

RIVET, HEPDATA and analysis preservation

Andy Buckley

University of Glasgow

The New HEPDATA,
CERN, 25 April 2016



THE ROYAL
SOCIETY



University
of Glasgow



RIVET: the MC analysis preservation framework

aka “Robust Independent Validation of Experiment & Theory”

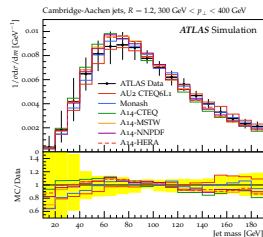
- ▶ Started 2006/7, following HERA HZTOOL experience: *lessons learned*
- ▶ **Aims:** collection of analysis tools and routines for collider data reproduction, in a generator-independent form
- ▶ **Implementation:** C++ (now C++11) with Python steering and scripts. Usable as library or programs
- ▶ **Distinctive features:** reconstruct from final state, automated calculation caching, “pluggability”
- ▶ **Emphasis** so far on “gold standard” of *unfolded data* – **not typical for BSM**
- ▶ **Users:** experiments for preservation, MC simulation developers, MC/data tuning efforts, analysis recasters...



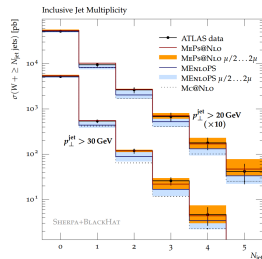
RIVET: the MC analysis preservation framework

aka “Robust Independent Validation of Experiment & Theory”

- ▶ Started 2006/7, following HERA HZTOOL experience: *lessons learned*
- ▶ **Aims:** collection of analysis tools and routines for collider data reproduction, in a generator-independent form
- ▶ **Implementation:** C++ (now C++11) with Python steering and scripts. Usable as library or programs
- ▶ **Distinctive features:** reconstruct from final state, automated calculation caching, “pluggability”
- ▶ **Emphasis** so far on “gold standard” of *unfolded data* – not typical for BSM
- ▶ **Users:** experiments for preservation, MC simulation developers, MC/data tuning efforts, analysis recasters...



ATLAS analysis encoded in RIVET

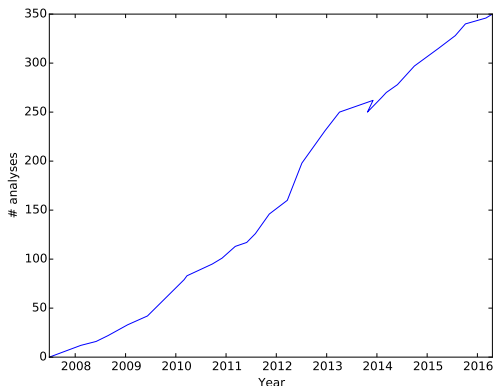


Sherpa MC development via RIVET

Analysis coverage

> **350 analyses!** Nearly 200 LHC – mainly ATLAS/CMS. Strong experiment communities and supported implementation efforts.

- ▶ Slow-up recently as LHC Run 1 SM analyses finish off / Run 2 SM not yet available
- ▶ But welcome new growth in top and Higgs measurements as unfolded differentials start to arrive



NB. glitch is Rivet 1.x → 2.x migration!

Nearly all SM so far: $\sim 320/350 \dots$

but **v2.5 introduces detector smearing machinery**

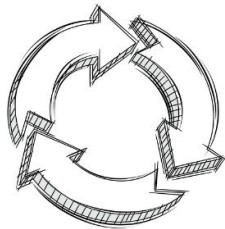
(beta release available for testing now)

HEPDATA integration

- ▶ **Links to HEPDATA unsurprising – we wrote HDv2 to support RIVET!** Certainly intend to continue that close relationship
- ▶ Multi-format output developed for ease of data lookup and bundling in RIVET: can't require a network connection for MC analysis!
- ▶ Links to/from HEPDATA via metadata system: can improve. Stable permalinks?
- ▶ Common in experiments for HEPDATA submission and RIVET submission to be coupled: similar scripts for ROOT conversion to both formats (and $\text{T}_\text{E}\text{X}$ tables, but different emphasis there)
- ▶ Analysis implementations and usage \Rightarrow fixes and data bug reports to HD
- ▶ `rivet-mkanalysis` “wizard” script uses HEPDATA and INSPIRE data to populate analysis metadata. . . **but currently screen-scrapey**
- ▶ **Need more regular bundled data updates and consistency checks – integrate with continuous validation & automatically couple Web page statuses**
- ▶ **Analysis distribution & maintenance via HEPDATA?!**

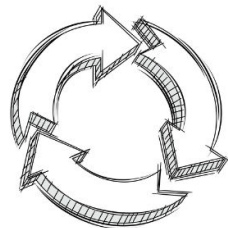
Development opportunities

- ▶ Extend established LHC SM preservation efforts to include BSM – cf. new smearing machinery
- ▶ Break strict core/analysis collection tie
⇒ distribute analyses separately
 - **BUT!**
 - Not trivial: still need some way to report/check version dependence, maintain/update core analyses, version code, ...
 - Unified publ/data/analysis record upload?
Classify “official” and “unofficial” implementations?
 - *Happy to discuss possible mechanisms!*



Development opportunities (2)

- ▶ Improved statistical interpretation, using deeper correlation info from HEPDATA— even inter-expt correlations? ⇒ Higher-level info.
- ▶ Interaction with CERN / LHC analysis data preservation? e.g. running via “MC view” of the data?
- ▶ Linked community databases for data, metadata and analysis preservation have endless potential



“Build it and they will come” – Field of Dreams

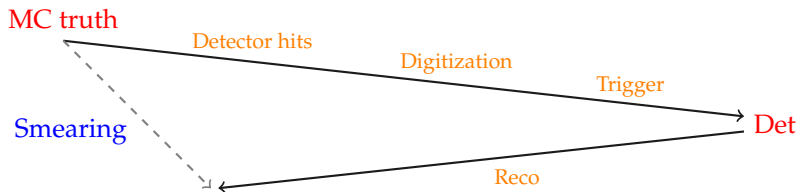
Summary & outlook

- ▶ **RIVET is the pre-eminent SM analysis preservation framework at the LHC** – and is flexible enough to apply far beyond
- ▶ HEPDATA integration since the beginning: common origins in CEDAR e-science programme. Data representation has marks in RIVET API e.g. `bookHisto` fns
- ▶ 10 years of development experience and iteration, with many lessons learned
- ▶ Now feels like the **right time to consciously expand to similar BSM coverage**, with e.g. *detector machinery*: combination with HEPDATA ⇒ **killer app for phenomenologists**
- ▶ But many potential benefits in closer integration for metadata, higher-level statistics, and maybe distribution
- ▶ **Great to see HEPDATA's success and new tech: looking forward to extended data types, and a richer ecosystem for LHC data reuse!**

Extras

A bit more on BSM & detector effects

A problem with explicit detector simulation for pheno:

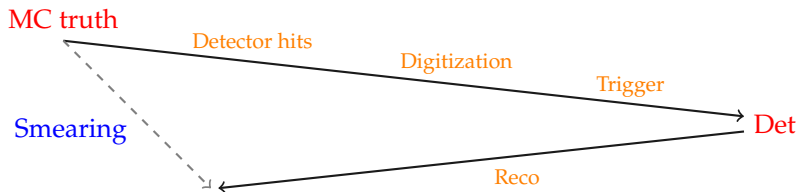


Reco already reverses most detector effects! Residual differences w.r.t. MC truth typically just a few percent, or trivially modellable via sampling of eff fns and Gaussian resolutions.

Efficiency & smearing functions are a very practical balance of accuracy vs. CPU vs. flexibility.

A bit more on BSM & detector effects

A problem with explicit detector simulation for pheno:



Example Rivet smearing interface:

```
FastJets fj(FastJets::ANTIKT, 0.6, JetAlg::ALLMUONS);
SmearedException sj(fj, JET_SMEAR_ATLAS_RUN2, JET_EFF_BTAG_ATLAS_RUN2);
addProjection(sj, "Jets");
...
Jets jets = applyProjection<JetAlg>(event, "Jets").jetsByPt(30*GeV);

// + similar for SmearedException
```