HepData status

Mike Whalley - IPPP Durham UK

DASPOS/DPHEP7 Workshop
21-22nd March 2012
CERN
Brief Introduction to HepData

• **Aim** - to compile published ‘cross section’ data and make them available in a computer database

• **Small group**, based at IPPP, Durham U. (UK) - DBmanager/physicist+non-physicist assistant

• **STFC(UK) funded** - just received funding to October 2016.

• **> 30 years**, began in collaboration with PDG - original DB management system BDMS

• **2009 moved** to more modern and long-term maintainable computing system based on MySQL and Java code - CEDAR
Java-Hibernate
Object-relational mapping library
relational database $\leftrightarrow$ OO model

MySQL
Database

Hibernate

Java OO Model

Apache-Tapestry
open-source component oriented Java web application framework

Aida
(Rivet)

User

plot
etc...

Tuesday, 19 March 2013
HepData - 'standard' record type

**Reaction Database Full Record Display**

**AAD 2011** — Measurement of inclusive jet and dijet production in pp collisions at \( \sqrt{s} = 7 \) TeV using the ATLAS detector

- **Experiment**: CERN-LHC-ATLAS (ATLAS)
- **Preprint as**: CERN-HEP-2011-192
- **Archived as**: ARXIV:1112.6297
- **Record in**: INSPIRE

CERN-LHC. Measurements of the inclusive jet and di-jet cross sections in proton-proton collisions at a centre-of-mass energy of 7 TeV. The analysis is based on the 2010 data sample with a total integrated luminosity of 37 pb⁻¹. The anti-KT algorithm is used for jet clustering with data given using both the parameters R=0.4 and R=0.6. Double-differential cross sections are given as a function of the inclusive jet PT in rapidity (y) bins and of the di-jet invariant mass in bins of half of the rapidity separation of the two jets (y*). Details of the systematic errors and their correlations are given in the link below. In the tables the first (sys) error is the correlated (in PT) systematic error and the second is the uncorrelated systematic error, both produced by combining the relevant errors in quadrature.

Details of systematic errors and their correlations

- View list of currently selected plots
- Total number of tables: 32. Displaying: 1 to 10. First | Previous | Next | Last

---

**Table 1**
as: plain text, AIDA, PyROOT, YODA, ROOT, mpl or jhepwork

<table>
<thead>
<tr>
<th>PT IN GEV</th>
<th>D2(SIG/DPTDYRAP IN PB/GEV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20. – 30.</td>
<td>4700000 ± 0.86% (stat) +20.1% ±20.3% (sys) +1.3% (sys)</td>
</tr>
<tr>
<td>30. – 45.</td>
<td>717000 ± 1.33% (stat) +17.0% ±16.3% (sys) +1.1% (sys)</td>
</tr>
<tr>
<td>45. – 60.</td>
<td>148000 ± 3.03% (stat) +12.8% ±11.5% (sys) +1.1% (sys)</td>
</tr>
<tr>
<td>60. – 80.</td>
<td>38100 ± 1.1% (stat) +10.4% ±10.1% (sys) +1.0% (sys)</td>
</tr>
<tr>
<td>80. – 110.</td>
<td>8520 ± 0.68% (stat) +10.5% ±11.5% (sys) +1.1% (sys)</td>
</tr>
<tr>
<td>110. – 160.</td>
<td>1480 ± 0.62% (stat) +9.9% ±9.3% (sys) +1.1% (sys)</td>
</tr>
</tbody>
</table>

---

**Reaction Database Single Dataset Plot**

**Reaction Database Composite Plot (Advanced)**

**Output formats:**
- html
- plain text
- AIDA - for RIVET
- PyRoot
- YODA
- mpl
- jhepwork
- plot (simple & advanced)

Tuesday, 19 March 2013
SUSY/Exotics ‘non-standard’ record types

At the beginning of 2011 we were asked (by the ATLAS SUSY group) if HepData could handle data sets other than the standard (2-D) ‘cross section’ type data.

Things like:
   - SLHA files
   - Tables of Acceptances & Efficiencies
   - 3-D tables of signal cross sections

We agreed to do this by creating a ‘resource’ area on the main HepData server which was linked to the specific HepData main record.
Addition of ‘resource’ file system on HepData server - linked into the specific data record

Paper

Dataset.....Dataset....Dataset.....

Bins
Points
Errors

xaxis.. yaxis..

Java OO Model

Hibernate

MySQL Database

Tapestry

Resource Area

SLHA
Acc.
Eff.
etc..

linked using InspireId to record

Aida etc......
(Rivet)

Tuesday, 19 March 2013
Beyond the standard record type

Searches for New Physics: Les Houches Recommendations for the Presentation of LHC Results


Recommendation 1a: Provide a clear, explicit description of the analysis in publications. In particular, the most crucial information such as basic object definitions and event selection should be clearly displayed in the publications, preferably in tabular form, and kinematic variables utilised should be unambiguously defined. Further information necessary to reproduce the analysis should be provided, as soon as it becomes available for release, on a suitable common platform.

Recommendation 1b: The community should identify, develop and adopt a common platform to store analysis databases, collecting object definitions, cuts, and all other information, including well-encapsulated functions, necessary to reproduce or use the results of the analyses, and as required by other recommendations.

Recommendation 2a: Provide histograms or functional forms of efficiency maps wherever possible in the auxiliary information, along with precise definitions of the efficiencies, and preferably provide them in standard electronic forms that can easily be interfaced with simulation or analysis software.

We note that it is already common practice in the LHC experiments to provide useful auxiliary information for the longer papers³⁰ e.g., in Rivet [29], on HEPdata [30] and/or collaboration twiki pages [1,2]. The inSPIRE [31] project may help to build a coherent information system, with detailed searchable and citeable entries. The ultimate goal

As mentioned, Rivet and HEPdata provide examples of such a platform, possibly supported by the inSPIRE indexing and searching infrastructure. Their functionality could be adapted to accommodate further needs, emerging from the discussions on the implementation of Recommendation 1b. The continued development of such tools should

These standard electronic forms could rely on a platform similar to that discussed Recommendation 1b, for example Rivet/HEPdata, data and routines.
Hunt for new phenomena using large jet multiplicities and transverse momentum with ATLAS in 4.7 fb^{-1} of sqrt(s) = 7 TeV proton-proton collisions

**Acceptances, efficiencies, errors and CLs values**

<table>
<thead>
<tr>
<th>mSUGRA interpretation</th>
<th>gluino(\rightarrow)t\bar{t}+\chi_0 interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>6/80 signal region</td>
<td>6/80 signal region</td>
</tr>
<tr>
<td>7/55 signal region</td>
<td>7/55 signal region</td>
</tr>
<tr>
<td>7/80 signal region</td>
<td>7/80 signal region</td>
</tr>
<tr>
<td>8/55 signal region</td>
<td>8/55 signal region</td>
</tr>
<tr>
<td>8/80 signal region</td>
<td>8/80 signal region</td>
</tr>
<tr>
<td>9/55 signal region</td>
<td>9/55 signal region</td>
</tr>
</tbody>
</table>

**Distribution of the variable ETmiss/sqrt(ht) for events with \(\geq 7\) jets each having transverse momentum**

<table>
<thead>
<tr>
<th>m12</th>
<th>Cls.exp</th>
<th>Cls.obs</th>
<th>NSig</th>
<th>Acc (%)</th>
<th>Acc x Eff (%)</th>
<th>Exp Unc (%)</th>
<th>NMC Stat Unc (%)</th>
<th>Th Unc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.046</td>
<td>0.042</td>
<td>47.0</td>
<td>0.053</td>
<td>0.053</td>
<td>36.0</td>
<td>39.0</td>
<td>6.2</td>
</tr>
<tr>
<td>210</td>
<td>0.023</td>
<td>0.028</td>
<td>74.0</td>
<td>0.099</td>
<td>0.16</td>
<td>32.0</td>
<td>28.0</td>
<td>9.3</td>
</tr>
<tr>
<td>220</td>
<td>0.24</td>
<td>0.25</td>
<td>35.0</td>
<td>0.099</td>
<td>0.16</td>
<td>32.0</td>
<td>28.0</td>
<td>9.3</td>
</tr>
<tr>
<td>230</td>
<td>0.082</td>
<td>0.083</td>
<td>61.0</td>
<td>0.31</td>
<td>0.11</td>
<td>16.0</td>
<td>14.0</td>
<td>10.0</td>
</tr>
<tr>
<td>240</td>
<td>0.035</td>
<td>0.036</td>
<td>37.0</td>
<td>0.28</td>
<td>0.12</td>
<td>18.0</td>
<td>13.0</td>
<td>12.0</td>
</tr>
<tr>
<td>250</td>
<td>0.017</td>
<td>0.018</td>
<td>37.0</td>
<td>0.072</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>260</td>
<td>0.008</td>
<td>0.009</td>
<td>15.0</td>
<td>0.12</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>270</td>
<td>0.018</td>
<td>0.021</td>
<td>14.0</td>
<td>0.08</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>280</td>
<td>0.022</td>
<td>0.025</td>
<td>13.0</td>
<td>0.05</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>290</td>
<td>0.044</td>
<td>0.051</td>
<td>5.0</td>
<td>0.11</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>300</td>
<td>0.046</td>
<td>0.051</td>
<td>5.0</td>
<td>0.11</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>310</td>
<td>0.026</td>
<td>0.030</td>
<td>4.0</td>
<td>0.11</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>320</td>
<td>0.026</td>
<td>0.030</td>
<td>4.0</td>
<td>0.11</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>330</td>
<td>0.046</td>
<td>0.051</td>
<td>5.0</td>
<td>0.11</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>340</td>
<td>0.026</td>
<td>0.030</td>
<td>4.0</td>
<td>0.11</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>350</td>
<td>0.046</td>
<td>0.051</td>
<td>5.0</td>
<td>0.11</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>360</td>
<td>0.026</td>
<td>0.030</td>
<td>4.0</td>
<td>0.11</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>370</td>
<td>0.046</td>
<td>0.051</td>
<td>5.0</td>
<td>0.11</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
<tr>
<td>380</td>
<td>0.026</td>
<td>0.030</td>
<td>4.0</td>
<td>0.11</td>
<td>0.12</td>
<td>13.0</td>
<td>15.0</td>
<td>12.0</td>
</tr>
</tbody>
</table>

**Note:** The following numbers are at detector level. No unfolding for detector resolution has taken place.
CERN-LHC. Search for SUSY in final states containing one isolated lepton (electron or muon), jets and missing transverse momentum in proton-proton collisions at a centre-of-mass energy of 7 TeV. The data sample, collected during 2010, has a total integrated luminosity of 35 pb⁻¹. No excess above the standard model is found. This record contains the distributions in missing ET, the transverse mass (MT) between the lepton and the missing transverse momentum vector, and the effective mass defined as the scalar sum of the three leading jets, the PT of the lepton and the missing ET. Also tabulated are the 95 PCT exclusion limits on m₀ and mₐ(1/2) for the MSUGRA/CMSSM model.

Link to the tables of MSUGRA/CMSSM SLHA parameters
Link to the combined 0 and 1 lepton analysis

View list of currently selected plots

Table 1
as: plain text, AIDA, PyROOT, YODA, ROOT, mpl or jhepwork

Distribution of ET(\textunderscore C\textunderscore M\textunderscore ISS\textunderscore ING) in GEV for data and background MC calculation.

Location: F 1
Contents of HepData

7882 records (=papers)

191 LHC  ATLAS:95
         CMS: 47
         ALICE: 34
         LHCB: 11
         TOTEM: 3
         LHCF: 1

ATLAS:
  38/57 STDM
  39/44 SUSY
  10/56 EXOT
  5/7 HION
  1/28 TOPQ
  1/8 BPHYS

CMS:
  25/63 STDM
  1/32 SUSY
  2/65 EXOT
  5/20 HION
  1/21 TOPQ
  13/17 BPHYS
HepData - finding the record

HepData has basic search facilities
- EXP, RE, OBS, FSP, REF etc...
+ Inspire type searches (eg. title:xxx)
+ individual records linked from Inspire....

+ working with Elsevier to place banner flag on their web page when paper has HepData record.
Inspire & HepData

There are Inspire<===>HepData links in the records
Plus, now:    ( thanks to Piotr Praczyk - Inspire Group)

* HepData data within and displayable in Inspire
* Inspire search terms in HepData   (eg keyword:supersymmetry)
Inspire & HepData

This data comes from the Durham HEPData project

Summary:

Comments:

Table

<table>
<thead>
<tr>
<th>ET_{3} (GeV)</th>
<th>d(\sigma)/dET_{3}/dETARAP_{3} (P_{3}/GeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.0 − 34.0</td>
<td>123 ± 1 (stat) +15.5%,−14.5% (sys)</td>
</tr>
<tr>
<td>34.0 − 39.0</td>
<td>62.1 ± 0.3 (stat) +10.8%,−9.8% (sys)</td>
</tr>
<tr>
<td>39.0 − 44.0</td>
<td>31.0 ± 0.2 (stat) +9.8%,−8.4% (sys)</td>
</tr>
<tr>
<td>44.0 − 50.0</td>
<td>17.2 ± 0.2 (stat) +10.2%,−8.1% (sys)</td>
</tr>
<tr>
<td>50.0 − 60.0</td>
<td>7.93 ± 0.08 (stat) +10.1%,−8.4% (sys)</td>
</tr>
<tr>
<td>60.0 − 70.0</td>
<td>3.54 ± 0.05 (stat) +9.8%,−8.5% (sys)</td>
</tr>
<tr>
<td>70.0 − 80.0</td>
<td>1.76 ± 0.03 (stat) +10.0%,−9.1% (sys)</td>
</tr>
<tr>
<td>80.0 − 90.0</td>
<td>0.908 ± 0.014 (stat) +9.3%,−7.9% (sys)</td>
</tr>
<tr>
<td>90.0 − 110.0</td>
<td>0.441 ± 0.005 (stat) +8.8%,−8.7% (sys)</td>
</tr>
<tr>
<td>110.0 − 130.0</td>
<td>0.168 ± 0.003 (stat) +8.6%,−8.7% (sys)</td>
</tr>
<tr>
<td>130.0 − 150.0</td>
<td>0.0725 ± 0.0016 (stat) +7.6%,−8.0% (sys)</td>
</tr>
<tr>
<td>150.0 − 170.0</td>
<td>0.0341 ± 0.0008 (stat) +8.8%,−10.0% (sys)</td>
</tr>
<tr>
<td>170.0 − 200.0</td>
<td>0.0146 ± 0.0004 (stat) +8.8%,−9.1% (sys)</td>
</tr>
<tr>
<td>200.0 − 230.0</td>
<td>0.00566 ± 0.00024 (stat) +9.0%,−10.6% (sys)</td>
</tr>
<tr>
<td>230.0 − 300.0</td>
<td>0.00138 ± 0.00008 (stat) +10.0%,−10.7% (sys)</td>
</tr>
<tr>
<td>300.0 − 400.0</td>
<td>0.000149 ± 0.000021 (stat) +15.2%,−13.4% (sys)</td>
</tr>
</tbody>
</table>

ETARAP_{3} \in (-1.0,1.0)

\bar{p} p \rightarrow \text{GAMMA} X

\sqrt{s} = 1960.0 \text{ GeV}
HepData - entering data

At present all data entry is done at Durham by either myself or my assistant.

Data files are sent to us by the experimenters which we convert into the required input format.

Authors validate before data is transferred to the public database.

Before the LHC we had to ask...but now we are getting several request/week to enter data from the LHC experimenters!

In the long term this situation needs to be changed.

Need to get external encoders or experiments themselves to help by inputting direct into data records with Durham providing an overarching management role controlling the actual entry into the public database.
**HepData - entering the data**  
**D.I.Y.**

We have been developing a web entry form  
+ simplified entry language

---

**DATA INPUT FORM**

1. Please specify a file to upload  
   - Choose File  
   - no file selected  
   - Upload

2. Uploading the file 6011.input  
   - Process  
   - the Input Data - then [Display] the Final Record

3. Extracts the ‘bibliographic’ metadata from arXiv to get started

---

**Example Data**

*author: AAD*  
*title: Search for new phenomena in the WW to J nu L prime nu prime final state in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector*  
*comment: CERN-LHC  
*siresid: 89999999*  
*author: AAD*  
*title: Search for new phenomena in the WW to J nu L prime nu prime final state in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector*  
*comment: CERN-LHC  
*siresid: 89999999*

---

Tuesday, 19 March 2013
HepData Usage

As a metric to measure the use of HepData we have continued to count the number of distinct internet nodes, excluding robots etc., accessing per month.

This remained steady from ~500 up to 2009, the start of the LHC, then has increased three-fold to the present.
HepData Summary

Continue to compile cross sections based on publications
>30 years (3+1) years funded by STFC from Oct 2012

2009 moved from BDMS to new maintainable computing system
based on Java OO data model & MySQL database.

Expanded to include ‘resource’ area for LHC extra ‘data’
for eg. LHC acceptances, efficiencies SLHA files

Inspire ↔ Hepdata collaboration greatly increased

HepData ‘data’ in Inspire + Inspire fields searchable from HepData

Record discovery (beyond direct keyword searching) expanded.

URL link to/from Inspire

URL link from LHC experiments publication pages
Banner page URL link from Elsevier publications (Phys.Lett..)

Proposed data input direct from experiments needed for future
trial web based system being used/assessed by outside users

Use figures increase 3-fold since 2009
now ~1500 distinct internet nodes per month.