

Lukas Heinrich (NYU), Eamonn Maguire (CERN) CHEP 2016, San Francisco, USA <u>hepdata.net</u>





HEP Scattering experiments going back to the 1950s

Each group of scientists will analyse particular signals by processing large numbers of collisions.

## HEPData

## What is it?



The resulting analysis will be published as a paper.

But where does the processed data go?



HEPData is the go to place for physicists to get access to the data underlying plots and tables in a publication.

It also links to the scripts and ROOT files for instance used in the analysis (for reproducibility).



## HEPData

## What is it?

ATA	N N					
		<u>s</u>				
			Table descri	ption		
	Table 1	Н	RE SQRT(S) IN GEV 7000	<b>P P&gt; X</b> SIG IN MB 95.35 ± 0.38 (stat) ± 1.25 (sys,experimental) ± 0.37	I	
=	Table 2 (F1	1)		(sys,extrapolation)		





## **The Durham HepData Project**

**REACTION DATABASE** • DATA REVIEWS • PDF PLOTTER

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

## **HEP**Data

## What is it?



#### ABOUT HEPDATA • SUBMITTING DATA

#### 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

Нер Data

HepData @HepData 24 Apr Added @ATLASpapers data on "Simultaneous measurements of the tt, W<sup>+</sup>W<sup>-</sup>, and Z/ $\gamma^* \rightarrow \tau\tau$  cross sections in pp at 7 TeV" to hepdata.cedar.ac.uk/view/ins1304455

HepData @HepData Hep Data

24 Apr

Added @CMSpapers data on "Search for SM production of four top quarks in the lepton + jets channel in pp at 8 TeV" to hepdata.cedar.ac.uk/view/ins1318946

🈏 Follow @HepData

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





Science & Technology Facilities Council



## New platform built upon Invenio (version 3)

Whether you're a data provider, or consumer, the new HEPData has many functionalities

## **Data Providers**

- 1. A Simplified Submission Process
- 2. A standard entry data format
- 3. Full review management system
- 4. Versioning
- 5. DOI minting
- 6. Sandbox

# HEPData

## **Data Consumers**

- 1. Publication Driven Search
- 2. Semantic Publishing
- 3. Data Conversion
- 4. Access in Analysis Environments



## **Data Providers**

- 2. A standard entry data format
- 4. Versioning
- 5. DOI minting
- 6. Sandbox

# HEPData

1. A Simplified Submission Process 3. Full review management system



- 2. A standard entry data format
- 4. Versioning
- 5. DOI minting
- 6. Sandbox

# HEPData

**Data Providers** 

1. A Simplified Submission Process

3. Full review management system





## **Submission Process**

↑ Submit ⓒ Sandbox ۞ Help 🎍 Dashboard 🙂 Log out





## **Data Providers**

- 2. A standard entry data format
- 4. Versioning
- 5. DOI minting
- 6. Sandbox

# HEPData

1. A Simplified Submission Process 3. Full review management system





## HEPData

## HEPData submission archive

#### submission.yaml

links the submission together by detailing the data files to be loaded, their name and description, and their assocated analysis files and code.

#### data records

YAML (or JSON) representation of the underlying data files including value errors in a verbose format.

#### external data files & links

analysis files, code, links to code repositories, etc.



Tables rendered from

JSON

Ð		ATA
	Table 1	Table descri
		Scripts
	RE	P P> X
	SQRT(S) IN GEV	SIG IN MB
	7000	95.35 ± 0.38 (stat) ± 1.25 (sys,ex (sys,extrapolation)

Processes YAML file, inserts records in to database and links publication record with data and files.

## HEPDATA



#### Plots rendered automatically using a custom library built upon D3.js

{JSON} Tables and plots



## Web Server





## **Data Providers**

- 2. A standard entry data format
- 4. Versioning
- 5. DOI minting
- 6. Sandbox

# HEPData

1. A Simplified Submission Process 3. Full review management system



# Table 2 📥 Download All 🗸



#### Abstract (data abstract)

CERN-LHC. This paper reviews and extends searches for the direct pair production of the scalar supersymmetric partners of the top and bottom quarks in proton--proton collisions collected by the ATLAS collaboration during the LHC Run 1. Most of the analyses use 20 fb $^{-1}$  of collisions at a centre-of-mass energy of  $\sqrt{s} = 8$  TeV, although in some case an additional 4.7 fb<sup>-1</sup> of collision data at  $\sqrt{s} = 7$  TeV are used. New analyses are introduced to improve the sensitivity to specific regions of the model parameter space. Since no evidence of third-generation squarks is found, exclusion limits are derived by combining several analyses and are presented in both a simplified model framework,

Copyright ~1975-Present, HEPData | Powered by Invenio, funded by STFC, UK, developed and hosted at CERN, supported by IPPP Durham.

## **HEP**Data

## **Comprehensive Review System**



## **Comprehensive Review System**

<b>HEPData Q</b> Search HEP Data	Search		
Q Browse all 🖉 Aad, Georges et al.			
Hide Publication Information Information	🗘 Upload New Files	Table 5	
ATLAS Run 1 searches for direct pair production of third-generation squarks at the Large Hadron Collider	✓ Download All  ✓ Filter 8 data tables	Combined excl through $\tilde{t}_1 \rightarrow$	usion $b + \tilde{\chi}$
Aad, Georges , Abbott, Brad , Abdallah, Jalal , Abdinov, Ovsat , Aben, Rosemarie , Abolins, Maris , AbouZeid, Ossama , Abramowicz, Halina , Abreu, Henso , Abreu, Bisarda	production cross sections for different values of the branching ratios for the decays	- Observed lim - Expected limi	it x=B t x=BF
ATLAS	<pre>&gt; passed review</pre>	<u>null</u>	
Eur.Phys.J. C75 (2015) 510, 2015		cmenergies	
http://dx.doi.org/10.17182/hepdata.71384	Table 4	♥ 8000.0	
DOI       View paper in Inspire       View old HepData         Additional Resources         Abstract (data abstract)         CERN-LHC       This paper reviews and extends searches for	Data from Figure 6 None Combined exclusion limits assuming that the stop decays through $\tilde{t}_1 \rightarrow t + \tilde{\chi}_1^0$ with branching ratio x and	Showing 50 of 182 v	
the direct pair production of the scalar supersymmetric partners of the top and bottom quarks in protonproton	Table 5	RE	P P CH
collisions collected by the ATLAS collaboration during the LHC Run 1. Most of the analyses use 20 fb <sup>-1</sup> of collisions at a centre-of-mass energy of $\sqrt{s} = 8$ TeV, although in some	Data from Figure 6	SQRT(S)	800
case an additional 4.7 fb <sup>-1</sup> of collision data at $\sqrt{s} = 7$ TeV are used. New analyses are introduced to improve the	Combined exclusion limits assuming that the stop decays through $\tilde{t}_1 \rightarrow t + \tilde{x}_1^0$ with	M(STOP) [GEV]	M(
Since no evidence of third-generation squarks is found, exclusion limits are derived by combining several analyses	branching ratio x and	246.08	55.
and are presented in both a simplified model framework,	• passed review	247.47	56.
	Table 6	251.73	59.
	Data from Figure 6 None	255.59	61
	Combined exclusion limits		

## **HEP**Data





Management System





No name provided Add Name

⊡ eamonnmag@gmail.com

#### SUBMISSIONS IN PROGRESS



HEPData Administration CReindex Database Cdit Profile

#### **WATCH LIST**

You can watch records for updates. Nothing has been watched ye

#### ▲ PERMISSIONS

 $\mathbf{O}_{\alpha}^{\alpha}$ 

You are a HEPData coordinator.
You are able to create and manage HEPData submi

You can see papers you have coordinator, uploader, or reviewe here:

coordinator uploader reviewer

Exclusive  $\rho^0$  meson photoproduction with a leading neutron at HERA

Eur.Phys.J. C76 (2016) 41

coordinator

Measurement of the forward-backward asymmetry in the distribution of leptons in  $t\bar{t}$ events in the lepton+jets channel Phys.Rev. D90 (2014) 072001

coordinator

ථ Log out
Log out
et.
ssions.
er rights to



## HEPData

- **Submissions Overview**
- Interactive dashboard providing stats on all submissions





No name provided Add Name

#### 🖂 eamonnmag@gmail.com

A Back to Dashboard

#### Submissions By Last Update Time



#### **Data Records Per Submission**



#### Collaboration TOP 10



#### Submission Status



Version













No name provided Add Name

#### 🖂 eamonnmag@gmail.com

K Back to Dashboard

#### Submissions By Last Update Time



#### Data Records Per Submission





#### Submission Status

Version

NASH (Encoded) NASH (Modified) OPY (Encoded) PRS (Encoded)

SIA (Encoded) SIS (Encoded)





## **Data Providers**

- 2. A standard entry data format
- 4. Versioning
- 5. DOI minting
- 6. Sandbox

# HEPData

1. A Simplified Submission Process 3. Full review management system



#### 🕹 Upload New Files Viewing version 2 -Version 1 Version 2 Table 1 Page 17 of preprint The measured fiducial cross sections. The first systematic uncertainty is the combined systematic uncertainty excluding luminosity, the second is the ... passed review Table 2 > Auxiliary Figure 9b. Signal acceptance for the GGM model with $tan(\beta)=30$ in the combined electron and muon SR-Z. passed review Table 3 > Figure 8A Normalized ZZ fiducial cross

section (multiplied by 10^6 for readability) in values of the leading reconstructed dilepton

The observed and expected EmissT distribution in the dielectron SR-Z. The negigible estimated contribution from Z+jets is distributions. The last bin contains the overflow.

#### energies 8000

#### Data

SQRT(S)	8000.0 GeV					
EVENTS	25 GE\	25 GEV				
ETMISS [GEV]	Data	Expected Background	GGM 700 200 1.5	GGM 900 600 1.5		
200 - 225	0	0	0	0		
225 - 250	6	0.95 -0.51, 0.41 stat	6.46	0.97		
250 - 275	1	0.9 -0.26, 0.41 stat	6.82	1.07		
275 - 300	1	0.42 -0.19, 0.12 stat	2.82	1.17		
300 - 325	1	0.34 -0.15, 0.16 stat	2.41	1.05		
325 - 350	2	0.07 -0.16, 0.19 stat	3.11	1.08		
350 - 375	1	0.68 -0.55, 0.56 stat	0.7	1.13		
375 - 400	1	0.17 -0.15, 0.1 stat	0.9	1.2		
400 - 425	0	0.24 -0.1, 0.11 stat	0.69	1.01		
425 - 450	1	0.01 ±0.08 stat	0.72	0.94		
450 475	0		0	0.00		

## **HEP**Data

## Versioning



Deselect variables or hic error bars by clicking on

#### Variables



## **Data Providers**

- 2. A standard entry data format
- 4. Versioning
- 5. DOI minting
- 6. Sandbox

# HEPData

1. A Simplified Submission Process 3. Full review management system



# All HEPData records get DOIs.

Each data table gets a versioned <u>DOI</u>.

The whole HEPData record is also given a <u>DOI</u> to encompass the whole collection.

## HEPData

## DOIs

When using this data, please cite the original publication:	
The ATLAS collaboration (2016).	
Measurements of four-lepton production in $pp$ collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. Phys.Lett. B753 (2016) 552-572.	nfo
http://dx.doi.org/10.1016/j.physletb.2015.12.048	ie
Additionally, you should also cite the $$ HEPData record:	nra
The ATLAS collaboration (2016).	•
Measurements of four-lepton production in $pp$ collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. HEPData.	
http://dx.doi.org/10.17182/hepdata.69768	
	> >
You can also cite the 5 data tables individually:	
The ATLAS collaboration (2016).	
Table 1 of Measurements of four-lepton production in $pp$ collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. HEPData.	ł
http://dx.doi.org/10.17182/hepdata.69768.v1/t1	
The ATLAS collaboration (2016).	
Table 2 of Measurements of four-lepton production in $pp$ collisions at $\sqrt{s}=$ 8 TeV with the ATLAS detector.	



## **Data Providers**

- 2. A standard entry data format
- 4. Versioning
- 5. DOI minting
- 6. Sandbox

# HEPData

1. A Simplified Submission Process 3. Full review management system



## **Data Consumers** Get access to the data in many environments

- 1. Search
- 2. Semantic Publishing
- 3. Data Conversion

# HEPData

4. Access in Analysis Environments



#### **Q** Search HEPdata Search **Reset search**

🕽 Max results 🗸

↓ Sort by -

 $\downarrow_{A}^{z}$  Reverse order

Showing 25 of 8194 results



#### Production of K<sup>\*</sup> (892)<sup>0</sup> and $\phi$ (1020) in p-Pb collisions at $\sqrt{s_{\rm NN}}$ = 5.02 TeV

Adam, Jaroslav; Adamova, Dagmar; Aggarwal, Madan Mohan; et al. The ALICE collaboration. No Journal Information, 2016.

% Inspire Record 1418181 % DOI 10.17182/hepdata.72807

The production of K\* (892)<sup>0</sup> and  $\phi$  (1020) mesons has been measured in p-Pb collisions at  $\sqrt{s_{\rm NN}}$  = 5.02 TeV. K\*<sup>0</sup> and  $\phi$  are reconstructed via their decay into charged hadrons with the ALICE detector in the rapidity range -0.5 < y < 0. The transverse momentum spectra, measured as a function of the multiplicity, have p<sub>T</sub> range from 0 to 15 GeV/c for K<sup>\*0</sup> and from 0.3 to 21 GeV/c for  $\phi$ . Integrated yields, mean transverse momenta and particle...

#### **30** data tables

Table 1	Average charged particle pseudo-rapidity density, $\langle dN_{ch}/d\eta \rangle$ defined using the VOA estimator; values for $\langle dN_{ch}/d\eta_{lab} \rangle$ are
Table 2	$p_{ m T}$ -differential yield of (K $^{*0}$ + $\overline{K}{}^{*0}$ )/2 in p-Pb collisions with
Table 3	$p_{ m T}$ -differential yield of (K $^{*0}$ + $\overline{K^{*0}}$ )/2 in p-Pb collisions with
More	

#### INVESTIGATION OF INCLUSIVE PROCESSES pi- A ---> pi- X AND pi- A ---> p (backwards) X AT 40-GeV/c

Abrosimov, A.T.; Albini, E.; Antipov, V.V.; et al. Conference Paper, 2016. % Inspire Record 209961 % DOI 10.17182/hepdata.39782 None

#### **3** data tables

 $\eta_{
m lab}$ , measured at mid-rapidity in visible cross section event classes and average number of colliding nucleons,  $\langle {
m N}_{
m coll} 
angle$ . Multiplicity classes are e corrected for vertexing and trigger efficiency. Since statistical uncertainties are negligible, only total systematic uncertainties are reported.

centre-of-mass energy/nucleon=5.02 TeV (NSD). Additional systematic error: +- 3.1% (normalization).

centre-of-mass energy/nucleon=5.02 TeV (0-20% multiplicity class).







## Detailed view with quick access to all information



## **HEP**Data



## Access all supplementary resources (images, slha files, RIVET analyses, etc. )

## HEPData

Q Browse all 🛛 🖉 Aad, Georges et al

#### Hide Publication Information

Search for supersymmetry with photons, bottom qua missing transverse momer proton-proton collisions a of-mass energy of 7 TeV w ATLAS detector

#### The ATLAS collaboration

Aad, Georges, Abajyan, Tatevik, Abbo Abdallah, Jalal, Abdel Khalek, Samah, Ahmed Ali, Abdinov, Ovsat, Aben, Ros Babak, Abolins, Maris

Phys.Lett. B719 (2013) 261-279, 201

http://dx.doi.org/10.5072/hepdata.9

INSPIRE Record HepData

#### Abstract (data abstract)

CERN-LHC. Study of events produced in collisions at a centre-of-mass energy of 7 have a final state consisting of an isolated high PT, at least on jet identified as origin bottom quark, and high missing PT. The used, collected in 2011, has a total integr luminsoity of 4.7 fb-1. The results of the interpreted in the context of general gau (GGM).

# ♥ filter Common Resources Table 1 Table 2 Table 3 Table 4 Table 5 Table 6

Saarak

Additional Publication Resources

#### Add Resource for Tabl

Туре	APPLgrid
URL	URL
escription	Short de



## Image Fil

Image file

Download

## **HEP**Data

			Dashboard ()	og out
	×			
		<b>99</b> Cite	• Watch Record	JSON
e 4				ISON
Analysis	◆	pdata.net/r		<u>130N</u>
		Franker, 155 Carl	10 10 10 10 10 10 10 10 10 10	7LAS 1 + 4.7 10 <sup>-1</sup> -7 TeV 10.7500 GeV
scription (optional)		Data MC		
	Add new Resource	reac	tions P> GAMMA BJET I	MM
μ(-Φ 	1100       GGM higgsino like neutralino, tard=15, b) =0         1000       900         900       000     <			
e	Image File			
	Thumbnail image file Download			



## **Data Consumers** Get access to the data in many environments

- 1. Publication Driven Search
- 2. Semantic Publishing
- 3. Data Conversion

# HEPData

4. Access in Analysis Environments



## **Semantic Publishing**

## Every article is tagged with <u>schema.org</u> vocabulary.

Makes it possible for Google and other search engines to understand our content.

https://hepdata.net/search

https://hepdata.net/record/ins1397180



## **HEP**Data

<u>Google's View</u>

Google's view



## **Data Consumers** Get access to the data in many environments

- 1. Publication Driven Search
- 2. Semantic Publishing
- 3. Data Conversion

# HEPData

4. Access in Analysis Environments



## Converter Convert from YAML to ROOT, YODA, CSV

Install via PIP, use as a web service, and contribute to more conversions with our extendable API!

## HEPData





## Conversion to the most common formats

nergy correlation function (TEEC).

0.0342, 0.0334	4 sys,jes	±0.0094	4 sys,jer	±0.0374 sys	,shower	
8 sys,unfoldin	g					
004, 0.0044 sy	s,jes ±	:0.0011 sy	s,jer ±	0.0044 sys,sh	ower	
5 sys,unfoldin	g					
0026, 0.0029 s	sys,jes	±0.0006 s	sys,jer	±0.0028 sys,s	hower	
2 sys,unfoldin	g					
0022, 0.0024 s	sys,jes	±0.0004 s	sys,jer	±0.0023 sys,s	hower	
1 sys,unfoldin	g					
.0022 sys,jes	±0.000	3 sys,jer	±0.0022	2 sys,shower	±0.0018	} sy

002 svs ies +0.0002 svs ier +0.002 svs shower +0.0015 svs nileun

## HEPData





## **Data Consumers** Get access to the data in many environments

- 1. Publication Driven Search
- 2. Semantic Publishing
- 3. Data Conversion

# HEPData

4. Access in Analysis Environments





## Every content page has a JSON equivalent...

So all of HEPData and its content can be accessed programmatically.

## HEPData





No need to leave the software environment you like.

The same can be said for use in ROOT or any analysis platform with a file parser :)

## HEPData

Use case: search, access, and get data directly from Mathematica







## Example File <u>here</u>

### Everything on Github & Open Source <u>http://www.github.com/hepdata</u>



hepdata-docker

Dockerfiles for HEPData/hepdata application Updated 12 days ago

invenio

----

Explore Features Enter	rprise Pricing Sign-up Sign in		
			100
	People 4.2		
Pythen #0 20			
Python # 0 1/ 0 leb Services)			
Python #4 20			
** >*			
mework for			
** 2*			
Python #0 9142			
-			
-			
	State of the local division of the local div		
	and the second division of the second divisio		
	-		
	and the second		
		and the second s	







## HEPData @ CERN

Eamonn Maguire

## HEPData @ Durham

Graeme Watt Michael Whalley Frank Kraus

## HEPData

## Acknowledgements

## HEPData @ NYU

Lukas Heinrich Kyle Cranmer

## HEPData @ Salamanca Juan Luis Boya Garcia

