HEPData Sync Status

Christian Gütschow

Lunga workshop
Craobh Haven

28 May 2019

Overview

```
            Of 416 Rivet analyses in Rivet-2.7.2,
                158 (38.0%) were compatible
            and 258 (62.0%) were incompatible.
```

➡ compare this with June 2018:

```
            Of 359 Rivet analyses in Rivet-2.6.0,
                66 (18.4%) were compatible
            and 293 (81.6%) were incompatible.
```

➡ imcompatible ones composed of 53 % LHC
(23 % ATLAS, 22 % CMS, 4 % ALICE, 2 % LHCb)
and 15 % Tevatron (9 % CDF and 6 % D0)

Common trivial issues

➡ different numerical precision

    ➡ paper quotes a cross-section of (0.350 ± 0.010) pb and this value goes on HEPData, while Rivet reference data file was prepared separately using full floating point precision

    ➡ action: fully sync against whatever is on HEPData

Common trivial issues

→ different numerical precision

  → paper quotes a cross-section of (0.350 ± 0.010) pb and this value goes on HEPData, while Rivet reference data file was prepared separately using full floating point precision

  → action: fully sync against whatever is on HEPData

→ additional tables on HEPData missing from Rivet reference data file

  → routine does not necessarily need all HEPData tables, so tables have been ommitted to save space

  → action: fully sync against whatever is on HEPData

  → future sync option: check which histograms are actually booked and only keep those?

Common trivial issues

→ different numerical precision

  → paper quotes a cross-section of (0.350 ± 0.010) pb and this value goes on HEPData, while Rivet reference data file was prepared separately using full floating point precision

  → action: fully sync against whatever is on HEPData

→ additional tables on HEPData missing from Rivet reference data file

  → routine does not necessarily need all HEPData tables, so tables have been ommitted to save space

  → action: fully sync against whatever is on HEPData

  → future sync option: check which histograms are actually booked and only keep those?

→ histograms have different names in HEPData vs Rivet ref data

  → semi-trivial: need to change cc and plot files as part of the sync

  → action: go through them one by one and sync against HEPData

Common non-trivial issues

→ common action: skip and discuss at Lunga

Common non-trivial issues

➜ common action: skip and discuss at Lunga

➜ about a third of the mismatches are due to lack of support for discrete binning

   ➜ e.g. integrated cross-sections or multiplicities are written out as `Scatter2D`s with zero bin width

   ➜ Rivet hist booking fails when zero bin widths are involved, hence dummy bin width manually put into Rivet reference data file to avoid crash

   ➜ possible action: provide discreet histogram class? need flag for individual HEPData tables?

   ➜ possible action: overload hist booking?

   ➜ perhaps book `Counter` if only one point, otherwise assign half distance between points as dummy uncertainty?

   ➜ what about data? plotting scripts cannot handle a point with ± 0 uncertainty in one dimension

Common non-trivial issues

➜ occasionally corrections applied to Rivet reference data (10 %ish of mismatches)

Common non-trivial issues

→ occasionally corrections applied to Rivet reference data (10 %ish of mismatches)

→ updated luminosity calibration, e.g.
```
"Note added (24 JUN 2014):
The cross section values reported in Tables X, Y, Z below should be multiplied
by a factor of 1.0187 to take into account the updated value of the integrated
luminosity for the ATLAS 2010 data taking period.  The uncertainty on the
global normalisation ("Lumi") increases slightly from 3.4% to 3.5%.
See Eur.Phys.J. C73 (2013) 2518 for more details."
```

→ proposal: keep original value and just sync against HEPData

Common non-trivial issues

➡ occasionally corrections applied to Rivet reference data (10 %ish of mismatches)

➡ updated luminosity calibration, e.g.
```
"Note added (24 JUN 2014):
The cross section values reported in Tables X, Y, Z below should be multiplied
by a factor of 1.0187 to take into account the updated value of the integrated
luminosity for the ATLAS 2010 data taking period.  The uncertainty on the
global normalisation ("Lumi") increases slightly from 3.4% to 3.5%.
See Eur.Phys.J. C73 (2013) 2518 for more details."
```

   ➡ proposal: keep original value and just sync against HEPData

➡ model-/generator-dependent extrapolations,
e.g. only published 'Born'-level cross-sections
and separate tables for 'Born'-to-dressed level predictions,
or sometimes even parton-to-particle-level correction factors

   ➡ option: use `finalize` to post-process reference data and write out to YODA file?

   ➡ option: extend plot file capabilities to allow on-the-fly differential ref data scaling?

Common non-trivial issues

➔ occasionally corrections applied to Rivet reference data (10 %ish of mismatches)

➔ updated luminosity calibration, e.g.
```
"Note added (24 JUN 2014):
The cross section values reported in Tables X, Y, Z below should be multiplied
by a factor of 1.0187 to take into account the updated value of the integrated
luminosity for the ATLAS 2010 data taking period.  The uncertainty on the
global normalisation ("Lumi") increases slightly from 3.4% to 3.5%.
See Eur.Phys.J. C73 (2013) 2518 for more details."
```

  ➔ proposal: keep original value and just sync against HEPData

➔ model-/generator-dependent extrapolations,
e.g. only published 'Born'-level cross-sections
and separate tables for 'Born'-to-dressed level predictions,
or sometimes even parton-to-particle-level correction factors

  ➔ option: use `finalize` to post-process reference data and write out to YODA file?

  ➔ option: extend plot file capabilities to allow on-the-fly differential ref data scaling?

➔ tables missing on HEPData

  ➔ first action: get in touch with analysis team

  ➔ failing that . . . upload them ourselves? remove from Rivet?