Integration of HEPData in INSPIRE

Piotr Praczyk CERN, 07.02.2012

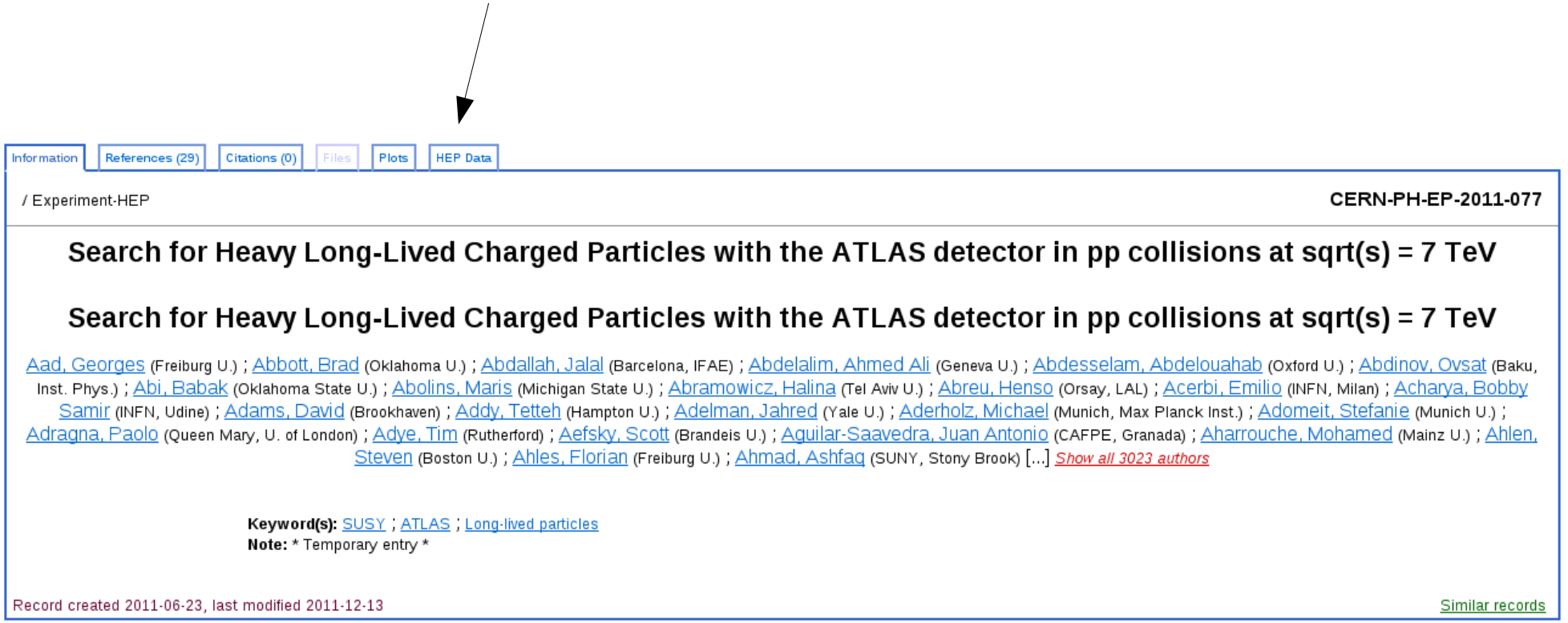
Where are we ?

- HEPData records are harvested and uploaded into a demo site containing complete INSPIRE
- Data is displayed together with records (ugly)

Where do we want to be ?

- HEPData record should be harvested and INSPIRE records updated once a day in the production environment
- Indexes should be establisehd to enable search
- The data should be displayed in a pretty way
- There should be format for displaying standalone HEPData records

An additional HEPData tab is displayed for records having data attached



HEPData entries as separate records

```
001100000 001__ 1100000
001100000 245__ $$9HEPDATA$$a<br >Charged particle MEAN(PTOUT**2) distribution<br >Location: F 9,T 12
001100000 336__ $$tDATASET
001100000 520__ $$9HEPDATA
001100000 6531_ $$c0$$k$$v 
001100000 6531_ $$c1$$kSQRT(S)$$v52 TO 57 GeV
001100000 6531_ $$c0$$k$$v 
001100000 6531_ $$c1$$kSQRT(S)$$v52.0-57.0 GeV
001100000 710__ $$gAMY Collaboration
001100000 786__ $$q0$$rKEK-PREPRINT-89-149$$w283337
001100000 8564_ $$uhttp://inspirehepdev2.cern.ch/record/1100000/files/Data$$ydata extracted from the table
001100000 910__ $$dMEAN(PTOUT**2) IN GEV**2$$n0$$t 
001100000 910__ $$d(1/N) * D(MEAN(NAME=PTOUT**2)) IN GEV**-2$$n1$$tRE : E+ E- --&gt; CHARGED X
001100000 980__ $$aDATA
```

The harvesting process

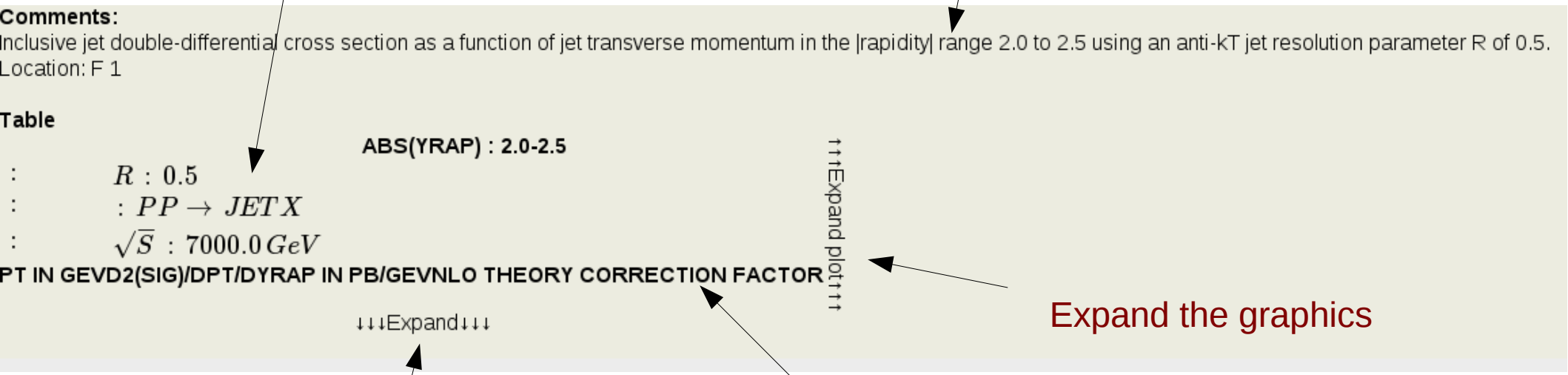
- List of all HEPData records is exposed

- Harvesting script reads the page once every day and checks which records have changed

The backend

The interface

http://inspirehepdev2.cern.ch/record/902309/hepdata



Comments:
Inclusive jet double-differential cross section as a function of jet transverse momentum in the |rapidity| range 2.0 to 2.5 using an anti-kT jet resolution parameter R of 0.5.
Location: F 1

Table

ABS(YRAP) : 2.0-2.5

↑↑↑Expand plot↑↑↑

: $R : 0.5$
: $: PP \to JETX$
: $\sqrt{S} : 7000.0\,GeV$

PT IN GEV	D2(SIG)/DPT/DYRAP IN PB/GEV	NLO THEORY CORRECTION FACTOR

↑↑↑Collapse↑↑↑

19.4 (bin: 18. – 21.)	10600000 ± 0.7% (stat) +20.6%,-17.0% (sys)	1.28 ± 27.9%
22.4 (bin: 21. – 24.)	5550000 ± 1.0% (stat) +18.3%,-15.4% (sys)	1.23 ± 25.4%
25.9 (bin: 24. – 28.)	2770000 ± 1.3% (stat) +16.6%,-14.2% (sys)	1.19 ± 22.1%
29.9 (bin: 28. – 32.)	1410000 ± 1.8% (stat) +15.3%,-13.2% (sys)	1.15 ± 18.8%
34.3 (bin: 32. – 37.)	748000 ± 2.3% (stat) +14.3%,-12.5% (sys)	1.11 ± 15.8%
39.8 (bin: 37. – 43.)	351000 ± 1.3% (stat) +13.5%,-11.9% (sys)	1.09 ± 12.9%
45.8 (bin: 43. – 49.)	169000 ± 1.9% (stat) +13.0%,-11.6% (sys)	1.07 ± 10.6%
52.3 (bin: 49. – 56.)	90100 ± 2.4% (stat) +12.7%,-11.4% (sys)	1.06 ± 8.8%
59.7 (bin: 56. – 64.)	44000 ± 1.0% (stat) +12.6%,-11.3% (sys)	1.05 ± 7.2%
68.6 (bin: 64. – 74.)	21500 ± 1.3% (stat) +12.6%,-11.4% (sys)	1.04 ± 5.9%
78.7 (bin: 74. – 84.)	10300 ± 1.9% (stat) +12.9%,-11.6% (sys)	1.04 ± 4.8%
90.0 (bin: 84. – 97.)	5020 ± 0.6% (stat) +13.2%,-11.8% (sys)	1.03 ± 3.8%
104.8 (bin: 97. – 114.)	2190 ± 0.8% (stat) +13.6%,-12.2% (sys)	1.03 ± 3.0%
122.7 (bin: 114. – 133.)	896 ± 0.4% (stat) +14.3%,-12.8% (sys)	1.03 ± 2.3%
142.2 (bin: 133. – 153.)	384 ± 0.6% (stat) +15.2%,-13.5% (sys)	1.03 ± 1.7%
162.7 (bin: 153. – 174.)	171 ± 0.6% (stat) +16.2%,-14.2% (sys)	1.03 ± 1.3%
184.2 (bin: 174. – 196.)	78.7 ± 0.8% (stat) +17.3%,-15.1% (sys)	1.03 ± 1.0%
207.1 (bin: 196. – 220.)	36 ± 0.8% (stat) +18.7%,-16.1% (sys)	1.03 ± 0.7%
231.6 (bin: 220. – 245.)	16.5 ± 1.1% (stat) +20.5%,-17.4% (sys)	1.03 ± 0.4%
257.5 (bin: 245. – 272.)	7.26 ± 1.2% (stat) +22.5%,-18.9% (sys)	1.03 ± 0.2%
285.0 (bin: 272. – 300.)	3.29 ± 1.7% (stat) +25.0%,-20.6% (sys)	1.03 ± 0.00%
313.9 (bin: 300. – 330.)	1.4 ± 2.5% (stat) +27.9%,-22.5% (sys)	1.03 ± 0.2%
344.8 (bin: 330. – 362.)	0.582 ± 3.7% (stat) +31.3%,-24.7% (sys)	1.03 ± 0.4%
377.3 (bin: 362. – 395.)	0.221 ± 5.9% (stat) +35.8%,-27.4% (sys)	1.03 ± 0.7%
411.2 (bin: 395. – 430.)	0.078 ± 9.4% (stat) +40.5%,-30.0% (sys)	1.02 ± 1.0%
508.1 (bin: 430. – 737.)	0.00555 ± 11.4% (stat) +52.7%,-36.1% (sys)	1.02 ± 1.5%

Graphical display axis

Level of detail axis

Plot summarising all the data

Comments:
Inclusive jet double-differential cross section as a function of jet transverse momentum in the |rapidity| range 2.0 to 2.5 using an anti-kT jet resolution parameter R of 0.5.
Location: F 1

Table

ABS(YRAP) : 2.0-2.5

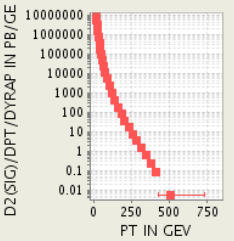
: $R : 0.5$
: $: PP \rightarrow JETX$
: $\sqrt{S} : 7000.0\,GeV$

↑↑↑Collapse plot↑↑↑

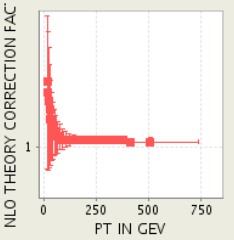
this is the place for a plot

PT IN GEV	D2(SIG)/DPT/DYRAP IN PB/GEV	NLO THEORY CORRECTION FACTOR
	↑↑↑Collapse↑↑↑	
19.4 (bin: 18. – 21.)	10600000 ± 0.7% (stat) +20.6%,-17.0% (sys)	1.28 ± 27.9%
22.4 (bin: 21. – 24.)	5550000 ± 1.0% (stat) +18.3%,-15.4% (sys)	1.23 ± 25.4%
25.9 (bin: 24. – 28.)	2770000 ± 1.3% (stat) +16.6%,-14.2% (sys)	1.19 ± 22.1%
29.9 (bin: 28. – 32.)	1410000 ± 1.8% (stat) +15.3%,-13.2% (sys)	1.15 ± 18.8%
34.3 (bin: 32. – 37.)	748000 ± 2.3% (stat) +14.3%,-12.5% (sys)	1.11 ± 15.8%
39.8 (bin: 37. – 43.)	351000 ± 1.3% (stat) +13.5%,-11.9% (sys)	1.09 ± 12.9%
45.8 (bin: 43. – 49.)	169000 ± 1.9% (stat) +13.0%,-11.6% (sys)	1.07 ± 10.6%
52.3 (bin: 49. – 56.)	90100 ± 2.4% (stat) +12.7%,-11.4% (sys)	1.06 ± 8.8%
59.7 (bin: 56. – 64.)	44000 ± 1.0% (stat) +12.6%,-11.3% (sys)	1.05 ± 7.2%
68.6 (bin: 64. – 74.)	21500 ± 1.3% (stat) +12.6%,-11.4% (sys)	1.04 ± 5.9%
78.7 (bin: 74. – 84.)	10300 ± 1.9% (stat) +12.9%,-11.6% (sys)	1.04 ± 4.8%
90.0 (bin: 84. – 97.)	5020 ± 0.6% (stat) +13.2%,-11.8% (sys)	1.03 ± 3.8%
104.8 (bin: 97. – 114.)	2190 ± 0.8% (stat) +13.6%,-12.2% (sys)	1.03 ± 3.0%
122.7 (bin: 114. – 133.)	896 ± 0.4% (stat) +14.3%,-12.8% (sys)	1.03 ± 2.3%
142.2 (bin: 133. – 153.)	384 ± 0.6% (stat) +15.2%,-13.5% (sys)	1.03 ± 1.7%
162.7 (bin: 153. – 174.)	171 ± 0.6% (stat) +16.2%,-14.2% (sys)	1.03 ± 1.3%
184.2 (bin: 174. – 196.)	78.7 ± 0.8% (stat) +17.3%,-15.1% (sys)	1.03 ± 1.0%
207.1 (bin: 196. – 220.)	36 ± 0.8% (stat) +18.7%,-16.1% (sys)	1.03 ± 0.7%
231.6 (bin: 220. – 245.)	16.5 ± 1.1% (stat) +20.5%,-17.4% (sys)	1.03 ± 0.4%
257.5 (bin: 245. – 272.)	7.26 ± 1.2% (stat) +22.5%,-18.9% (sys)	1.03 ± 0.2%
285.0 (bin: 272. – 300.)	3.29 ± 1.7% (stat) +25.0%,-20.6% (sys)	1.03 ± 0.00%
313.9 (bin: 300. – 330.)	1.4 ± 2.5% (stat) +27.9%,-22.5% (sys)	1.03 ± 0.2%
344.8 (bin: 330. – 362.)	0.582 ± 3.7% (stat) +31.3%,-24.7% (sys)	1.03 ± 0.4%
377.3 (bin: 362. – 395.)	0.221 ± 5.9% (stat) +35.8%,-27.4% (sys)	1.03 ± 0.7%
411.2 (bin: 395. – 430.)	0.078 ± 9.4% (stat) +40.5%,-30.0% (sys)	1.02 ± 1.0%
508.1 (bin: 430. – 737.)	0.00555 ± 11.4% (stat) +52.7%,-36.1% (sys)	1.02 ± 1.5%

ABS(YRAP) : 2.0-2.5



ABS(YRAP) : 2.0-2.5



Plots of separate data columns

Improved display of data headers

RE : P P --> SLEPTON X

SQRT(S) : 7000.0 GeV

$: PP \rightarrow SLEPTON\,X$

$\sqrt{S}\ : 7000.0\,GeV$

· The structure of header is parsed and translated into LaTeX

· Symbols are replaced using a dictionary (that should be maintained on the HEPData side and harvested together with data)

· Javascript plugin renders formulas in the browser

TODO

- Establish way of harvesting the dictionaries of symbols (for improved data headers display)
- Invent nicer way of breaking long lines of headers/titles
- Improve the general look&feel
- Create BibFormat for standalone HEPData records
- Establish indexes for HEPData search
- Invent way to display HEPData entries not connected to publications (virtual records?)