HEPData experience from CEDAR+OpenMAPP

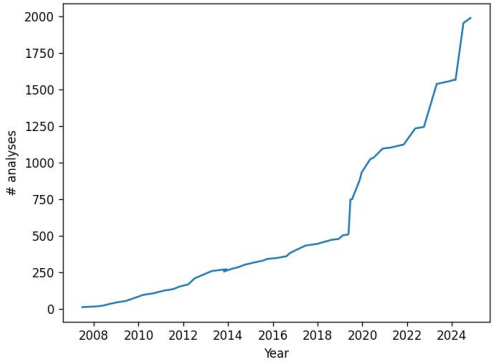
Andy Buckley,
University of Glasgow

HEPData Users Committee
15 January 2025

University of Glasgow

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 ⇒ JSON format for Rivet-HD linking being generalised
- **DOI minting:** Graeme advising on mechanisms for analysis DOIs, discoverability and citeability

HS³

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 querying 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)

HS3

pyhf
differentiable
likelihoods