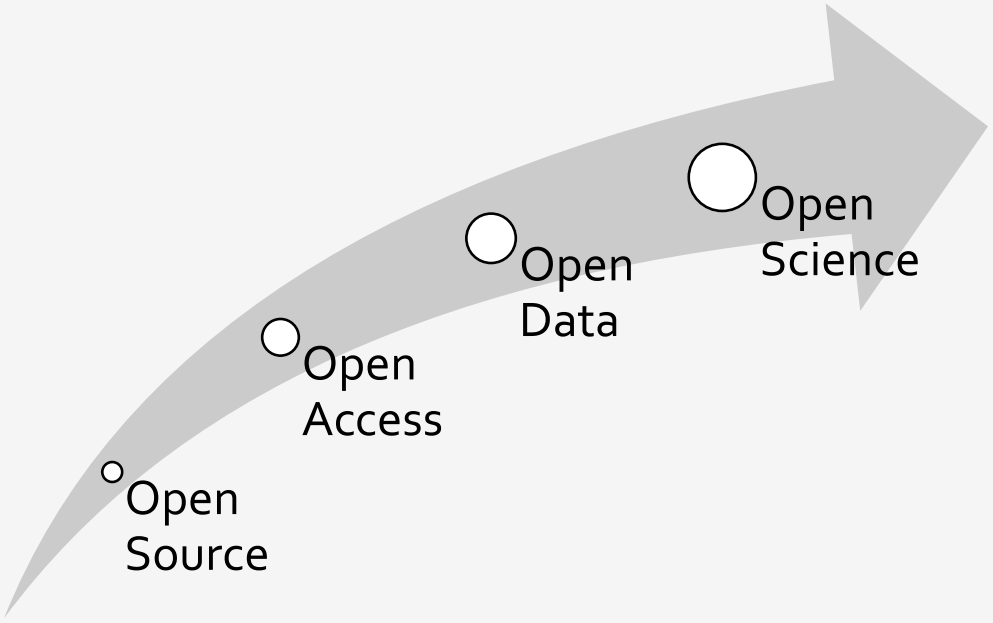
OPEN DATA

Patricia Herterich

16.04.2015

On the way to Open Science…

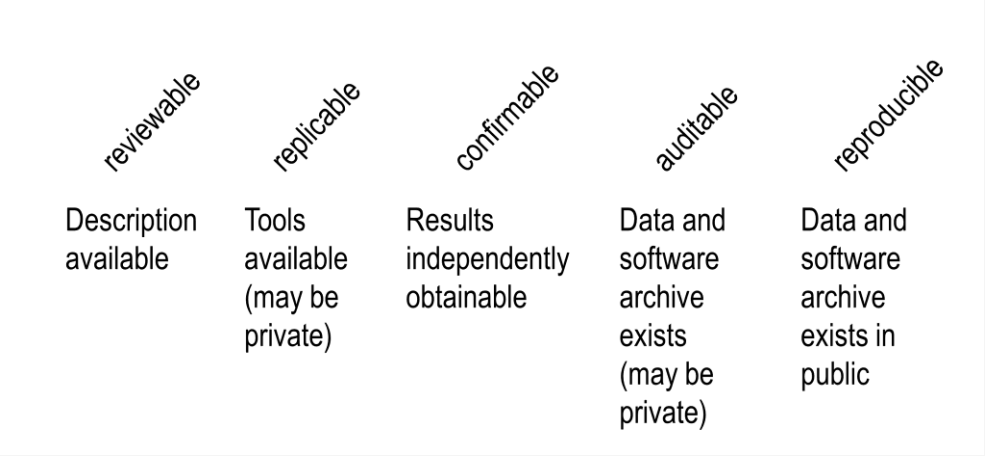


Benefits of Open Science

- For society:
  - Public availability & reusability of scientific data
  - Public accessibility & transparency of scientific communication

- For scientific communities:
  - Reproducibility of research results
  - Leveraging web-based tools to facilitate scientific collaboration

Re-producible research

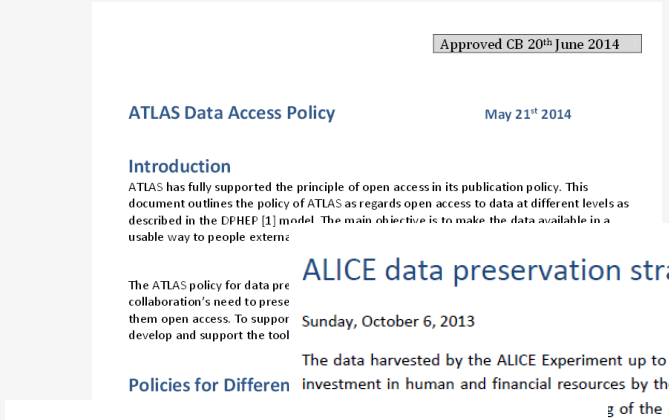


Open Data: incentives

- Funder policies

Open Data: initiatives

HEP Open Data



Approved CB 20th June 2014

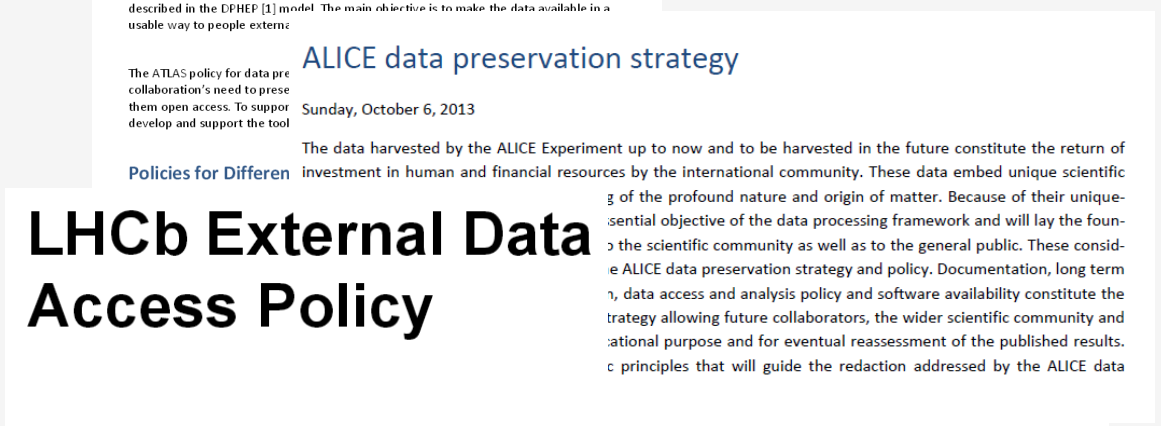
ATLAS Data Access Policy May 21st 2014

Introduction

ATLAS has fully supported the principle of open access in its publication policy. This document outlines the policy of ATLAS as regards open access to data at different levels as described in the DPHEP [1] model. The main objective is to make the data available in a usable way to people externa

The ATLAS policy for data pre
collaboration's need to prese
them open access. To suppor
develop and support the tool

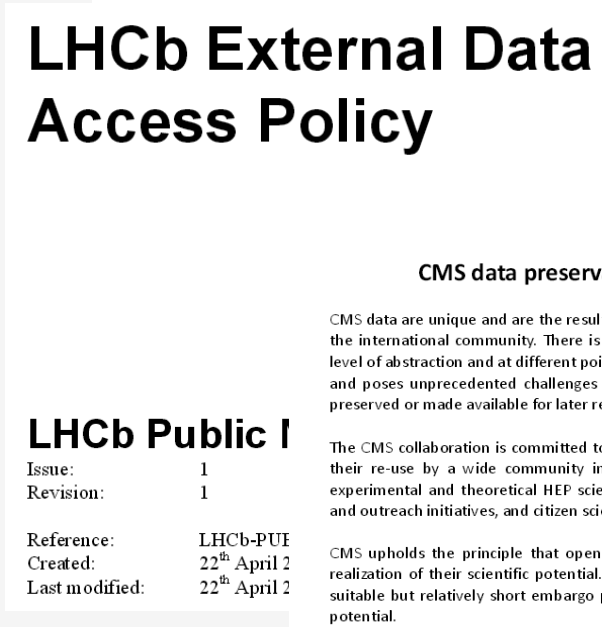
Policies for Differen



ALICE data preservation strategy

Sunday, October 6, 2013

The data harvested by the ALICE Experiment up to now and to be harvested in the future constitute the return of investment in human and financial resources by the international community. These data embed unique scientific g of the profound nature and origin of matter. Because of their unique- sential objective of the data processing framework and will lay the foun- o the scientific community as well as to the general public. These consid- e ALICE data preservation strategy and policy. Documentation, long term n, data access and analysis policy and software availability constitute the trategy allowing future collaborators, the wider scientific community and ational purpose and for eventual reassessment of the published results. c principles that will guide the redaction addressed by the ALICE data

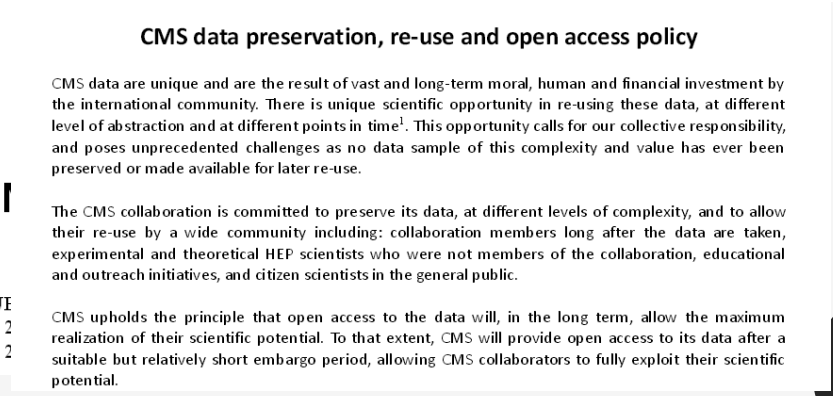


LHCb External Data Access Policy

LHCb Public I

Issue: 1
Revision: 1

Reference: LHCb-PUI
Created: 22th April 2
Last modified: 22th April 2



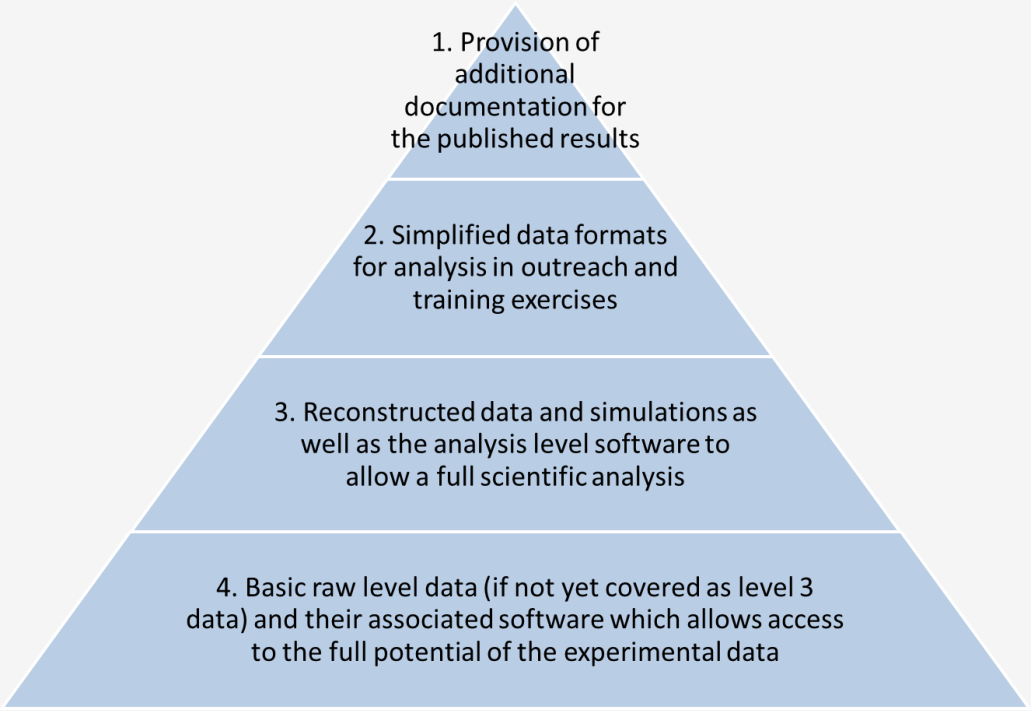
CMS data preservation, re-use and open access policy

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 time[1]. 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.

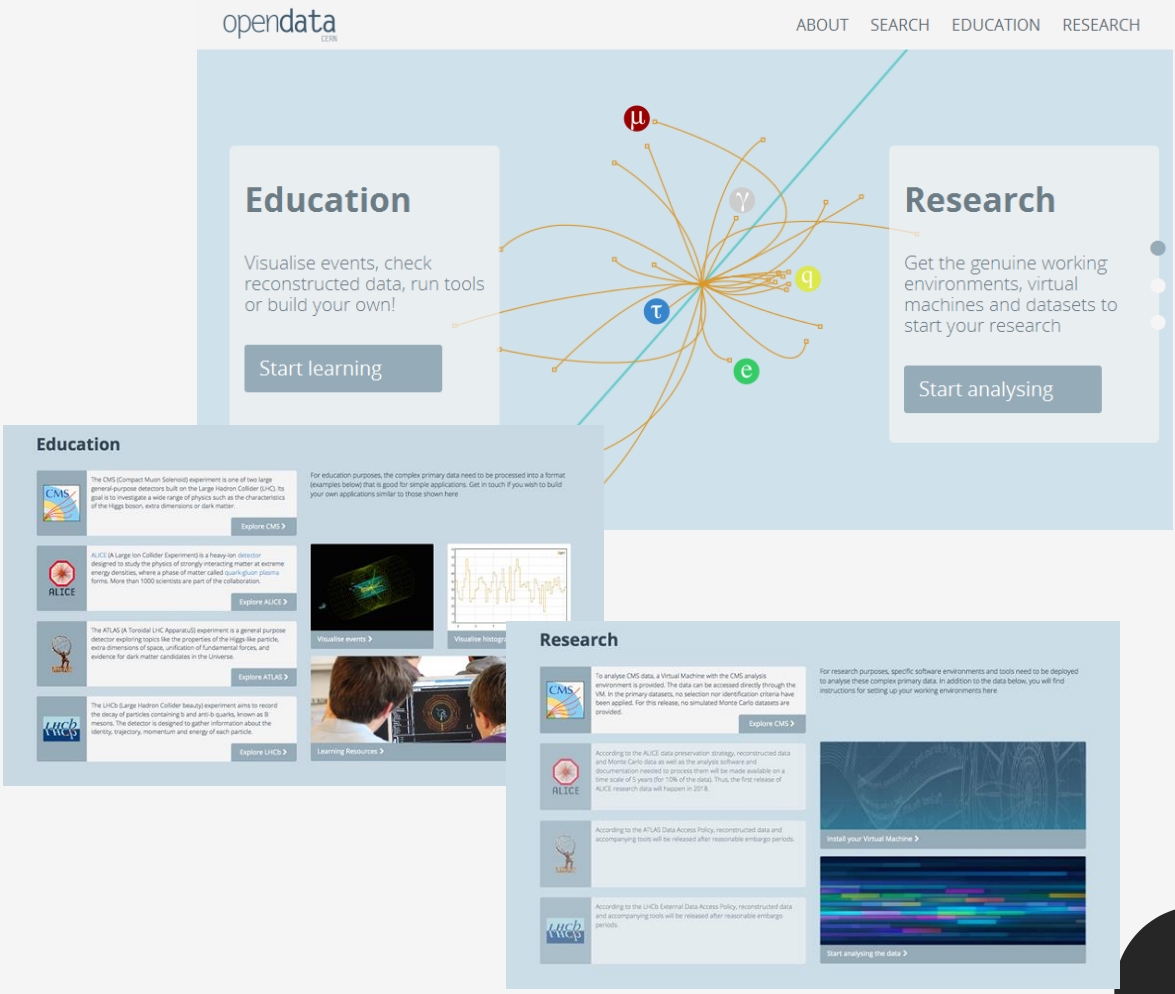
The CMS collaboration is committed to preserve its data, at different levels of complexity, and to allow their re-use by a wide community including: collaboration members long after the data are taken, experimental and theoretical HEP scientists who were not members of the collaboration, educational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

Data in High-Energy Physics



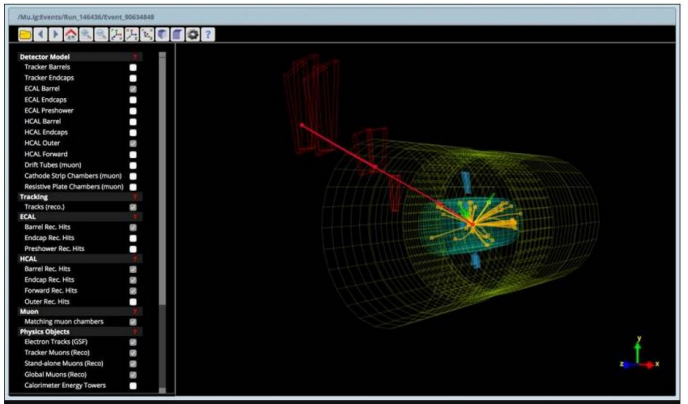
But how to make them open?



Release

CERN makes public first data of LHC experiments

Posted by Cian O'Luanaigh on 20 Nov 2014. Last updated 20 Nov 2014, 16.59.
Voir en français

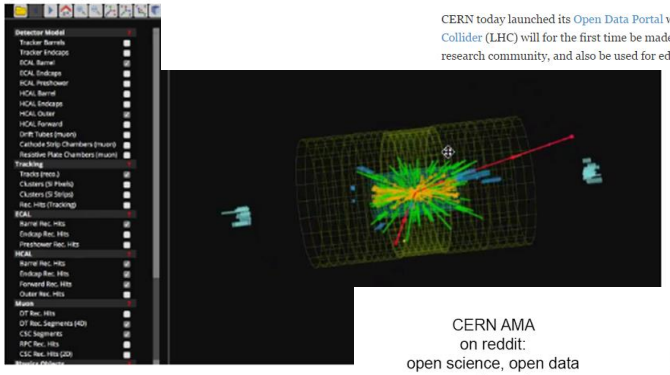


The web-based CMS event display, accessible through the CERN Open Data Portal, shows a proton-collision event recorded by the CMS detector (Image: CMS/Open Data Portal)

CERN today launched its Open Data Portal where data from real collision events, produced by experiments at the Large Hadron Collider (LHC) will for the first time be made openly available to all. It is expected that these data will be of high value for the research community, and also be used for education purposes.

CERN
@CERN

CERN laun
public the
/go/tN15T



RETWEETS 768 FAVORITES 450

6:43 AM - 20 Nov 2014

CERN AMA
on reddit:
open science, open data
and more!

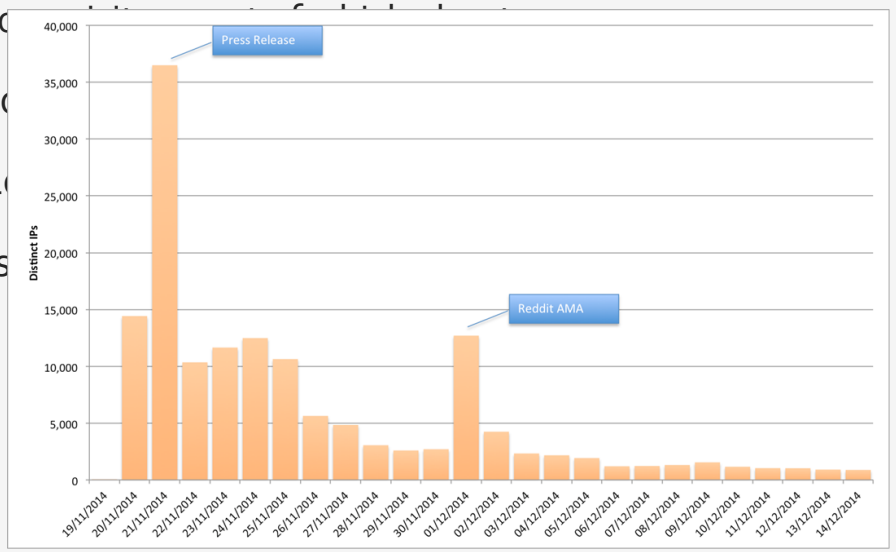
Monday, 1 December
15:30 CET
(GMT +01)

/u/askCERN
#cernopendata

Some numbers…

- At the public release:
  - Serving ~15 GB per hour [usage ~50 times now]
  - After a day or two was about ~4 GB per hour [~20 times now]

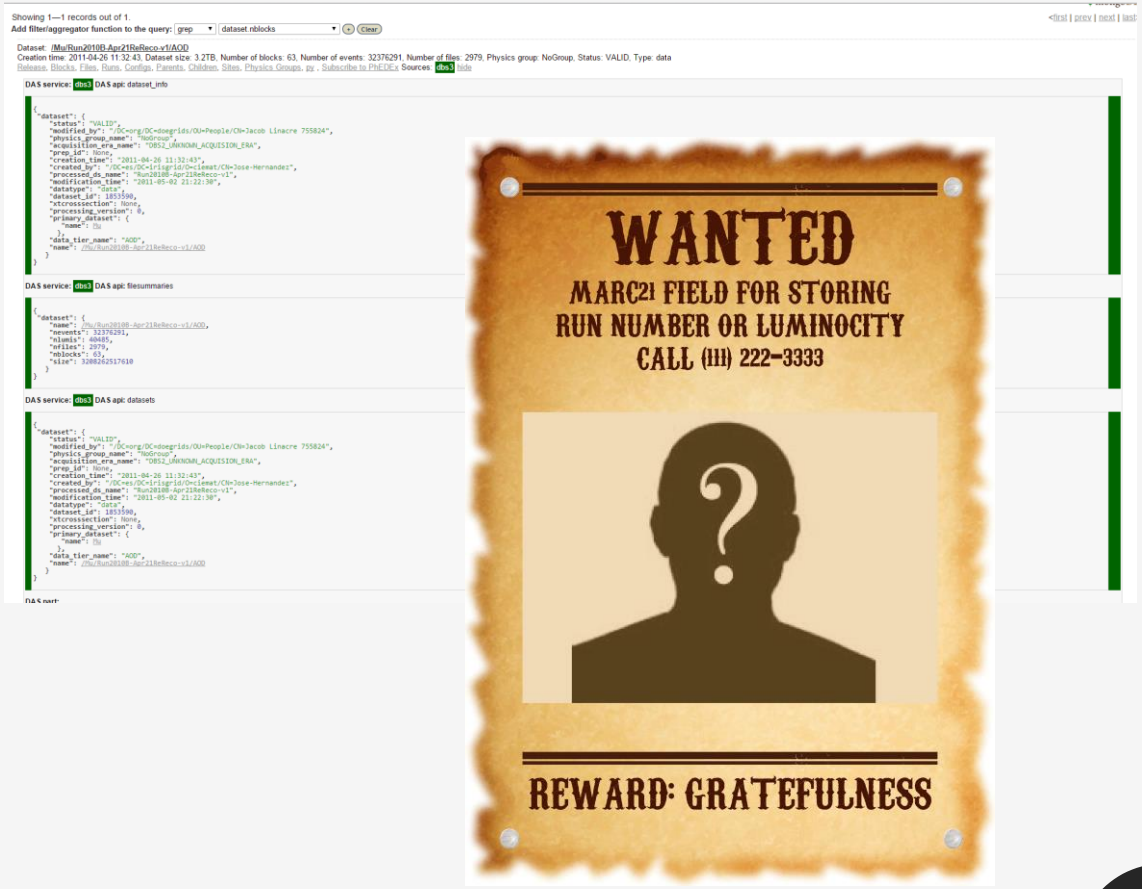
- Typical day now:
  - ~10
- ~10
- ~4(
- res



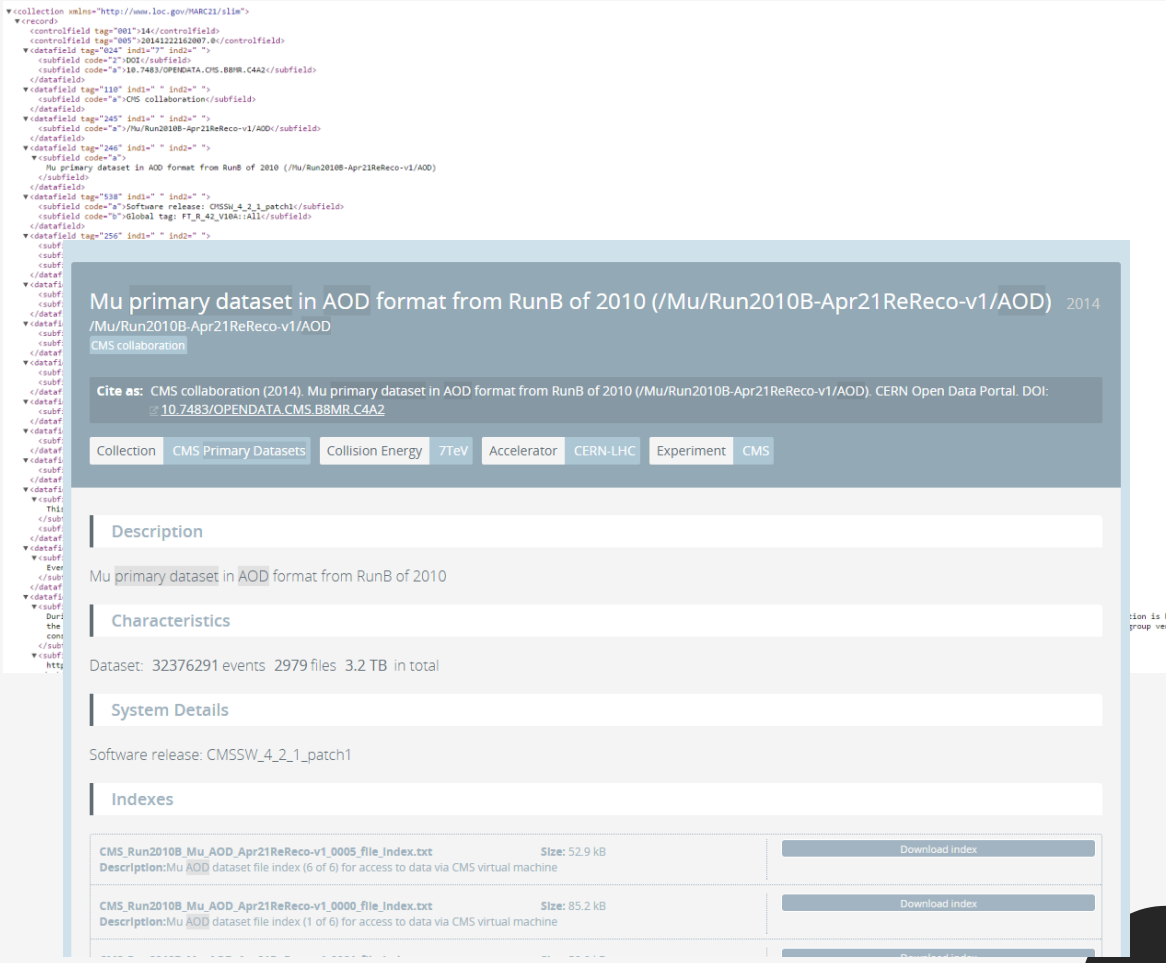
…and other impact

- We know that the CODP release resulted in:
- New collaborations
- Re-use of primary datasets for machine learning and “real physics” analysis
- New data “mash-ups”
- Adaption of code examples for new analysis

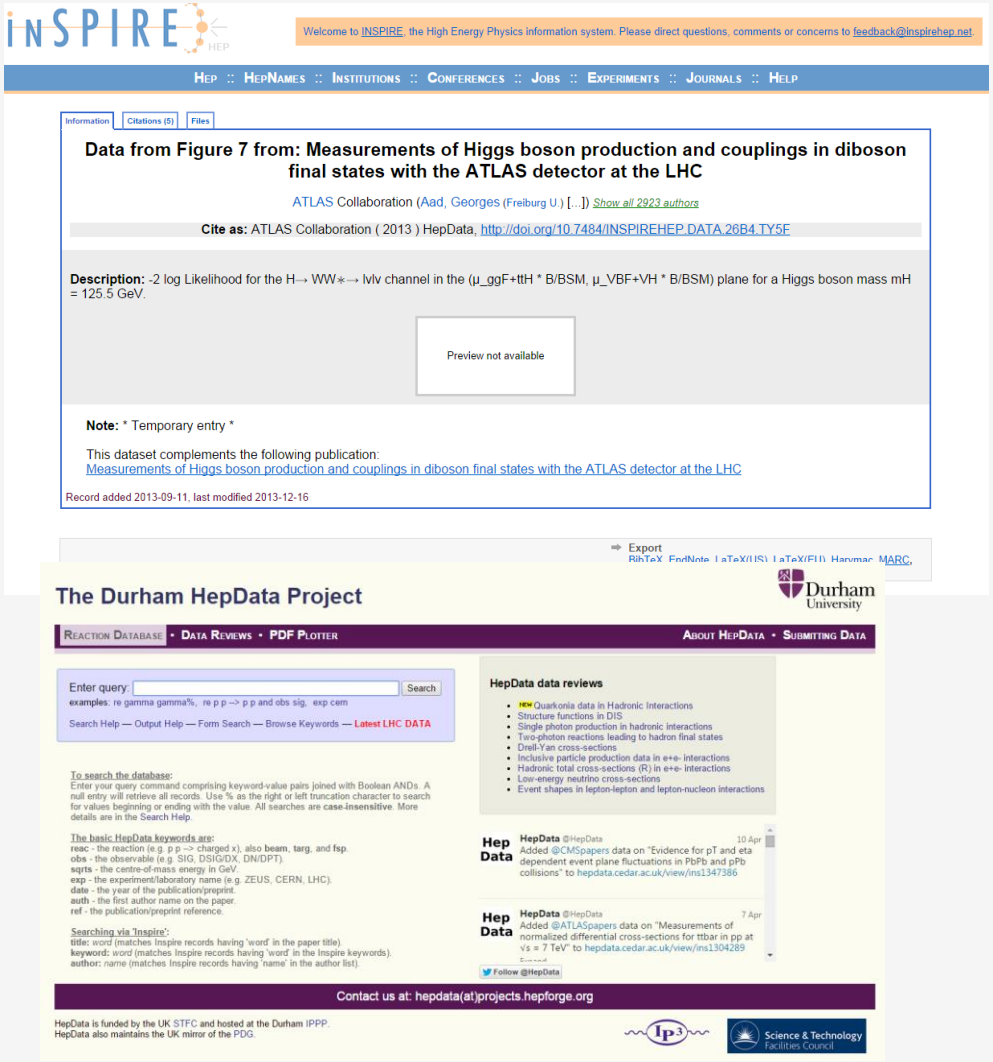
Metadata challenges



Our solution



Small scale data



INSPIRE HEP

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

Information | Citations (5) | Files

Data from Figure 7 from: Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC

ATLAS Collaboration (Aad, Georges (Freiburg U.) [..]) *Show all 2923 authors*

Cite as: ATLAS Collaboration (2013) HepData, http://doi.org/10.7484/INSPIREHEP.DATA.26B4.TY5F

Description: -2 log Likelihood for the H→ WW∗→ lνlν channel in the (μ_ggF+ttH * B/BSM, μ_VBF+VH * B/BSM) plane for a Higgs boson mass mH = 125.5 GeV.

Preview not available

Note: * Temporary entry *

This dataset complements the following publication:
Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC

Record added 2013-09-11, last modified 2013-12-16

Export
BibTeX EndNote LaTeX(US) LaTeX(EU) Harvmac MARC

The Durham HepData Project

Durham University

REACTION DATABASE • DATA REVIEWS • PDF PLOTTER ABOUT HEPDATA • SUBMITTING DATA

Enter query: Search
examples: re gamma gamma%, re p p --> p p and obs sig, exp cern
Search Help — Output Help — Form Search — Browse Keywords — Latest LHC DATA

To search the database:
Enter your query command comprising keyword-value pairs joined with Boolean ANDs. A null entry will retrieve all records. Use % as the right or left truncation character to search for values beginning or ending with the value. All searches are **case-insensitive**. More details are in the Search Help.

The basic HepData keywords are:
reac - the reaction (e.g. p p --> charged x), also **beam**, **targ**, and **fsp**.
obs - the observable (e.g. SIG, DSIG/DX, DN/DPT).
sqrts - the centre-of-mass energy in GeV.
exp - the experiment/laboratory name (e.g. ZEUS, CERN, LHC).
date - the year of the publication/preprint.
auth - the first author name on the paper.
ref - the publication/preprint reference.

Searching via 'Inspire':
title: *word* (matches Inspire records having 'word' in the paper title).
keyword: *word* (matches Inspire records having 'word' in the Inspire keywords).
author: *name* (matches Inspire records having 'name' in the author list).

HepData data reviews
- NEW Quarkonia data in Hadronic Interactions
- Structure functions in DIS
- Single photon production in hadronic interactions
- Two-photon reactions leading to hadron final states
- Drell-Yan cross-sections
- Inclusive particle production data in e+e- interactions
- Hadronic total cross-sections (R) in e+e- interactions
- Low-energy neutrino cross-sections
- Event shapes in lepton-lepton and lepton-nucleon interactions

Hep Data HepData @HepData 10 Apr
Added @CMSpapers data on "Evidence for pT and eta dependent event plane fluctuations in PbPb and pPb collisions" to hepdata.cedar.ac.uk/view/ins1347386

Hep Data HepData @HepData 7 Apr
Added @ATLASpapers data on "Measurements of normalized differential cross-sections for ttbar in pp at √s = 7 TeV" to hepdata.cedar.ac.uk/view/ins1304289

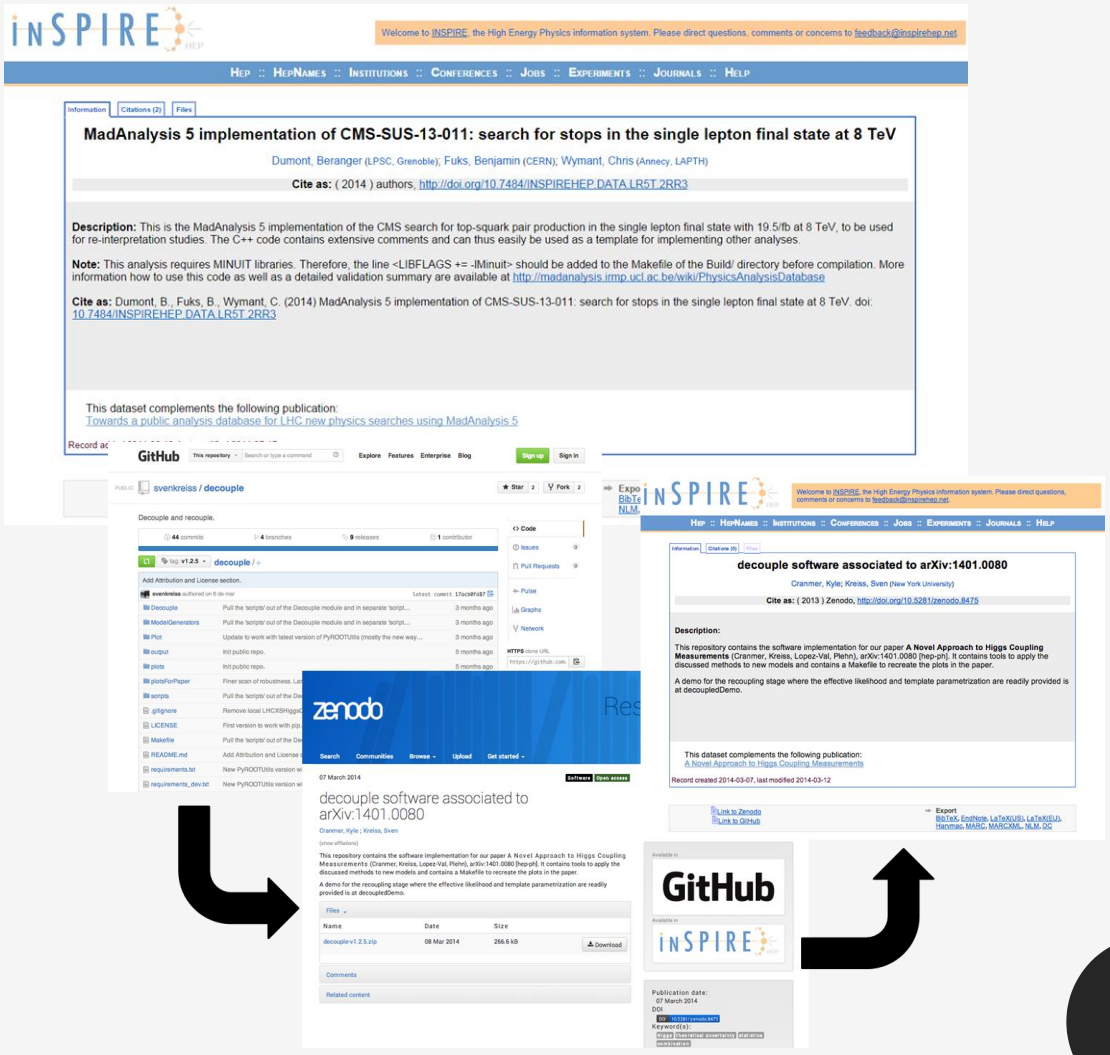
Follow @HepData

Contact us at: hepdata(at)projects.hepforge.org

HepData is funded by the UK STFC and hosted at the Durham IPPP.
HepData also maintains the UK mirror of the PDG.

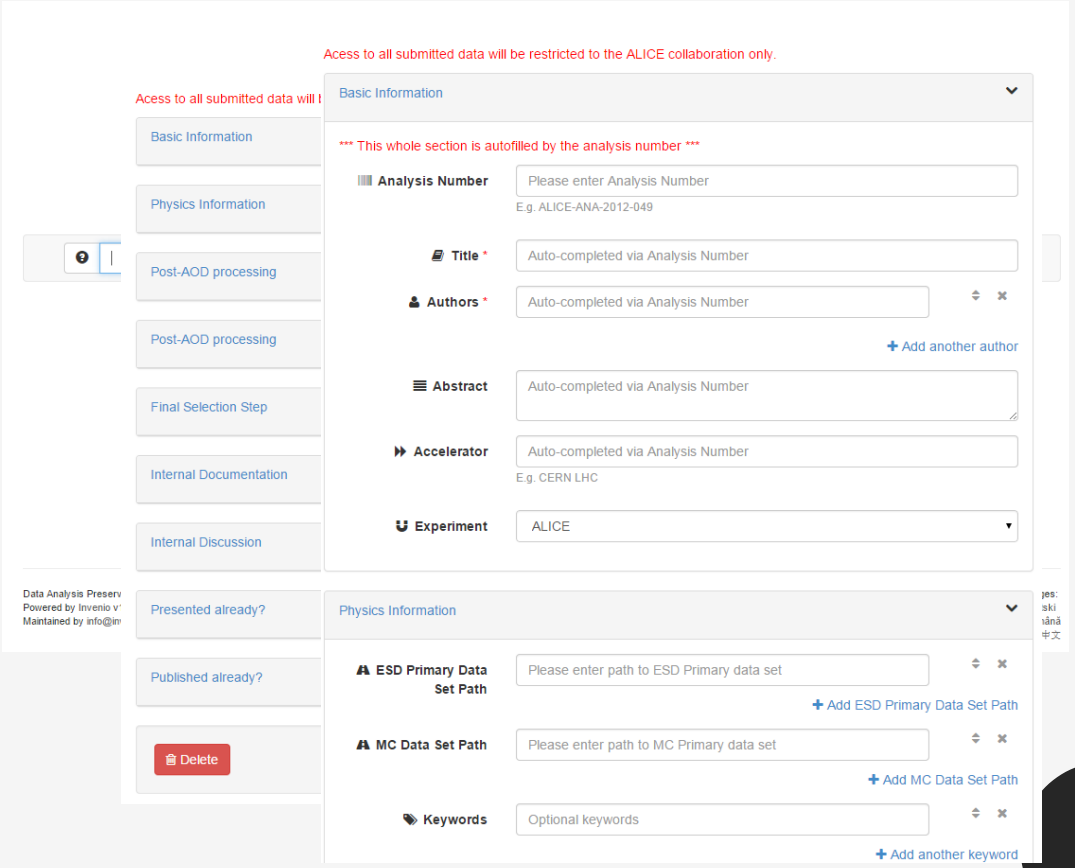
lp3 Science & Technology Facilities Council

Code


INSPIRE
MadAnalysis 5 implementation of CMS-SUS-13-011: search for stops in the single lepton final state at 8 TeV
Dumont, Beranger (LPSC, Grenoble); Fuks, Benjamin (CERN); Wymant, Chris (Annecy, LAPTH)
Description: This is the MadAnalysis 5 implementation of the CMS search for top-squark pair production in the single lepton final state with 19.5/fb at 8 TeV, to be used for re-interpretation studies. The C++ code contains extensive comments and can thus easily be used as a template for implementing other analyses.
Note: This analysis requires MINUIT libraries. Therefore, the line <LIBFLAGS += -lMinuit> should be added to the Makefile of the Build/ directory before compilation. More information how to use this code as well as a detailed validation summary are available at http://madanalysis.irmp.ucl.ac.be/wiki/PhysicsAnalysisDatabase
Cite as: Dumont, B., Fuks, B., Wymant, C. (2014) MadAnalysis 5 implementation of CMS-SUS-13-011: search for stops in the single lepton final state at 8 TeV. doi: 10.7484/INSPIREHEP.DATA.LR5T.2RR3
This dataset complements the following publication:
Towards a public analysis database for LHC new physics searches using MadAnalysis 5
GitHub
svenkreiss / decouple
zenodo
decouple software associated to arXiv:1401.0080
Cranmer, Kyle; Kreiss, Sven (New York University)
Cite as: (2013) Zenodo, http://doi.org/10.5281/zenodo.8475
Description:
This repository contains the software implementation for our paper A Novel Approach to Higgs Coupling Measurements (Cranmer, Kreiss, Lopez-Val, Plehn), arXiv:1401.0080 [hep-ph]. It contains tools to apply the discussed methods to new models and contains a Makefile to recreate the plots in the paper.
A demo for the recoupling stage where the effective likelihood and template parametrization are readily provided is at decoupledDemo.
This dataset complements the following publication:
A Novel Approach to Higgs Coupling Measurements
GitHub
INSPIRE

Open Data is just the tip of the RDM iceberg…

- An analysis capturing and management tool for HEP



Data Analysis Preservation

- Capture
  - Entire workflow
  - With data, code, statistical models, documentation
  - Environment, Virtual Machines
  - OAIS compatible

- Interoperability with
  - Experiments' databases
  - Existing platforms such as the CERN Open Data Portal, INSPIRE

Sources

- V. Stodden, J. Borwein, and D.H. Bailey. "Setting the default to reproducible". In: computational science research. SIAM News 46 (2013), pp. 4-6.
- ATLAS Collaboration (2014). ATLAS Data Access Policy. CERN Open Data Portal. DOI: 10.7483/OPENDATA.ATLAS.T9YR.Y7MZ
- ALICE Collaboration (2013). ALICE data preservation strategy. CERN Open Data Portal. DOI: 10.7483/OPENDATA.ALICE.54NE.X2EA
- CMS Collaboration (2012). CMS data preservation, re-use and open access policy. CERN Open Data Portal. DOI: 10.7483/OPENDATA.CMS.UDBF.JKR9
- LHCb Collaboration (2013). LHCb External Data Access Policy. CERN Open Data Portal. DOI: 10.7483/OPENDATA.LHCb.HKJW.TWSZ

Websites

- http://opendata.cern.ch/
- http://analysis-preservation.cern.ch/
- http://home.web.cern.ch/about/updates/2014/11/cern-makes-public-first-data-lhc-experiments
- https://cmsweb.cern.ch/das/
- https://www.datacite.org/
- https://rd-alliance.org/
- http://www.re3data.org/
- http://www.openscience.org/blog/?p=269
- https://inspirehep.net/
- http://hepdata.cedar.ac.uk/
- https://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/
- https://actu.epfl.ch/news/data-management-plan-at-epfl/
- http://www.nsf.gov/bfa/dias/policy/dmp.jsp
- https://www.stfc.ac.uk/1386.aspx

Acknowl-edgements

My colleagues from the CODP and DAPF team

Work sponsored by the Wolfgang Gentner Programme of the Federal Ministry of Education and Research