

Recall: CMS + ATLAS Higgs->4Lepton demonstrator project for PUNCH platform



A.Geiser, DESY, 21.8.24, Cross-TA use case meeting

Main purpose: **demonstrate practical feasibility of a PUNCH use case on the PUNCH Science Data Platform going significantly beyond what is already available outside PUNCH** (i.e. not just an import of things already available elsewhere), **using PUNCH resources** already now wherever possible.

PUNCH goal stated in fall 2021: **“to set up a working prototype within the first year”** as part of the **TA4** activities

-> see reports at 2022 Göttingen general meeting

Transformation of data from different projects/sources to common analysis data format, TA4/WP3.

(Current prototype still limited to HEP as starting point, extension to other PUNCH4NFDI communities conceptually started).

Actual usage of storage4punch resources, TA2/WP1, (including **test of the corresponding access procedures, TA4/WP3**).

PUNCH DESY/Bonn/KIT/GSI storage tested successfully.

-> breakthrough! use this!

Usage of ReAna and C4P resources to run scripts, TA4/WP4, TA2 (including **test of the corresponding access procedures, TA4/WP3**).

Access execution via ReAna tested successfully.

-> breakthrough! use this!

Documentation on PUNCH gitlab, TA4/WP4, TA4/WP3

Direct Contributors:

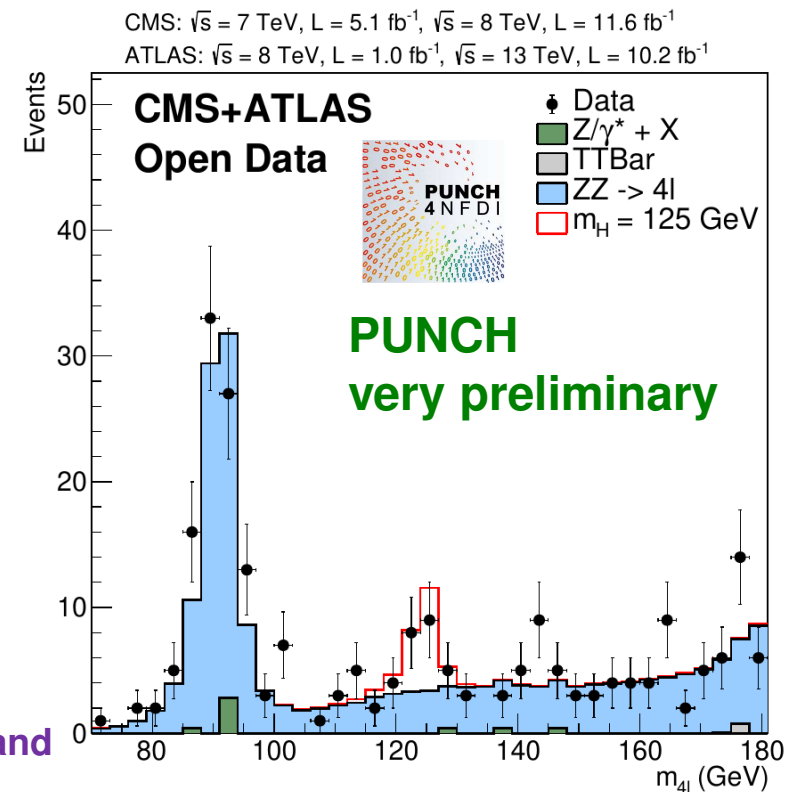
Lucas Karwatzki, Ting Chan (2024)
Tilde Bonnevier-Wallstedt (2023)
Matheus Costa Reis (2022)
Aritra Bal, Murillo Rebuzzi (2021)
+ many more!

Technical contributors:

Manuel Giffels, Benoit Roland
Oliver Freyermuth
Christoph Wissing, Harry Enke
reanasupport@aip.de

21.08.24

A. Geiser, Cross-TA use case meeting



Reminder of the example

LHC Open Data

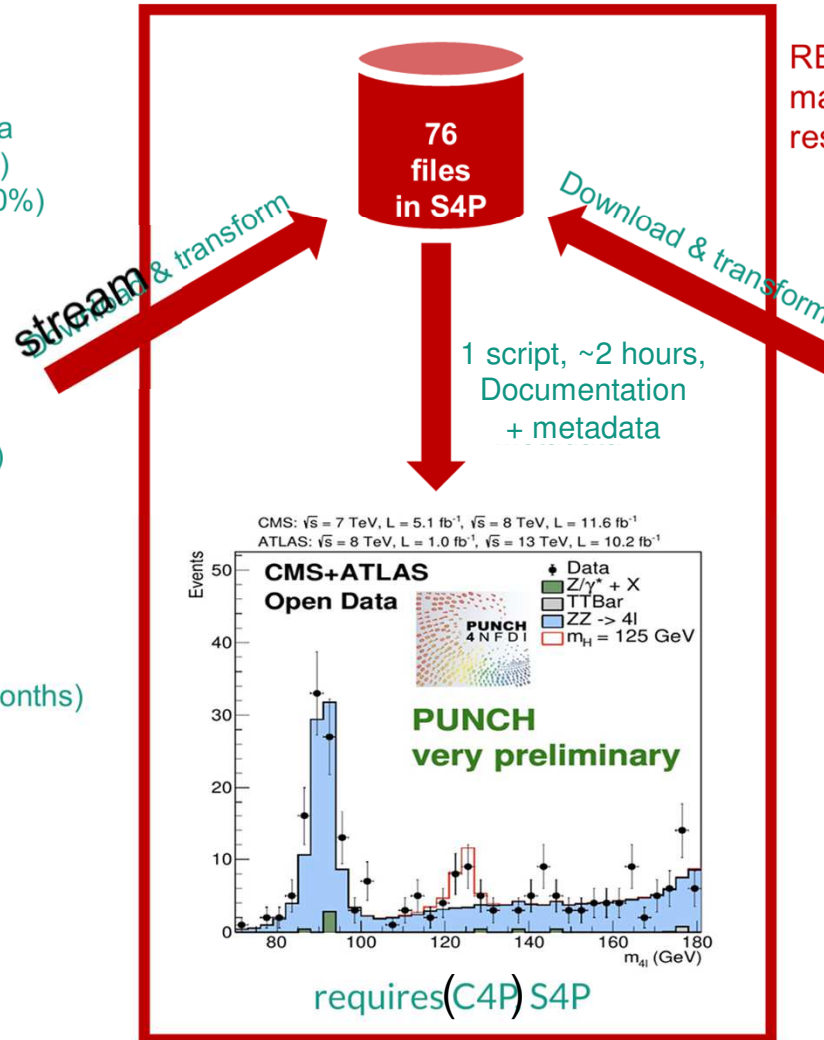
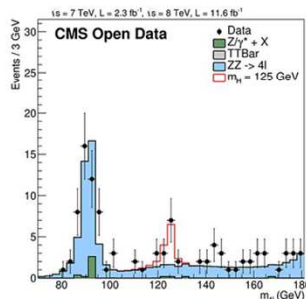
Original CMS legacy research data
(2 PB via CERN Open Data portal)
(2010 data 100%, 2011/12 data 70%)



CMS legacy software from public
github via CERN Open Data portal)



Produce histograms (many CPU months)



or local CPU

REANA for workflow
management and
resource organisation

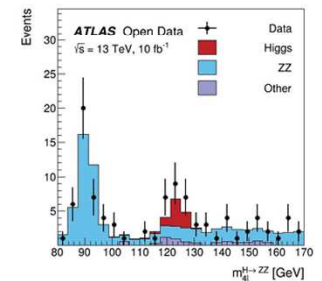
Non-public ATLAS legacy data



Simplified educational data sets 2012 /
2016 via CERN Open Data portal or
ATLAS Open Data portal



VM with dedicated software
package or Jupyter notebook



For real: CMS + ATLAS Higgs->4Lepton demonstrator project for PUNCH platform



S4P	C4P	PUNCH ReAna	ReAna dev	Unique to PUNCH	Doc in READ ME	Name/Path
-	-	-	-	-	partial	https://gitlab-p4n.aip.de/punch/tutorials/CMS_nanoaod_usage_examples various examples
DESY	-	-	-	-	yes	https://gitlab-p4n.aip.de/punch/tutorials/punch_nanoaod_usage_examples MuHistos_punch_DESY.cxx
Bonn	-	-	-	-	yes	MuHistos_punch_Bonn.cxx
KIT	-	-	-	-	yes	MuHistosuniversal_KIT_2012.cxx
GSI	-	-	-	-	yes	MuHistos_punch_GSI.cxx
-	-	yes	-	-	yes	reana.yaml
-	yes	-	-	-	yes	MuHistos_eospublic.cxx
DESY	yes	-	yes	-	yes	reana_c4p_desy.yaml
Bonn	yes	-	yes	-	yes	reana_c4p_bonn.yaml
KIT	yes	-	yes	-	yes	reana_c4p_KIT.yaml
DESY / KIT	-	-	-	yes !	levels 1,2,3,4	https://gitlab-p4n.aip.de/punch/usecase_demonstrators/hep_higgsto4l_punch/ README.md
DESY / KIT	yes	-	yes	yes !	Levels 1,2,3,4	README.md (level 1), reana-c4p.yaml (level 2), reana-c4p-level3.yaml, reana-c4p-level4.yaml.

Higgs → 4Lepton

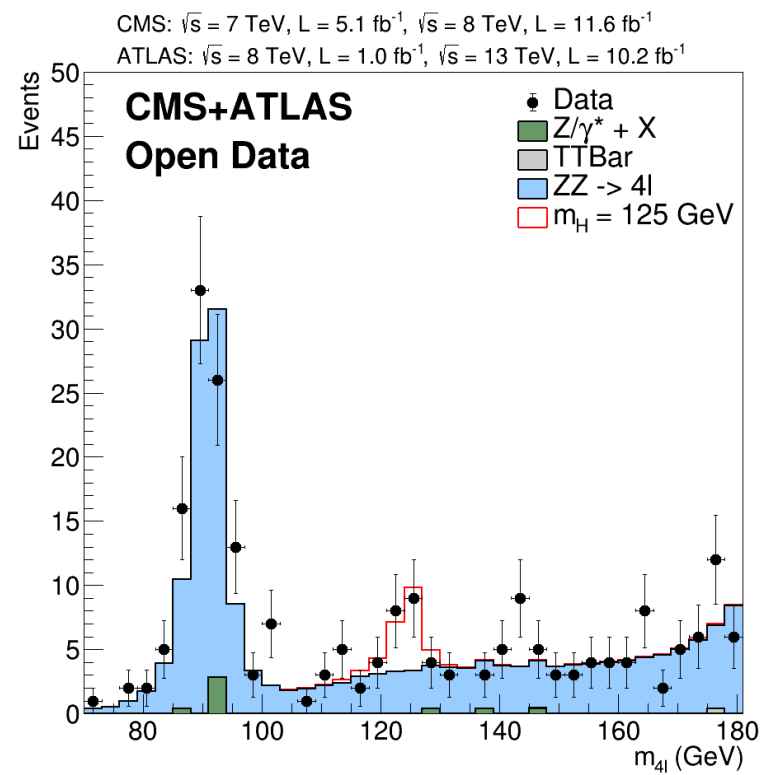
Create the 4Lepton mass spectrum:

Using 2010, 2011 and 2012 CMS open data and 2012 and 2016 ATLAS open data.

Four complexity levels are implemented:

- 1. **Observe** the final output plot
- 2. **Reproduce** the final plot from predefined histograms
- 3. **Produce** a root data input from parts of the full analysis
- 4. **Produce** the full analysis (WIP)

 Ready to be run, by other PUNCH users
[\[here\]](https://gitlab-p4n.aip.de/punch/usecase_demonstrators/hep_higgsto4l_punch)



For real: CMS + ATLAS Higgs->4Lepton demonstrator project for PUNCH platform



https://gitlab-p4n.aip.de/punch/usecase_demonstrators/hep_higgsto4l_punch

eting (29-Marc...

punch / UseCase_Demonstrators / HEP_Higgsto4L_PUNCH

README.md

HEP_Higgsto4L_PUNCH

This is a use case demonstrating the ability within PUNCH to analyse Open Data from different HEP experiments at research level within a single workflow, using low level datasets transformed to a common so-called nanoAOD format (PUNCH flavour). Although the example is meant to serve as a starting point for users to learn how to set up their own publishable analysis on the data being provided, all transformed input files and the software and hardware running thereon are provided on a best effort basis, and no liability is taken by PUNCH or the collaborations providing the original (pre-transformed) Open Data for their correctness.

For a documentation of the content of the input NanoAOD ntuple files, see <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAODRun1>. This documentation strictly applies to the CMS ntuples directly provided through the CERN Open Data store <https://opendata.cern.ch>, which have been produced/transformed with the code available in <https://github.com/cms-opendata-analyses/NanoAODRun1ProducerTool>. The data in this format are designed to be able to reproduce or conceptually extend about 50% of the published CMS physics results on these datasets. Essentially the same code has also been used to transform some additional CMS Open Data sets needed for this example, for which the derived sets have not yet been provided directly by the CMS Open Data team, and which are available for PUNCH users through the PUNCH store. Some variables (not needed in this analysis example) are still missing on the additional ATLAS ntuples derived from educational Open Data and also provided through the PUNCH store through dedicated data transformations. Additional documentation of these ATLAS ntuples will be linked here soon.

...



Your workflows

Refreshed at 21:59

Search...

Status



Show deleted runs

Latest first

✓ **hep_higgsto4l_punch_level4** #2

11 MiB

Finished 6 days ago

finished in 1h 43m 59s

step 3/3

✓ **hep_higgsto4l_punch_level4** #1

10.99 MiB

Finished 6 days ago

finished in 2h 42m 11s

step 3/3

✓ **hep_higgsto4l_punch_level3** #4

392 KiB

Finished 13 days ago

finished in 12 min 19 sec

step 3/3

✗ **hep_higgsto4l_punch_level3** #3

100 KiB

Finished 13 days ago

failed after 0 seconds

step 0/3

Resource usage: Higgs → 4Lepton

All tests are run on DESY naf-cms24.

Level	Storage4punch (Read only)	Eospublic (Read only)	Real time	Local CPU time	Storage	Real time	ReAna CPU time	Storage
1	/	/	/	/	/	/	/	/
2 - Local	1.9 MB	/	6 min*	30 sec*	3.2 MB	/	/	/
2 - ReAna	1.9 MB	/	>10 sec*	>10 sec*	3.2 MB	19 min*	12 sec*	120 KB
3 - Local	1.3 GB	/	7 min*	1 min 15 sec*	3.4 MB	/	/	/
3 - ReAna	1.3 GB	/	>10 sec*	>10 sec*	3.2 MB	20 min*	55 sec*	340 KB
4 - Local	206 GB	668 GB	2h 35 min*	1 h 30 min*	7.1 MB	/	/	/
4 - ReAna	206 GB	668 GB	>10 sec*	>10 sec*	3.2 MB	1 h 37 min*	53 min*	5.9 MB

*: Local times are only approximate and will vary from system to system.

*: ReAna times are approximate only and will vary depending on cluster occupancy.

Thanks a lot to Baida Achkar for testing! More testers very welcome

For real: CMS + ATLAS Higgs->4Lepton demonstrator project for PUNCH platform



Proposed plan:

Get more feedback from testing within PUNCH (so far: authors + 1 test user, Baida ☺)

- * **members of PUNCH applicant institutions already have access:**

https://gitlab-p4n.aip.de/punch/usecase_demonstrators/hep_higgsto4l_punch

- * **e.g. still technical issues with linux 8, while linux 7 and linux 9 seem to work fine**

Make the example publicly available through a “filtered” rollout:

- * **Make gitlab documentation public as a pilot project** (e.g. by mid-September?)
- * **Allow access also to members of all participant institutions (includes DPG, CERN, ...) on request**
-> this covers essentially the whole German PUNCH community, plus part of the international one
- * **Define appropriate limits for resource use; attach license/DOI**

Integrate into mid-term report as an actual PUNCH achievement, together with other examples

Extend to include metadata (Bielefeld repository), and mapping to DRP registry if possible

Make it more explicit that the demonstrator setup is applicable to about 50% of LHC-related physics topics

(at least for the CMS part, since ATLAS part is still semi-educational)

Prepare a public document

-> expect/hope for first user physics publications starting from this setup within ~2-3 years