CRIS-MAC2024 IOP Publishing
Journal of Physics: Conference Series **3053** (2025) 012039 doi:10.1088/1742-6596/3053/1/012039

The Virtual Observatory in the age of multimessenger Astroparticle Physics

Claudio Galelli[1], Bruno Khélifi[2], Catherine Boisson[1], and Mathieu Servillat[1]

[1]LUTh, Observatoire de Paris, Meudon, France
[2]APC, Université Paris Cité, Paris, France

E-mail: claudio.galelli@obspm.fr

Abstract. The International Virtual Observatory Alliance (IVOA) plays a pivotal role in advancing the FAIR principles within the domain of astrophysics, ensuring that scientific data is Findable, Accessible, Interoperable, and Reusable (FAIR). The IVOA establishes standardized models for data and metadata, and data access protocols. Using Virtual Observatory (VO) compatible tools, enables seamless retrieval of data from different datasets and catalogs, fostering the connection and interoperability between different observatories, surveys, and research groups. Until recent times the very-high-energy astrophysics and astroparticle physics domains have not been specifically integrated into the IVOA initiatives. However, a growing effort is present in the community to delineate standards and data formats that can adapt the IVOA guidelines to the field's specificities. This effort has begun to pay off with the first integration of high-level data products from H.E.S.S. into the VO environment, and the creation of the common data format for gamma-ray astronomy (GADF), which is currently being reworked into the very-high-energy open data format (VODF), a data and metadata format for gamma-ray and multimessenger astrophysics as much compliant as possible with the future VHE features of the IVOA standards.

1 The open data landscape in astrophysics and efforts in the VHE domain

The field of Very-High-Energy (VHE) gamma-ray astronomy is experiencing a period of significant transformation. With the ever-increasing number of detected sources exceeding 300 [1], the commissioning and operation of powerful detector arrays like CTAO [2] and LHAASO [3] are pushing the boundaries of sensitivity and discovery. These observations study the most energetic processes in the universe, revealing objects like supernova remnants, pulsar wind nebulae, and active galactic nuclei.

VHE gamma-ray astronomy is becoming increasingly intertwined with multi-wavelength and multi-messenger astronomy. Observations at different wavelengths, including neutrinos and gravitational waves, offer a more complete understanding of astrophysical phenomena. To facilitate these combined analyses, the VHE community is actively promoting open data practices [4][5] and open-source analysis tools [6], to promote collaboration between different experiments and adjacent research fields. This movement towards open science is driven by several factors, including national and international roadmaps [1][2] and the inherent open nature of the next-generation CTA observatory. Indeed, CTAO will be operated as a

---

[1] `https://science.nasa.gov/open-science/`

[2] `https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en`

traditional astronomical observatory, with open proposal-based observational campaigns, for which the data retrieved will undergo a proprietary period before being published openly for the scientific community and the wider public. Therefore, It is paramount to have a cohesive and effective open data policy.

The key aspects of an effective open data policy, for any scientific field, are summarized in the FAIR keywords [7]:

- **Findability:** Data and software should be easy to discover for humans and machines through standardized naming conventions, comprehensive metadata descriptions, and registration in public repositories. This allows researchers to efficiently locate relevant resources for their work.
- **Accessibility:** Data and software should be readily available to anyone with the necessary permissions and technical capabilities. This often involves removing access barriers like licensing restrictions or complex download procedures. Open access, with clear guidelines, promotes collaboration and facilitates broader scientific inquiry.
- **Interoperability:** Data and software should be designed to work seamlessly with each other, regardless of their source or format. This might involve using standardized file formats, data models, and communication protocols. Interoperability allows researchers to integrate data and tools from different sources, enabling more comprehensive analyses.
- **Reusability:** Data and software should be well-documented and easy to understand, allowing researchers to reuse them for different purposes beyond their initial creation. This includes providing clear usage instructions, code examples, and detailed descriptions of data formats. Reusability maximizes the value of research outputs and avoids unnecessary duplication of effort.

The International Virtual Observatory Alliance (IVOA) [3] is a driving force in the push towards FAIR data in astrophysics, providing standards and guidelines for data and metadata models within the field. The collection of guidelines and tools provided by the IVOA comprise the Virtual Observatory (VO). Following these guidelines helps achieve FAIR compliance and improves interoperability between data from various experiments. The scope of these guidelines is wide, including:

- **Metadata Recommendations**: IVOA provides recommendations for the correct format of metadata, which is used to describe the data in detail. This metadata helps researchers understand the context, provenance, and quality of the data, facilitating its findability and reusability. A central example of a guideline about metadata and provenance is the Observatory Core Component (ObsCore) data model [4] which, as the name suggests, defines the core components of the Observation data model that are necessary to perform data discovery when querying data centers for astronomical observations of interest.
- **Registry Services**: IVOA promotes the development and use of data registries that list and describe astronomical datasets. These registries make data discoverable for researchers searching for specific information. The RegTAP protocol [5] was defined to allow users to search and retrieve metadata from various data registries using a standardized approach, enabling researchers to discover data across multiple archives without needing to learn different query methods. Examples of registry services built according to IVOA recommendations are the ESAC science archive [6], which lists and describes datasets generated by ESA missions, and the Vizier online catalog [7], which provides access to a vast collection of astronomical catalogs and tables, categorized by astronomical object type and research area.
- **Standardized Data Models and Formats**: IVOA defines common data models, describing the necessary information carried by the data files to be published, and file formats for astronomical data, such as FITS (Flexible Image Transport System) [9]. These standards ensure that data from different telescopes and instruments can be easily understood and combined for multi-wavelength or multi-instrument analyses.

---

[3] `https://www.ivoa.net/`

[4] `https://www.ivoa.net/documents/ObsCore/`

[5] `https://wiki.ivoa.net/twiki/bin/view/IVOA/RegTAP12RFC`

[6] `https://www.esa.int/About_Us/ESAC`

[7] `https://vizier.cds.unistra.fr/`

While the IVOA guidelines and recommendations offer broad applicability across various astrophysical subfields, applying them to high-energy (HE) and very-high-energy (VHE) data presents a greater challenge. This complexity stems from the fundamental difference in the data itself. At higher energies, all instruments rely on counting photons or particles, which often leads to key concepts from lower-energy astrophysics becoming inapplicable at these extremes. An illustrative example is the concept of an "event," which is central to HE, VHE, and multi-messenger astronomy but absent or very different in other fields.

For these reasons, adapting VHE and multi-messenger data to adhere to FAIR (Findable, Accessible, Interoperable, Reusable) principles and IVOA guidelines requires careful consideration. Success stories already exist, exemplified by the H.E.S.S. and ANTARES data releases. These releases, accessible via VO-integrated tools like Aladin and TopCat, demonstrate the potential for achieving findability and accessibility. However, they represent isolated contributions. A more systematic approach is necessary to ensure all data and scientific products become FAIR. The High-Energy Interest Group (HEIG) is the recently established working group in the IVOA that focuses on finding, adapting, or describing new guidelines for VHE data.

The most evident form of standardization in astrophysics data, and science data in general, is a universal data format. The VHE community developed the Gamma-ray Open Data Format (GADF) [8] as the first common data format, inspired by practices in X-ray astronomy. GADF is widely used in experiments like H.E.S.S., MAGIC, HAWC, and VERITAS. However, the limitations of GADF and the push towards stricter adherence to IVOA standards and FAIR principles led to the creation of its successor: the Very-High Energy Open Data Format (VODF) [8] [9] [10]. VODF prioritizes multi-messenger aspects and explicit IVOA compliance while introducing specific guidelines for high-energy data.

2 The Very-High Energy Open Data Format initiative

Launched in early 2023, the Very-High Energy Open Data Format (VODF) initiative has the official support of eleven VHE experiments: ASTRI, CTAO, FACT, Fermi-LAT, HAWC, H.E.S.S., IceCube, KM3NeT, MAGIC, SWGO, and VERITAS (listed alphabetically). The Fermi-LAT experiment's data format served as the foundation for the earlier GADF format and has been established for a significant period. Therefore, Fermi-LAT's participation in VODF brings valuable expertise and ensures the format caters to both ground-based and satellite experiments.

The initiative operates under a governance document developed in collaboration with official delegates from each participating experiment. This document outlines a clear and straightforward structure:

- Steering Committee: Comprised of one official delegate from each experiment.
- Lead Editors: Three editors oversee technical aspects, one for each experimental technique (IACT, WCD, Neutrino).
- Conveners: Two individuals responsible for overall coordination.

The development and improvement of the VODF format will be driven by open contributions from the VHE community (users and experiments) as well as experts from other astronomical fields. Users and experts are encouraged to participate by providing input on features, documentation, or associated tools. Lead editors will guide discussions, validate accepted proposals through consensus, and escalate any unresolved issues to the Steering Committee for arbitration.

Lead editors are responsible for maintaining supporting tools and reporting progress to the Steering Committee. Major decisions regarding future direction, roadmaps, and significant improvements will be formulated by the lead editors and evaluated by the Steering Committee.

The VODF aims to act as a standardized data model and format specifically designed for VHE gamma-ray and neutrino astronomy. VHE experiments exhibit significant variations in their detection techniques, instrumentation, and raw data types. Ground-based gamma-ray detectors fall into two main categories: Imaging Atmospheric Cherenkov Telescopes (IACTs) and Water Cherenkov Detectors (WCDs). Neutrino detectors, on the other hand, employ optical modules submerged in vast volumes (cubic kilometers) of ice or water.

Despite these differences, VHE experiments share some key properties after undergoing initial data processing (calibration, reconstruction, and background reduction):

---

[8] `https://vodf.readthedocs.io/en/latest/index.html`

[9] `https://github.com/vodf`

- All experiments provide a list of gamma-ray or neutrino candidates, characterized by their arrival time, energy, and celestial direction.
- The instrumental response to these events can be described using the same quantities and exhibits similar observational dependencies. This allows for a consistent factorization of the associated Instrument Response Functions (IRFs) across instruments.
- Since these instruments investigate the same astrophysical phenomena, the resulting scientific products are similar: sky maps, spectra, and light curves, accompanied by related statistical information like significance or likelihood profiles.

This shared nature of high-level data products allows for a unified data format across instruments. The success of this approach has been demonstrated through its ability to generate scientific results and promote true interoperability between instruments

As the primary goal of VODF is to establish a standard set of data models and file formats these standards must encompass reconstructed event-level data (science-ready data) and higher-level products like N-dimensional binned data cubes (encompassing sky images, light curves, and spectra) and source catalogs.

In parallel, several facilities will publicly open their data (e.g. CTAO, KM3NeT) or their archive, and the use of certified repositories is strongly recommended. As a consequence, the data should be correctly curated such that they respect as closely as possible the FAIR principles. Metadata describing the associated data is mandatory. VODF aims to standardize the format of these metadata for each level of data.

2.1 Data levels in the VODF

VODF aims to settle standards of different levels of VHE data resulting from the calibration, reconstruction, and background reduction of the instrument analysis pipelines. Data from all the different adhering instruments share many common structures such as:

- the time formats, following the FITS standards [12]
- the coordinate formats, following the IAU resolutions [13]
- the N-dimensional maps, with regular or sparse axis that can contain either bins or points and that handle physical units,
- the metadata, with standard keywords associated to e.g. the instrument, maintainer, data release identifier, data format version, VO standards;
- the provenance information, following the IVOA data model [10];

The data types and data levels to format are illustrated in the figure 1 and described below.

1. **Science-Ready Data**: This level provides data directly usable for scientific analysis. It includes a selection of events (likely containing both signal and background) with a final set of reconstructed parameters for each event. Additionally, it incorporates the instrument's response function (IRF), currently divided into four key components: effective area, energy dispersion, point spread function, and background model. It also stores auxiliary data describing observation conditions, such as the stable observation period, pointing position(s), and livetime. Finally, to adhere to FAIR principles, particularly findability and accessibility, this level should include index tables for efficient data retrieval.
2. **Binned Science Data**: This level is generated by binning the spatial, temporal, and/or spectral information within the science-ready data. This binning is performed over a specific target region or area of interest on the celestial sphere, as chosen by the user. The resulting data is typically presented in instrumental units, such as counts per bin.
3. **Advanced Science Data**: This level represents astrophysical products derived from the binned science data. It involves combining the binned events with the binned IRFs to generate products like spectral energy distributions, light curves, sky maps, and phasograms. These products are typically presented in physical units relevant to astrophysics, such as flux, energy, or angles.

---

[10] `https://www.ivoa.net/documents/ProvenanceDM/`

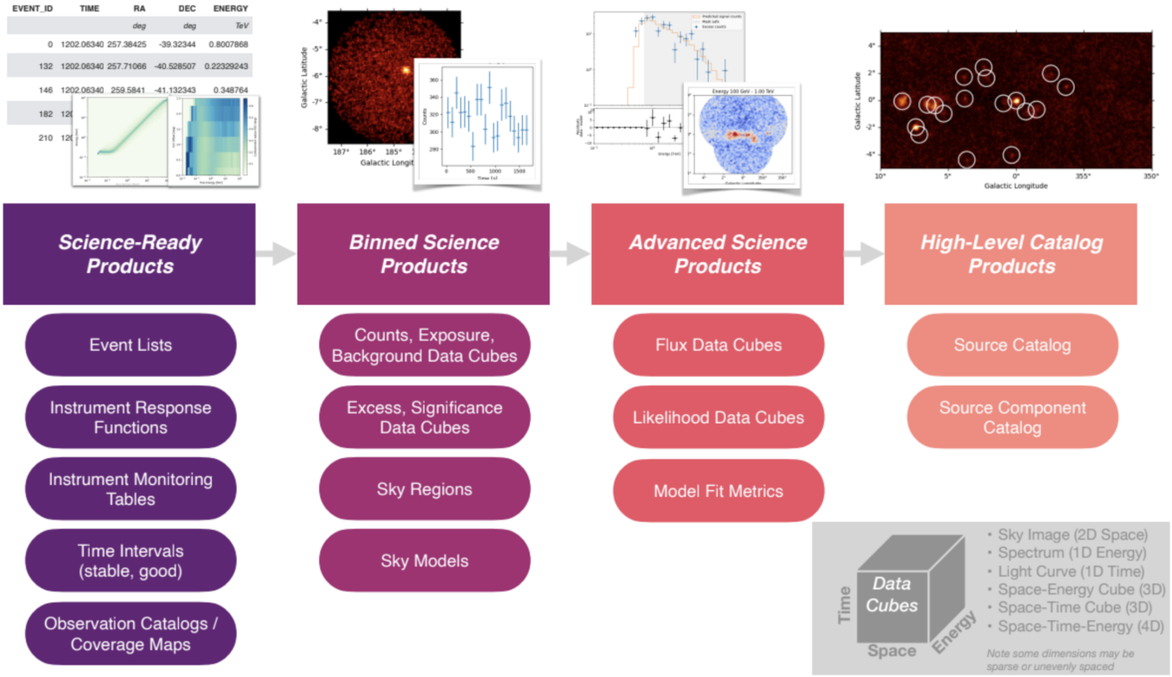

Figure 1: Diagram of the different data levels. Each rounded box stands for data and their associated metadata.

4. **High-Level Catalog Data**: This level contains a collection of astrophysical objects or VHE sources. It provides detailed information on their properties and associations, including source components, upper limits on size or flux, observation periods, and scientific alerts.

The definition of such a format will be preceded by a data modeling analysis. The data models of each data level could be considered to be released in addition to the data formats.

3 Conclusions

In conclusion, the field of VHE gamma-ray astronomy is undergoing a transformative period characterized by a surge in data volume, advancements in instrumentation, and a growing emphasis on open science practices. The adoption of standardized data formats like VODF and open-source software like Gammapy are crucial steps in facilitating collaboration, fostering multi-messenger analyses, and driving new discoveries in the high-energy universe.

The VODF initiative, in particular, is a new effort aiming to establish data standards for VHE data generated by ground-based gamma-ray and neutrino instruments. Backed by eleven prominent astroparticle projects, VODF seeks to define a new data format for high-level VHE instrument products, ensuring adherence to FAIR principles and close alignment with IVOA standards.

Data formatting plays a critical, yet often unnoticed, role in Open Science. It allows open data to comply with FAIR principles, just as open software aligns with FAIR4RS principles [11]. The VODF initiative, therefore, represents a significant undertaking that supports both existing and under-construction VHE experiments, along with the development of VHE Science Analysis Tools. By establishing a standardized format, VODF will pave the way for true interoperability between different VHE datasets, facilitating joint multi-wavelength and multi-messenger analyses.

References

[1] S. P. Wakely and D. Horan, 2016, *TeVCat: An online catalog for Very High Energy Gamma-Ray Astronomy*, in proceedings of the 30th International Cosmic Ray Conference (2007), [ads:2008ICRC....3.1341W], TeVCaT catalog.

[2] W. Hofmann, R. Zanin, 2023, *The Cherenkov Telescope Array, Handbook of X-ray and Gamma-ray Astrophysics* [arXiv:2305.12888].

[3] X.-H. Ma, et al., 2022, *LHAASO Instruments and Detector technology*, Chinese Phys. C 46 030001 [doi:10.1088/1674-1137/ac3fa6].

[4] H.E.S.S. Collaboration, 2018, *H.E.S.S. first public test data release*, [arXiv:1810.04516].

[5] ANTARES Collaboration, *Data sets for searches for cosmic neutrino point sources with ANTARES*, open data release.

[6] A. Donath et al., 2023, *Gammapy: A Python package for gamma-ray astronomy*, 10.1051/0004-6361/202346488, A&A.

[7] M. D. Wilkinson, et al., 2016, *The FAIR Guiding Principles for scientific data management and stewardship*, Scientific Data 3.

[8] C. Deil et al., 2016, *Open high-level data formats and software for gamma-ray astronomy*, http://dx.doi.org/10.1063/1.4969003, AIP Conference Proceedings

[9] D. C. Wells, E. W. Greisen, and R. H. Harten, 1981, *FITS - a Flexible Image Transport System*, Astronomy and Astrophysics Supplement 44.

[10] B. Khélifi, et al., 2023, *The Very-high-energy Open Data Format: towards a shared, open data format in very-high-energy astronomy*, Proc. of the 28th International Cosmic Ray Conference 2023.

[11] M. Barker, et al., 2022, *Introducing the FAIR Principles for research software*, Scientific Data 9, 622.

[12] A. H. Rots et al., 2015, *Representations of time coordinates in FITS. Time and relative dimension in space*, A&A 574 A36 [arXiv:1409.7583].

[13] G. H. Kaplan, 2005, *The IAU Resolutions on Astronomical Reference Systems, Time Scales, and Earth Rotation Models*, U.S. Naval Observatory Circular No. 179 [arXiv:astro-ph/0602086]