

Data pipelines discussion

Andy Buckley
University of Glasgow

BSM Reinterpretation Forum
Durham, 30 August 2023



University
of Glasgow

Data pipelines: theory, metadata, automation

Possible topics for discussion:

❖ Database interoperability

- linking between dbs: HepData, Inspire, Rivet/MadAnalysis/GAMBIT/SModelS/CheckMATE/... Zenodo?
- via DOIs, or simpler JSON-based index URLs like Rivet/HD? (How does Inspire/HD work?)
- do experiments/CERN have any plans to operate outside the community dbs?

❖ Automation, discoverability, semantics

- REST querying: ability to retrieve dataset and aux-file lists, links, direct-access URLs via HTTP
- conventions: standard variable naming in likelihood models? systematics naming/versioning?
- to what extent do the HEP data repositories need to *understand* their (aux) contents?

❖ Theory predictions (particularly precision-SM, for BSM injection)

- linking between datasets now possible: how to structure multiple possible theory predictions for each dataset? only the original experiment MC lines in the experiment analysis record?
 - a free-for-all of community predictions with different gens, orders, systematics on expt records?
 - or as aux/HepData datasets for theory papers?
- breakdowns of different backgrounds, signals, $S + \sum_i B_i$ sum, ...
- Need good meta-metadata: informative title, summary, run cards, versioning, date, personnel, ...