

HepData Present and Future

Graeme Watt (incoming Database Manager, since October 2013)

HepData advisory board kickoff meeting
IPPP Durham, 12th November 2014

- Evolution of data submission procedure
- Other recent/future developments

<http://hepdata.cedar.ac.uk>

Data “input” format

```
*dataset:  
*location: Page 20 of preprint  
*dscomment: The measured total cross sections.  
The first systematic uncertainty is the combined  
systematic uncertainty excluding luminosity, the  
second is the luminosity  
*reackey: P P --> Z0 Z0 X  
*obskey: SIG  
*qual: RE : P P --> Z0 Z0 X  
*yheader: SIG(total) IN FB  
*xheader: SQRT(S) IN GEV  
*data: x : y  
7000; 6.7 +- 0.7 (DSYS=+0.4,-0.3,DSYS=0.3:lumi);  
*dataend:
```

<http://hepdata.cedar.ac.uk/view/ins1203852/d2/input>

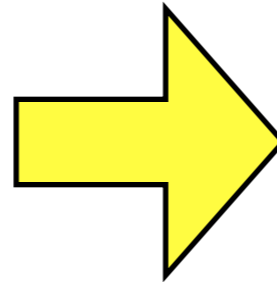


Table 2 (Page 20 of preprint.) or as: input,
plain text, AIDA, PyROOT, YODA, ROOT, mpl,
ScaVis or MarcXML

The measured total cross sections. The first systematic
uncertainty is the combined systematic uncertainty
excluding luminosity, the second is the luminosity.

RE	P P -> Z0 Z0 X
SQRT(S) IN GEV	SIG(total) IN FB
7000.	6.7 ± 0.7 (stat) +0.4,-0.3 (sys) ± 0.3 (sys,lumi)

Plot
SelectPlot

- All data needs to be transformed into a standard “input” text format.
- Metadata for each table followed by data points in a structured format.
- *reackey and *obskey keywords can be used for searching database.

Data extraction from arXiv source

DB Manager

(Mike Whalley, Graeme Watt)

Look at hep-ex (and nucl-ex) listings.
Identify relevant papers containing numbers.
Instruct DB Assistant how to organise data.

DB Assistant

(Joanne Bentham, non-physicist, full-time)
(Stephen Webster, 2014 summer student)

Extract numbers from .tex source.
Format as required "input" file.

Check, edit, add record to public DB.

- DB Manager can also perform entire tedious procedure.
- Only small fraction of papers can be manually processed.
- Most papers do not contain numerical values of data points.

Manual processing of submitted text files

Experiment

Submit data points by email in arbitrary text format.

Check record on test database.
Email corrections if necessary.

DB Manager

(Mike Whalley, Graeme Watt)

Process numbers into “input” format.
Add metadata for each table by hand.
Enter record on test database.

Add record to public database.

- Processing each paper often took hours/days if many tables.
- Tedious, error-prone and unnecessary procedure.
- Experiments were unaware of existence of “input” format.

Direct submission in “input” format

- Earlier this year we started to encourage use of “input” format.
- Links were added to export each table to the “input” format.
- An annotated sample input file was also provided.
- More comprehensive documentation of format still needed.

Experiment

Prepare text file in “input” format.
Test record by uploading via online form.
Submit completed input file by email.

DB Manager

(Mike Whalley, Graeme Watt)

Check, edit, add record to public DB.

- Almost all submissions in “input” format within a few months.
- Third-party tools developed using “input” format:
 - Script by Mike Flowerdew (ATLAS) to convert objects in ROOT files to “input” format.
 - ROOT class by Christian Holm Christensen (ALICE) to extend TGraph to support data points with multiple errors, that can export to and import from HepData “input” format.

Automated submission system

(Procedure agreed in HepData meeting at CERN on 26th June 2014)

Administrator

(e.g. convener of physics group)

<http://hepdata.cedar.ac.uk/manage>

Enter Inspire ID to allocate paper.
Password sent by email.

Receive email notification that
record is 'Ready' on test DB.

OK?

No

Yes

Add record to public DB.

Encoder

(e.g. primary author of physics analysis)

<http://hepdata.cedar.ac.uk/input>

Upload text file in "input" format
(+ optional supplementary files).
Record is added to test DB.
Flag as 'Ready' once satisfied.

- System launched two weeks ago. Active Administrators so far: **ATLAS** (EXOT, STDM, SUSY), **CMS** (B2G, SMP), **LHCb** [Future: **ALICE**]

<http://hepdata.cedar.ac.uk/manage>

Administrator Interface

Administrator ID:

Admin Password:

ATLASSUSY Group

The administrator is: **atlas-phys-susy-hepdata@cern.ch**

Assign a new Paper

Inspire ID:

De-assign a Paper

Inspire ID: (from list below and test database)

View Assigned Papers (for ATLASSUSY Group)

Paper	Public Database (hepdata)	Test Database (hepdatabasetest)	Status	Add/Remove (public db)
1304458	28/10/14 16:46:39 [input] [normal]	28/10/14 16:41:39 [input] [normal]	Added	<input type="button" value="Remove"/>

- First paper added by ATLAS SUSY group:
<http://hepdata.cedar.ac.uk/view/ins1304458>

<http://hepdata.cedar.ac.uk/input>

DATA INPUT FORM

Inspire ID:

Password:

Refresh

Status=Added:

Select a file to upload using no file selected
then to the text box below for processing.

OR

with arXiv:

the Input Data - then - or the Processed Record

```
*author: AAD
*reference: ARXIV:1407.0603 : 2014
*reference: CERN-PH-EP-2014-144 : 2014
*reference: JHEP 1409,103 (2014) : 2014
*reference: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/SUSY-2013-10/ : 2014
*doi: 10.1007/JHEP09(2014)103
*status: Encoded 28 OCT 2014 by ATLAS
*experiment: CERN-LHC-ATLAS
*detector: ATLAS
*inspireId: 1304458
*cdsId: 1728108
*durhamId: 6259
*title: Search for supersymmetry in events with large missing transverse momentum, jets, and at least one tau lepton in 20 fb^-1 of sqrt(s)=8 TeV proton-proton collision data with the ATLAS detector
*comment: CERN-LHC. A search for supersymmetry (SUSY) in events with large missing transverse momentum, jets, at least one hadronically decaying tau lepton
```

The following allows uploading of files to the area resource/1304458

- Encoder can upload up to four supplementary files of any format.

<http://hepdata.cedar.ac.uk/view/ins1304458>

AAD 2014 – Search for supersymmetry in events with large missing transverse momentum, jets, and at least one tau lepton in 20 fb⁻¹ of sqrt(s)=8 TeV proton-proton collision data with the ATLAS detector

Experiment: [CERN-LHC-ATLAS \(ATLAS\)](#)

Published in [JHEP 1409,103 \(2014\)](#) (DOI:10.1007/JHEP09(2014)103)

Preprinted as [CERN-PH-EP-2014-144](#)

Archived as: [ARXIV:1407.0603](#)

Auxiliary Material: <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/SUSY-2013-10/>

Record in: [INSPIRE](#)

Record in: [CERN Document Server](#)

CERN-LHC. A search for supersymmetry (SUSY) in events with large missing transverse momentum, jets, at least one hadronically decaying tau lepton and zero or one additional light leptons (electron/muon), has been performed using 20.3 fb⁻¹ of proton-proton collision data at sqrt(s)=8 TeV recorded with the ATLAS detector at the Large Hadron Collider. No excess above the Standard Model background expectation is observed in the various signal regions and 95% confidence level upper limits on the visible cross section for new phenomena are set. The results of the analysis are interpreted in several SUSY scenarios, significantly extending previous limits obtained in the same final states. In the framework of minimal gauge-mediated SUSY breaking models, values of the SUSY breaking scale Lambda below 63 TeV are excluded, independently of tan(beta). Exclusion limits are also derived for an mSUGRA/CMSSM model, in both the R-parity-conserving and R-parity-violating case. A further interpretation is presented in a framework of natural gauge mediation, in which the gluino is assumed to be the only light coloured sparticle and gluino masses below 1090 GeV are excluded.

[slha file for one example point from the bRPV grid. Grid parameters are m₀=600 GeV and m_{1/2}=600 GeV.](#)

[slha file for one example point from the nGM grid. Grid parameters are m_{gluino}=940 GeV and m_{stau}=210 GeV.](#)

[slha file for one example point from the GMSB grid. Grid parameters are Lambda=60 TeV and tan\(beta\)=30.](#)

[slha file for one example point from the mSugra grid. Grid parameters are m₀=800 GeV and m_{1/2}=400 GeV.](#)

Total number of tables: 113. Displaying: 1 to 10. [First](#) | [Previous](#) | [Next](#) | [Last](#) | [All](#)

- Links to supplementary files appear below the abstract.

@HepData on Twitter

- Twitter feed launched almost one year ago.
- Widget with Twitter feed embedded on homepage.
- Now sent ~200 tweets with over 100 followers.

 Pinned Tweet
HepData @HepData · Nov 21
This Twitter feed will announce new records added and other updates to the HepData database for High Energy Physics data.
← ↻ 3 ★ 2 ...

HepData @HepData · Nov 6
Added @LHCbPhysics data on "Measurement of the cross-section for $Z \rightarrow e^+ e^-$ production in pp collisions at 7 TeV" to hepdata.cedar.ac.uk/view/ins1208102
← ↻ 2 ★ ...

 **Frank Close** @closefrank · Feb 14
Anyone interested in particle physics data (tech) #ff @HepData for updates and state of the art.
← ↻ 1 ★ 2 ...

Kyle Cranmer @KyleCranmer · 4 hrs
Waking up at 2AM for @HepData advisory board meeting

Juan Rojo @JuanRojoC · Dec 14
Just discovered that @HepData has a twitter feed - very useful to get updates when new data becomes available!

LaTeX support via MathJax

```

*dataset:
*location: Fig 3
*dscomment: Transverse  $\langle \sum p_T^{\text{ch}} / \delta\eta\delta\phi \rangle$  vs.  $p_T^{\text{lead}}$  in  $|\eta| < 2.5$  in incl jet / excl dijet events
*reackey: P P --> JET X
*reackey: P P --> JET JET X
*qual: PT(LEAD) IN GEV : > 20
*qual: RE : P P --> JET X : P P --> JET JET X
*qual: Region : Transverse
*qual: SQRT(S) IN GEV : 7000.0
*qual: Selection : Inclusive jet : Exclusive dijet
*yheader:  $\langle \sum p_T^{\text{ch}} / \delta\eta\delta\phi \rangle$  (GeV)
*xheader:  $p_T^{\text{lead}}$  (GeV)
*data: x : y : y
20.0 TO 30.0; 1.184 +- 0.002133 (DSYS=+0.05209,-0.04949); 1.156 +- 0.007098 (DSYS=+0.1071,-0.1025);

```

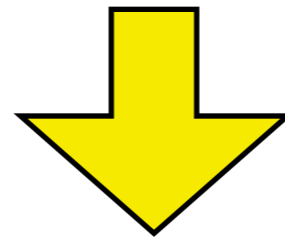


Table 1 (Fig 3.) or as: [input](#), [plain text](#), [AIDA](#), [PyROOT](#), [YODA](#), [ROOT](#), [mpl](#), [ScaVis](#) or [MarcXML](#)

Transverse $\langle \sum p_T^{\text{ch}} / \delta\eta\delta\phi \rangle$ vs. p_T^{lead} in $|\eta| < 2.5$ in incl jet / excl dijet events.

PT(LEAD)	> 20 GeV	
RE	P P → JET X	P P → JET JET X
Region	Transverse	
SQRT(S)	7000.0 GeV	
Selection	Inclusive jet	Exclusive dijet
p_T^{lead} (GeV)	$\langle \sum p_T^{\text{ch}} / \delta\eta\delta\phi \rangle$ (GeV)	
20.00 – 30.00	1.184 ± 0.002 (stat) +0.052,-0.049 (sys)	1.156 ± 0.007 (stat) +0.107,-0.102 (sys)

<http://hepdata.cedar.ac.uk/view/ins1298811/d1>

Possible future developments

- Experiments can now **upload their own data**. Need to encourage greater use, including upload of older data.
- **Data format** is already flexible with any number of x and y axes in a table. But could provide better support for correlation matrices, 3-D plots, exclusion contours, etc.
- Also want better integration of **auxiliary files** in DB.
- Strengthen links with **Rivet** (<http://rivet.hepforge.org>). (**Rivet** allows event generators to reproduce details of an experimental analysis.)
- Improve **searchability** by revamping search indices.
- **Public API** to allow interface to HepData in user code?
- Greater integration with **Inspire**, possibly by using **Invenio** framework to display HepData records.