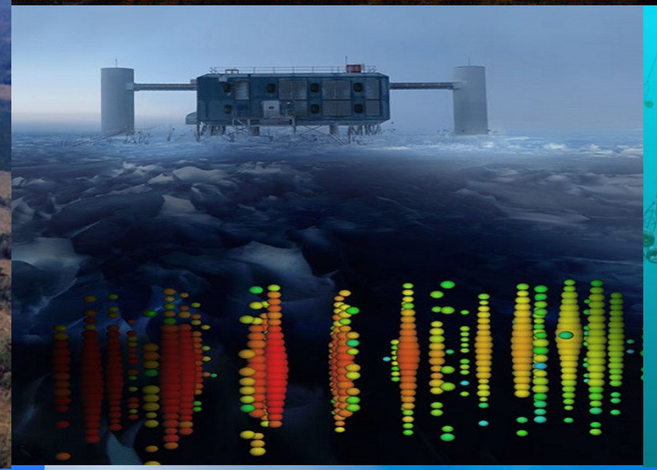
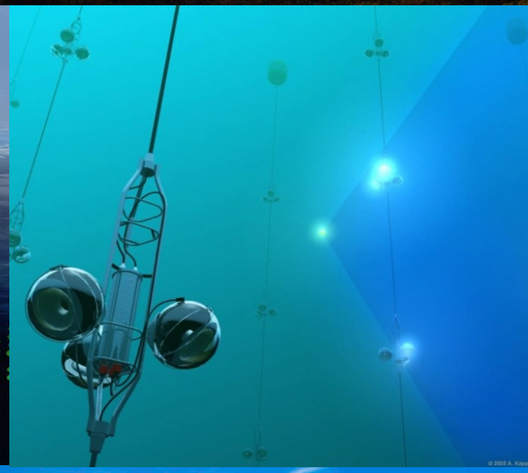
Open data from astroparticle experiments: a snapshot of the present and a glimpse of the future







Outline

OPEN DATA : **Introduction**

How are ASTROPARTICLE EXPERIMENTS*
doing with OPEN DATA?

Gamma-rays

Cosmic-rays

Neutrinos

Gravitational waves

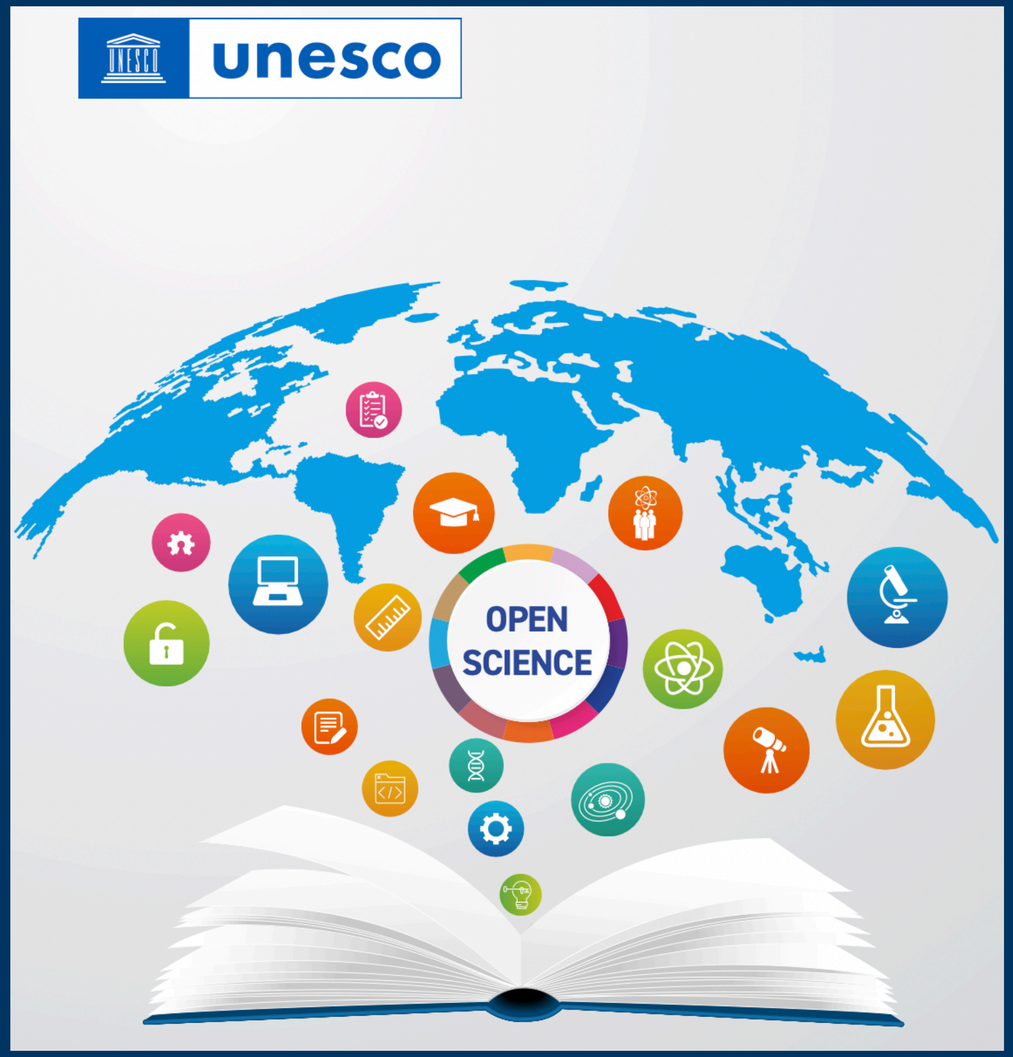
OPEN DATA **in astroparticle physics**:
PRESENT vs FUTURE

* ground based-only

Introduction

Open Research Data: a modern definition…

Open Data: a hot topic in modern science



WHAT

- Digital and analogue data, both raw and processed
- Metadata (data on the the data)
- Numerical scores, textual records, images
- Analysis codes and workflows

HOW

- Viable to be openly used, reused, retained and redistributed by anyone, subject to acknowledgement
- Available in a timely and user-friendly, human- and machine-readable and actionable format,
- In accordance with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles
- Supported by regular curation and maintenance.

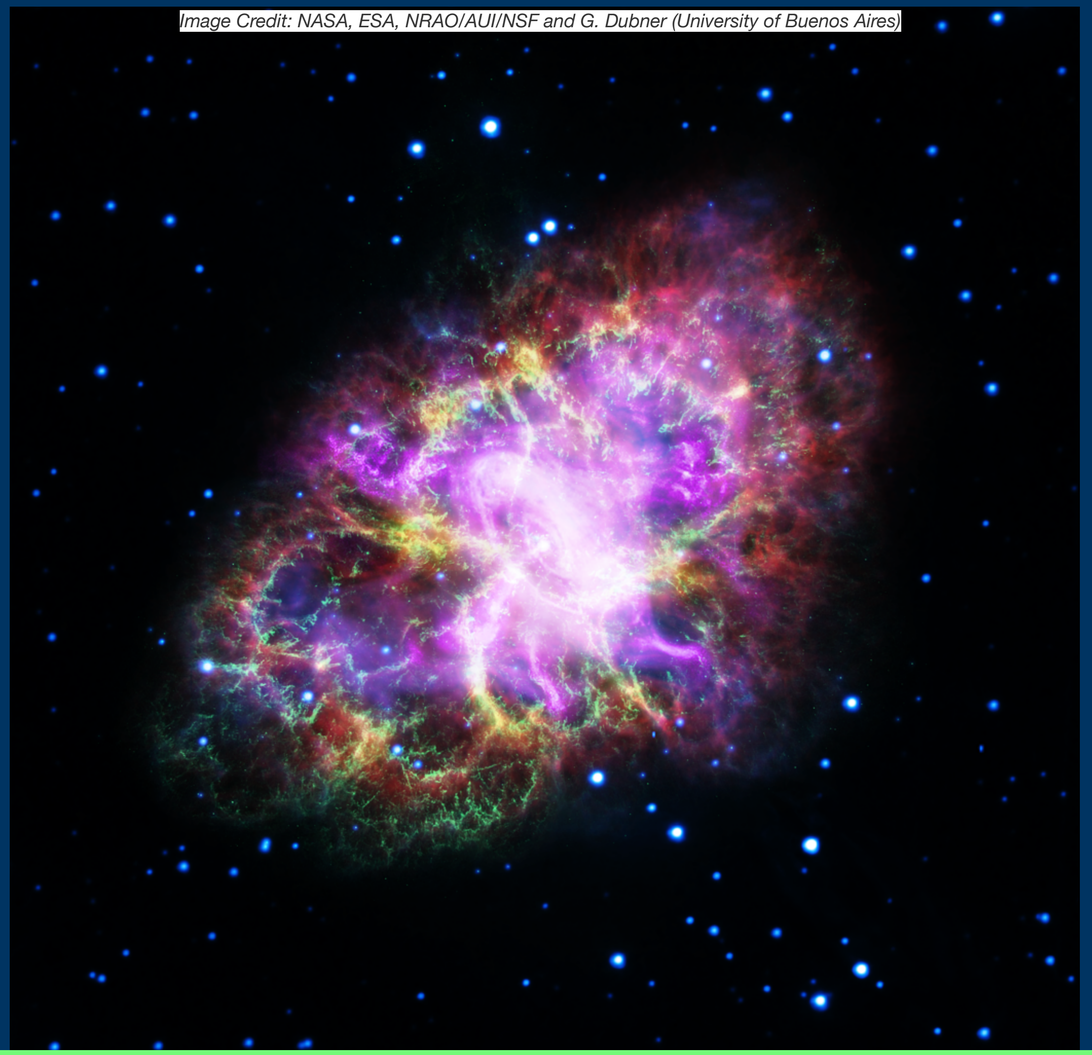
…but not a modern idea

The Crab Nebula remnant of SN1054: an ancient success of open data

Year 1054: a Chinese astronomer sees a strange and brilliant new star appear in the sky, in the stellar division of Peih (now known as constellation of Taurus). He notes its position and monitor it for 6 months, until it disappears.

A Japanese astronomer notes too the same strange star. Besides the position, he notes that it was as bright as the planet Jupiter.

The two records remain buried in the Oriental chronicles for ≈ 900 y



Year 1731: John Bevis identifies for the first time the Crab Nebula

Year 1921: Carl Otto Lampland discovers that the Crab Nebula is "expanding". Around that time, Lundmark publishes a list of apparently temporary objects, as extracted from translations of ancient Chinese chronicles.

Year 1928: Edwin Hubble makes the calculation that allows him to conclude that "…the nebula is expanding rapidly, at a such rate that it must have required 900 years to reach its present dimension"

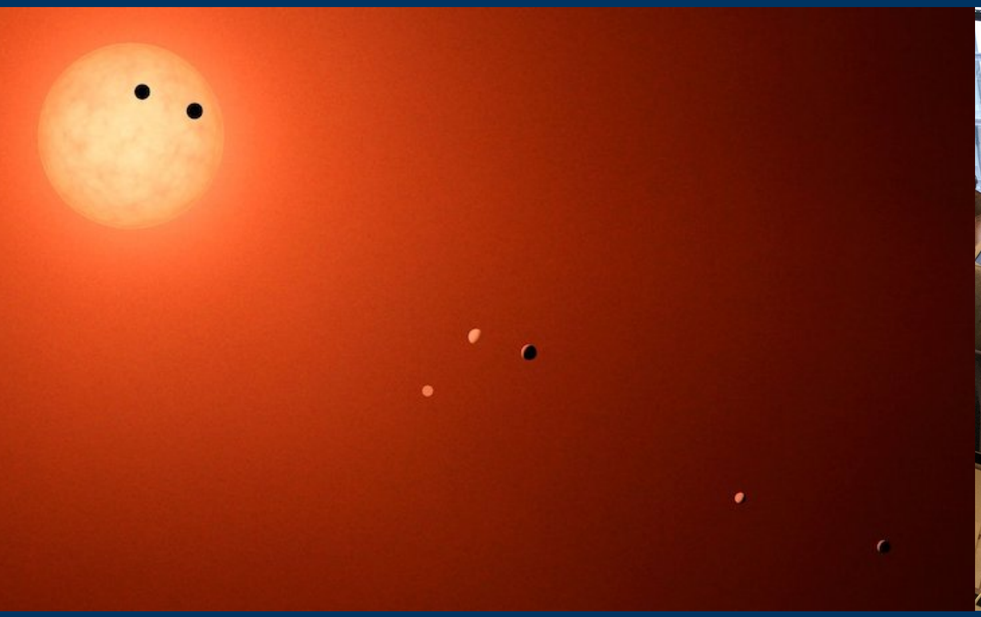
Open data maximise the scientific output of data that experimentalists have to work to obtain

Open Research Data: why?

Aren't chronicles and publications enough?

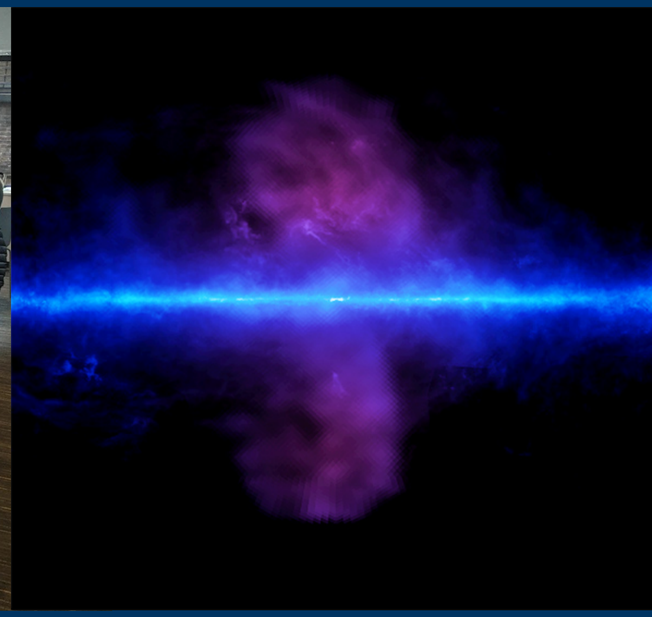
Publications make research results publicly available, to allow the rest of the academic audience to apprehend the results of research.

Open research data allow the rest of the academic audience either to verify the results of research or to conduct new analyses.



Public data of the Kepler Space Telescope (NASA) ended in 2018: discovery of new exoplanets (2021)

18 scientific papers based on public data of the GAIA mission (ESA) resulted from a 5-day research marathon hosted at the Flatiron Institute (2018)



Public data of the Fermi gamma-ray space telescope (NASA): discovery of the Fermi bubbles (2010)

Open data maximise the scientific output of data that experimentalists have to work to obtain

Open Research Data: why? (Or, why not?)

Yes, benefits, but also concerns and challenges

Benefits

Maximisation of the experiment output

Production of new science by the community so to progress in a more efficient and powerful way

More accountability (verification of results obtained)

New collaborations (not only among experimentalists but also between experimentalists and theoretician)

Automatic data preservation

Science development in places that do not have much science developed yet

Reach for the broader non-academic community that, after all, is the one funding public research

Concerns and challenges

Ownership (a collaboration might not want to release data that they could use for another project or might worry about proper credit and attribution)

Data will be available to those who are not involved in the “collection and curation”: what is the incentive to take part in these important activities? If the incentive to collect complex data is removed, the complex data may never be collected.

Can time delays (embargo/latency) really protect the incentive to work on data collection and curation?

Will external people be able to perform accurate analyses as they are not familiar with the intricate details of the data? Huge effort to make the data really understandable

Open data only is not enough: they need to be accompanied by software, workflows and explanations, all of which need to be readapted throughout the experiment lifecycle

Open Data and astroparticle physics

A community in-between astronomy and particle physics

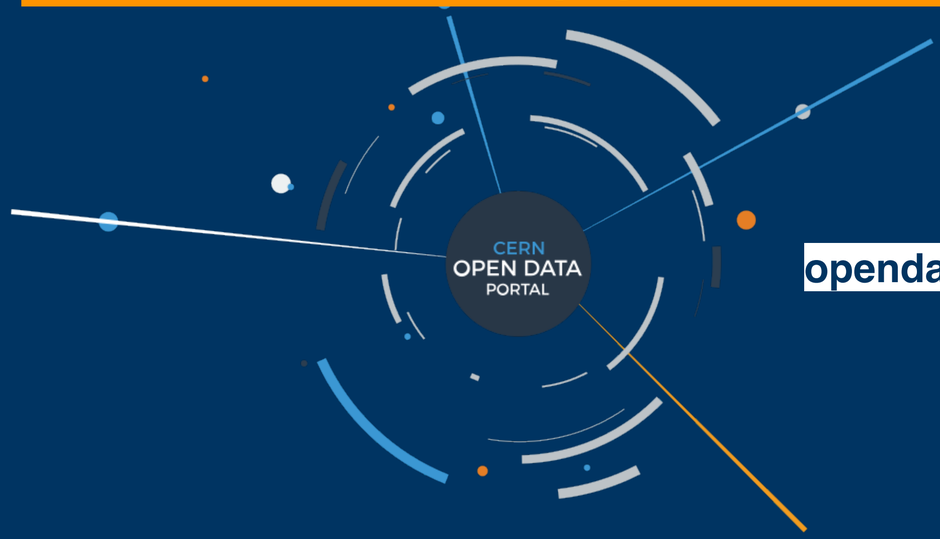
Astronomy

Astronomy pioneered Open Access to publications and data.

Data from large sky surveys are archived by public institutions (e.g., NASA, ESA), and made publicly available (e.g., http://archive.stsci.edu/, http://ned.ipac.caltech.edu/, http://skyview.gsfc.nasa.gov/, http://simbad.u-strasbg.fr/ simbad/).

Also, astronomical data have been integrated globally through the Virtual Observatory (VO), set up in 2002. The VO enable access to a multitude of on-line resources through a framework of interoperable tools.

Particle physics



opendata.cern.ch

Public access to (some, and growing) data produced at CERN (ALICE, ATLAS, CMS, LHCb, OPERA).

Not only data: software and documentation needed to understand and analyse them.

Aimed at outreach (all experiments) and research (only CMS at the moment)

The products are shared under open licenses; issued with a digital object identifier (DOI) to make them citable objects.

Open Data and astroparticle physics

A community in-between astronomy and particle physics

Astroparticle Physics, like astronomy, **seeks to understand the Universe and its components**.

But, much younger community, it has a **culture on data and software distribution that is closer to particle physics.**

Very large collaborations, in which member duties include detector design, construction, operation and maintenance, simulation development, calibration, that are a significant part of the **collective work and integrated with data analysis**.

Based on MoUs that defines the access to data and software (shifts, common funds…)

Large variety of locations in the world (no “central” CERN-like structure) and of experimental methods

“Old” experiments (≈ 20 years old): open data were not a topic as hot today, but awareness growing and boosted by (and needed for) the natural multi-messenger approach

Open data: how have the current experiments adhered?

VHE gamma-rays: H.E.S.S.

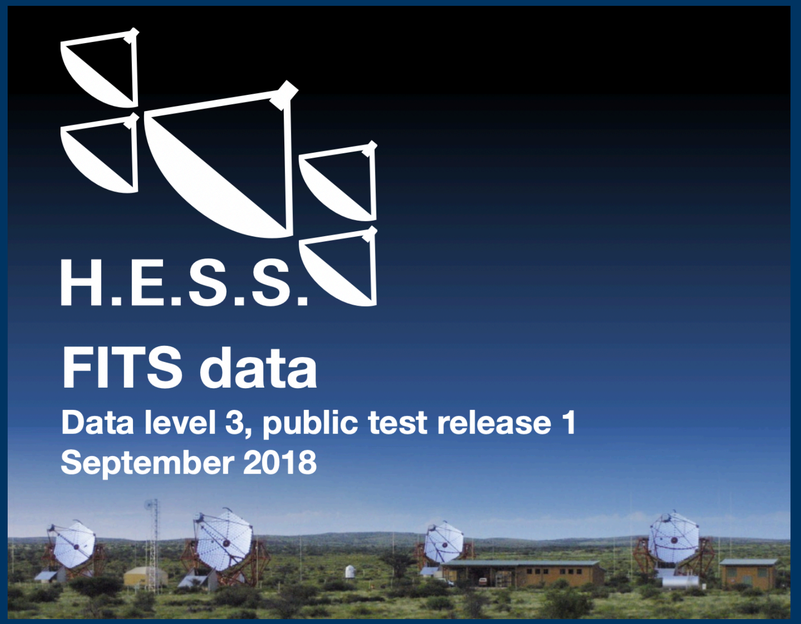
www.mpi-hd.mpg.de/hfm/HESS/pages/dl3-dr1/

H.E.S.S.: 5 Cherenkov telescopes in Namibia, operating since 2002 to study gamma rays in the energy range from tens of GeV to tens of TeV.

End of operation foreseen in 2025 (possibly after)

More than 200 scientists, ≈ 40 institutions, more than 10 countries

More than 100 publications (arXiv and/or Open Access journals)



Data release in September 2018 (una tantum).

High-level data (DL3): photon-candidate lists and instrument response functions (IRF), and relative documentation

Format: Conversion into Flexible Image Transport System (FITS) format, widely used in astronomy (http://fits.gsfc.nasa.gov/).

Small release: Includes observations of some gamma-ray sources (Crab, PKS 2155-304, MSH 15-52, RX J1713.7-3946) and of empty fields (“off-source” fields) for background modelling.

Complies with the open format specifications developed for the CTA Observatory

Goal: to have real VHE data publicly available for software and analysis method testing, to judge whether it properly specifies all data necessary for high-level science analysis. Not (yet) for science analysis.

VHE gamma-rays: MAGIC

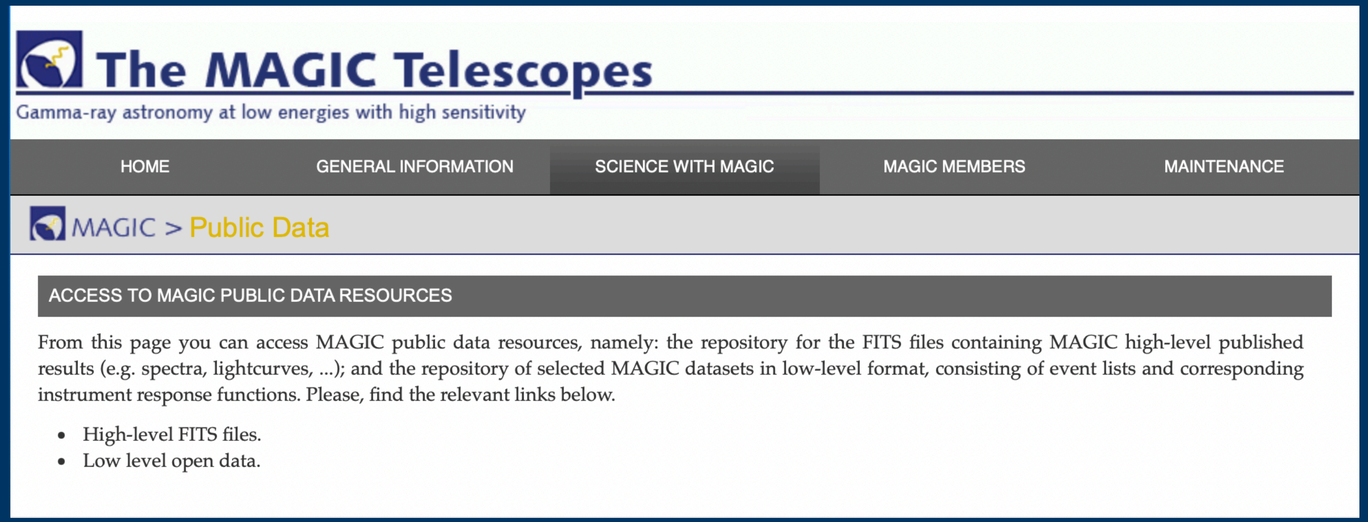
magic.mpp.mpg.de/public/public-data/

MAGIC: 2 Cherenkov telescopes in La Palma (Spain), operating since 2003 to study gamma rays in the energy range from ≈ 30 GeV to ≈ 100 TeV.

End of operation foreseen in 2024 (possibly after)

About 200 scientists, ≈ 20 institutions, more than 10 countries

About 200 publications (arXiv and/or OA journals)



Data release since 2019 (regularly updated since then).

High-level data (DL4: sky-maps, histograms, spectra, light curves and, in general, any set of data included in published papers) for all publications, and DL3 data (photon list + IRF) for 3 of them (TXS 0506+056 and Crab).

Format DL4 data: FITS, and interface with the Virtual Observatory. Searchable by source.

Complies with the open format specifications developed for the CTA Observatory

Goal: increment the quantity of DL3 data, and prepare the MAGIC "legacy" (archive). Include additional, "easy-to-use" format (i.e., ASCII) for a wider distribution.

VHE gamma-rays: MAGIC

magic.mpp.mpg.de/public/public-data/

"To download a MAGIC fits file, right-click on the corresponding "FITS" link and choose "Save As". To open the file, you will need a fits viewer, like fv.

Search the database by the source name, article title, reference or filter by the year."

I searched for "Crab" :-)

Crab					
Source	**Article**	**Year**	**Reference**	**Download**	**Other**
PKS 1510-089	Observation of a sudden cessation of a very-high-energy gamma-ray flare in PKS 1510-089 with H.E.S.S. and MAGIC in May 2016	2020	(MAGIC collaboration, A&A 648, A23(2021))	FITS	H.E.S.S data
Crab Nebula	Measurement of the Crab nebula spectrum over three decades in energy with the MAGIC telescopes	2015	(J. Aleksić et al., JHEA 5, 30)	FITS	
Crab Pulsar	Detection of bridge emission above 50 GeV from the Crab Pulsar with MAGIC Telescopes	2014	(J. Aleksić et al., A&A 565, L12)	FITS	
Crab Pulsar	Phase-resolved energy spectra of the Crab Pulsar in the range of 50-400 GeV measured with the MAGIC Telescopes	2012	(J. Aleksić et al., A&A 540, A69)	FITS	
Crab Nebula	Performance of the MAGIC stereo system obtained with Crab Nebula data	2012	(J. Aleksić et al., APh 35, 7)	FITS	
Crab Pulsar	Observations of the Crab Pulsar between 25 and 100 GeV with the MAGIC I Telescope	2011	(J. Aleksić et al., ApJ 742, 43)	FITS	
Crab Pulsar	Observation of Pulsed Gamma-Rays Above 25 GeV from the Crab Pulsar with MAGIC	2008	(E. Aliu et al., Science 322, 1221)	FITS	
Crab Pulsar	VHE Gamma-Ray Observation of the Crab Nebula and Pulsar with MAGIC	2008	(J. Albert et al., ApJ 674, 1037)	FITS*	

VHE gamma-rays: VERITAS (VTSCat)

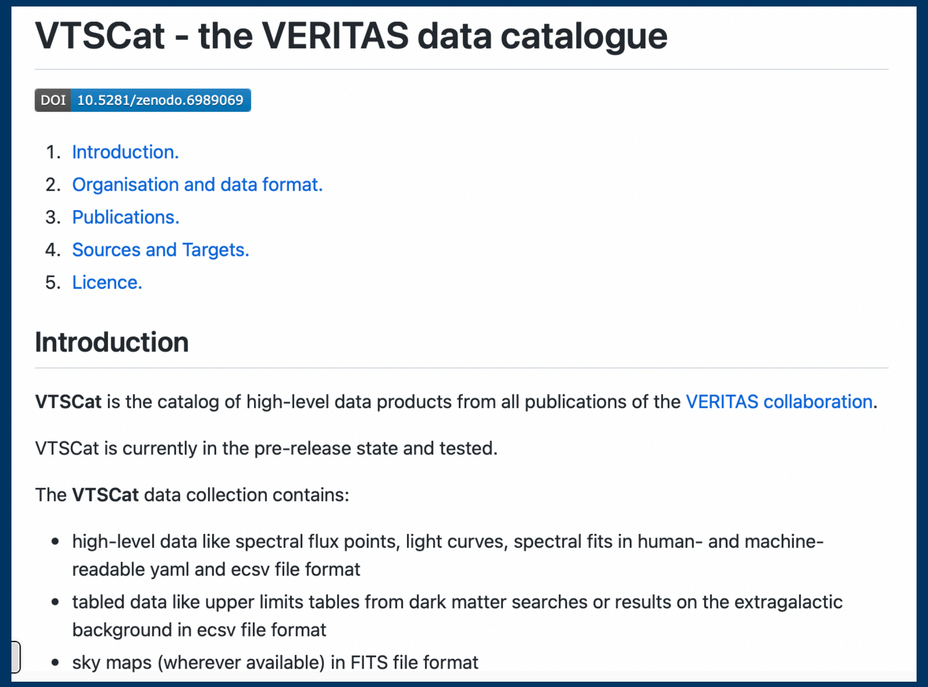
github.com/VERITAS-Observatory/VERITAS-VTSCat

VERITAS: 4 Cherenkov telescopes in Arizona, USA, operating since 2007 to study gamma rays in the energy range from ≈ 80 GeV to ≈ 30 TeV.

End of operation foreseen in 2025

About 80 scientists, mostly from USA and Canada, with members from Germany and Ireland too

More than 100 publications (arXiv and/or OA journals)



VTSCat - the VERITAS data catalogue

DOI 10.5281/zenodo.6989069

1. Introduction.
2. Organisation and data format.
3. Publications.
4. Sources and Targets.
5. Licence.

Introduction

VTSCat is the catalog of high-level data products from all publications of the VERITAS collaboration.

VTSCat is currently in the pre-release state and tested.

The **VTSCat** data collection contains:

- high-level data like spectral flux points, light curves, spectral fits in human- and machine-readable yaml and ecsv file format
- tabled data like upper limits tables from dark matter searches or results on the extragalactic background in ecsv file format
- sky maps (wherever available) in FITS file format

Data release in 2021 (regularly updated).

VTSCat is the catalogue of all VERITAS published observations.

High-level data (DL4: sky-maps, histograms, spectra, light curves and, in general, any set of data included in published papers) for all publications

Format DL4 data: FITS, ECSV, YAML.

Available on GitHub, Zenodo and HEASARC @ NASA.

Goal: to make VERITAS data more easily accessible, so that they can be easily integrated and help to enhance the work needed for future publications within the community.

VHE gamma-rays: VERITAS (VTSCat)

github.com/VERITAS-Observatory/VERITAS-VTSCat

2008	Update VER-000058.yaml	2 months ago
2009	update to v0.8.0	29 days ago
2010	update to v0.8.0	29 days ago
2011	add 2021 papers; yml and ecsv syntax corrections	8 months ago
2012	update to v0.8.0	29 days ago
2013	add 2021 papers; yml and ecsv syntax corrections	8 months ago
2014	update to v0.8.0	29 days ago
2015	update to v0.8.0	29 days ago
2016	update to v0.8.0	29 days ago
2017	update to v0.8.0	29 days ago
2018	update to v0.8.0	29 days ago
2019	add 2021 papers; yml and ecsv syntax corrections	8 months ago
2020	update to v0.8.0	29 days ago
2021	update to v0.8.0	29 days ago
2022	update to v0.8.0	29 days ago
sources	update to v0.8.0	29 days ago

- 2018A&A...620A.181A
- 2018ApJ...856...95A
- 2018ApJ...857...33A
- 2018ApJ...861..134A
- 2018ApJ...861L..20A
- 2018ApJ...862...41A
- 2018ApJ...866...24A
- 2018ApJ...867L..19A
- 2018PhRvD..98b2009A
- 2018PhRvD..98f2004A

Data files organised by year and publication, using ADS bibcodes as reference identifiers, and by sources.

The description files for sources can be found in the sources subdirectory and include the most relevant names for a given object (common name in the field, VERITAS object identifier, primary identifier by SIMBAD), and the object coordinates.

VHE gamma and cosmic rays: HAWC

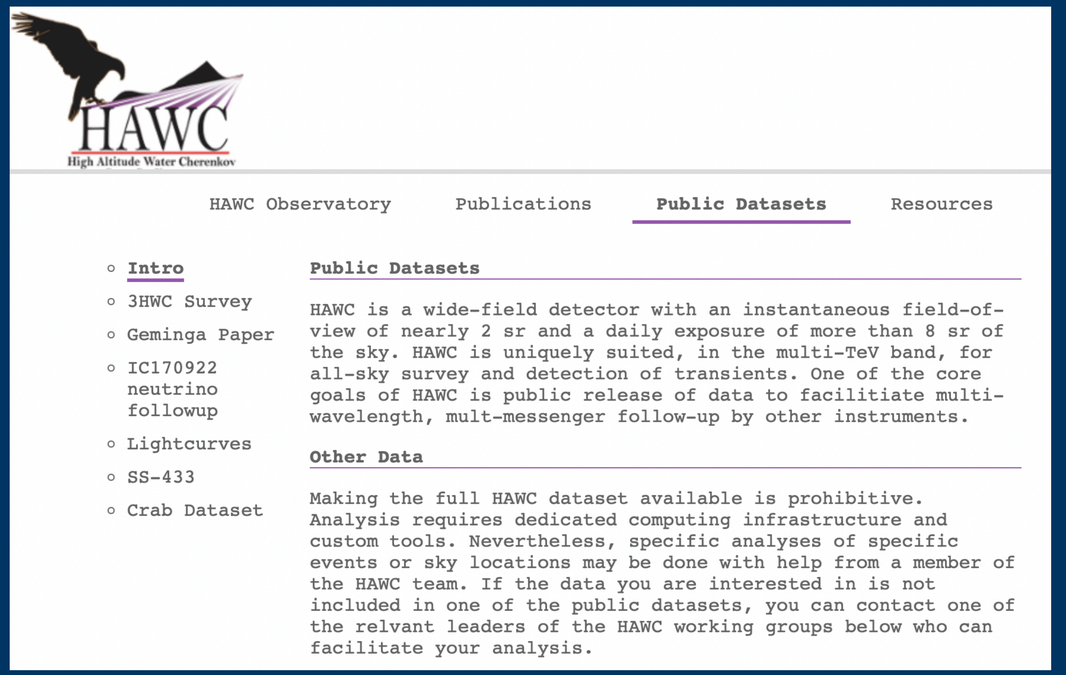
data.hawc-observatory.org

HAWC: extensive air-shower array of more than 600 water-Cherenkov tanks at Pico de Orizaba, Mexico, operating since 2013 to study gamma rays in the energy range hundreds of GeV to tens of TeV, and cosmic rays in the TeV range.

End of operation foreseen in 2025

More than 100 scientists, ≈ 80 institutions, 8 countries

More than 40 publications (arXiv and/or Open Access journals)



Data release in 2021, regularly updated

Data and codes, plus interactive tools.

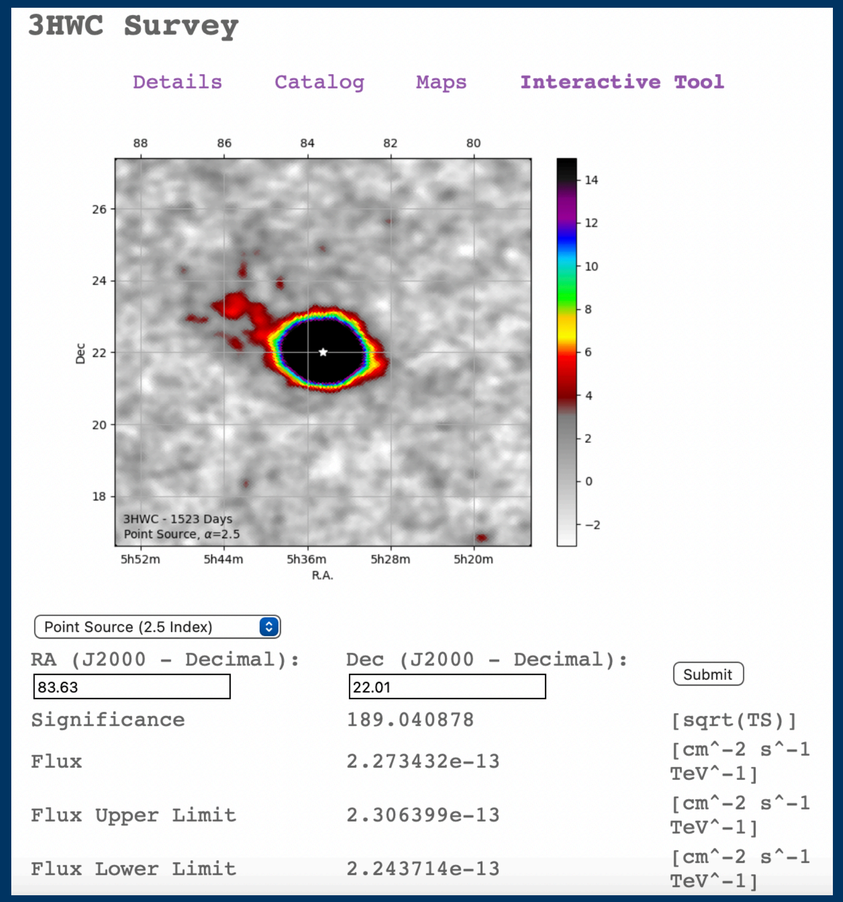
3HWC Catalog (1523 days): source list and interactive tool for significance maps and fluxes + daily light curves of Mrk 421 and Mrk 501 + Follow-up on IceCube-170922A neutrino alert (data) + Gamma-ray counts/background maps and analysis scripts for the Crab, Geminga and SS 433

Format: FITS, ECSV, YAML.

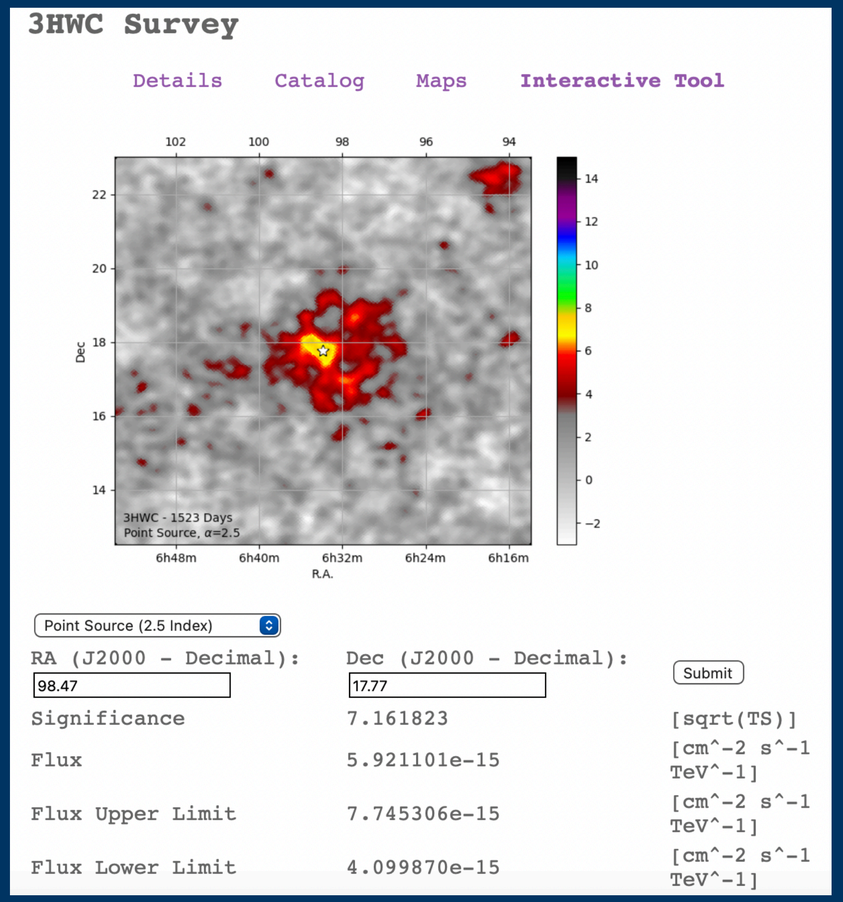
Goal: facilitate multi-wavelength, multi-messenger follow-up by other instruments

VHE gamma and cosmic rays: HAWC

data.hawc-observatory.org



Crab Nebula (default)



Geminga (my selection :-)

e.g., 3HWC Catalog: interactive tool

UHE cosmic rays: Pierre Auger Observatory

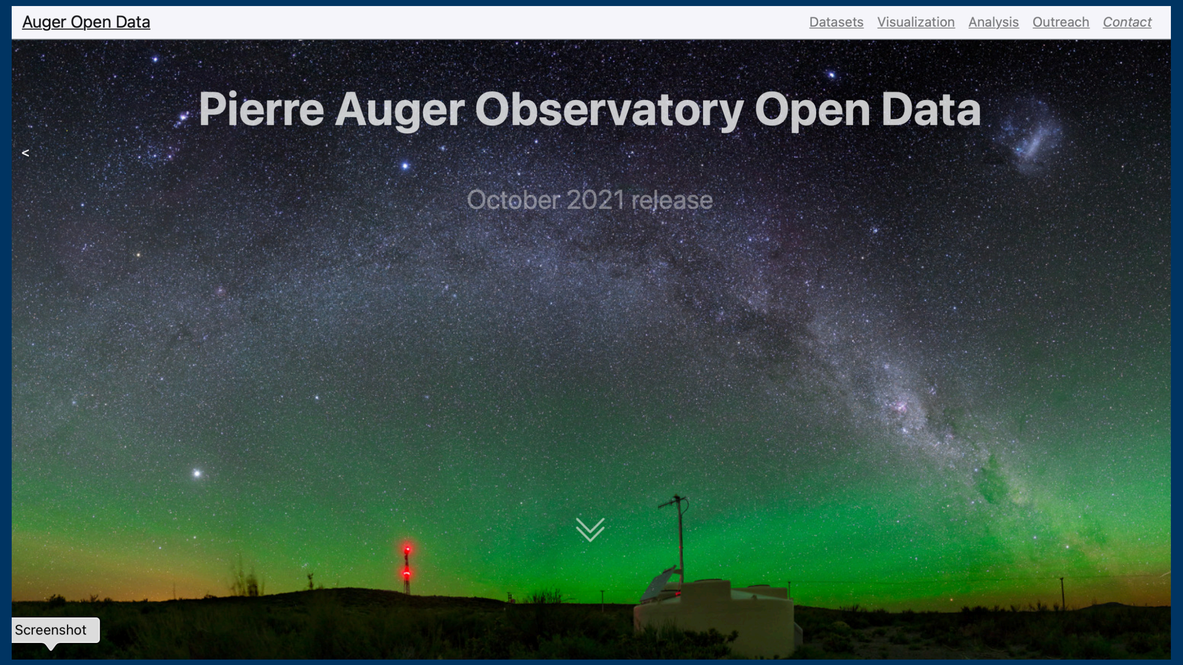
opendata.auger.org

Auger: 3000 km2 extensive air-shower array of ≈1600 water-Cherenkov tanks and 5 fluorescence telescopes in Argentina, operating since 2004 to study cosmic rays (and search for gammas and neutrinos) above 100 PeV.

End of operation in 2025 (to be extended to at least 2030, upgrade on-going)

More than 400 scientists, more than 90 institutions, 18 countries

More than 100 publications (arXiv and/or OA journals)



First data release in February 2021, regularly updated since then.

High-level and calibrated raw data plus tutorial scripts to learn reading the data + Python notebooks to build the cosmic ray energy spectrum, sky-map and elongation rate based on publications + event display 2D and 3D + extended documentation.

Quantity: 10% of shower data, randomly extracted (last digit of event ID). 100% of scaler and atmospheric data.

Format: JSON for data, Python for scripts. Interactive tools for running and modifying scripts

Goal: Increase regularly the released data set, either adding new types of events or enlarging the fraction.

UHE cosmic rays: Pierre Auger Observatory

opendata.auger.org

Visualization

This page provides an Event Browser, that can be used to display any cosmic-ray event in the data release. Events can be selected by providing their id or by selecting the value of some of their main properties. Some example events can also be selected from the menu below. Once an event is selected its components can be browsed in different tabs.

Visualize some example events

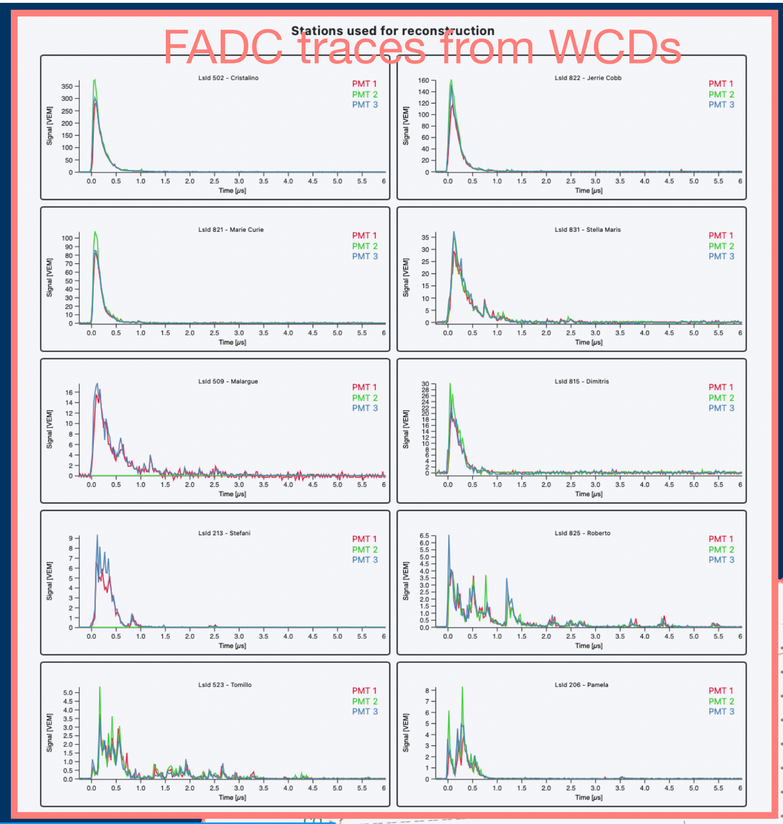
	Nb of stations	Energy [EeV]	Zenith Angle [deg]	Time [gps]
Min.	1	0	0	756950413
Max.		1000	60	1261872018

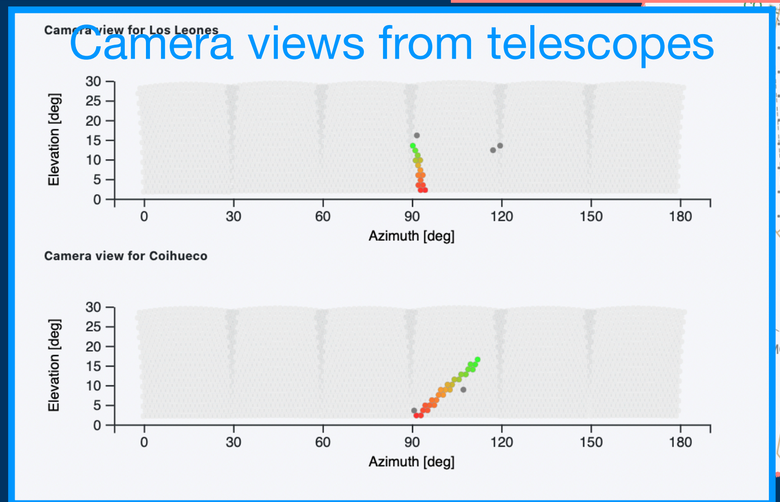
Event type: Select event type — Select

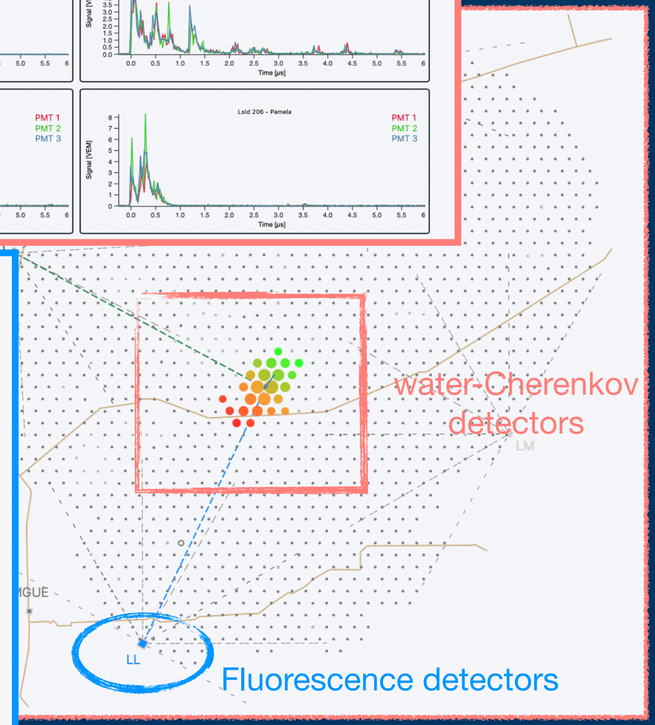
Select an event by id:

Enter specific event ID — Go

e.g., the event browser







UHE neutrinos: ANTARES

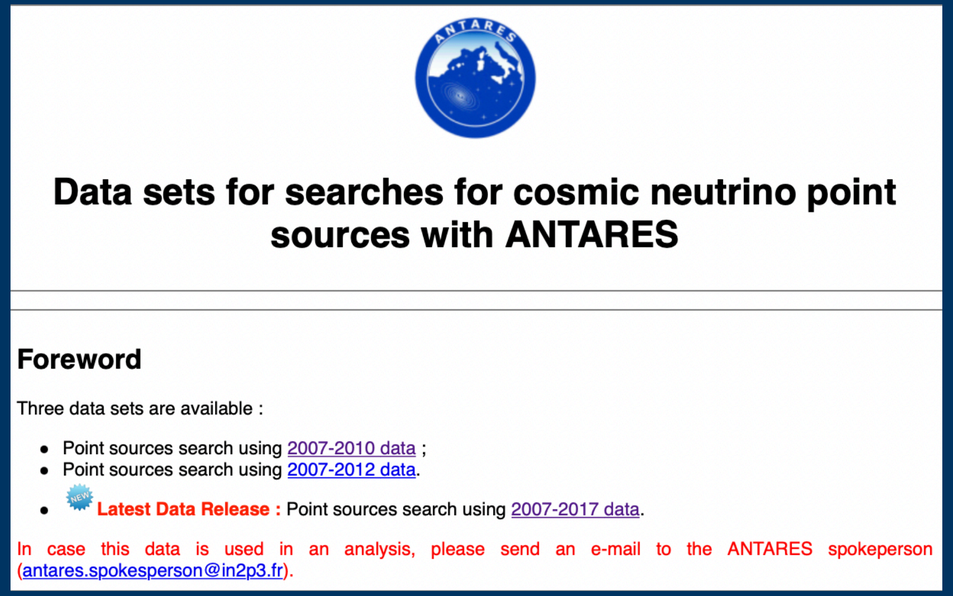
antares.in2p3.fr/publicdata.html

ANTARES: deep-sea detector, 40 km off Toulon, France, made of 12 lines of 900 optical modules, operating since 2007 to observe neutrinos with energies above 100 GeV.

End of operation in February 2022

More than 100 scientists, ≈ 30 institutions, 8 countries

More than 80 publications (arXiv and/or OA journals)



Data sets for searches for cosmic neutrino point sources with ANTARES

Foreword

Three data sets are available :

- Point sources search using 2007-2010 data ;
- Point sources search using 2007-2012 data.
- **Latest Data Release :** Point sources search using 2007-2017 data.

In case this data is used in an analysis, please send an e-mail to the ANTARES spokeperson (antares.spokesperson@in2p3.fr).

First data release in 2016, updated in 2019

High-level data: declination, right ascension, detected number of hits and angular resolution for each released event

Gradual releases: three larger and larger data samples (corresponding to different periods and publications): muon tracks, used for the search of point sources (released up to 2017)

Format: ASCII

Effective area, acceptance, cumulative angular resolution distribution and the point spread function for an E-2 source spectrum are also provided

Goal: Increase the released data set (include neutrino oscillation results) up to release all data in two years

UHE neutrinos: IceCube

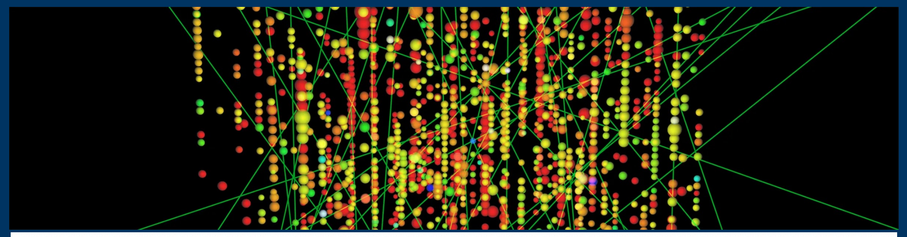
icecube.wisc.edu/science/data-releases/

IceCube: cubic-kilometer particle detector in ice near the Amundsen-Scott South Pole Station, consisting of 84 lines of more than 5000 optical modules, operating since 2011 to observe neutrinos from ≈ 100 GeV to 10 EeV

Upgrade on-going

About 300 scientists, ≈ 50 institutions, 12 countries

Almost 200 publications (arXiv and/or Open Access journals)



Data Releases

Featured Releases

On July 12, 2018, the IceCube Collaboration together with partner telescopes and observatories announced the first evidence of a source of high-energy cosmic neutrinos. The results were presented in two papers published in Science. The first paper provided an analysis of the neutrino alert event IceCube-170922A in coincidence with electromagnetic observations of the blazar TXS 0506+056. The second paper presented an analysis of IceCube neutrino point source data in the direction of TXS 0506+056 prior to the alert event. Together these results allowed the identification of this blazar as the first likely source of high-energy neutrinos and cosmic rays.

Regular data release since 2011

High-level data: direction (right ascension, declination), time, reconstructed energy, and quality information of each event (e.g., number of hits, angular resolution, effective area)

Release corresponding to the two Science publications on TXS 0506+056 + 25 other sets used for the point and diffuse source search

Includes data from Amanda

Goal: Increase regularly the released data set, after the main analyses are completed and results are published by the international collaboration.

UHE neutrinos: IceCube

icecube.wisc.edu/science/data-releases/

HESE 7.5 year data release

Posted on December 15, 2021

Introduction

This data release accompanies results published in *Physical Review D* describing a revisited analysis of the HESE (high-energy starting events) sample with an additional 4.5 years of data, newer glacial ice models, and improved systematics treatment.

For further details, refer to the IceCube publication doi:10.1103/PhysRevD.104.022002 / arXiv:2011.03545.

Data release

Suggested citation for this dataset:

IceCube Collaboration (2021): HESE 7.5 year data release. Dataset. DOI: https://doi.org/10.21234/4EQJ-BB17

Click here to download (.zip, 75 MB)

Provided in this release are several json files. The first contains the 102 data events that pass the HESE selection criterion. The others file contain information for the MC events used to compute the expected data event rates. The contents of these files are described below.

As an instructive example, we provide a python3 script which reproduces the fit of the data to a single power-law astrophysical flux in the same manner as described in the text. The primary goal of these scripts are to provide a working example utilizing the information provided in the data files, and we encourage readers to use these files as a jumping point into their own analyses.

Data files

The zipped data download contains the following:

- README: A starting point file, describing the full contents of the data release files, including file locations, descriptions of variables provided for the analyzed events, and summaries of provided scripts.
- LICENSE: The GNU Lesser General Public License.
- Python scripts: Individual files for scripts described in the README file.
- “resources” folder: Contains the “data” subfolder, among others, that holds the four json files mentioned above.
- “test” folder: Contains a single test script.

For additional detail, see the GitHub repository.

HESE events

All-sky point-source IceCube data: years 2008-2018

Posted on January 26, 2021

Introduction

IceCube has performed several searches for point-like sources of neutrinos. The events contained in this release make up the sample used in IceCube’s 10-year time-integrated neutrino point source search [1]. Events in the sample are track-like neutrino candidates detected by IceCube between April 2008 and July 2008.

The data contained in this release of IceCube’s point source sample shows 3.3σ evidence of a cumulative excess of events from a catalogue of 110 potential sources, primarily driven by four sources (NGC 1068, TXS 0506+056, PKS 1424+240, and GB6 J1542+6129). NGC 1068 gives the largest excess and appears in spatial coincidence with the hottest spot in the full Northern sky search [1].

IceCube’s 10-year neutrino point source event sample includes updated processing for events between April 2012 and May 2015, leading to differences in significances of some sources, including TXS 0506+056. For more information, please refer to [2].

This release contains data beginning in 2008 (IC40) until the spring of 2018 (IC86-VII). In order to standardize the release format of IceCube’s point source candidate events, this release duplicates and supplants previously released data from 2012 and earlier. Events from this release cannot be combined with other IceCube public data releases.

Data release

Suggested citation for this dataset:

IceCube Collaboration (2021): All-sky point-source IceCube data: years 2008-2018. Dataset. DOI: http://doi.org/DOI:10.21234/sxvs-mt83

Click here to download (.zip, 35 MB)

Included in the download are the following files:

Data files

- Read Me: A file describing each field of the data, uptime, and tabulated response files.
- The “events” subfolder contains the events observed in the 10-year sample of IceCube’s point source neutrino selection. Each file corresponds to a single season of IceCube data taking, including roughly one year of data. For each event, reconstructed particle information is included.
- In order to properly account for **detector uptime**, IceCube maintains “good run lists,” These contain information about “good runs,” periods of data taking useful for analysis.

Muon tracks

Gravitational waves: LVK (Ligo Virgo Geo600 Kagra)

https://www.gw-openscience.org

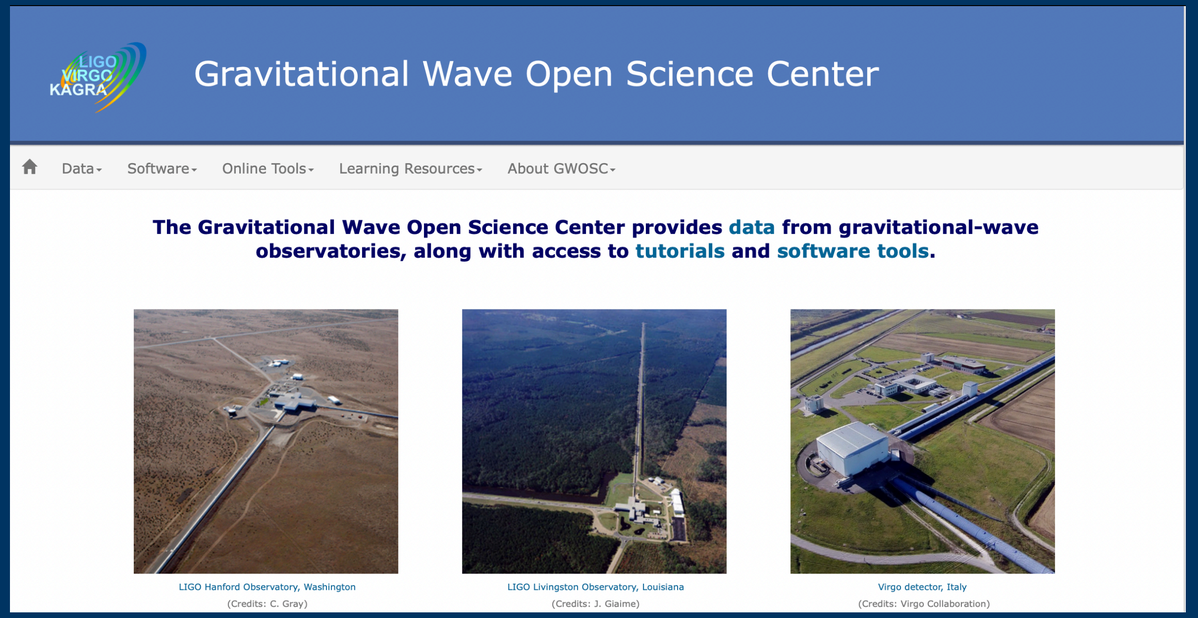
VIRGO: Michelson interferometer with 3 km arms, in Italy.

LIGO: two laser interferometers with 4 km arms, in USA

GEO: a British–German interferometric GW detector with 600 m arms, in Germany.

KAGRA: laser interferometer with 3 km arms, in Japan

KAGRA joined LIGO, Virgo, and GEO towards the end of the third observing run O3 to form the LVK collaboration.



Regular data release since 2016

All GW data are publicly accessible, up to O3.

Main data are the gravitational-wave strain arrays, released as time series sampled at 16 KHz (original), or 4 KHz (down-sampled).

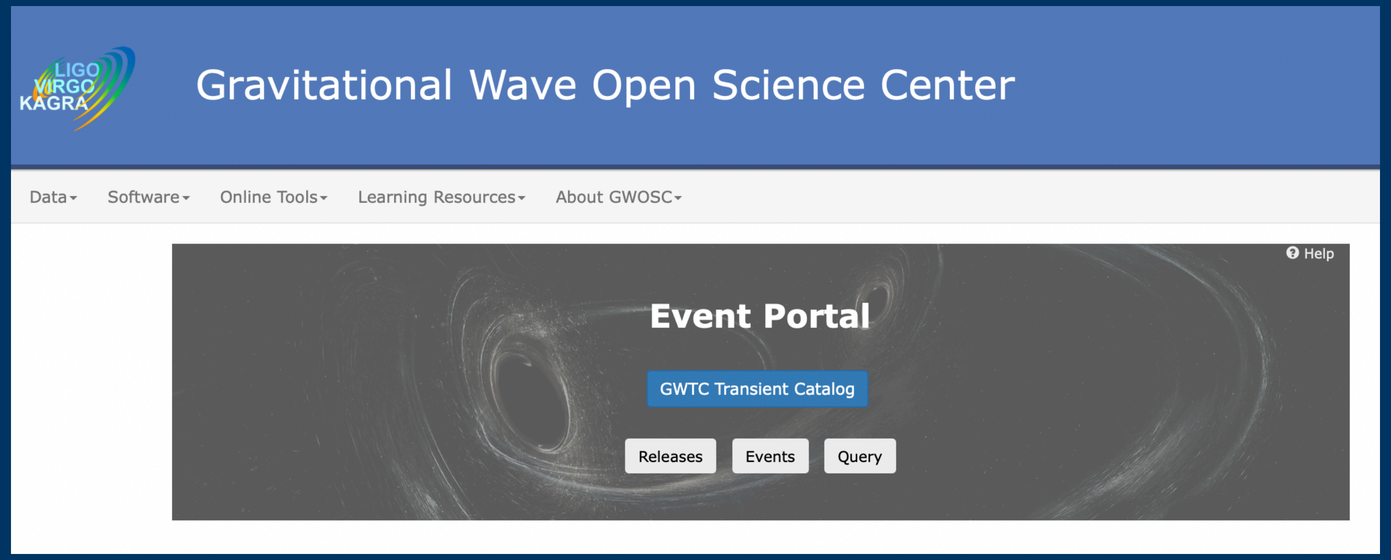
Data-quality information, documentation, usage guidelines, tutorials, and supporting software are also available

Content of the release: data relative to validated discoveries and the entire datasets of an observation run.

Periodicity and latency: every 6 months, in blocks of 6 months of data, with a latency of 18 months from the end of acquisition of each observing block.

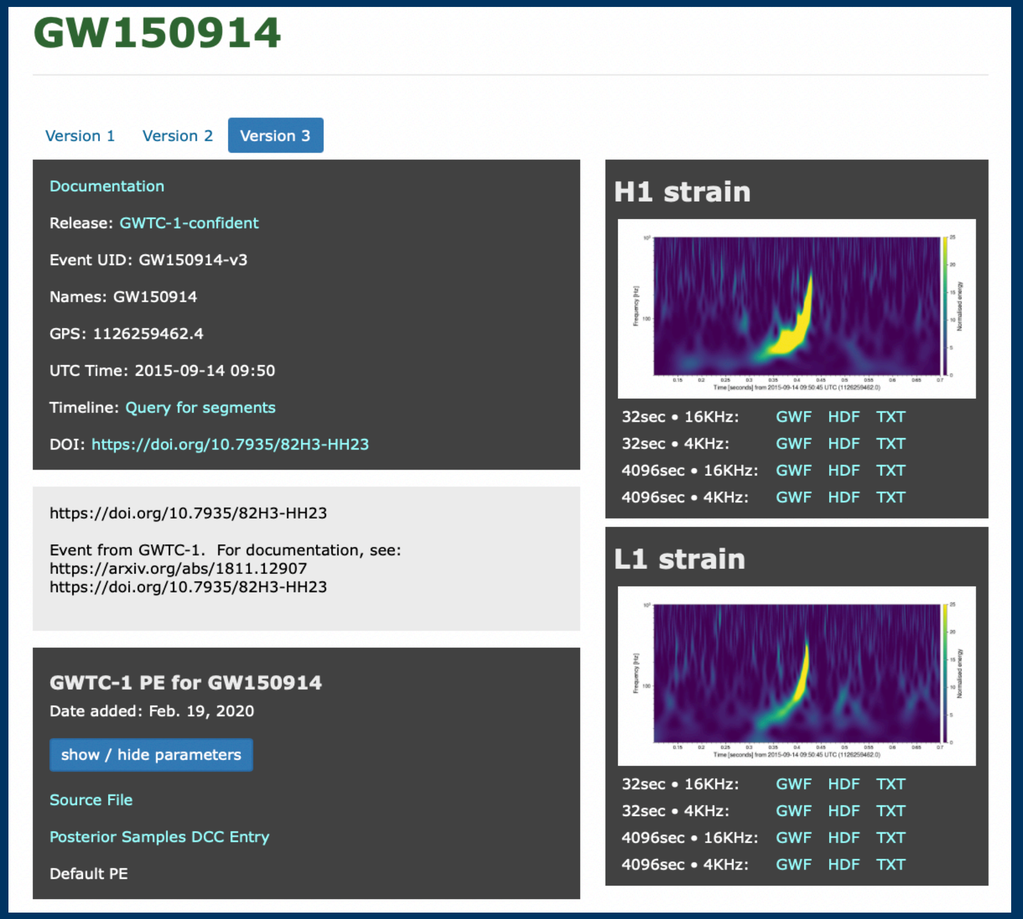
Gravitational waves: LVK (Ligo Virgo Geo600 Kagra)

https://www.gw-openscience.org



All data

All events



The discovery event :-)

Open data in astroparticle physics: present vs future

Astroparticle experiments: how are they doing with open data?

A very large variety of releases at very varied degree

Open data in principle (ex-UNESCO)

1. Digital and analogue data, both raw and processed
2. Metadata (data on the the data)
3. Numerical scores, textual records, images and sounds
4. Analysis codes and workflows
5. Viable to be openly used, reused, retained and redistributed by anyone
6. Available in a timely and user-friendly, human- and machine-readable and actionable format,
7. FAIR (Findable, Accessible, Interoperable, and Reusable) principles
8. Supported by regular curation and maintenance.

Open data in practice

1. All experiments release processed (high-level) data, few of them raw data too
2. All released data include metadata
3. All released data include numerical and textual records, images
4. Very few experiments provide codes
5. Released data are downloadable and usable
6. Not all released data are user-friendly (some need installation of specific software)
7. Released data are well-findable and accessible, Interoperation is still to be developed
8. All released data appear to be regularly curated and well-maintained

Astroparticle experiments: how are they doing with open data?

Present vs future

20 years ago (i.e., old experiments):
open data were a vague idea

Internal-only data and software management
(Data and codes access for the collaborations only)

Open data conceived “a posteriori”
(during the experiment):
Additional work to render data and codes accessible by anyone (even by companion collaborations)

Open data realised by “volunteers” collaborators:
a variety of format, styles and tools

Available open data “independent” (“un-coordinated”):
No single place where one can find all of them

New experiments (i.e,. CTA, Km3Net…):
open data are designed since the project start

Data and software management for external users too
(Data and codes external-access already in the projects)

Open data conceived “a priori”
(planned before the experiment):
Work on data format and tools as open-source as possible

Open data realised by “specialised” collaborators:
effort towards common format and tools

Coordinated effort to make open data interfaced in a Virtual Observatory (e.g., ESCAPE infrastructure)

Astroparticle experiments: will data still usable in 900 y? ;-)

Care about the longevity of data and science

Change of mentality (also in our daily work):

e.g.,

- spend more time to familiarise new members with ongoing and past analyses in our teams
- ask ourself if our next PhD student can build on the work of our current one
- ensure verification and validation of our codes before running new analyses.
- discuss with collaborators if our and their results are preserved and reusable in the long term.
- think about publishing code, data and notes on each analyses in at least internal repositories
- …

Change of organisation:

- Spend time at the beginning of a project to plan the output data of the experiment, how they can be organised not only for collaborators but also to be shared (during operation and at the end of it)

- Need for a new type of physicist that is neither an “experimentalist” or a “theorist”, but a “data physicist” that need to have not only an astroparticle physics (or high-energy physics in general) background, but also training in statistics/data science/machine learning and scientific computing.

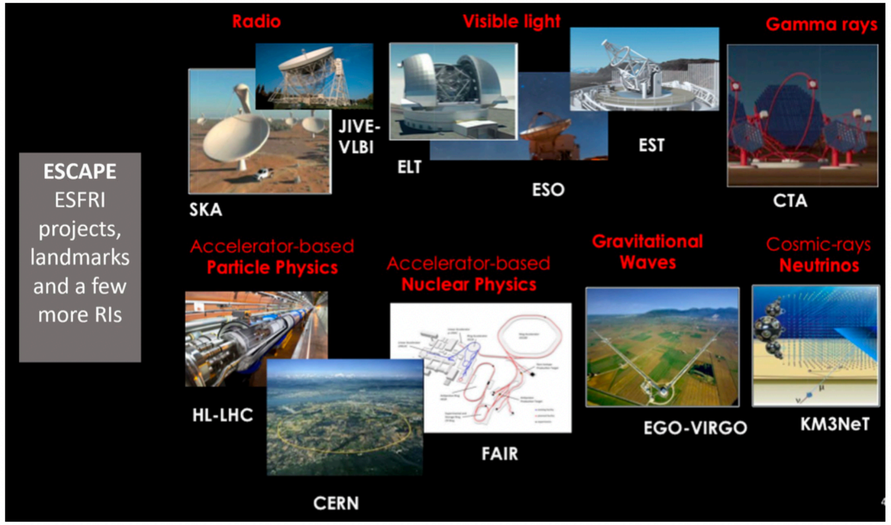
- And need to envision new career paths (and sensibilise evaluators) for data physicists.

Astroparticle experiments: how are they doing with open data?

Towards a coordinated open-data effort

ESCAPE

- EOSC is the European Commission action in response to EU member states' shared policy about the uptake of Open Science
- ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructure (48 months; 1/2/2019; lead CNRS-LAPP)



ESCAPE Work Programme

Data Lake:
- Build a scalable, federated, data infrastructure as the basis of open science for the ESFRI projects within ESCAPE. Enable connection to compute and storage resources.

Software Repository:
- Repository of "scientific software" as a major component of the "data" to be curated in EOSC. Implementation of a community-based approach for the continuous development of shared software and for training of researchers and data scientists.

Virtual Observatory:
- Extend the VO *FAIR* standards, methods and to a broader scientific context; prepare the VO to interface the large data volumes of next facilities.

Science Platforms:
- Flexible science platforms to enable the open data analysis tailored by and for each facility as well as a global one for transversal workflows.

Citizen Science:
- Open gateway for citizen science on ESCAPE data archives and ESFRI community

G.Lamanna