

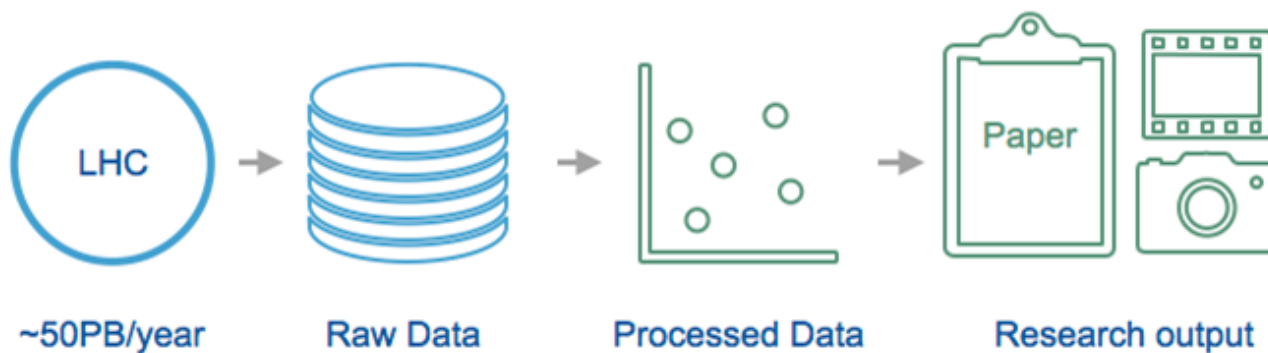
Data at CERN

Services for publishing, linking and preservation

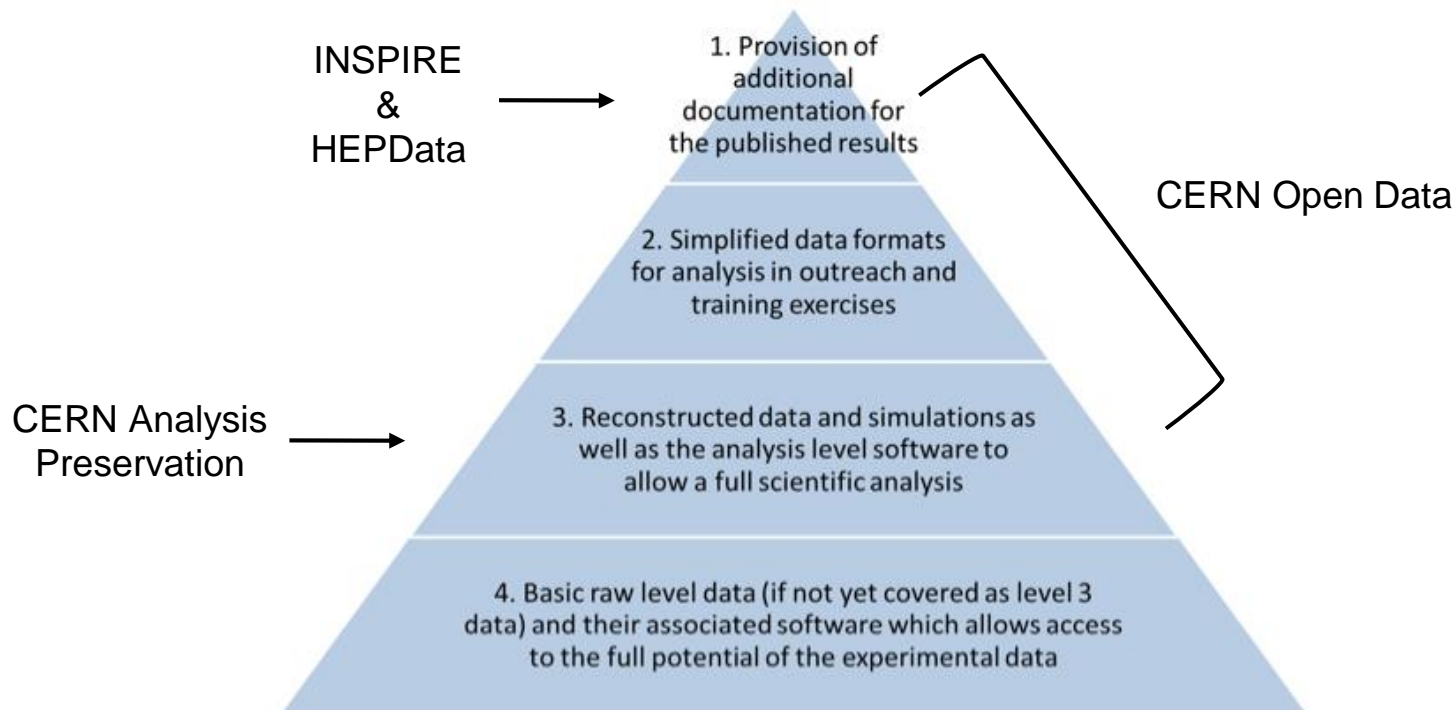
Slides prepared by Artemis Lavasa and Sünje Dallmeier-Tiessen

Presentation at the CERN-UNESCO School on Digital Libraries, Nairobi, 10th October 2018 given by Jens Vigen

Data in HEP



Data in HEP



Access to HEP data: Data Policies

LHCb External Data Access Policy

ALICE data preservation strategy

Sunday, October 6, 2013

The data harvested by the ALICE investment in human and financial information for the in depth understanding of the collision process.

Approved CB 20th June 2014

ATLAS Data Access Policy

May 21st 2014

CMS data preservation, re-use and open access policy

CMS data are unique and are the result of vast and long-term moral, human and financial investment by the international community. There is unique scientific opportunity in re-using these data, at different level of abstraction and at different points in time¹. This opportunity calls for our collective responsibility, and poses unprecedented challenges as no data sample of this complexity and value has ever been preserved or made available for later re-use.

The CMS collaboration is committed to preserve its data, at different levels of complexity, and to allow their re-use by a wide community including: collaboration members long after the data are taken, experimental and theoretical HEP scientists who were not members of the collaboration, educational and outreach initiatives, and citizen scientists in the general public.

CMS upholds the principle that open access to the data will, in the long term, allow the maximum realization of their scientific potential. To that extent, CMS will provide open access to its data after a suitable but relatively short embargo period, allowing CMS collaborators to fully exploit their scientific potential.

publication policy. This is to data at different levels as take the data available in a

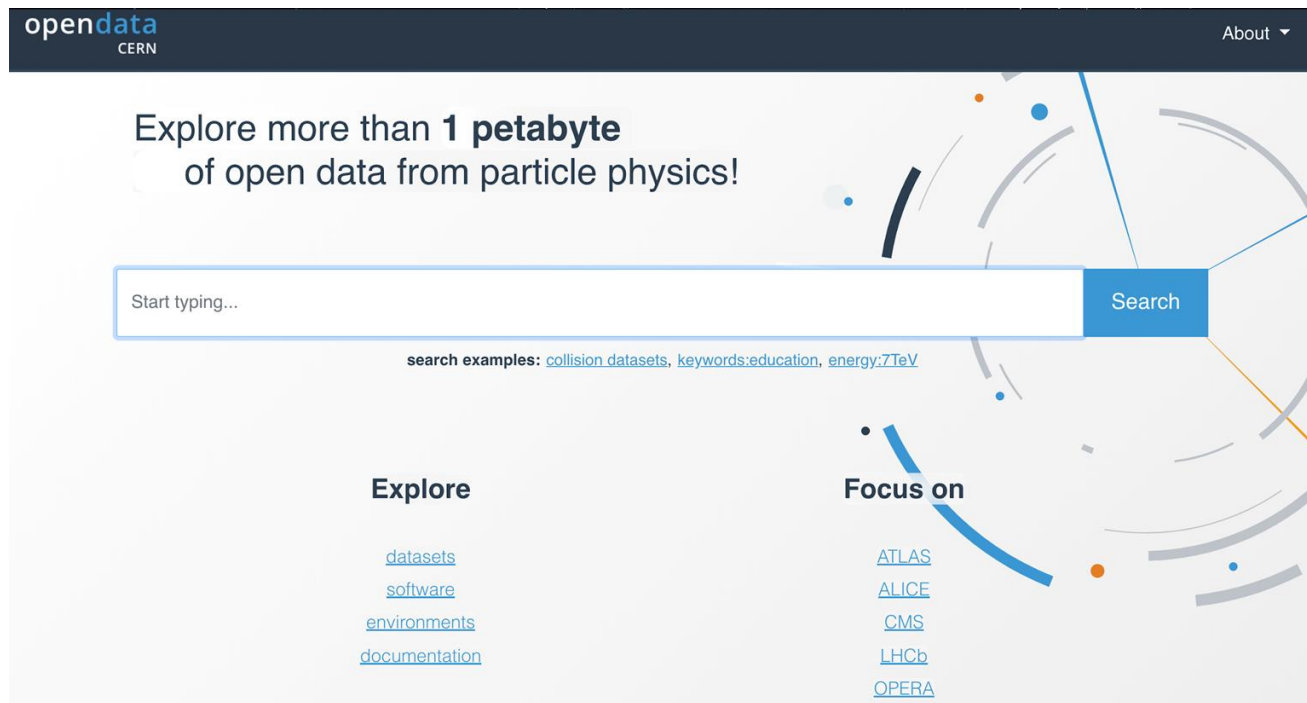
rate document. The some requirements with making | resources will be required to

on can be considered at four conditions: see Def (1). This

- ❑ Open data for three levels
- ❑ Immediate release for level 1
- ❑ Embargo periods for level 3
- ❑ CC0 – public domain dedication

Services at CERN

CERN Open Data portal (opendata.cern.ch)



Platform for curated releases of CERN data sets, software, supplementary materials, etc., over 1PB of data

CERN Analysis Preservation

Welcome to the CERN Analysis
Preservation Portal.

Our mission is to preserve physics analyses to facilitate
their future reuse

Do you want to know more? Check out what the service is about

 Log in with CERN

Restricted-access
service,
preservation of
knowledge and
assets of physics
analyses

INSPIRE-HEP (inspirehep.net)



Welcome to [INSPIRE](#), the High Energy Physics information system. Please direct questions, comments or concerns to feedback@inspirehep.net.

[HEP](#) :: [HEPNAMES](#) :: [INSTITUTIONS](#) :: [CONFERENCES](#) :: [JOBS](#) :: [EXPERIMENTS](#) :: [JOURNALS](#) :: [HELP](#)

HEP Search

High-Energy Physics Literature Database

Use "find" for SPIRES-style search ([other tips](#))

Brief format Search [Easy Search](#) [Advanced Search](#)

[find "Phys.Rev.Lett., 105" :: more](#)

HOW TO SEARCH

SPIRES syntax is (mostly) supported (requires "find")

[find a richter, b and t quark and date > 1984](#)
[find j phys.rev., D50, 1140 or j hep.0903.112](#)
[find eprint arxiv:1007.5048](#) (Note the plots available on the detailed record)
[find fulltext "quark-gluon plasma"](#) (Note new "fulltext" operator)
[find a ellis and refersto a witten](#) (Note "refersto")
[find a kane and citedby title SUSY and topcite 200+](#) (Note "citedby")

New techniques:

[1985 richter quark multiplicity](#)
[arXiv:1007.5048](#)
[citedby:author:ellis -refersto:author:witten](#)
[author:randall j author:sundrum cited:450->1350](#)

Additional Help:

[More search tips and full help](#)

INSPIRE UPDATES

See our blog at blog.inspirehep.net for updates on new features and other news. You can also follow us at [@inspirehep](https://twitter.com/inspirehep) on twitter. To send us feedback use feedback@inspirehep.net. The data in INSPIRE is updated daily. To request corrections to data in INSPIRE, write us at help@inspirehep.net. INSPIRE superseded SPIRES in 2012.

HEP

[Additions](#)
[Corrections](#)
[Search Tips](#)
[FAQ](#)
[Topcites: annual | recent](#)
[Reviews](#)
[HEP Citesummary](#)
[Tools](#)

INSPIRE

[About INSPIRE](#)
[Content Policy](#)
[INSPIRE Help Central](#)
[Blog](#)
[Twitter](#)
feedback@inspirehep.net

RESOURCES

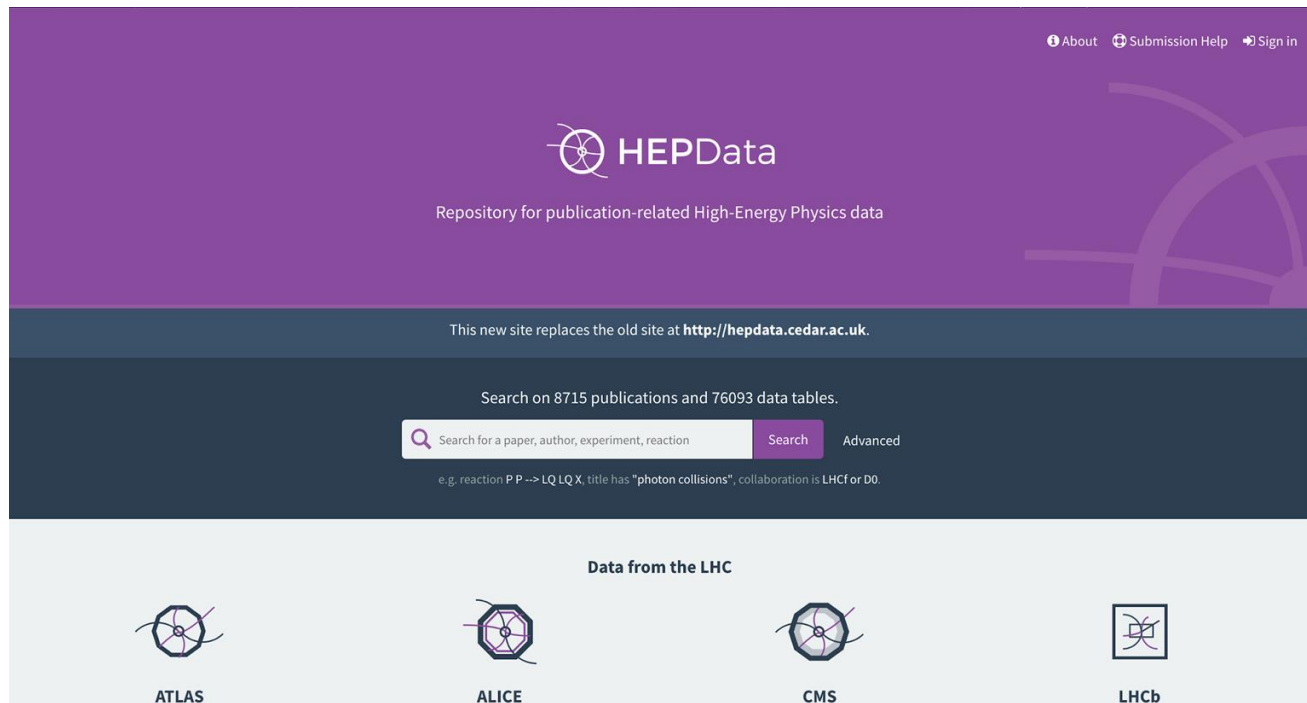
[ADS](#)
[arXiv](#)
[HepData](#)
[INSIS](#)
[PDG](#)
[PDG review of online resources](#)

INSPIRE News

2018-10-03 The latest papers to reach 1,000 citations in INSPIRE:
[#topcites](https://lit.co/YDnMzh2vXs)
<https://lit.co/iXm0Qsz6LN>
2018-10-03 Congratulations to Ashkin-Mourou-Strickland on the 2018 Nobel Prize in Physics


Core HEP literature aggregator, "long tail" of data

HEPData (hepdata.net)



The screenshot shows the HEPData website homepage. At the top right, there are links for 'About', 'Submission Help', and 'Sign in'. The main header features the HEPData logo and the text 'Repository for publication-related High-Energy Physics data'. Below this, a dark blue banner states: 'This new site replaces the old site at <http://hepdata.cedar.ac.uk>.' The search section includes a search bar with the placeholder text 'Search for a paper, author, experiment, reaction', a 'Search' button, and an 'Advanced' link. Below the search bar, an example search query is provided: 'e.g. reaction P P --> LQ LQ X, title has "photon collisions", collaboration is LHCf or D0.' The bottom section, titled 'Data from the LHC', displays four logos for ATLAS, ALICE, CMS, and LHCb.


About Submission Help Sign in

 HEPData

Repository for publication-related High-Energy Physics data


This new site replaces the old site at <http://hepdata.cedar.ac.uk>.


Search on 8715 publications and 76093 data tables.


 Search for a paper, author, experiment, reaction **Search** Advanced


e.g. reaction P P --> LQ LQ X, title has "photon collisions", collaboration is LHCf or D0.

Data from the LHC

 ATLAS



 ALICE

 CMS

 LHCb

Repository for
tabular data
associated with
publications

Zenodo (zenodo.org)

[Upload](#)[Communities](#)

[Log in](#)[Sign up](#)

Recent uploads





October 3, 2018 (v1.0.3)

Dataset

Open Access

View

PsPM-SC4B: SCR, ECG, EMG, PSR and respiration measurements in a delay fear conditioning task with auditory CS and electrical US

 Staib, Matthias; Castegnetti, Giuseppe;  Tzovara, Athina;  Paulus, Philipp C.; Hofer, Nicolas;  Bach, Dominik R.

This dataset includes pupil size response (PSR), skin conductance response (SCR), electrocardiogram (ECG), electromyogram (EMG) and respiration measurements. Also included are CS and US information, keypress responses, keypress response times, key correctness and shock ratings for each of 21...

Uploaded on October 3, 2018
3 more version(s) exist for this record




October 1, 2018 (v1.0)

Preprint

Open Access


View

DOIBoost - Boosting CrossRef for Research

 La Bruzzo, Sandro;  Manghi, Paolo;  Mannocci, Andrea


Research in information science and scholarly communication strongly relies on the availability of openly accessible datasets of scholarly entities metadata and, where possible, their relative payloads.

Zenodo now supports usage statistics!



[Read more](#) about it, in our newest blog post.

Using GitHub?



Just [Log in](#) with your GitHub account and [click here](#) to start preserving your repositories.

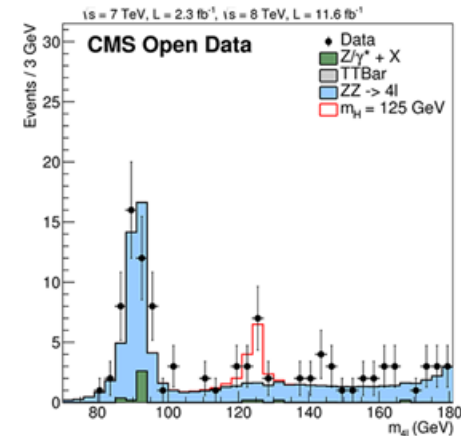
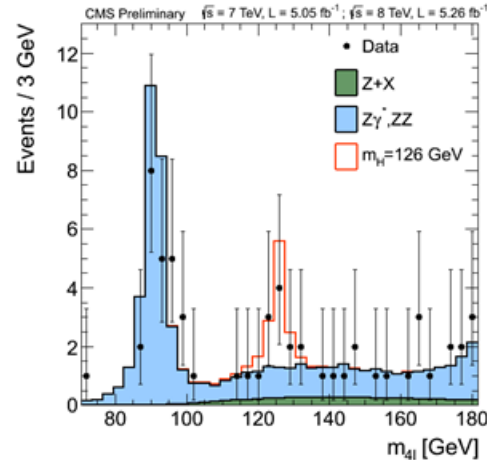
Zenodo in a nutshell

- **Research. Shared.** — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- **Citeable. Discoverable.** — uploads

Catch-all open repository, variety of resource types, interdisciplinary

Examples of impact using CERN Open Data

- Thaler et al. used CMS open data that were released on the portal in 2014 to confirm their jet substructure model predictions
- Nur Zulaiha Jomhari et al. used CMS open data from 2011 and 2012 from the portal to run a particle physics analysis example about the Higgs decay physics using the same data formats, the same software tools and computational techniques used by CMS experimental physicists
- On the left is the official CMS analysis of the "Higgs-to-four-lepton" decay channel that led to the Higgs boson experimental discovery in 2012 and the plot on the right was the result of the analysis by Nur Zulaiha Jomhari et al.



<https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.119.132003>
<https://journals.aps.org/prd/abstract/10.1103/PhysRevD.96.074003>
<http://opendata.cern.ch/record/5500>

Features

- Developed using open-source software
- Open access licenses
- Persistent identifiers (DOIs, ORCID iDs)
- Versioning
- Machine-readable, high-quality metadata
- Provisions for big data and high complexity
- Publishing less “traditional” resources (Jupyter notebooks, Virtual Machines..)
- Enabling reuse

