digits of precision and exponents ranging from $\sim 10^{-5}$ – 10^{-22} . Since specification of the spectral range is largely fundamental to data discovery in the HEA regime, we propose to add attributes *energy_min* and *energy_max* that specify the minimum and maximum spectral range values in units of eV. Note that the sense of these attributes is *opposite* that of *em_min* and *em_max* because of the inverse wavelength relationship between energy and wavelength, so numerical comparisons must be transposed (*e.g.*, $E > E_{\text{thresh}}$ becomes $\lambda < hc/E_{\text{thresh}}$). (An alternate approach would be to add attributes *em_min_energy* and *em_max_energy* that represent the energies corresponding to *em_min* and *em_max* in units of eV. This is less desirable since queries on an energy would need to be specified as $em_max_energy \leq E < em_min_energy$, which is likely confusing.)

4.5 *obs_mode*

Many HEA instruments may be configured using multiple observing modes and these observing modes may significantly impact the structure and characteristics (*e.g.*, calibration accuracy) of the resulting observation datasets. For example, the *Chandra* ACIS instrument typically produces **event-lists** with 2-dimensional spatial coordinates (*i.e.*, imaging) but has an observation mode that continuously reads-out the detector, effectively producing an **event-list** with a single spatial dimension (and the other spatial dimension is collapsed); users looking for imaging data would want to restrict their queries to exclude the latter observing mode.

We propose to add an optional attribute **obs_mode** that allows the data provider to specify the observation mode for an observation. Constraints on observation mode can provide a simple way to discover data sets for a specific facility/instrument combination. We note that permissible **obs_mode** values will vary from facility to facility and from instrument to instrument.

4.6 *tracking_mode*

Since HEA telescopes are event-counting instruments, data can be accumulated even if the pointing moves or rotates relative to the sky during an observation. However, the way in which this happens has an impact on how the data are later processed. *Tracking* is defined by what the center of the field-of-view moves with during an observation: fixed to a celestial coordinate (the most common case for ground telescopes with movable mounts or those in space), fixed to a horizontal coordinate (also called a "drift scan", and the standard case for telescopes without movable mounts), fixed to a celestial body position like a planet or moon, or with no particular fixed

coordinate system such as data taken when the instrument is slewing (*i.e.*, where the field of view moves with arbitrary direction and speed while the instrument is repositioning).

To distinguish these, we propose to add an optional attribute **tracking_mode**. Constraints on tracking mode can provide a simple way to discover data sets for a specific facility/instrument combination. We note that permissible **tracking_mode** values may vary from facility to facility and from instrument to instrument.

BKH: where to put in addition convergent/divergent/parallel pointing? should we replace scan_mode by pointing_mode?
KK: any opinion?

4.7 *analysis_mode*

Most **HEA** instruments employ significant software processing to transform raw data into the **event-bundle** data exposed to users, including algorithms for calibration and event property reconstruction. The way in which this processing is configured therefore has a potentially large impact the content of the reduced datasets; indeed the same observation processed with two different configurations may result in different scientific performance. In some cases, multiple processing configurations within the same observation collection are used to provide users with a wider range of scientific coverage.

We propose to add an optional attribute **analysis_mode** that allows the data provider to specify the data reduction/analysis mode for an observation, in case more than one is applied. Constraints on analysis mode can provide a simple way to discover data sets for a specific facility/instrument combination. We note that permissible **analysis_mode** values may vary from facility to facility and from instrument to instrument.

4.8 *event_type*

Some **HEA** instruments allow data to have event partitioning based on a data analysis quality associated with the reconstruction and the discrimination. Some analyses can flag each event by a quality label, partitioning the dataset into strictly disjoint event subsets. For each quality label, a set of **IRFs** should be derived and can be rendered public⁸.

We propose to add an optional attribute **event_type** that specifies the data quality flag for an observation. It will allow the data provider to split the event list into several event lists labelled by a unique **event_type** for a given observation, and to distribute their associated **IRFs**. Constraints on

⁸As made by the Fermi-LAT collaboration, **IRFs** are produced for each event type: see https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_LAT_IRFs/IRF_overview.html.

event type can provide a simple way to discover data sets for a specific facility/instrument combination and to reduce the downloaded data volume. We note that permissible **event_type** values may vary from facility to facility and from instrument to instrument.

4.9 Additional Columns

Like for the ObsCore Recommendations, it is allowed to add any additional columns that a data provider aims to add. This should be done as for the additional columns in ObsCore:

Service providers may include additional columns in the *ivoa*. ObsCore table to expose additional metadata. These columns must be described in the *TAP_SCHEMA.columns* table and in the output from the VOSI-tables resource (Graham and Rixon et al., 2017). Users may access these columns by examining the column metadata for individual services and then using them explicitly in queries or by selecting all columns in the query (*e.g.* “select * from ivoa.ObsCore ...” in an ADQL query). In order to provide homogeneity in the keywords used as optional fields, we recommend where possible to use the items defined in the full data model of ObsCore and flagged as optional. ObsTAP compliant services will support all columns defined as mandatory and possibly some of the optional ones. Queries built up using additional columns defined specifically for a given archive might not be portable.

5 Vocabulary Enhancements

5.1 Evolution of the Data Product Type vocabulary

The IVOA Data Product Type Vocabulary⁹ provides terms, labels, and descriptions for many types of astronomical data products. However, there are some additions and changes that are appropriate to better support HEA datasets.

5.1.1 Event list

We first propose to better define an **event-list**:

event-list: A dataset containing a collection of observed particle-detection events, such as incoming HEA particles, where an event is typically characterized by a spatial position, a time, and a spectral value (*e.g.*, an energy, a channel, a pulse height)

⁹See <http://www.ivoa.net/rdf/product-type>.

5.1.2 Response functions

We then propose to add the following product types to define the different response functions (e.g. **IRFs**) generally used in **HEA**:

response-function: A dataset that maps a physical quantity to an observable. Narrower terms can be used to indicate more precisely the response function.

aeff: A dataset that records the “effective area” of a telescope and/or instrument. The effective area is the geometric area of the telescope and/or instrument reduced by efficiency factors such as reflectivity and vignetting, among other effects (Deil and Wood et al., 2022).

edisp: A dataset that records the probability density of detecting an event with an energy estimator (proxy) given the true energy of the event (Deil and Wood et al., 2022).

bkgrate: A dataset that models the rate of residual events that are not from the expected source type (*e.g.* for gamma-ray instrument it measures residual non-gamma-ray events coming from charged cosmic rays) (Deil and Wood et al., 2022).

psf: A dataset that records the probability density function of spatial/angular spreading of incident photons from a point source caused by the instrument (detector and/or mirror and/or analysis) (Deil and Wood et al., 2022; George and Yusaf, 2011).

arf: A dataset that records the combined telescope/instrument effective area and detector quantum efficiency as a function of energy (George and Arnaud et al., 1998).

rmf: A dataset that records the probability density function mapping from energy space into detector pulse height (or position) space (George and Arnaud et al., 1998).

5.1.3 Event bundle

Some use cases require access to an **event-bundle**, a bundle of datasets that includes the **event-list** and associated data:

event-bundle: compounded dataset 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 calibrated in physical units when containing **IRFs** or other data products that can be used to construct **IRFs**.

An **event-bundle** might for example consist of an **event-list** and the associated **response-functions** used to calibrate the dataset. It may also contain provenance information, data quality time-series, and preview images or plots.

5.1.4 Advanced data products

In addition to product types that focus on event data, we note that existing ObsCore definitions do not adequately span the breadth of advanced data products (with *calib_level* ≥ 3) that may be generated from astronomical observations. The computational complexity of analyzing **HEA** data robustly in the extreme Poisson regime (*e.g.*, Bayesian X-ray aperture photometry applied simultaneously to multiple overlapping detections and observations) means that data providers may choose to provide such analysis products directly to the end user.

For example, the *Chandra* Source Catalog includes 38 types of advanced data products (for a total of ~ 90 million files) and $\sim 50\%$ of these data product types are not well represented by a *dataprodut_type* value that allows for meaningful data discovery. Users will certainly want to discover these data products independently from the associated observation data (and many of these data products combine data from multiple observations). We therefore propose the following additional *dataprodut_types* for these advanced data products, and note that these *dataprodut_types* will certainly be useful independent of waveband (*i.e.*, they can be equally applicable to UV/optical, IR, and radio datasets):

draws: A dataset that represents draws computed from a probability distribution, for example the Markov chain Monte Carlo (MCMC) draws used when computing the Bayesian marginal probability density function for a random variable. The draws can be interpreted to provide a robust estimation of the probability distribution of variable, and correlations between the draws provide information about how well the draws converge to the parent probability distribution.

pdf: A dataset that represents the probability density function of a quantity, for example the Bayesian marginal probability density function for a random variable. The probability density function provides a robust estimation of the variable and allows arbitrary confidence intervals to be computed directly from the distribution.

region: A dataset that includes an encoding of (one or more) regions of parameter space, for example a spatial region or a region of phase space covered by a dataset. The set of dimensions represented by the region can be arbitrary.

The **measurements** data product type is quite useful for many different types of advanced data products (that may be derived from multiple observations) but users of those products often may not be interested the progenitor datasets, especially if many advanced data products are extracted from a single or a few progenitors (*e.g.*, measurements associated with sources detected in a single observation field). We propose to delete the caveat associated with **dataprodect_type** = “measurements” in the ObsCore IVOA Recommendation (§4.1.1) that requires the derived data products be exposed “together with the progenitor observation dataset”.

5.1.5 Summary table

The proposed vocabulary entries are listed in Table 1 with their label and parents.

Term	Label	Description	Parent
aeff	Effective Area	A dataset that records the “effective area” of a telescope and/or instrument. The effective area is the geometric area of the telescope and/or instrument reduced by efficiency factors such as reflectivity and vignetting, among other effects	#response-function
arf	Ancillary Response File	A dataset that records the combined telescope/instrument effective area and detector quantum efficiency as a function of energy	#response-function
bkgrate	Background Rate	A dataset that models the rate of residual events that are not from the expected source type (<i>e.g.</i> for gamma-ray instrument it measures residual non-gamma-ray events coming from charged cosmic rays)	#response-function
draws	Draws	A dataset that records statistical draws computed from a probability distribution, for example Markov chain Monte Carlo (MCMC) draws used when computing the Bayesian marginal probability density function for a random variable.	
edisp	Energy Dispersion	A dataset that records the probability density of detecting an event with an energy estimator (proxy) given the true energy of the event	#response-function, #pdf

event-bundle	Event Bundle	An event-bundle dataset is a compounded dataset containing an event-list and multiple files or other substructures that are products necessary to analyze the event-list .	
event-list	Event list	A dataset containing a collection of observed particle-detection events, such as incoming HEA particles, where an event is typically characterized by a spatial position, a time, and a spectral value (<i>e.g.</i> , an energy, a channel, a pulse height)	#temporally-resolved-dataset
pdf	Probability Density Function	A dataset that records the probability density function of a quantity, for example the Bayesian marginal probability density function for a random variable.	#measurements
psf	Point Spread Function	A dataset that records the probability density function of spatial/angular spreading of incident photons from a point source caused by the instrument (detector and/or mirror and/or analysis)	#response-function, #pdf
region	Region	A dataset that encodes (one or more) regions of parameter space, for example a spatial region or a region of phase space covered by a dataset. The set of dimensions represented by the region can be arbitrary.	#measurements
response function	Response Function	A dataset that maps a physical quantity to an observable. This term is mainly intended for retrieval. To annotate datasets, use a narrower term.	

rmf	Redistribution Matrix File	A dataset that records the probability density function mapping from energy space into detector pulse height (or position) space	#response-function, #pdf
------------	----------------------------	--	-----------------------------

Table 1: IVOA Data Product Type Vocabulary extension

5.1.6 Clarification of “flux” in some term definitions

We propose to clarify the terms that use the word “flux” in the description of some terms, currently **light-curve**, **polarization-resolved-dataset**, **polarized-spectrum**, and **spectrum**, to be applicable to **HEA** data.

The issue is that the term “flux” is not well defined, and the standard astronomical definition of “flux” is an energy flux (with SI units W m^{-2}). This interpretation is bolstered by the statements “flux or magnitude” applied to several of the descriptions since optical/IR magnitude and energy flux density are tightly related. However, with this definition, many **HEA** data products (that have also units of counts) would not satisfy the descriptions of a light-curve or a spectrum. We often use a few definitions of flux:

- particle flux, units=[counts/m2/s],
- particle flux per steradian, units=[counts/m2/s/str],
- particle flux density (or differential particle flux) units=[counts/m2/s/TeV],
- energy flux, units=[W/m2],
- energy flux per steradian (*i.e.* radiance), units=[W/m2/str],
- energy flux density (or differential energy flux), units=[W/m2/str/TeV],

These are not all covered by the UCD vocabulary. This would be particularly useful for cases where we actually should have the same UCDs as *e.g.* a Radio measurement, but where we use very different units. For example: Jansky and $\text{erg/cm}^2/\text{s}/\text{TeV}$ are both types of “spectral flux density”, but the unit analysis doesn’t agree due to frequency-energy equivalency.

Restating the descriptions to explicitly state “Particle or energy flux or magnitude” where appropriate would resolve this ambiguity associated to unit confusions. These different type of fluxes will be proposed to be added in the UCD list.

5.2 DataLink vocabularies

For some use cases, we proposed to expose the different associated datasets via Datalink. Each Datalink is described by several attributes, including the mandatory *semantics* attribute and a *content_qualifier*.

The terms defined for response functions (see §5.1.2) may thus be used to fill the *content_qualifier* attributes, with *semantics* = **#calibration**.

5.3 UCD Enhancements

5.3.1 Pulse Height

For many X-ray and gamma-ray instruments the signal observed in a given detector spectral channel is the result of event counting and would typically be recorded as a Pulse Height Amplitude (PHA), or perhaps a Pulse Invariant (PI) value that is calculated from PHA by applying an appropriate gain calibration. The PHA (or PI) can be related to the incident particle energy by applying the appropriate **response-function**, and higher data calibration level products may replace or augment these values with quantities such as energy, or perhaps particle or energy flux.

There is currently no UCD defined for a raw pulse height amplitude measure like PHA (or PI). PHA is such an important quantity to **HEA** datasets that we propose adding a new UCD *phys.pulseHeight* for these raw data values. We note that the background signal (both of instrumental and cosmological origin) may be significant for many **HEA** detectors and so the detected events are unrelated to any observed source on the sky.

A proposed solution suggests using *src.var.amplitude;src.var.pulse;stat.uncalib* for PHA, but this is not really appropriate since the connection to *src* (“observed source viewed on the sky”) is misleading and *src.var.amplitude* is defined as the “amplitude of variation” of the source which is a completely separate concept from an astronomical perspective.

5.3.2 Electromagnetic spectrum description in UCD

The current definitions for the gamma-ray domains should be corrected and extended to include the **HEA**, **Very High Energy (VHE)** and **Ultra High Energy (UHE)** gamma-ray domains.

5.3.3 Event Type

For **VHE** (and GeV) data there is the notion of event type that can be mandatory for some data releases. We propose to add a new UCD *instr.evt-type* that identifies these data values.

5.3.4 Particles

Observations may concern other particles than the one currently described in the UCD list. The following particles could be added: electron/positron, cosmic rays, as well as their energy

5.3.5 Statistical UCDs

We suggest restricting *stat.max/stat.min* to mean the maximum/minimum statistic and adding new terms *stat.upperlimit/stat.lowerlimit* for upper/lower limits. For upper/lower limits, one expects a confidence level to be provided, which could be described by a UCD *stat.confidenceLevel*.

One proposes also to add a likelihood profile as UCD

5.3.6 Evolution of UCD list

The proposed UCD entries are listed in Table 2 with their descriptions.

Label	UCD word	Description
S	em.gamma.hard	Hard gamma ray (500 keV - 100 MeV)
S	em.gamma.he	High-Energy gamma ray (100 MeV - 10 GeV)
S	em.gamma.vhe	Very-High-Energy gamma ray (10 GeV - 100 TeV)
S	em.gamma.uhe	Ultra-High-Energy gamma ray (100 TeV - 10 PeV)
S	phys.pulseHeight	Pulse height amplitude measure
S	phys.particle.electron	Related to electron/positron
S	phys.particle.cosmicray	Related to cosmic rays particles
P	stats.error.negative	Negative statistical error
P	stats.error.positive	Positive statistical error
P	stat.upperlimit	Upper limit
P	stat.lowerlimit	Lower limit
P	stat.confidenceLevel	Level of confidence for a upper/lower limit computation
P	stat.likelihood.profile	Profile of likelihood values
E	phot.strcount	Particle flux per steradian (dimension: $[L^{-2} T^{-1} sr^{-1}]$)
E	phot.diffcount	Particle flux density (or differential particle flux) (dimension: $[L^{-2} T^{-1} E^{-1}]$)
E	phot.strflux	Energy flux per steradian or radiance (dimension: $[W L^{-2} sr^{-1}]$)
E	phot.enflux	Energy flux density (dimension: $[W L^{-2} sr^{-1} E^{-1}]$)

Table 2: UCD words proposed extension

And here is the proposals of update of the existing flux terms:

Label	UCD word	Description
E	phot.count	Flux expressed in counts (dimension: $[L^{-2} T^{-1}]$)
E	phot.flux	Photon flux or irradiance or energy flux (dimension: $[W L^{-2}]$)

Table 3: UCD words proposed upgrade

BKH: Karl, should we add the excess number? `stat.excess` or `stat.fit.excess`? Any idea for the reco/true physical parameters?

5.4 MIME-types Enhancements

Data files used in the [HEA](#) domain should have appropriate MIME-types, so that they can be shown in ObsCore tables or elsewhere.

Formats based on FITS could thus be declared as:

- **x-fits-gadf:** for FITS files following the GADF specification ([Deil and Wood et al., 2022](#))
- **x-fits-vodf:** for FITS files following the VODF specification ([Khélifi and Zanin et al., 2023](#))

6 Proposal of the *ivoa.obscore_he* attributes

This section summarizes the proposal for the [HEA](#) extension of ObsCore. Note that the name of the extension could be *ivoa.obscore_he*, *ivoa.obscore-hea*, *ivoa.obscore_he* or *ivoa.obscore-he*. We use here the first term in the [Appendix A](#).

Column Name	UType	Unit	Type	Description	MAN
ev_xel	TBD	unitless	int	Number of events in an event_list	NO

s_ref_energy	TBD	eV	float	Energy at which the ObsCore spatial characterisation attributes s_fov , s_region, s_resolution are defined	NO
em_ref_energy	TBD	eV	float	Energy at which the ObsCore spatial characterisation attributes em_res_power, em_resolution are defined	NO
s_ref_oaa	TBD	deg	float	Off-axis angle (<i>i.e.</i> , the angular separation of the target or source from the telescope optical axis) at which the ObsCore spatial characterisation attributes s_fov , s_region, s_resolution are defined	NO
em_ref_oaa	TBD	deg	float	Off-axis angle (<i>i.e.</i> , the angular separation of the target or source from the telescope optical axis) at which the ObsCore spectral characterisation attributes em_res_power, em_resolution are defined	NO
t_intervals	TBD	unitless	TMOC	List of observation intervals or stable/good time intervals describing the exact observation time coverage	NO
energy_min	TBD	float	eV	Energy associated to the ObsCore attribute em_max, describing the minimal energy of the dataset	NO
energy_max	TBD	float	eV	Energy associated to the ObsCore attribute em_min, describing the maximal energy of the dataset	NO
obs_mode	TBD	unitless	string	Observation mode of the observation (e.g. TBU)	NO
tracking_mode	TBD	unitless	string	Tracking mode of an observation (e.g., sidereal rate, moving target [solar system] tracking, drift scans)	NO
analysis_mode	TBD	unitless	string	Data reduction/analysis mode	NO
event_type	TBD	unitless	string	Data quality flag of the events (e.g. “good psf”, “good rejection”, “Nhit (100,200)”)	NO

Table 4: Attributes for the [HEA](#) extension of ObsCore

7 Summary

BKH: Ian, could you start writing the final section?

Glossary

GTI	Good Time Interval
HE	High Energy
HEA	High Energy Astrophysics
HEIG	High Energy Interest Group
IRF	Instrument Response Function
IVOA	International Virtual Observatory Alliance
MOC	Multi-Order-Coverage
STI	Stable Time Interval
UHE	Ultra High Energy
VHE	Very High Energy
VO	Virtual Observatory

References

- Bradner, S. (1997), ‘Key words for use in RFCs to indicate requirement levels’, RFC 2119. <http://www.ietf.org/rfc/rfc2119.txt>.
- Deil, C., Wood, M., Hassan, T., Boisson, C., Contreras, J. L., Knödlseider, J., Khelifi, B., King, J., Mohrmann, L., Linhoff, M. and Donath, A. (2022), ‘Data formats for gamma-ray astronomy - version 0.3’. doi:10.5281/zenodo.7304668.
- Demleitner, M., Gray, N. and Taylor, M. (2021), ‘Vocabularies in the VO Version 2.0’, IVOA Recommendation 25 May 2021. doi:10.5479/ADS/bib/2021ivoa.spec.0525D, <https://ui.adsabs.harvard.edu/abs/2021ivoa.spec.0525D>.
- Demleitner, M., Gray, N. and Taylor, M. (2023), ‘Vocabularies in the VO Version 2.1’, IVOA Recommendation 06 February 2023. doi:10.5479/ADS/bib/2023ivoa.spec.0206D, <https://ui.adsabs.harvard.edu/abs/2023ivoa.spec.0206D>.
- Donath, A., Terrier, R., Remy, Q., Sinha, A., Nigro, C., Pintore, F., Khelifi, B., Olivera-Nieto, L., Ruiz, J. E., Brügge, K., Linhoff, M., Contreras, J. L., Acero, F., Aguasca-Cabot, A., Berge, D., Bhattacharjee, P., Buchner, J., Boisson, C., Carreto Fidalgo, D., Chen, A., de Bony de Lavergne, M., de Miranda Cardoso, J. V., Deil, C., Fülling, M., Funk,

- S., Giunti, L., Hinton, J., Jouvin, L., King, J., Lefaucheur, J., Lemoine-Goumard, M., Lenain, J.-P., López-Coto, R., Mohrmann, L., Morcuende, D., Panny, S., Regeard, M., Saha, L., Siejkowski, H., Siemiginowska, A., Sip"ocz, B. M., Unbehaun, T., van Eldik, C., Vuillaume, T. and Zanin, R. (2023), ‘Gammapy: A python package for gamma-ray astronomy’, *A&A* **678**, A157. doi:[10.1051/0004-6361/202346488](https://doi.org/10.1051/0004-6361/202346488).
- 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. doi:[10.5479/ADS/bib/2022ivoa.spec.0727F](https://doi.org/10.5479/ADS/bib/2022ivoa.spec.0727F), <https://ui.adsabs.harvard.edu/abs/2022ivoa.spec.0727F>.
- Fruscione, A., McDowell, J. C., Allen, G. E., Brickhouse, N. S., Burke, D. J., Davis, J. E., Durham, N., Elvis, M., Galle, E. C., Harris, D. E., Huenemörder, D. P., Houck, J. C., Ishibashi, B., Karovska, M., Nicastro, F., Noble, M. S., Nowak, M. A., Primini, F. A., Siemiginowska, A., Smith, R. K. and Wise, M. (2006), CIAO: Chandra’s data analysis system, in D. R. Silva and R. E. Doxsey, eds, ‘Observatory Operations: Strategies, Processes, and Systems’, Vol. 6270 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 62701V. doi:[10.1117/12.671760](https://doi.org/10.1117/12.671760), <https://ui.adsabs.harvard.edu/abs/2006SPIE.6270E..1VF>.
- George, I., Arnaud, K. and Tennant, A. (1998), ‘The calibration requirements for spectral analysis’, Technical Report OGIP/92-002, NASA/GSFC, Dec 1998. https://heasarc.gsfc.nasa.gov/docs/heasarc/caldb/docs/memos/cal_gen_92_002/cal_gen_92_002.pdf.
- George, I. and Yusaf, R. (2011), ‘The ogip format for 2-d (image) point spread function datasets’, Technical Report OGIP/92-027, NASA/GSFC, Nov 2011. https://heasarc.gsfc.nasa.gov/docs/heasarc/caldb/docs/memos/cal_gen_92_027/cal_gen_92_027.pdf.
- Graham, M., Rixon, G., Dowler, P., Major, B., Grid and Web Services Working Group (2017), ‘IVOA Support Interfaces Version 1.1’, IVOA Recommendation 24 May 2017. doi:[10.5479/ADS/bib/2017ivoa.spec.0524G](https://doi.org/10.5479/ADS/bib/2017ivoa.spec.0524G), <https://ui.adsabs.harvard.edu/abs/2017ivoa.spec.0524G>.
- Khélifi, B., Zanin, R., Kosack, K., Olivera-Nieto, L. and Schnabel, J. (2023), ‘The Very-high-energy Open Data Format: towards a shared, open data format in very-high-energy astronomy’, arXiv:2308.13385. doi:[10.48550/arXiv.2308.13385](https://doi.org/10.48550/arXiv.2308.13385), <https://ui.adsabs.harvard.edu/abs/2023arXiv230813385K>.
- 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>.

Servillat, M., Boisson, C., Bonnarel, F., Cresitello-Dittmar, M., Cristofari, P., Evans, I., Evans, J., Fuessling, M., Jaffe, T., Khélifi, B., Kosack, K., Louys, M., Michel, L., Nebot, A., Schnabel, J., Schussler, F. and the HE discussion group at IVOA (2024), ‘Virtual Observatory and High Energy Astrophysics Version 1.0’, IVOA Note 12 November 2024. <https://www.ivoa.net/documents/Notes/VOHE/20241112/index.html>.

A Detailed Science Use Cases for ObsCore

A.1 Event-List Data and Responses

A.1.1 Use Case — Search for event lists surrounding Sgr A*, for example for an X-ray morphological study

Identify all HEA event lists encompassing Sgr A for initial selection for subsequent X-ray morphological studies. Since the focus is on X-ray morphological studies, only the event lists and not the event bundles are desired.*

Find all datasets satisfying:

- (i) Target name = “Sgr A*” or position inside 30 arcmin from (266.4168, -29.0078)
- (ii) dataproduct_type = “event-list”

```
SELECT * FROM ivoa.obscore
NATURAL JOIN ivoa.obscore_healpix
WHERE
(target_name = 'Sgr A*' OR
CONTAINS(POINT(s_ra, s_dec), CIRCLE, 266.4168, -29.0078, 0.5) = 1)
AND (dataproduct_type = 'event-list')
```

A.1.2 Use Case — Search for event bundles that include Cas A for X-ray spectrophotometric evolution studies

Identify all event bundles that include the Cas A SNR and have at least 1 million events for subsequent spectrophotometric studies of the SNR expansion. Since only a few observations are expected to match this request and because the focus is on X-ray spectrophotometric studies, the event bundles that include the responses or the ancillary products used to make the responses are required.

Find all datasets satisfying:

- (i) Target name = “Cas A” or position inside 10 arcmin from (350.8584, +58.8113)
- (ii) dataproduct_type = “event-bundle”
- (iii) ev_xel $\geq$ 1,000,000

```
SELECT * FROM ivoa.obscore
NATURAL JOIN ivoa.obscore_healpix
WHERE
(target_name = 'Cas A' OR
CONTAINS(POINT(s_ra, s_dec), CIRCLE, 350.8584, +58.8113, 0.16667) = 1)
AND (dataproduct_type = 'event-bundle')
AND (ev_xel  $\geq$  1000000)
```

A.1.3 Use Case — Search for event bundles via DataLink that include Cas A for a TeV spectromorphology study

*Identify all event lists and their associated *IRFs* that include the Cas A SNR for subsequent TeV spectromorphology studies from a VERITAS data release. Since the instrumental responses are mandatory to remove instrumental effects, the event bundles that include the *IRFs* are required.*

Find all VERITAS datasets satisfying:

- (i) Target name = “Cas A” or position inside 4 deg from (350.8584, +58.8113)
- (ii) dataproduct_type = “event-bundle”
- (iii) obs_collection = “VERITAS-DR1”
- (iv) access_format = “datalink”

```
SELECT * FROM ivoa.obscore
NATURAL JOIN ivoa.obscore_healpix
WHERE
(target_name = 'Cas A' OR
CONTAINS(POINT(s_ra, s_dec), CIRCLE, 350.8584, +58.8113, 4.0) = 1)
AND (dataproduct_type = 'event-bundle')
AND (obs_collection = 'VERITAS-DR1')
AND (access_format = 'application/x-votable+xml;content=datalink')
```

Then, for each row of the output, we get the “access_url” of the DataLink and access to the data.

A.1.4 Use Case — Identify PSF response-functions for further analysis of previously downloaded data products

Identify all Chandra Source Catalog point spread functions for source detections that fall within 2 arcmin radius of (83.84358, -5.43639) in the Orion star-forming complex for Chandra observation 4374. These PSFs will be used to analyze previously downloaded catalog data products for the same field.

Find all datasets satisfying:

- (i) Position inside 3 arcmin from (83.84358, -5.43639)
- (ii) dataproduct_type = “response-function”
- (iii) dataproduct_subtype = “psf”
- (iv) obs_id = “4374”
- (v) obs_collection = “CSC2”

```
SELECT * FROM ivoa.observatory
NATURAL JOIN ivoa.observatory_heav
WHERE
  (CONTAINS(POINT(s_ra, s_dec), CIRCLE, 83.84358, -5.43639, 0.033333) = 1)
  AND (dataproduct_type = 'response-function')
  AND (dataproduct_subtype = 'psf')
  AND (obs_id = '4374')
  AND (obs_collection = 'CSC2')
```

A.1.5 Use Case — Get all the IRFs for a given CTAO observation, for simulation purposes

Simulations are frequently used to estimate the science performance for a given astrophysical use case. To realise such simulations, IRFs are required.

Find the CTAO datasets satisfying:

- (i) dataproduct_type = “aeff” or dataproduct_type = “psf” or dataproduct_type = “edisp” or dataproduct_type = “bkgrate”
- (ii) obs_id = “4374”
- (iii) obs_collection = “CTAO-DR1”

```
SELECT * FROM ivoa.observatory
NATURAL JOIN ivoa.observatory_heav
WHERE
  (dataproduct_type = 'event-list' OR dataproduct_type = 'aeff')
```

```

OR dataproduct_type = 'edisp'
OR dataproduct_type = 'psf' OR dataproduct_type = 'bkgrate')
AND (obs_id = '4374')
AND (obs_collection = 'CTAO-DR1')

```

A.1.6 Use Case — Search for event lists that include a fully calibrated spectral axis for BL Lac for rapid X-ray spectrophotometric evaluation

Identify all event lists that include the BL Lac, have a fully calibrated spectral axis (i.e., spectral responses have already been applied), and have at least 10,000 events. These data will be used to prepare slides for a presentation. Note that since `calib_status = 2` may not specify that the spectral axis is fully calibrated in physical units (HEA event lists are often considered “calibrated” even if the spectral axis is in pulse height units) the calibration status of the spectral axis must be checked explicitly.

Find all datasets satisfying:

- (i) Target name = “BL Lac” or position inside 5 arcmin from (330.680338, +42.27777)
- (ii) `dataproduct_type` = “event-list”
- (iii) `calib_level` ≥ 2
- (iv) `em_calib_status` = “calibrated”
- (v) `ev_xel` ≥ 10000

```

SELECT * FROM ivoa.observ
NATURAL JOIN ivoa.observ_he
WHERE
(target_name = 'BL Lac' OR
CONTAINS(POINT(s_ra, s_dec), CIRCLE, 330.680338, +42.27777, 0.083333) = 1)
AND (dataproduct_type = 'event-list')
AND (calib_level >= 2)
AND (em_calib_status = 'calibrated')
AND (ev_xel >= 10000)

```

A.1.7 Use Case — Search for spatially resolved spectropolarimetric observation of the Crab with spectral resolution $R > 100$

Identify all event bundles for observations of the Crab that intersect the 1.0–100.0 keV energy range, have calibrated spatial and time axes, are spatially resolved in 2 dimensions in equatorial coordinates, have spectral resolution

$R > 100$, and include polarimetry measurements. Note that ObsCore specifies that the axes lengths — s_xel1 , s_xel2 , em_xel , t_xel , pol_xel — should be set to -1 for non-pixelated data like event lists, so these quantities are not useful for this query.

Find all datasets satisfying:

- (i) Target name = “Crab” or position inside 5 arcmin from (83.6324, +22.0174)
- (ii) dataproduct_type = “event-bundle”
- (iii) calib_level ≥ 2

```
SELECT * FROM ivoa.obscore
NATURAL JOIN ivoa.obscore_healpix
WHERE
(target_name = 'Crab' OR target_name = 'M1' OR
CONTAINS(POINT(s_ra, s_dec), CIRCLE, 83.6324, +22.0174, 0.083333) = 1)
AND (dataproduct_type = 'event-bundle')
AND (calib_level >= 2)
AND (s_resolution > 100)
AND (energy_min < 100000.0) AND (energy_max >= 1000.0)
AND (o_ucd LIKE '%phys.polarization%')
AND (o_ucd LIKE '%pos.eq%')
```

A.1.8 Use Case — Search for SWGO event lists and their IRFs for the event type ‘very-good’ in the region of Cygnus loop for a TeV spectromorphology study

Identify all event lists and their associated IRFs of the region of Cygnus loop (3° diameter) taken with SWGO. Only data of the event type ‘very-good’ are selected, in order to limit the amount of downloaded data.

Find all SWGO datasets satisfying:

- (i) s_region position intersects with the source modeled by a circle of 1.5 deg around (312.775, 30.683)
- (ii) dataproduct_type = “event-list” or “aeff” or “edisp” or “psf” or “bkgrate”
- (iii) obs_collection = “SWGO-DR1”
- (iv) event_type = “very-good”

First, run the ObsCore query:

```

SELECT * FROM ivoa.obscore
NATURAL JOIN ivoa.obscore_healpix
WHERE (INTERSECTS(s_region, CIRCLE(312.775, 30.683, 1.5)) = 1)
AND (dataprod_type = 'event-list' OR dataprod_type = 'aeff' OR dataprod_type = 'psf' OR dataprod_type = 'bkgrate')
AND (obs_collection = 'SWG0-DR1')
AND (event_type = 'very-good')

```

Then, for each row of the output, we identify the nature of the data and get it from its “access_url”.

A.1.9 Use Case — Search for event lists and their IRFs of CTAO South observations at energies above 10 TeV for blind search of PeV-atrons from a data release using DataLink

Identify all event lists and their associated IRFs taken by CTAO South that contains events above 10 TeV. Data taken with the Small Size Telescopes or Medium Size Telescopes can be then selected.

Find all CTAO datasets satisfying:

- (i) dataprod_type = “event-list”
- (ii) obs_collection = “CTAO-DR1”
- (iii) access_format = “datalink”
- (iv) “CTA-S” in instrument_name
- (v) energy_max $\geq 10e+12$

First, run the ObCore query:

```

SELECT * FROM ivoa.obscore
NATURAL JOIN ivoa.obscore_healpix
WHERE (dataprod_type = 'event-list')
AND (obs_collection = 'CTAO-DR1')
AND (access_format = 'application/x-votable+xml;content=datalink')
AND (instrument_name LIKE 'CTAO-S')
AND (ENERGYMAX GE 10e+12)

```

The query output is a VOTable that follows the DALI specification.

Then, process this VOTABLE to access to the data:

- (i) for each row of the query output, get the “obs_id” and the “access_url” of the DataLink
- (ii) get the VOTABLE associated with the “access_url”

- (iii) for each row of the VOTABLE, get the “content_qualifier” and the “accessURL”
- (iv) download the data associated to each “accessURL”

```

FOR EACH ROW OF OUTPUT_VOTABLE:
  OBS_ID = OUTPUT_VOTABLE['ID']
  DATALINK_TABLE = GET OUTPUT_VOTABLE['access_url']
  FOR EACH ROW OF DATALINK_TABLE:
    - IF ROW['content_qualifier'] = 'aeff'
      AEFF_FILE['OBS_ID'] = GET RAW['accessURL']
    - IF ROW['content_qualifier'] = 'edisp'
      EDISP_FILE['OBS_ID'] = GET RAW['accessURL']
    - IF ROW['content_qualifier'] = 'psf'
      PSF_FILE['OBS_ID'] = GET RAW['accessURL']
    - IF ROW['content_qualifier'] = 'bkgrate'
      BKGRATE_FILE['OBS_ID'] = GET RAW['accessURL']
    - IF ROW['content_qualifier'] = 'event-list'
      EVENT_FILE['OBS_ID'] = GET RAW['accessURL']

```

To Be Checked - Mathieu, Mireille?

A.2 Very-High-Level Data Products

A.2.1 Use Case — Search for Chandra Source Catalog position error MCMC draws for X-ray detections in the vicinity of Gaia DR3 486718823701242368

Identify all Chandra Source Catalog position error MCMC draws for source detections that fall within 5 arcsec radius of (54.036061, +61.907633). The MCMC draws will be evaluated to establish whether there are potentially unresolved X-ray sources that may coincide with the white dwarf for observation planning.

Find all datasets satisfying:

- (i) Position within 5.0 arcsec from (54.036061, +61.907633)
- (ii) dataproduct_type = “draws”
- (iii) dataproduct_subtype = “poserr”
- (iv) obs_collection = “CSC2”

```

SELECT * FROM ivoa.obscore
NATURAL JOIN ivoa.obscore_healpix

```

```

WHERE
(CONTAINS(POINT(s_ra, s_dec), CIRCLE, 54.036061, +61.907633, 0.0013888) = 1)
AND (dataprodect_type = 'draws')
AND (dataprodect_subtype = 'poserr')
AND (obs_collection = 'CSC2')

```

A.2.2 Use Case — Search for flux maps for CTAO-North observations between two observations ID

Identify all flux maps from the observation of CTAO-North between selected observation identifiers selected by the user .

Find all datasets satisfying:

- (i) dataprodect_type = “image”
- (ii) dataprodect_subtype = “flux-map”
- (iii) obs_collection = “CTAO-DR1”
- (iv) int(obs_id) >= 4374
- (v) int(obs_id) <= 4379
- (vi) “CTAO-N” in instrument_name

```

SELECT * FROM ivoa.observ
NATURAL JOIN ivoa.observ_he
WHERE
(dataproduct_type = 'image')
AND (dataprodect_subtype = 'flux-map')
AND (obs_collection = 'CTAO-DR1')
AND (instrument_name LIKE 'CTAO-N')
AND (CAST(obs_id AS INTEGER) > 4374)
AND (CAST(obs_id AS INTEGER) <= 4379)

```

A.2.3 Use Case — Search for M31 source light curves and aperture photometry probability density functions that intersect a specific time interval

Identify all light curves and aperture photometry probability density functions of X-ray sources detected in the field of M31 covering the energy range 0.3–7.0 keV that include observation data in the interval MJD 56320–56325 TT during which interval a transient event was thought to have occurred. Because the data products are expected to include extremely sparse time axes, the t_intervals TMOC must be used for the query.

Find all datasets satisfying:

- (i) Position within 1.5 degrees from (10.6847, +41.2688)
- (ii) dataproduct_type = “timeseries” or “pdf”
- (iii) calib_level = 4
- (iv) energy_min $\leq$ 0.3 and energy_max $\geq$ 7.0
- (v) t_intervals TMOC intersects¹⁰ MJD 56320–56325 TT

```
SELECT * FROM ivoa.obscore
NATURAL JOIN ivoa.obscore_healpix
WHERE
  (CONTAINS(POINT(s_ra, s_dec), CIRCLE, 10.6847, +41.2688, 1.5) = 1)
  AND ((dataproduct_type = 'timeseries') OR (dataproduct_type = 'pdf'))
  AND (calib_level = 4)
  AND (energy_min <= 300.0) AND (energy_max >= 7000.0)
  AND (INTERSECTS(TMOC(17, t_intervals), ...) = 1)
```

A.2.4 Use Case — Search for the CTAO flux light curves of PKS 2155-304 in 2030

Identify all light curves obtained on the source PKS 2155-304 in 2030 with the CTAO observatory.

Find all datasets satisfying:

- (i) dataproduct_type = “timeseries”
- (ii) dataproduct_subtype = “flux”.
- (iii) obs_collection = “CTAO-DR1”.
- (iv) tmin $\geq$ 62502 (*e.g.* 2030-01-01)
- (v) tmax $\geq$ 62866 (*e.g.* 2030-12-31)
- (vi) target_name = “PKS 2155-304”

```
SELECT * FROM ivoa.obscore
NATURAL JOIN ivoa.obscore_healpix
WHERE
  (dataproduct_type = 'timeseries')
  AND (dataproduct_subtype = 'flux')
  AND (obs_collection = 'CTAO-DR1')
  AND (target_name = 'PKS 2155-304')
  AND (tmin >= 62502)
  AND (tmax <= 62866)
```

¹⁰We note that this function does not yet exist in ADQL

B Changes from Previous Versions

No previous versions yet.

C Contributions to the Note

The authors of this Note contributed to write and structure the text. However, the note was initiated and elaborated in several dedicated workshops, Interop meetings, and in specific **IVOA HEIG** group meetings, involving more people. The **IVOA HEA** group keeps track of its activities on an **IVOA** web page: <https://wiki.ivoa.net/twiki/bin/view/IVOA/HEGroup>.

Further material can be found with those links:

- 2025-04-29: French ASOV workshop “High Energy in the VO”, <https://indico.obspm.fr/event/2674/>
- 2024-11-16: IVOA Malta meeting, DM session with 2 High Energy presentations (B. Khelifi/I. Evans), <https://wiki.ivoa.net/twiki/bin/view/IVOA/InterOpNov2024DM>
- 2024-11-15: IVOA Malta Plenary, CSP Plenary session, <https://wiki.ivoa.net/twiki/bin/view/IVOA/InterOpNov2024CSPPlenary>
- 2024-05-21: IVOA Sydney meeting, DM Session High Energy focus, <https://wiki.ivoa.net/twiki/bin/view/IVOA/InterOpMay2024DM>
- 2023-06-28: IVOA standards for High Energy Astrophysics (French VO Workshop), <https://indico.obspm.fr/event/1963/>
- 2023-05-11: IVOA Bologna meeting: presentation ("DM for High Energy astrophysics", M. Servillat) and first IVOA HE group meeting, https://wiki.ivoa.net/internal/IVOA/IntropMay3023DM/2023-05-11_IVOA_meeting_-_VOHE.pdf
- 2022-10-11: Virtual Observatory and High Energy Astrophysics (French VO Workshop), <https://indico.obspm.fr/event/1489/>