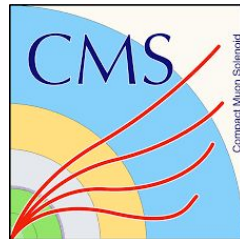


Preservation of SM analyses in CMS

Clemens Lange, Markus Seidel
PSI, Riga Technical University



Main tool: Rivet

- Preserves complete particle-level analyses in form of code
 - No text→physics ambiguities anymore for fiducial regions
... because our papers do not always contain all the necessary information :(
- Allows other scientists to compare new MC predictions to our data
 - MC tuning (Professor), new physics limits (Contur)
 - Not much additional work for original authors but additional citations!
- Already >1000 analyses from LEP, LHC and other experiments integrated
- Provides a lot of easy-to-use analysis “projections”, for example
 - DressedLeptons: leptons clustered with surrounding FSR photons
 - FastJets with ghost bottom/charm/tau tagging automatically included
 - Z/WFinder, InvariantMassFinalStates, etc., even PartonicTops (though against philosophy)
 - No need to browse through event record yourself

Rivet integration to CMSSW

- *RivetInterface* available to run Rivet on CMS datasets
- Problem: lacking connection between Rivet and original analyzers' code
 - Often had small differences between objects in Rivet and CMS analysis
 - Several CMS Rivet analyses with *SpecialDressedLeptons* classes
- *ParticleLevelProducer*: Rivet algos within analysis chain [[github](#), [CMS-NOTE-2017-004](#)]
- Runs on CMS (Mini)Aod format and adds Rivet-compatible output collections
- Output also integrated in flat NanoAod format

Output objects

```
<reco::GenParticleCollection>("neutrinos");  
<reco::GenParticleCollection>("photons");  
<reco::GenJetCollection>("leptons");    --> Charged leptons dressed with surrounding photons  
<reco::GenJetCollection>("jets");      --> jets with flavor information  
<reco::GenJetCollection>("fatjets");   --> fat jets with flavor information  
<reco::GenParticleCollection>("consts"); --> jet and dressed lepton constituents  
<reco::GenParticleCollection>("tags");  --> tag particles (charm/bottom hadrons, taus)  
<reco::METCollection>("mets");        --> MET in fiducial volume abs(eta)<5
```

Important default settings (see below for configuration options):

- Hadron decay products are ignored for finding dressed leptons, photons, and neutrinos.
- Neutrinos dressed leptons are excluded from jet clustering.

NANOAOD

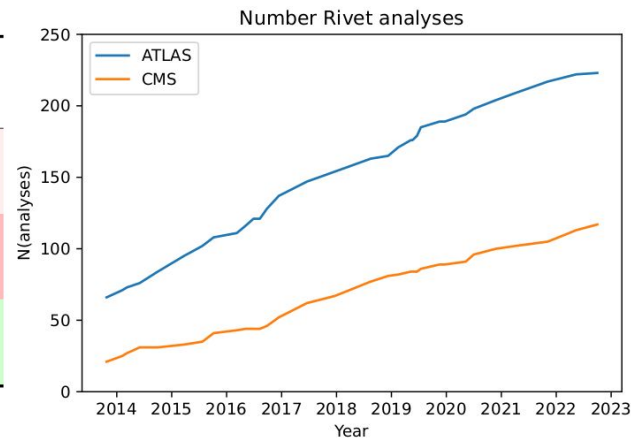
Rivet-compatible objects are integrated in NANOAOD as well. Please use the following collections:

```
GenJet (with hadronFlavour)  
GenJetAK8 (with hadronFlavour)  
GenDressedLepton  
MET_fiducialGenPt/Phi (abs(eta)<5, compatible with detector volume)  
GenMET (MET in full phase space)
```

CMS Rivet coverage

Key	ALICE	ATLAS	CMS	LHCb	Forward	HERA	e ⁺ e ⁻ (≥ 12 GeV)	e ⁺ e ⁻ (≤ 12 GeV)	Tevatron	RHIC	SPS	Other
Rivet wanted (total):	290	337	465	180	17	476	703	513	1118	477	56	3
Rivet REALLY wanted:	36	39	92	15	0	15	1	1	9	2	5	1
Rivet provided:	29/319 = 9%	190/527 = 36%	104/569 = 18%	17/197 = 9%	8/25 = 32%	34/510 = 7%	192/895 = 21%	347/860 = 40%	58/1176 = 5%	8/485 = 2%	4/60 = 7%	131/134 = 98%

[\[Link to full version\]](#)



Key	TOP	SMP	HIG	SUS	HIN	EXO	BPH	B2G	UNKNOWN
Rivet wanted (total):	63	84	65	54	86	89	14	33	10
Rivet REALLY wanted:	15	36	7	6	16	5	2	2	0
Rivet provided:	18/81 = 22%	78/162 = 48%	1/66 = 2%	3/57 = 5%	3/89 = 3%	4/93 = 4%	3/17 = 18%	0/33 = 0%	0/10 = 0%

[\[Link to full version, accessible for CMS members\]](#)



CMS Rivet coverage (no searches)

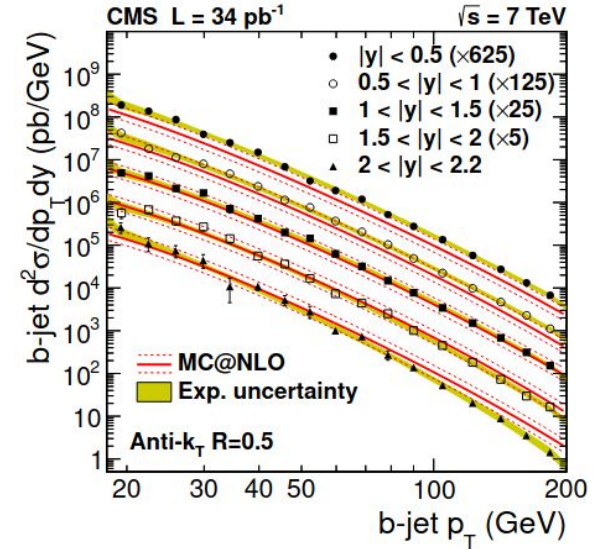
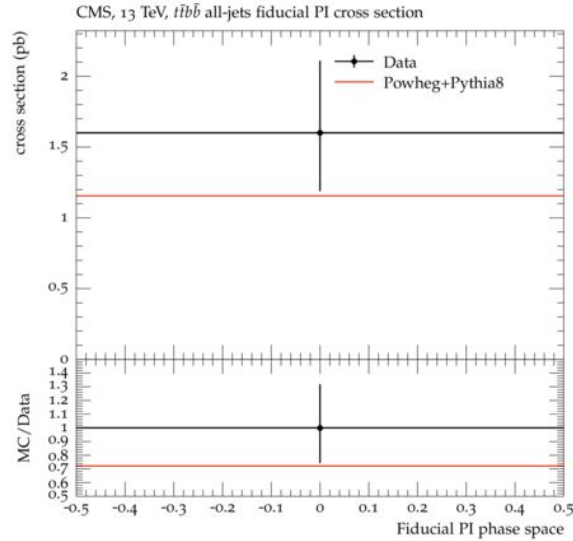
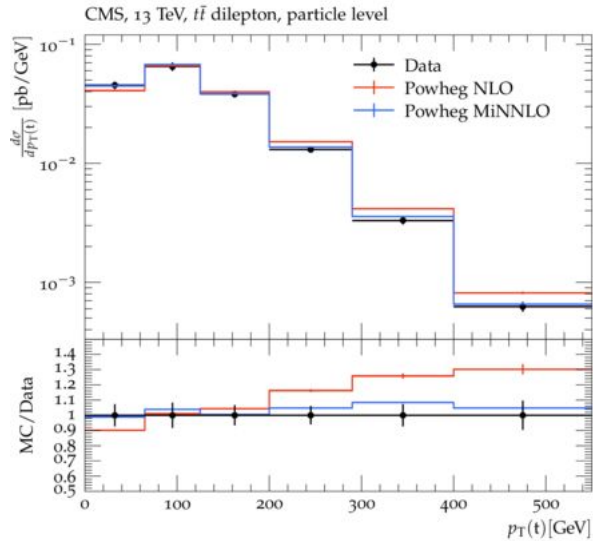
- Complicated situation for searches: many competing frameworks!
 - see many other talks at this workshop and [RecastingTools twiki](#)
- Rivet situation looks more manageable if we concentrate on measurements
- Not much change for TOP and SMP, considerably lowers the bar for HIG!
- What are the needs of the heavy-ion community?

Key	TOP	SMP	HIG	SUS	HIN	EXO	BPH	B2G	UNKNOWN
Rivet wanted (total):	54	76	12	0	85	1	10	0	8
Rivet REALLY wanted:	15	36	7	0	16	0	2	0	0
Rivet provided:	18/72 = 25%	76/152 = 50%	1/13 = 8%	0	3/88 = 3%	0/1 = 0%	3/13 = 23%	0	0/8 = 0%

[\[Link to full version, accessible for CMS members\]](#)

TOP and BPH contributions in 2022

- CMS_2018_I1703993 (ttbar diff xs in dilepton at 13 TeV)
- CMS_2019_I1753720 (13 TeV ttbb)
- CMS_2012_I1089835 (7 TeV incl b-jets)

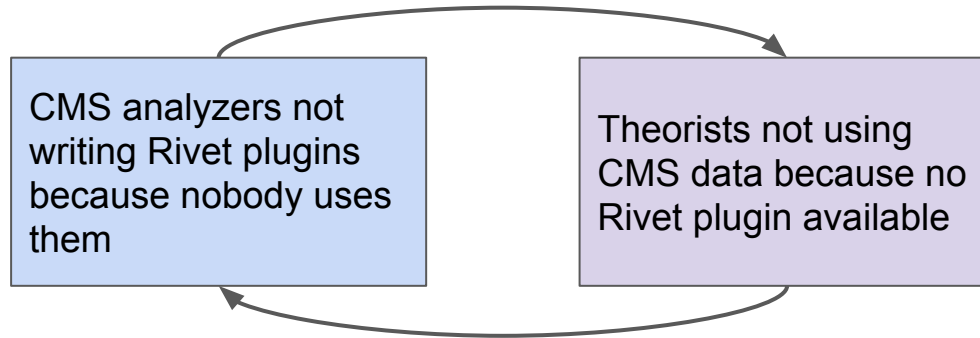


SMP contributions in 2022

- CMS_2022_I2079374 (Z pT over a wide mass range at 13 TeV)
- CMS_2021_I1866118 (Z boson plus jets (DPS) at 13 TeV)
- CMS_2021_I1963239 (Inclusive & Mueller-Navelet dijet production at 2.76 TeV)
- CMS_2017_I1497519 (Z+jets at 8 TeV)
- CMS_2020_I1776758 (Z+HF at 13 TeV)
- CMS_2021_I1972986 (Inclusive jets at 13 TeV)
- CMS_2021_I1978840 (W+gamma at 13 TeV)
- CMS_2021_I1847230 (8/13 TeV 3-jet and Z-2jet)
- CMS_2021_I1920187 (13 TeV q/g substructure in Z+jet/dijet)
- CMS_2021_I1932460 (13 TeV 4-jet DPS)

→ vector bosons, jets, jet substructure, DPS

How to improve CMS Rivet situation?

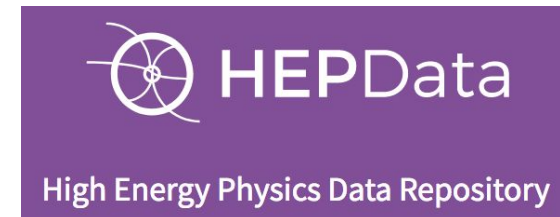


Break the cycle:

1. CMS → write more plugins
2. Theorists → please use both ATLAS and CMS plugins when available
3. Theorists → contact CMS analysis conveners
 - ~~A little push needed~~ an expression of interest highly appreciated and will be followed up
 - Convener contact information at <https://cms.cern.org/cms-scientific-results>

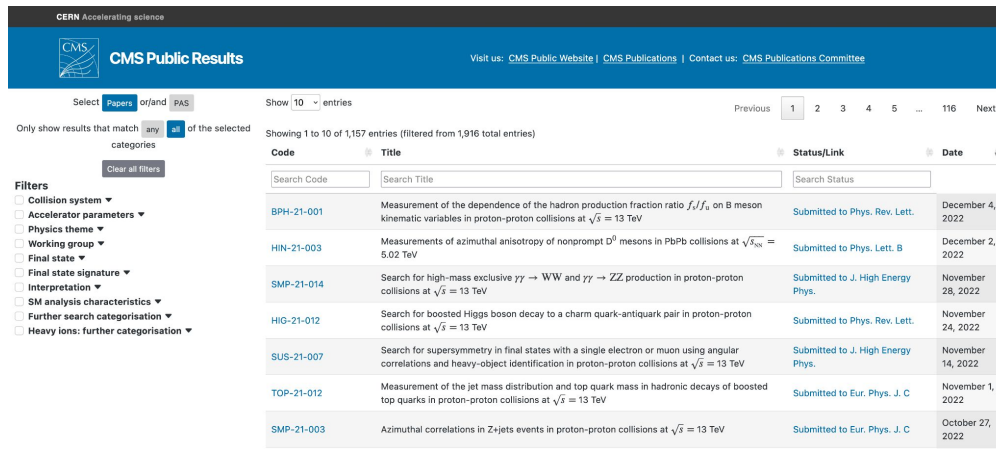
Analysis preservation efforts: HEPData

- Since about three years, creation of HEPData record mandatory as part of internal CMS approval process
 - → every CMS analysis should have a HEPData entry since then
 - Also some efforts made to create records for older analyses
- Recommended tool: `hepdata_lib`
 - See https://github.com/HEPData/hepdata_lib/
 - Used by large majority of analyses
 - Currently maintained by a single CMS member → looking for additional maintainer(s)



Additional material on analysis webpage

- Every analysis has its dedicated analysis webpage
- Allows download of all plots and tables in PDF and PNG format
- Often additional plots/tables made available
 - e.g. selection efficiencies
- Links to these webpages can be found from arXiv, <https://cms-results.web.cern.ch/cms-results/public-results/publications/> as well as a relatively new [CMS Results Search Webpage](#)



The screenshot shows the CMS Public Results search interface. At the top, it says "CERN Accelerating science" and "CMS Public Results". Below this, there are navigation links: "Visit us: CMS Public Website | CMS Publications | Contact us: CMS Publications Committee".

The main content area shows a search results table. On the left, there are filters for "Collision system", "Accelerator parameters", "Physics theme", "Working group", "Final state", "Final state signature", "Interpretation", "SM analysis characteristics", "Further search categorisation", and "Heavy ions: further categorisation".

The search results table has columns for "Code", "Title", "Status/Link", and "Date". The table shows 10 entries, with the first entry being "BPH-21-001: Measurement of the dependence of the hadron production fraction ratio f_1/f_0 on B meson kinematic variables in proton-proton collisions at $\sqrt{s} = 13$ TeV".

Code	Title	Status/Link	Date
BPH-21-001	Measurement of the dependence of the hadron production fraction ratio f_1/f_0 on B meson kinematic variables in proton-proton collisions at $\sqrt{s} = 13$ TeV	Submitted to Phys. Rev. Lett.	December 4, 2022
HIN-21-003	Measurements of azimuthal anisotropy of nonprompt D^0 mesons in PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV	Submitted to Phys. Lett. B	December 2, 2022
SMP-21-014	Search for high-mass exclusive $\gamma\gamma \rightarrow WW$ and $\gamma\gamma \rightarrow ZZ$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV	Submitted to J. High Energy Phys.	November 28, 2022
HIG-21-012	Search for boosted Higgs boson decay to a charm quark-antiquark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV	Submitted to Phys. Rev. Lett.	November 24, 2022
SUS-21-007	Search for supersymmetry in final states with a single electron or muon using angular correlations and heavy-object identification in proton-proton collisions at $\sqrt{s} = 13$ TeV	Submitted to J. High Energy Phys.	November 14, 2022
TOP-21-012	Measurement of the jet mass distribution and top quark mass in hadronic decays of boosted top quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV	Submitted to Eur. Phys. J. C	November 1, 2022
SMP-21-003	Azimuthal correlations in Z +jets events in proton-proton collisions at $\sqrt{s} = 13$ TeV	Submitted to Eur. Phys. J. C	October 27, 2022

Analysis preservation efforts

- Standard Model and Higgs groups archive and preserve their statistical models in a CMS-internal GitLab area
 - These are mostly used for combinations
- Similar efforts in search groups
- These models are currently not made public, but plans exist (see talk by A. Marini)

However, currently no structured approach for analysis code preservation

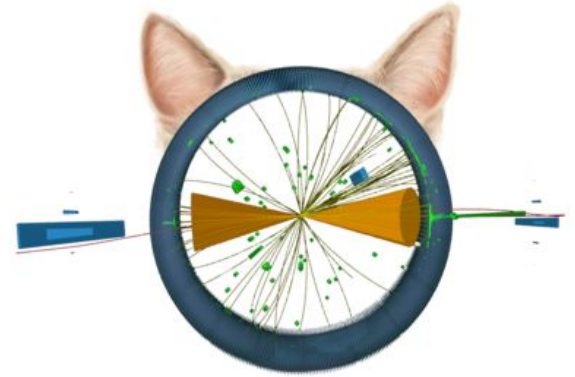
- Code resides mostly on CERN GitLab and GitHub, but location mostly only known to analysts themselves

New Common Analysis Tools group

CMS decided to create a new group within the Physics Coordination domain: Common Analysis Tools

- Conveners appointed, now forming subgroups
- One of the three subgroups dedicated to workflows and analysis reusability
- Plan to create common analysis code repository

The work of this group should significantly improve analysis preservation in CMS



Summary

Analysis preservation ~~strategy~~ efforts in CMS SM groups:

- HepData: mandatory
- Rivet: 18% to 40% (depending on definition)
- Additional material: sometimes
- Public analysis code: none (yet?)