



Feed back on HEPDATA usage from CMS

M. Gouzevitch
on behalf of CMS collaboration
and
HEPDATA contact people in CMS

OUTLOOK

Feed back from ~3 years experience with HEPDATA in CMS.

- 1) Statistics
- 2) Format conversion
- 3) Fields usage
- 4) Content
- 5) Additionnal material
- 6) Administrator wished
- 7) Socialogy

1) Statistics



42 entries

In 2014 CMS have shown a clear support for HEPADA project. The submission became part of the publication process and service points was provided to fill some old entries.

1) Statistics

Access frequency

		2010	2011	2012	2013	2014	2015	total	arXiv	title
845323	CMS	8	45	137	47	42	45	324	1002.0621	Transverse momentum and pseudorapidity d
855299	CMS	93	67	104	50	57	103	474	1005.3299	Transverse-momentum and pseudorapidity d
878118	CMS	0	12	184	68	70	29	363	1011.4193	(NOV 2010)
879315	CMS	0	152	574	270	211	253	1460	1011.5531	Charged particle multiplicities in pp in
882871	CMS	0	42	181	10	26	27	286	1012.5545	Measurement of the Inclusive Upsilon pro
885663	CMS	0	4	114	31	31	33	213	1101.5029	Dijet Azimuthal Decorrelations in pp Col
889175	CMS	0	78	182	39	21	11	331	1102.2020	Measurement of Dijet Angular Distributio
889807	CMS	0	8	69	15	13	51	156	1102.3194	Measurement of B anti-B Angular Correlat
890166	CMS	0	8	229	23	91	62	413	1102.4828	Strange Particle Production in pp Collis
891482	CMS	0	1	598	7	69	2	677		Search for Supersymmetry in \$pp\$ Collisi
896764	CMS	0	3	193	92	104	91	483	1104.3547	Charged particle transverse momentum spe
902309	CMS	0	0	298	147	51	71	567	1106.0208	Measurement of the Inclusive Jet Cross S
916908	CMS	0	37	125	34	29	29	254	1107.0330	Measurement of the Underlying Event Acti
921788	CMS	0	0	102	61	29	42	234	1108.0566	Measurement of the Drell-Yan Cross Secti
941555	CMS	0	0	88	55	118	41	302	1110.4973	Measurement of the Rapidity and Transver
944755	CMS	0	0	468	370	171	50	1059	1111.1557	\$J/\psi\$ and \$\psi_{2S}\$ production in \$
1088823	CMS	0	0	76	139	82	36	333	1202.2554	Study of high-pT charged particle suppre
1089835	CMS	0	0	529	71	38	55	693	1202.4617	Inclusive b-jet production in pp collisi
1107659	CMS	0	0	473	136	250	9	868	1204.1409	Measurement of the elliptic anisotropy o
1122847	CMS	0	0	0	0	237	48	285	1207.3973	Forward-backward asymmetry of Drell-Yan
1123117	CMS	0	0	443	429	347	563	1782	1207.4724	Study of the inclusive production of cha

1) Statistics

	2010	2011	2012	2013	2014	2015	total	arXiv	title
1185414	CMS	0	0	0	107	105	13	225	1209.2922 Measurement of the Y1S, Y2S and Y3S pola
1208923	CMS	0	0	0	267	341	304	912	1212.6660 Measurements of differential jet cross s
1224539	CMS	0	0	0	251	97	1	349	1303.4811 Studies of jet mass in dijet and W/Z+jet
1225274	CMS	0	0	0	174	118	140	432	1303.5900 Measurement of the Y(1S), Y(2S), and Y(3
1230937	CMS	0	0	0	36	177	28	241	1304.7498 Measurement of the ratio of the inclusiv
1242440	CMS	0	0	0	0	91	154	245	1307.3442 Study of the production of charged pions
1253367	CMS	0	0	0	72	206	24	302	1309.2030 Searches for anomalous ttbar production
1256938	CMS	0	0	0	0	373	18	391	1310.1138 Measurement of associated W + charm prod
1262319	CMS	0	0	0	35	1084	296	1415	1310.7291 DY
1263706	CMS	0	0	0	0	0	320	320	1502.05387 Nuclear effects on the transverse moment
1340084		0	0	0	0	0	260	260	1501.04198 Dijet bump hunt
1345023	CMS	0	0	0	0	0	227	227	1502.04155 Measurement of the prompt J/psi and psi(
1345354	CMS	0	0	0	0	0	264	264	1502.05664 Measurement of the Zgamma production cro
1347386	CMS	0	0	0	0	0	234	234	1503.01692 Evidence for transverse momentum and pse
TOTALS:		102	504	6194	3530	6110	6352		133 papers

- This table is courtesy from Mike Walley and Graeme Watt
- Just to not annoy you with everything I show only entries with more than 200 accesses (bias toward older entries).
- Soft QCD, Heavy-Ion, Heavy-Flavors, PDF and BSM communities are main users.

2) Format conversions

1) Yoda → HEPDATA

- This is useful for analyses optimized to MC tuning. They use internally RIVET to compare their results to Gen level MC. So sometimes it is the easiest to get HEPDATA from YODA.
- RIVET contains only $ERROR = STAT + SYS$. So this approach loses information but sometimes it is the only solution.

2) ROOT → HEPDATA

- This is the most common case.
- Most of time this is done privately tweaking one of the scripts kindly provided by <http://hepdata.cedar.ac.uk/submittingdata>
- But it shall be better automatized : kind of online tool where authors provides for example TH1F or TGraphErrors in root format and HEPDATA pre-shape them in text format.

3) Latex → HEPDATA

- A rather common case is to have a latex table in the paper that one shall convert to HEPDATA.
- Of course latex tables can be quite different. So this is not obvious.

3) Fields usage

*status: Checked 11 NOV 2014 by Aleph Bet (CMS)

*status: Encoded 08 NOV 2014 by Alpha Beta (CMS)

1) Many “encoders” do not realize that they shall put their full name. It is always a point I had to fix. May be good if it could be somehow automatised.

3) Fields usage

*obskey: D2SIG/DCOSTHETA/DPHI ; SIG/DET.

- 2) This is a very obscure field. I understand it is mainly historical. People tend to fill it randomly or copy paste from previous examples.
- Either we need a list of examples and a rule or it is better to remove.

reackey: P P --> GAMMA X

*reackey: P P --> Z0 X

*reackey: P P --> MU+ MU- GAMMA X

*reackey: P P --> JETS

*reackey: P P --> JET JET

*reackey: P P --> JET X

- 3) This is more clear what it serve to, but it is not obvious to find the right way to write it. Need more documentation, list of examples or rules.

4) CONTENT

1) HEPDATA normally contains “DATA + Uncertainties” :) About this we agree.

2) What about:

- MC or theory predictions for comparison?
- Limits from searches? Observed? Expected? ± 1 or 2 sigma?
- Background estimations?

5) Additional material

1) Number of files :

- it is limited to 4. Having more would be better.
- The solution of providing a *.tar.gz overcomes this problem but it is practical for very large number of files. For 5 files it may be better to be able to upload 5 files.

2) Format for histograms :

- Sometimes it is useful to provide histograms that are not strictly speaking measurements. Signal shapes for a search for example.
How shall we do it ?
- ROOT format ? YODA format ?

6) Administrator wishes

- 1) To save my time I would like to access the unfinished entries directly from administrator screen. Just by clicking on them. In 50% of cases I just need add 1-2 lines like "checked by Maxime Gouzevitch" or change a comma. This is much faster than to request authors to do it. So I have to come back to my mails, find right one, click etc... It is not so long by summed together those operations takes time.
- 2) It might be also useful to be able to change the flag from READY to In Progress from my radar, when the entry needs seriou work.
- 3) It might be also useful to be able to leave some comments for myself in front of each entry. Like mini: minutes, to know what have to be followed up.

=====

- Proposed to Graeme in May 2015.
- Answer : These are all good suggestions, but not so easy to implement.
- What about now ?

7) SOCIOLOGY

1) HEPDATA is now a privileged hub to access / implement data produced by the LHC. It is well supported by the management and resources are allocated within the experiments.

2) The HEPDATA entry is created after “Acceptance” by the journal of a paper.

- Sometimes it is considered as a “burden” by the authors. Less now than before since it is part of normal “approval process” now.
- Sometimes student who did the analysis are not there anymore.
- Most of the time the author who implement HEPDATA does it once in their life.
 - Easier it is better it is : more entries would be created.
 - More effort required from administrators : 2-4 iterations.

3) Support :

- Graeme is very responsive and helpful. Many thanks to him.

OUTLOOK

- CMS realize now that HEPDATA is one of the most practical way to spread out our data. It also increase their popularity.
- HEPDATA submission is already quite easy. Still some technical improvements are necessary to make HEPDATA submission part of a standard flow.
- We shall share experience with ATLAS colleagues who have a nice expertise and are offer ahead of us in the field. But we are also frequently “cross users” of HEPDATA entries. So it is our common interest.

PS: many people complained about the search tool on HEPDATA. It is not very practical and sometimes it is easier to find a paper on arXiv and then from this paper to find the HEPDATA entry.