

CERN Open Data Portal

Jake COWTON, Sunje DALLMEIER-TIESSEN, Pamfilos FOKIANOS, Laura RUEDA GARCIA, Patricia Sigrid HERTERICH, Jiri KUNCAR, Tibor SIMKO, Tim SMITH

DPHEP Workshop, June 2016



Open Data Portal

opendata
CMS

ABOUT SEARCH EDUCATION RESEARCH

Education

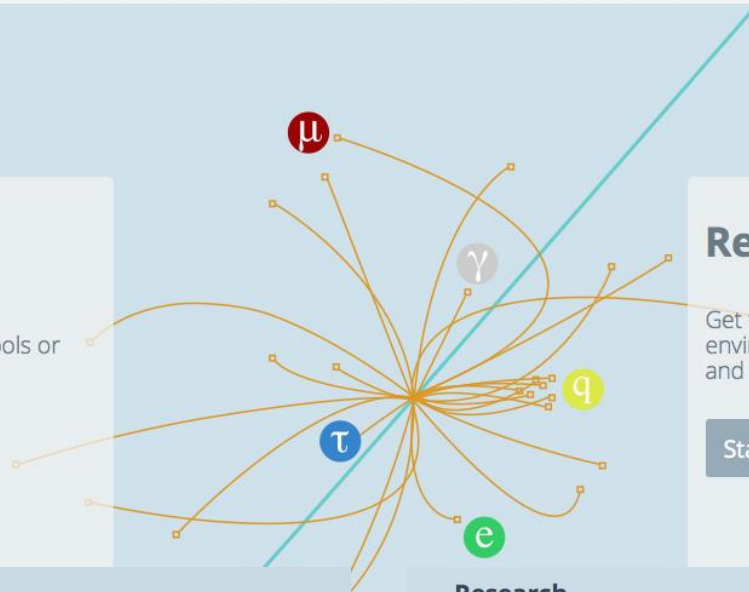
Visualise events, check reconstructed data, run tools or build your own!

Start learning

Research

Get the genuine working environments, virtual machines and datasets to start your research

Start analysing



Education



The CMS (Compact Muon Solenoid) experiment is one of two large general-purpose detectors built on the Large Hadron Collider (LHC). Its goal is to investigate a wide range of physics such as the characteristics of the Higgs boson, extra dimensions or dark matter.

Explore CMS >



ALICE (A Large Ion Collider Experiment) is a heavy-ion detector designed to study the physics of strongly interacting matter at extreme energy densities, where a phase of matter called quark-gluon plasma forms. More than 1000 scientists are part of the collaboration.

Explore ALICE >



The ATLAS (A Toroidal LHC Apparatus) experiment is a general purpose detector exploring topics like the properties of the Higgs-like particle, extra dimensions of space, unification of fundamental forces, and evidence for dark matter candidates in the Universe.

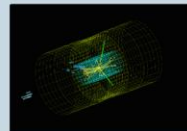
Explore ATLAS >



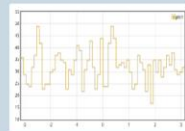
The LHCb (Large Hadron Collider beauty) experiment aims to record the decay of particles containing b and anti-b quarks, known as B mesons. The detector is designed to gather information about the identity, trajectory, momentum and energy of each particle.

Explore LHCb >

For education purposes, the complex primary data need to be processed into a format (examples below) that is good for simple applications. Get in touch if you wish to build your own applications similar to those shown here



Visualise events >



Visualise histograms >



Learning Resources >

Research



To analyse CMS data, a Virtual Machine with the CMS analysis environment is provided. The data can be accessed directly through the VM. In the primary datasets, no selection nor identification criteria have been applied. For this release, no simulated Monte Carlo datasets are provided.

Explore CMS >



According to the ALICE data preservation strategy, reconstructed data and Monte Carlo data as well as the analysis software and documentation needed to process them will be made available on a time scale of 5 years (for 10% of the data). Thus, the first release of ALICE research data will happen in 2018.



According to the ATLAS Data Access Policy, reconstructed data and accompanying tools will be released after reasonable embargo periods.



According to the LHCb External Data Access Policy, reconstructed data and accompanying tools will be released after reasonable embargo periods.

For research purposes, specific software environments and tools need to be deployed to analyse these complex primary data. In addition to the data below, you will find instructions for setting up your working environments here



Install your Virtual Machine >



Start analysing the data >

Timeline

■ **November 2014** Initial release



ALICE: educational datasets, modules, VM



ATLAS: masterclass datasets



CMS: primary datasets (30 TB of 2010), educational data sets, analysis example and software, VM, event display, histogramming



LHCb: masterclass datasets, VM

■ **February 2015** ATLAS update



ATLAS: addition of enhanced Kaggle ML challenge datasets, software, documentation

■ **May 2015** ALICE update



ALICE: addition of ESD datasets (10 TB of 2010)

L3: Reconstructed Data

opendata CERN

ABOUT SEARCH EDUCATION RESEARCH

Search

Home > CMS > CMS Primary Datasets

Mu primary dataset in AOD format from RunB of 2010 (/Mu/Run2010B-Apr21ReReco-v1/AOD) 2014

/Mu/Run2010B-Apr21ReReco-v1/AOD
CMS collaboration

Cite as: CMS collaboration (2014). Mu primary dataset in AOD format from RunB of 2010 (/Mu/Run2010B-Apr21ReReco-v1/AOD). CERN Open Data Portal. DOI: [10.7483/OPENDATA.CMS.B8MR.C4A2](https://doi.org/10.7483/OPENDATA.CMS.B8MR.C4A2)

Collection CMS Primary Datasets Collision Energy 7TeV Accelerator CERN-LHC Experiment CMS

Description

Mu primary dataset in AOD format from RunB of 2010

Characteristics

Dataset: 32376291 events 2979 files 3.2 TB in total

Issues & Limitations

This dataset contains all runs from 2010-01-06 to 2010-12-31. For more information, see the CMS list of validated runs Cert_13605.

Disclaimer

The open data are released under the [Creative Commons Attribution-NonCommercial-ShareAlike license](#). All releases will have a unique DOI that you can request to cite in any applications or publications.

Export MARCXML

© 2014 CERN Open Data Terms of Use Privacy Policy Contact

ALICE ATLAS CMS LHCb

CMS: 50% of 2010; 30TB
ALICE: 10%(5yr), 100%(10yr)
LHCb: 50%(5yr), 100%(10yr)

opendata CERN

ABOUT SEARCH EDUCATION RESEARCH

Search

Home > Virtual Machines

CERN Virtual Machines allow you to run Scientific Linux on any operating system and access the CERN Open Data Portal environments and software tools.

CMS Virtual Machines

ALICE Virtual Machines

LHCb Virtual Machines

CMS Virtual Machines: How to install

The CMS-specific VM includes the [ROOT](#) framework and [CMSSW](#). Follow the instructions below to setup a CERN Virtual Machine on your computer with CMS data.

1. How to install a CERN VM
2. Issues & Limitations

How to install a CERN Virtual Machine

Step 1: Installing VirtualBox

VirtualBox is a free, open source and multiplatform application to run virtual machines; you can [download](#) the package for your platform from the [VirtualBox website](#).

You will need administrative ("root") privileges on every platform to perform the installation of VirtualBox.

Note: the latest tested version of VirtualBox working with CernVM is 4.3.14. If you have troubles with the latest version of VirtualBox, pick that one: the previous version is available [on a different page](#).

Virtual Machines: CernVM

The screenshot displays a virtual machine environment with a desktop background. The top menu bar includes 'Finder', 'File', 'Edit', 'View', 'Go', 'Window', and 'Help'. The system status bar shows the date and time as 'Thu Dec 23 2:09 AM'. The dock at the bottom contains various application icons.

The main window is titled 'csmShow: /home/test/CMSSW_3_7_0/src/dijet mass events from V8reco Op'. It features a control panel with 'Run' (132601) and 'Event' (9369482) fields, and a 'Summary View' on the left listing particle collections like ECAL, HCAL, Jets, Tracks, Muons, etc.

The central 'Rho Phi' plot shows a circular detector cross-section with tracks. A '3D Lego' view to the right shows a 3D reconstruction of the detector. Below it is a table of jet properties:

Collection	Jets					
∇ Pt	eta	phi	ECAL	HCAL	emf	s
67.9	0.201	1.357	66.6	23.8	0.738	
65.4	-1.190	-1.707	96.2	21.6	0.817	
6.5	-2.686	-1.031	52.2	9.8	0.847	
6.3	-3.071	-2.791	45.5	26.5	0.618	

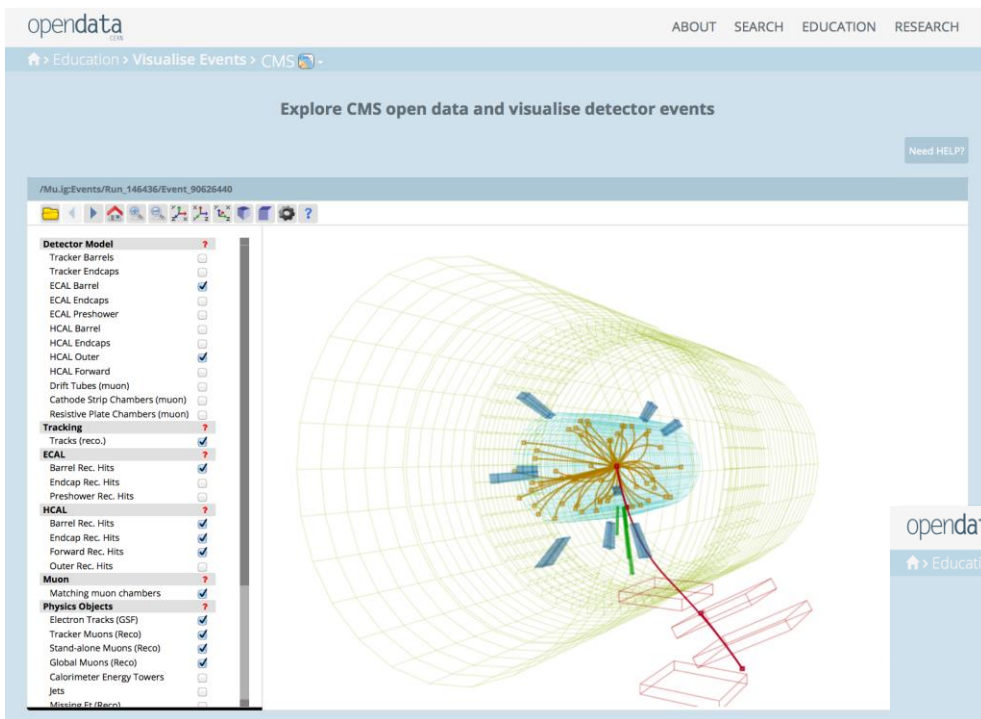
Below the table is a 'TriggerTable' section with a 'Filter Name' field and a list of triggers including 'HLT_PbPbV0LLTA', 'HLT_Activity_PixelClusters', and 'HLT_Activity_DT'.

On the right, the 'Eve Main Window' displays a 3D visualization of the detector and event tracks, along with 'RPhi View' and 'RhoZ View' plots. The 'RPhi View' shows a circular plot with tracks, and the 'RhoZ View' shows a 2D plot of tracks.

At the bottom, a terminal window shows the following commands and output:

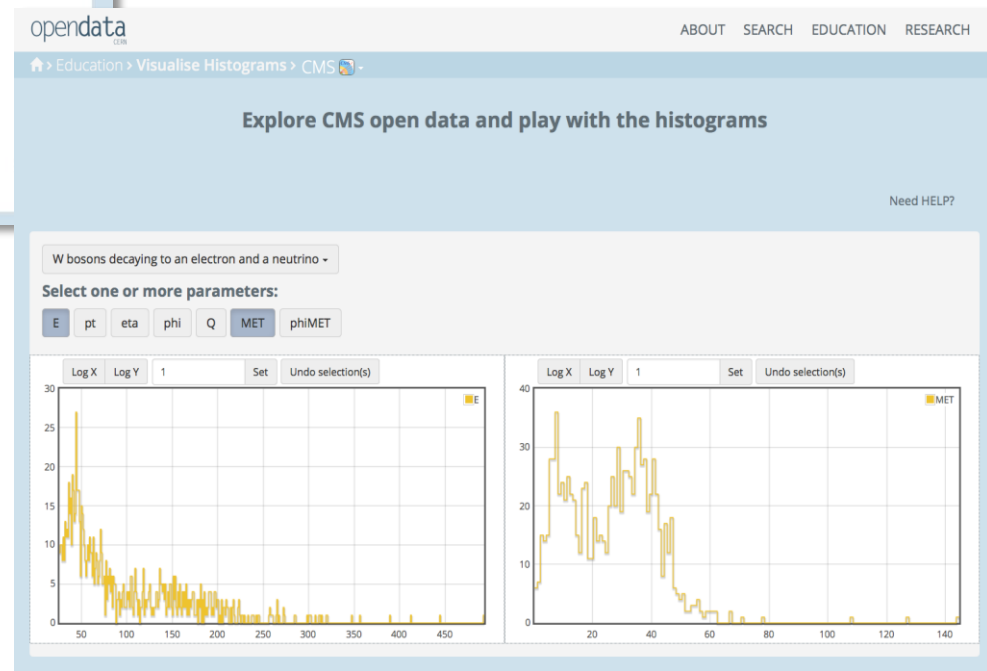
```
terminal - test@localhost:~  
[test@localhost ~]$ sudo du -hs /var/cache/cvmfs2/alice/  
373M /var/cache/cvmfs2/alice/  
[test@localhost ~]$ sudo du -hs /var/cache/cvmfs2/cms  
327M /var/cache/cvmfs2/cms  
[test@localhost ~]$
```

L2: Outreach and Education



Visualise detector events

Basic Histogramming



L2: Higgs Kaggle Challenge

kaggle

Host

Competitions

Scripts

Jobs

Community

Sign up

Login

Higgs challenge

Completed • \$13,000 • 1,785 teams

Higgs Boson Machine Learning Challenge

Mon 12 May 2014 – Mon 15 Sep 2014 (8 months ago)

Dashboard

Home

Data

Make a submission

Information

Description

Evaluation

Rules

Prizes

About the Sponsors

Timeline

Winners

Forum

Leaderboard

Public

Private

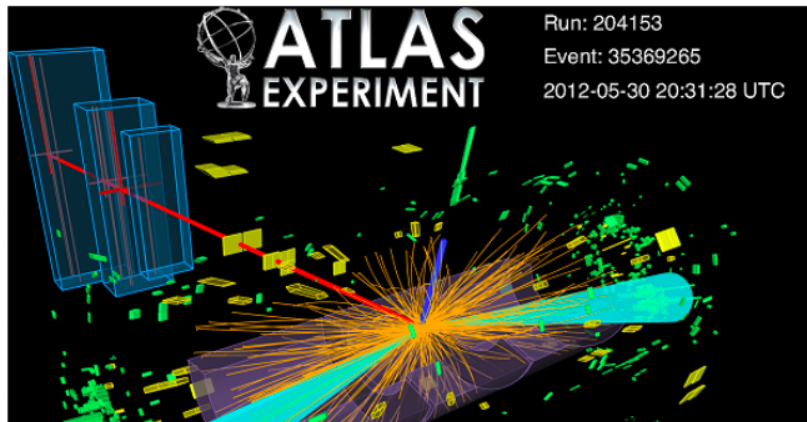
Leaderboard

1. Gábor Melis

2. Tim Salimans

Competition Details » [Get the Data](#) » [Make a submission](#)

Use the ATLAS experiment to identify the Higgs boson



opendata

ABOUT SEARCH EDUCATION RESEARCH

Search

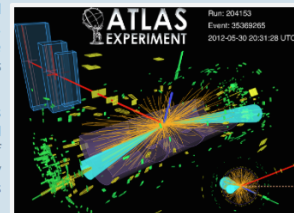
ATLAS Higgs Challenge 2014

The Higgs boson was announced July 4, 2012 and 13 saw a number of prestigious awards, physicists, the discovery of a new particle and difficult quest to measure its characteristics model of nature.

How often it decays into other particles. ATLAS is taking place at the Large Hadron Collider at CERN and processes using head-on collisions of energy. The ATLAS experiment has recently discovered a Higgs boson decaying into two tau particles, but this is often obscured by background noise.

The Higgs Boson Machine Learning Challenge is to explore the potential of advanced machine learning methods to analyze the data of the experiment. No knowledge of particle physics is required. Using simulated data with generated by ATLAS, your task is to classify events into "tau tau decay of a Higgs boson" versus "background".

The challenge was held on the Kaggle platform from May to September 2014, drawing more than 1700 participants. Very promising results were achieved.



Dataset Semantics

EventId: An unique integer identifier of the event.

DER_mass_MMC: The estimated mass m_H of the Higgs boson candidate, obtained through a particle identification integration.

DER_mass_transverse_met_lep: The transverse mass between the missing transverse energy and the lepton.

DER_mass_vis: The invariant mass of the hadronic tau and the lepton.

DER_pt_h: The modulus of the vector sum of the transverse momentum of the hadronic tau, the lepton and the missing transverse energy vector.

DER_deltaeta_jet_jet: The absolute value of the pseudorapidity separation between the two jets ($\eta \leq 1$).

DER_mass_jet_jet: The invariant mass of the two jets (undefined if $PRI_jet_num \leq 1$).

DER_prodelta_jet_jet: The product of the pseudorapidities of the two jets (undefined if $PRI_jet_num \leq 1$).

DER_deltar_tau_lep: The R separation between the hadronic tau and the lepton.

DER_pt_tot: The modulus of the vector sum of the missing transverse momenta and the transverse momenta of the hadronic tau, the lepton, the leading jet (if $PRI_jet_num \geq 2$) and the subleading jet (if $PRI_jet_num = 3$).

DER_met_phi_centralty: The centrality of the azimuthal angle of the missing transverse energy vector and the lepton.

DER_lep_eta_centralty: The centrality of the pseudorapidity of the lepton w.r.t. the two jets (undefined if $PRI_jet_num \leq 1$).

PRI_tau_pt: The transverse momentum $\sqrt{p_x^2 + p_y^2}$ of the hadronic tau.

PRI_tau_eta: The pseudorapidity η of the hadronic tau.

PRI_tau_phi: The azimuth angle ϕ of the hadronic tau.

PRI_lep_pt: The transverse momentum $\sqrt{p_x^2 + p_y^2}$ of the lepton (electron or muon).

PRI_lep_eta: The pseudorapidity η of the lepton.

PRI_lep_phi: The azimuth angle ϕ of the lepton.

PRI_met: The missing transverse energy \vec{E}_T^{miss} .

L1: Publication Data

CERN Accelerating science Sign in Directory

CERN Document Server

Search Submit Help Personalize

Home > Articles & Preprints > Published Articles > First observation of $\bar{B}^0 \rightarrow J/\psi K^+ K^-$ and search for $\bar{B}^0 \rightarrow J/\psi \phi$ decays > Comments

Information Discussion (0) Files

First observation of $\bar{B}^0 \rightarrow J/\psi K^+ K^-$ and search for $\bar{B}^0 \rightarrow J/\psi \phi$ decays - Aaij, R et al - arXiv:1308.5916

Main file(s):


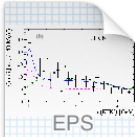
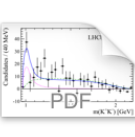
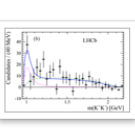

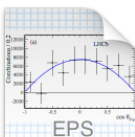
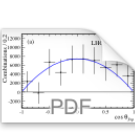
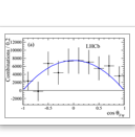

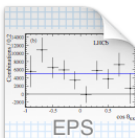
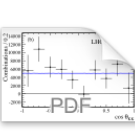
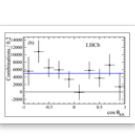
prd.88.e072005
version 1 [prd.88.e072005.pdf](#) [1.32 MB] 07 Nov 2013, 15:17 APS Open Access article

Additional file(s):

Related data file(s)
version 1 [Related data file\(s\).zip](#) [10.36 MB] 02 Sep 2013, 16:41

arXiv file(s):

arXiv:1308.5916
version 4 [arXiv:1308.5916.pdf](#) [4.36 MB] 26 Oct 2013, 04:15 (see previous)

 Fig15b.C	 Fig15b.eps	 Fig15b.pdf	 Fig15b.png
 Fig16a.C	 Fig16a.eps	 Fig16a.pdf	 Fig16a.png
 Fig16b.C	 Fig16b.eps	 Fig16b.pdf	 Fig16b.png



Hep Data

Interest



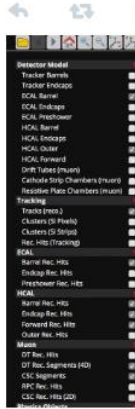
CERN launches Open Data Portal to public the data of LHC experiments
[/go/tN15T](#) #cernopendata

Follow

40,000

Press Release

Reddit AMA



NewScientist Physics & Math

Home News In-Depth Articles Opinion CultureLab Galleries Topic Guides Last Word Subscribe Dating

SPACE TECH ENVIRONMENT HEALTH LIFE **PHYSICS&MATH** SCIENCE IN SOCIETY

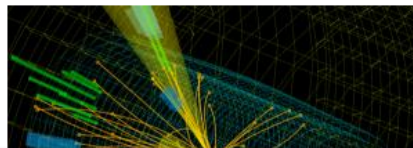
Home | Physics & Math | News

Run your own experiment using CERN's public LHC data

18:30 24 November 2014 by Jacob Aron
For similar stories, visit the [The Large Hadron Collider](#) Topic Guide

Why build your own particle accelerator when you can borrow CERN's? The home of the Large Hadron Collider near Geneva, Switzerland, has started putting data from its experiments online for anyone to use. They hope it could fuel education, art and perhaps even physics discoveries.

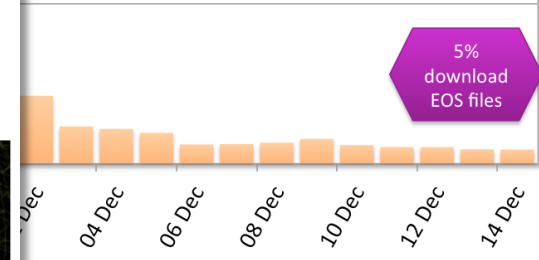
Like 821 Tweet 269 +1 29
Share 294



"It's very important that we keep this data open and usable," says Kati Lassila-Perini of the CMS experiment at the LHC, which has uploaded 27 terabytes of data to the new [CERN Open Data Portal](#). A web interface lets you visualise the paths of particles created by collisions at the LHC, or you can work directly with the data for more serious analysis.

Other LHC experiments have uploaded smaller data sets for educational purposes, allowing budding particle physicists to try their hand at massive-scale physics. The data could also be turned into art, as was done previously at a [CERN arts festival](#).

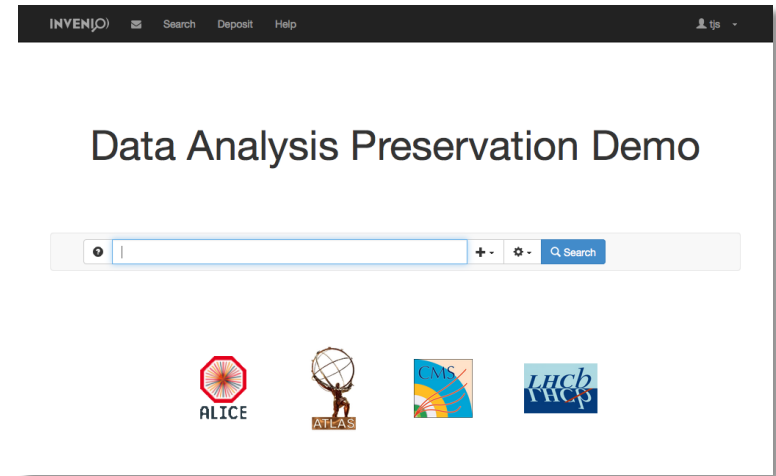
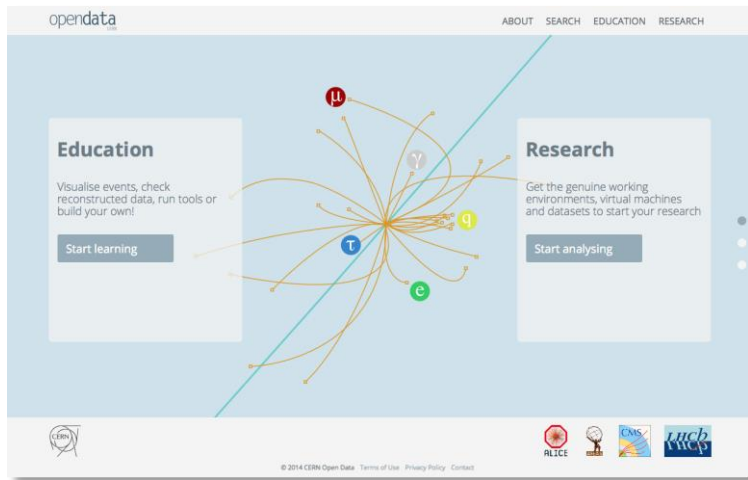
It's possible that physicists might search through the data for new discoveries, but most people with the necessary skills already work on LHC experiments, says Lassila-Perini. "There are not so many scientists around."



~16,000 users of event display
~3,000 users of histogramming
New education/theory connections
Re-use for statistical analyses, machine learning

Conclusion

Capture, Share, Preserve



In close collaboration with:

Kati Lassila-Perini, Tom McCauley, Achintya Rao

Alicia Calderon, Ana Rodriguez-Marrero, Adam Huffman, Jonatan Piedra

Silvia Amerio, Ben Couturier, Ana Trisovic

Mihaela Gheata, Costin Grigoras

Kyle Cranmer, Lukas Heinrich, Felix Socher, David Rousseau

Mike Hildreth, Jakob Blomer, Luca Mascetti

Frank Berghaus, Jamie Shiers





www.cern.ch