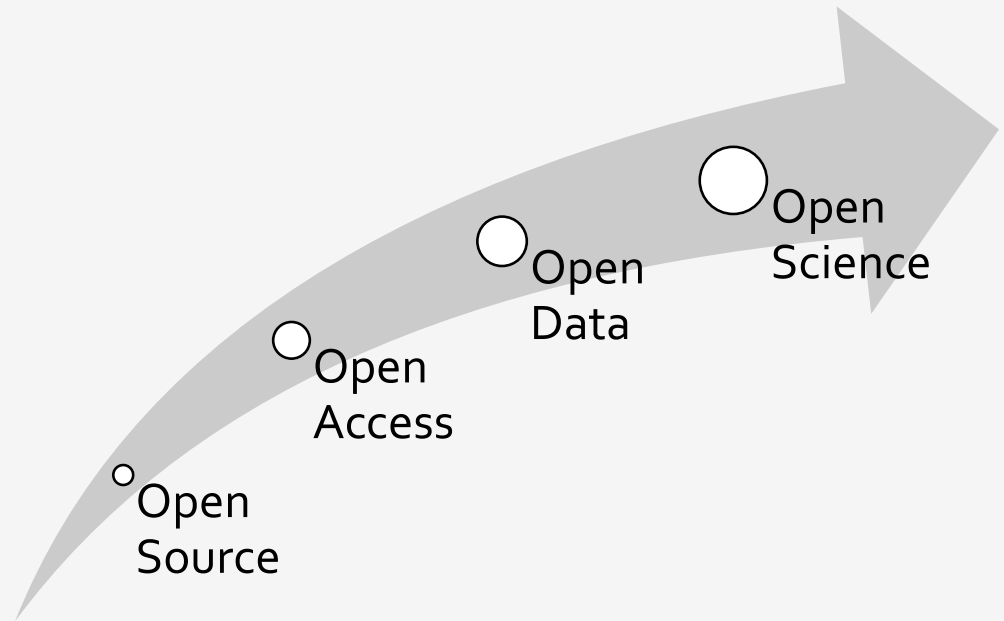


# *OPEN DATA*

*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*

*reviewable*

Description  
available

*replicable*

Tools  
available  
(may be  
private)

*confirmable*

Results  
independently  
obtainable

*auditable*

Data and  
software  
archive  
exists  
(may be  
private)

*reproducible*

Data and  
software  
archive  
exists in  
public

---



# Open Data: incentives

- Funder policies

The screenshot shows the NSF website with the following elements:

- Header:** National Science Foundation logo and tagline "WHERE DISCOVERIES BEGIN". A search bar and "QUICK LINKS" button are on the right.
- Navigation:** A horizontal menu with links: HOME, FUNDING, AWARDS, DISCOVERIES, NEWS, PUBLICATIONS, STATISTICS, ABOUT NSF, FASTLANE.
- Left Sidebar:** "Office of Budget, Finance and Award Management (BFA)" with a green circular graphic. Below it are links for "DIAS Home", "CAAR Branch", "Policy Office", "Systems Office", "View DIAS Staff", and a "Search DIAS Staff" box. At the bottom is "BFA Organization".
- Main Content:** "Dissemination and Sharing of Research Results" section. It includes "NSF Data Sharing Policy" and "NSF Data Management Plan Requirements".

The screenshot shows the PLOS website with the following elements:

- Header:** PLOS logo and navigation links: Publications, Innovation, Open Access, Newsroom, Community.
- Content:** "Data Access for the Open Access Literature: PLOS's Data Policy" article, dated December 13, 2013.

The screenshot shows the STFC website with the following elements:

- Header:** Science & Technology Facilities Council logo and navigation links: Funding, Research, Innovation, Skills, Public Engagement, News, Events and Publications, About Us.
- Content:** "STFC DATASHEET" section with a background image of a facility.

The screenshot shows the EPFL News Research Office website with the following elements:

- Header:** EPFL logo and navigation links: YOU ARE, BY SCHOOL, ABOUT EPFL.
- Section:** "NEWS RESEARCH OFFICE" with "Create an account" and "Help" links.
- Article:** "Data Management Plan at EPFL" dated 30.03.15. The article text includes: "You need to prepare a Data Management Plan for your proposal but don't know how to proceed? Read the article below, which will give you important information about the process at EPFL!" and "Funding agencies such as the European Commission with the Horizon 2020 Research and Innovation program now require, depending on specifics of the research project, the inclusion of a Data Management Plan (DMP) by which the management and safeguard of sensitive data (on humans notably) is".
- Image:** A graphic with the text "OPEN DATA" on a blue background.
- Footer:** "© 2015 EPFL"

The screenshot shows the STFC Scientific Data Policy page with the following elements:

- Section:** "Scientific Data Policy"
- Text:** "STFC has updated its Scientific Data policy. STFC's Executive Board recently approved the introduction of an over-riding data policy to provide guidance to its staff and communities. The policy consists of a set of general principles that cover the wide variety of scientific communities and existing practices that fall within STFC's remit. The key principle of the policy is that all funded activities are required to have a data management plan, which must be in line with recommended good practice. These individual plans will then have the added check of being subject to approval by the relevant STFC boards and panels. Although this has not been a critical issue, a single standardised policy has clear advantages in a single organisation. The policy was drawn up by an internal technical working group set up in 2009 and is in line with current thinking as well as improving transparency for those working for and with STFC."
- Links:** "STFC scientific data policy" link.
- Footer:** "Terms & conditions", "Cymraeg", "FOI", "Copyright", "Glossary", "Sitemap", "Accessibility", and social media icons.

# *Open Data: initiatives*



# HEP Open Data

Approved CB 20<sup>th</sup> June 2014

## ATLAS Data Access Policy

May 21<sup>st</sup> 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 external to the collaboration.

The ATLAS policy for data preservation is based on the collaboration's need to preserve their data in a way that allows them open access. To support this, we will develop and support the tool

## ALICE data preservation strategy

Sunday, October 6, 2013

### Policies for Different

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 information of the profound nature and origin of matter. Because of their unique scientific objective of the data processing framework and will lay the foundation for the scientific community as well as to the general public. These considerations are the ALICE data preservation strategy and policy. Documentation, long term preservation, data access and analysis policy and software availability constitute the strategy allowing future collaborators, the wider scientific community and the general public to have access to the data for their own research and for eventual reassessment of the published results. The following principles that will guide the redaction addressed by the ALICE data

# LHCb External Data Access Policy

## 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 levels of abstraction and at different points in time. 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.

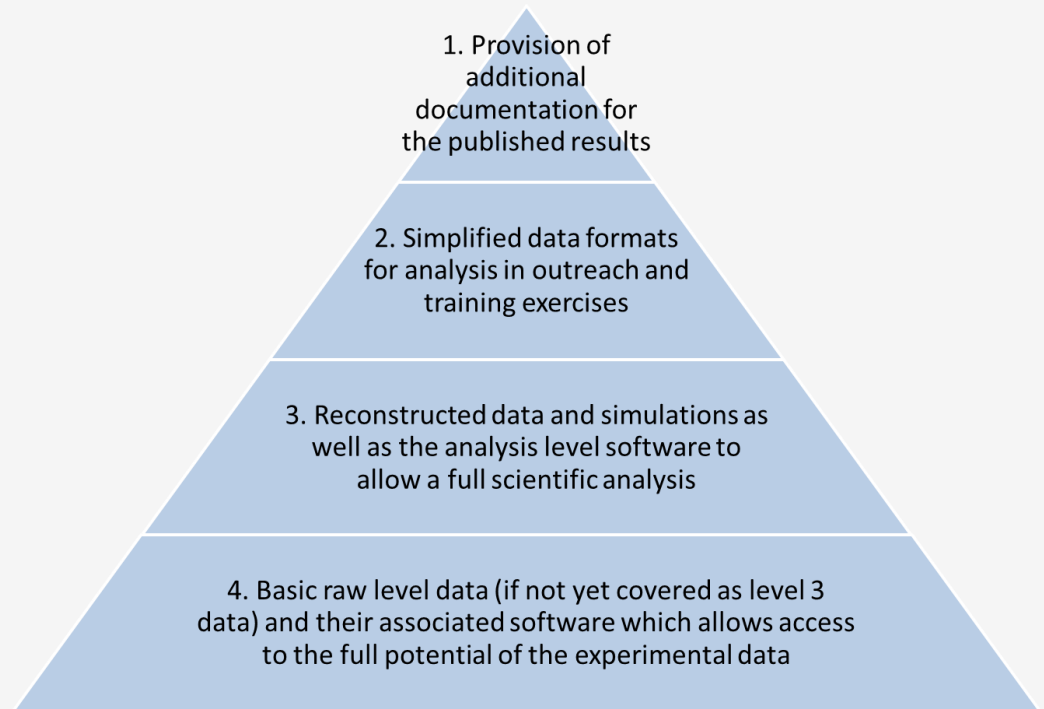
The CMS collaboration is committed to preserve its data, at different levels of complexity, and to allow their re-use by a wide community including: collaboration members long after the data are taken, experimental and theoretical HEP scientists who were not members of the collaboration, educational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

## LHCb Public Information

Issue: 1  
Revision: 1  
Reference: LHCb-PUI  
Created: 22<sup>nd</sup> April 2013  
Last modified: 22<sup>nd</sup> April 2013

# *Data in High- Energy Physics*





# But how to make them open?

opendata  
CERN

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 space 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.

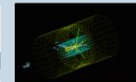
Explore ATLAS >



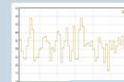
The LHCb (Large Hadron Collider beauty) experiment aims to record the decay of particles containing b and anti-b quarks, known as B mesons. The detector is designed to gather information about the identity, trajectory, momentum and energy of each particle.

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.



Visualization event >



Visualization histogram >



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 available.

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 after 70% 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 >

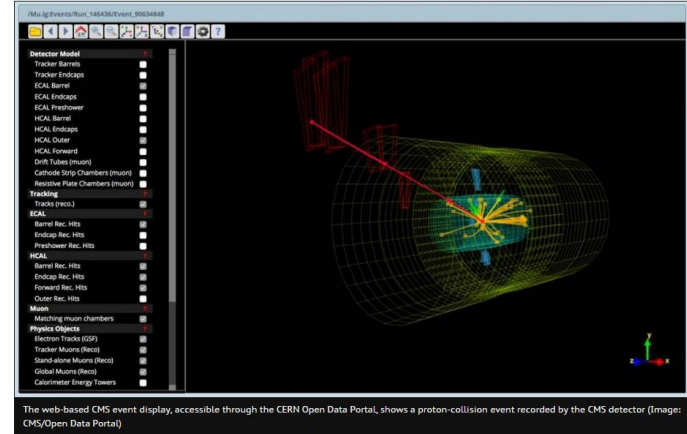


Start analysing the data >

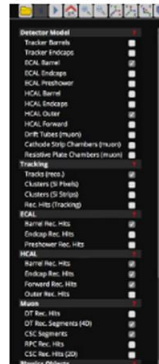
# Release

## CERN makes public first data of LHC experiments

Posted by Cian O'Lualaigh on 20 Nov 2014. Last updated 20 Nov 2014, 16:59.  
 Voir en français

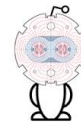


CERN launches public the CERN Open Data Portal  
[/go.tn15T](https://go.tn15T)



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



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

[/u/raskCERN](https://raskCERN)  
[#cermendata](https://cermendata)

RETWEETS 768 FAVORITES 450

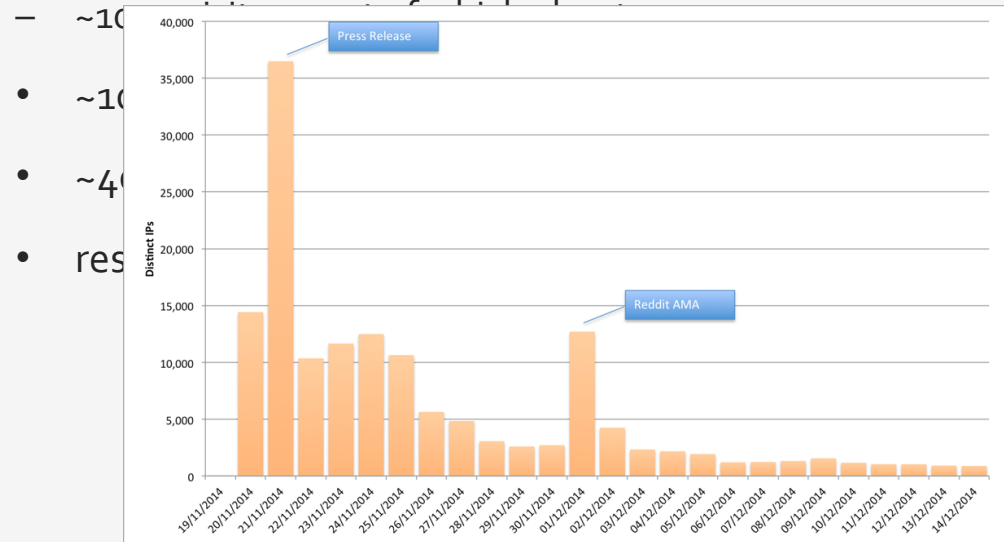


6:43 AM - 20 Nov 2014

# Some numbers...

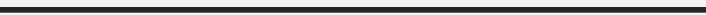
- At the public release:
  - Serving ~15 GB per hour [usage ~50 times now]
  - After a day or two was about ~4 GB per hour [~20 times now]

- Typical day now:



# *...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

Showing 1--1 records out of 1.  
Add filter/aggregator function to the query:

Dataset: [/Mun2010B-Apr21Recco-v1AOD](#)  
Creation time: 2011-04-26 11:32:43. Dataset size: 3.2TB. Number of blocks: 63. Number of events: 32376291. Number of files: 2979. Physics group: NoGroup. Status: VALID. Type: data  
[Dataset](#) [Blocks](#) [Files](#) [Stats](#) [Config](#) [Parents](#) [Children](#) [Stats](#) [Physics Groups](#) [By](#) [Subcollections](#) [TUE/EG](#) [Sources](#) [2979](#) [Info](#)

DAS service: [rblcks](#) DAS api: dataset\_info

```
{
  "dataset": {
    "status": "VALID",
    "modified_by": "jdc@org/DC-06ogrifa/04/People/04/Jacob Linecra 755824",
    "physics_group_name": "NoGroup",
    "acquisition_name": "08S2_LANOML_ACQUISITION_E8A",
    "prep_id": "None",
    "creation_time": "2011-04-26 11:32:43",
    "created_by": "jdc@org/DC-06ogrifa/04/People/04/Jacob Linecra 755824",
    "processed_ds_name": "Run2010B-Apr21Recco-v1",
    "modification_time": "2011-05-02 21:22:30",
    "datatype": "Data",
    "dataset_size": 32376291,
    "crosssection": "None",
    "processing_version": 0,
    "primary_dataset": {
      "name": "Data",
      "data_tier_name": "AOD",
      "name": "/Mun2010B-Apr21Recco-v1AOD"
    }
  }
}
```


DAS service: [rblcks](#) DAS api: filesummaries

```
{
  "dataset": {
    "name": "/Mun2010B-Apr21Recco-v1AOD",
    "events": 32376291,
    "files": 2979,
    "blocks": 63,
    "size": 3288262517618
  }
}
```

DAS service: [rblcks](#) DAS api: datasets

```
{
  "dataset": {
    "status": "VALID",
    "modified_by": "jdc@org/DC-06ogrifa/04/People/04/Jacob Linecra 755824",
    "physics_group_name": "NoGroup",
    "acquisition_name": "08S2_LANOML_ACQUISITION_E8A",
    "prep_id": "None",
    "creation_time": "2011-04-26 11:32:43",
    "created_by": "jdc@org/DC-06ogrifa/04/People/04/Jacob Linecra 755824",
    "processed_ds_name": "Run2010B-Apr21Recco-v1",
    "modification_time": "2011-05-02 21:22:30",
    "datatype": "Data",
    "dataset_size": 32376291,
    "crosssection": "None",
    "processing_version": 0,
    "primary_dataset": {
      "name": "Data",
      "data_tier_name": "AOD",
      "name": "/Mun2010B-Apr21Recco-v1AOD"
    }
  }
}
```

Flask error



# Our solution

```
<collection xmlns="http://www.loc.gov/MARC21/slim">
  <record>
    <controlfield tag="001">14</controlfield>
    <controlfield tag="005">2014122121007.0</controlfield>
    <datafield tag="024" ind1="1" ind2="1">
      <subfield code="2">0001</subfield>
      <subfield code="a">10.7483/OPENDATA.CMS.BBMR.C4A2</subfield>
    </datafield>
    <datafield tag="110" ind1="1" ind2="1">
      <subfield code="a">CMS collaboration</subfield>
    </datafield>
    <datafield tag="245" ind1="1" ind2="1">
      <subfield code="a">/Mu/Run2010B-Apr21ReReco-v1/AOD</subfield>
    </datafield>
    <datafield tag="246" ind1="1" ind2="1">
      <subfield code="a">
        Mu primary dataset in AOD format from RunB of 2010 (/Mu/Run2010B-Apr21ReReco-v1/AOD)
      </subfield>
    </datafield>
    <datafield tag="538" ind1="1" ind2="1">
      <subfield code="a">Software release: CMSSW_4_2_1_patch1</subfield>
      <subfield code="b">Global tag: FT_R_42_V10A1:All</subfield>
    </datafield>
    <datafield tag="256" ind1="1" ind2="1">
      <subfield code="a">
        Mu primary dataset in AOD format from RunB of 2010 (/Mu/Run2010B-Apr21ReReco-v1/AOD)
      </subfield>
      <subfield code="b">
        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.BBMR.C4A2
      </subfield>
    </datafield>
  </record>
</collection>
```

## 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.BBMR.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

### System Details

Software release: CMSSW\_4\_2\_1\_patch1

### Indexes

CMS_Run2010B_Mu_AOD_Apr21ReReco-v1_0005_file_index.txt	Size: 52.9 kB	Download index
Description: Mu AOD dataset file index (6 of 6) for access to data via CMS virtual machine		
CMS_Run2010B_Mu_AOD_Apr21ReReco-v1_0000_file_index.txt	Size: 85.2 kB	Download index
Description: Mu AOD dataset file index (1 of 6) for access to data via CMS virtual machine		

# Small scale data

**INSPIRE** HEP

Welcome to [INSPIRE](#), the High Energy Physics information system. Please direct questions, comments or concerns to [feedback@inspirehep.net](mailto: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 2823 authors](#)

**Cite as:** ATLAS Collaboration ( 2013 ) HepData, <http://doi.org/10.7484/INSPIREHEP.DATA.2684.TY5E>

**Description:**  $-2 \log$  Likelihood for the  $H \rightarrow WW \rightarrow l\nu l\nu$  channel in the  $(\mu_{ggF+tH} * B/BSM, \mu_{VBF+VH} * B/BSM)$  plane for a Higgs boson mass  $m_H = 125.5$  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

## The Durham HepData Project

Export  
RinTeX ForNode LaTeX(LaTeX) Harvmar MARC

REACTION DATABASE • DATA REVIEWS • PDF PLOTTER

ABOUT HEPDATA • SUBMITTING DATA

Enter query:  Search

examples: re gamma gamma%, re p -> p p and obs sig. exp cem  
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

**HepData** @HepData  
Added @CMSpapers data on "Evidence for pT and eta dependent event plane fluctuations in PbPb and pPb collisions" to hepdata.cedar.ac.uk/view/ms1347386 10 Apr

**HepData** @HepData  
Added @ATLASpapers data on "Measurements of normalized differential cross-sections for ttbar in pp at  $\sqrt{s} = 7$  TeV" to hepdata.cedar.ac.uk/view/ms1304289 7 Apr

Follow @HepData

Contact us at: [hepdata\(at\)projects.hepforge.org](mailto: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.

IP3  
Science & Technology  
Facilities Council

# Code

The image displays a workflow for finding and accessing code related to a physics publication. It starts with an INSPIRE record for a MadAnalysis 5 implementation, which includes a description, citation, and a link to a GitHub repository. The GitHub page shows the repository structure and a Zenodo link. A second Zenodo record is shown, which is linked to the GitHub repository and provides a direct download link for the code. The Zenodo record also includes a description of the software and its association with a specific arXiv preprint.

**INSPIRE** HEP :: HEP-NAMES :: INSTITUTIONS :: CONFERENCES :: JOBS :: EXPERIMENTS :: JOURNALS :: HELP

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

Information | Citations (2) | Files

### MadAnalysis 5 implementation of CMS-SUS-13-011: search for stops in the single lepton final state at 8 TeV

Dumont, Beranger (LPSG, Grenoble); Fuks, Benjamin (CERN); Wymant, Chris (Anney, LAPTH)

Cite as: ( 2014 ) authors, <http://doi.org/10.7484/INSPIREHEP.DATA.LR5T.2RR3>

**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 <code>LIBFLAGS += -JMinuit</code> 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.imp.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](http://doi.org/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](#)

Record at [INSPIRE](#)

**GitHub** This repository | Search or type a command | Explore Features Enterprise Blog | Sign up | Sign in

Public | svenkreis / decouple

Decouple and recouple.

- 44 commits
- 4 branches
- 9 releases
- 1 contributor

Code

- Issues
- 1 Pull Requests
- in-Pull
- in-Graphs
- Network

Files

- 1 README.md
- 1 LICENSE
- 1 Makefile
- 1 requirements.txt
- 1 requirements\_dev.txt
- 1 README.md
- 1 LICENSE
- 1 Makefile
- 1 requirements.txt
- 1 requirements\_dev.txt

zenodo

07 March 2014

### decouple software associated to arXiv:1401.0080

Cranner, Kyle; Kreiss, Sven

This repository contains the software implementation for our paper [A Novel Approach to Higgs Coupling Measurements](#) (Cranner, 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 decoupleDemo.

File	Date	Size
decouple-v1.2.5.zip	08 Mar 2014	256.6 kB

Download

Comments

Related content

Link to Zenodo | Link to GitHub

Export: BibTeX, EndNote, LaTeX, LaTeXBib, HTML, MARC, MATCOOL, NLM, DC

**INSPIRE** HEP :: HEP-NAMES :: INSTITUTIONS :: CONFERENCES :: JOBS :: EXPERIMENTS :: JOURNALS :: HELP

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

Information | Citations (2) | Files

### decouple software associated to arXiv:1401.0080

Cranner, 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](#) (Cranner, 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 decoupleDemo.

This dataset complements the following publication:  
[A Novel Approach to Higgs Coupling Measurements](#)

Record created 2014-03-07, last modified 2014-03-12

Link to Zenodo | Link to GitHub

Export: BibTeX, EndNote, LaTeX, LaTeXBib, HTML, MARC, MATCOOL, NLM, DC

**GitHub**

INSPIRE

Publication date: 07 March 2014  
DOI: [10.5281/zenodo.8475](https://doi.org/10.5281/zenodo.8475)  
Keywords (a): Higgs boson, Higgs boson production, Higgs boson decay



# *Open Data is just the tip of the RDM iceberg...*

- An analysis capturing and management tool for HEP

Access to all submitted data will be restricted to the ALICE collaboration only.

Access to all submitted data will I

Basic Information

\*\*\* This whole section is autofilled by the analysis number \*\*\*

**Analysis Number** Please enter Analysis Number  
E.g. ALICE-ANA-2012-049

**Title \*** Auto-completed via Analysis Number

**Authors \*** Auto-completed via Analysis Number [+ Add another author](#)

**Abstract** Auto-completed via Analysis Number

**Accelerator** Auto-completed via Analysis Number  
E.g. CERN LHC

**Experiment** ALICE

Presented already?

Published already?

[Delete](#)

Physics Information

**ESD Primary Data Set Path** Please enter path to ESD Primary data set [+ Add ESD Primary Data Set Path](#)

**MC Data Set Path** Please enter path to MC Primary data set [+ Add MC Data Set Path](#)


**Keywords** Optional keywords [+ Add another keyword](#)

Data Analysis Preservation  
Powered by Invenio v  
Maintained by info@in

es:  
siki  
sina  
中文

# *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](https://doi.org/10.7483/OPENDATA.ATLAS.T9YR.Y7MZ)
  - ALICE Collaboration (2013). ALICE data preservation strategy. CERN Open Data Portal. DOI: [10.7483/OPENDATA.ALICE.54NE.X2EA](https://doi.org/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](https://doi.org/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](https://doi.org/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>
-

# *Acknowledgements*

My colleagues from the CODP and DAPF team

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

