## HEPData experience from CEDAR+OpenMAPP

Andy Buckley, University of Glasgow

HEPData Users Committee 15 January 2025



## **HEPData for CEDAR/Rivet**

 ★ HEPData is the main data-source for Rivet analysis — the core tool in the MCnet CEDAR set ⇒ MC dev, tuning, etc.

⇒ General experience very positive: HEPData is essential!

- Major issue has been ~self-inflicted: years of experiments/individuals submitting different data files to Rivet and HD. Graeme has been very helpful with fixes.
- Other historic difficulties with e.g. "zero-width bins" used for inclusive cross-sections, or integer bins for multiplicities. Handled better by YODA2 histogramming, now integrated with the HD web
- ☆ Rivet also contributing to HD: several hundred analyses were only in Rivet ⇒ sent to HD for bulk inclusion (with warning), hopefully to go live soon





## **HEPData for BSM reinterpretation**

- HEPData is also an important source for BSM model \* interpretations against published data. Very pleasing to have search-data also in HD!
- Also now pushing (particularly from Contur) to store experiments' \* theory estimates, especially super-expensive precision SM backgrounds: very valuable for future-proofing
- \* Statistical models: pyhf, Spey, HS3, ONNX: reinterpretation increasingly needs to use more diverse data types than "just histograms". OpenMAPP to help with semantic awareness of aux-file meanings, and add REST querying of resources
- \* **Metadata and linking/discoverability:** combining reint frameworks requires identifying coverage and commonalities in dbs and tools  $\Rightarrow$ JSON format for Rivet-HD linking being generalised
- DOI minting: Graeme advising on mechanisms for analysis DOIs, \* discoverability and citeability

SQRT(S)(Q = 13 TEV)	13000.0 GeV		visualize
РТ_ТОРН	DSIG/DPT_TOPH	DSIG/DPT_TOPH (breakdown of systematics)	0.55
0 - 15	0.0406332907141 see, rgs, ent	0.0406331907141 ave.non. mt 4.100 outer-rpr 4.000 outer-rpr 4.000 outer-rpr	0.35-
15 - 30	0.119051626697 2+2.474% rox 42.00% on \$2.89% rox.beninesity securitarity	0.119051620097 p+2.4346 nm 44895 milescorps	0.28-
30 - 45	0.199555433894 av1.949% est 2000 est	0.199595433894 p+19496 dat 4500 (million PDF d. New, anti-heritati model + 57 more errors (2000 e1	0.39-
45 - 60	0.2623466657017 Int.682% And 1100% And \$2.82% syn,hemisenity supervisionly	0.262346667017 (*1.663% nor 1.552% (*1.562%) 2007 (*1.562%) *17 (507 (*1.55 %))	a záo záo záo záo záo cáo ráo sáo záo záo záo záo záo záo záo záo záo z
60 - 75	0.310152420733 #41882% ent 110000 ent	0.310152420733 ++182% our 2000 spinorer	Sum errors 🖬 Log Scale (X) 🗌 Log Scale (Y) 🗌
75 - 90	0.328356444443 (*1.449% me (*8.58% m) 53.89% (m).minority merching	0.325956444465 p+Lease and annex solution-pp alares solution and stress core and solution	Dasalart variablas er hitle rifferent annr hars hv rörlens
90 - 105	0.347746054803 ±1.383% end 225% ve EL89% en.Juniceds constaints	0.347740064500 2*1.353% star 2.58% wy.inter-Par 148% wy.inter-Par 148% wy.inter-Par	Variables
105 - 120	0.330544907341 av1.872% end 10.00% eve 12.85% exclusionity scortainty	0.330644907341 s+1.372% and ALBEN sectors for	DSIG/DPT_TOPH Summed error
120 - 135	0.282109219545 s+L455% and "\$100% av	0.282109219545 prilation on Aller university	DSIG/DPT_TOPH (breakdown of systematics) Summed error
135 - 150	0.2311646045 #+1.819% was there up and a second sec	0.2311646045 selation and 2000 system for	
150 - 165	0.189267423695 p+1.715% and \$4.68% av \$2.89% syn, landersky successing	0.189207423986 2*1715% INC 2440% INCIDENT	
165 - 180	0.151896835452 ±1.656% etc. 2000 ers 12.65% ep.).eniceth.constitute	0.151896835452 2*64666 444 2100 (10140747 6100 (1014074) (101707775) (101707	HAR
	Ő		
			-IS <sup>3</sup>
		<b>.</b>	and the second s

PT TOPH

## HEPData tasks in OpenMAPP work-plan

- OpenMAPP is a CHIST-ERA international project to build "e-infrastructure" for meta-analysis in particle physics
- HEPData is a key part!
- IPPP and Glasgow to lead on HEPData extensions and integrations with reinterpretation frameworks (WP1):

Tasks T1.1 Extension of HEPData functionalities (M1–M6: responsible: 7; involved: 4,5,7) Add backend code to the HEPData system to store, retrieve, and be able to query data types beyond primary data-tables and uncertainties. Build a HEPData mechanism for linking between data objects via internal DOIs. e.g. for associating theory predictions to measurement tables, or reference simulated event samples to their validation datasets. Add a REST guerying mechanism to HEPData, allowing interrogation of what resources of which types are associated with an analysis, and allowing each to be retrieved by a distinct URL. Provide a programmatic interface to the REST API via the hepdata cli Python package. The data types concerned by T1.1 include statistical models in JSON format, serialised machine-learning models (e.g., ONNX), binned and symbolic detector-response functions, combined sets of event-generator steering files for reference configurations and their corresponding reference outputs, etc. To be carried out by the UK partners, coordinated by partner 7 (Graeme Watt) T1.2 Interfacing to HEPData (M1–M12: responsible: 4; involved: 1,2,4,5,6,7) Build interfaces to HEPData for the various reinterpretation frameworks to retrieve analysis results and associated results data and auxiliary data. e.g. statistical models, ML models, etc. (see list of data types in T1.1) Upgrade the YODA data type exported by HEPData to use the next generation YODA API (with capabilities for representing high-dimensional data objects) and its new HDF5-based data format. Use this machinery to populate / update data contents of all the analysis frameworks. Partners involvement 1: SModelS; 2: MadAnalysis; 4,8: ADL; 5,7: Rivet/Contur; coordinated by partner 4 (Andy Buckley)









