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

A.Geiser, DESY, 17.7.23, TA4/WP3 meeting   **Outline of summer student project (Tilde)**

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

Representation of the corresponding metadata in a common PUNCH scheme based on XML and datacite, **TA4/WP2.**
(Current prototype: preliminary practical starting point, **ongoing** including visualization.) **Details being implemented** by Ding-Ze (Lisa).

Actual usage of storage4punch resources, TA2/WP1,
(including **test of** the corresponding **access procedures**, TA4/WP3).
PUNCH DESY pilot dcache storage tested successfully:
~~https://intra.punch4nfdi.de/?md=/docs/TA2/WP1/StoragePrototyping.md~~
https://**hifis-storage.desy.de**/punch/HEP-OpenData/
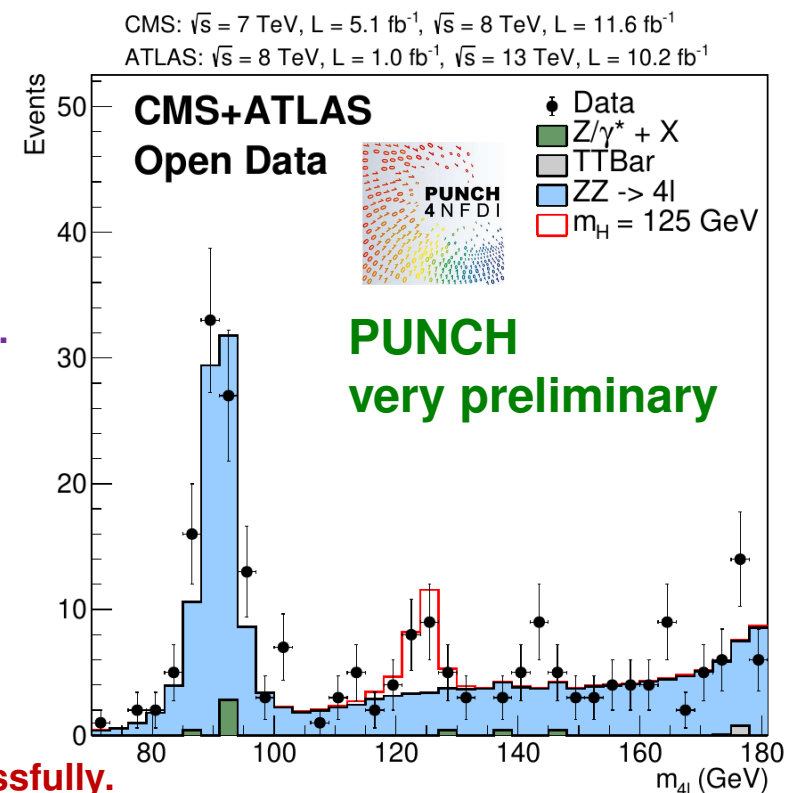**Bonn storage still to be tested.** **-> use this!**

Usage of compute4punch resources to run scripts, TA2/WP2
(including **test of** the corresponding **access procedures**, TA4/WP3).
Access to and usage of KIT PUNCH computing (re)tested successfully.
Containerization into docker container still to be done. -> Summer student project started July 18
for the moment switched to ReAna

CMS: $\sqrt{s}$ = 7 TeV, L = 5.1 fb^{-1}, $\sqrt{s}$ = 8 TeV, L = 11.6 fb^{-1}
ATLAS: $\sqrt{s}$ = 8 TeV, L = 1.0 fb^{-1}, $\sqrt{s}$ = 13 TeV, L = 10.2 fb^{-1}

CMS+ATLAS
Open Data

- Data
- Z/γ^* + X
- TTBar
- ZZ -> 4l
- m_H = 125 GeV

PUNCH very preliminary

m_{4l} (GeV)

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

A.Geiser, DESY, 17.7.23, TA4/WP3 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

Usage of portal resources for **scripts and documentation, TA4/WP4**
(including e.g. "automatic" transformation Twikis -> MarkDown, **TA4/WP3**).
description of transformed data format available on AIP gitlab:
https://intra.punch4nfdi.de/?md=/docs/TA4/WP3/Workbook.md
- **Textual description/documentation of analysis workflow to be added on AIP gitlab -> Summer student project starting July 18**
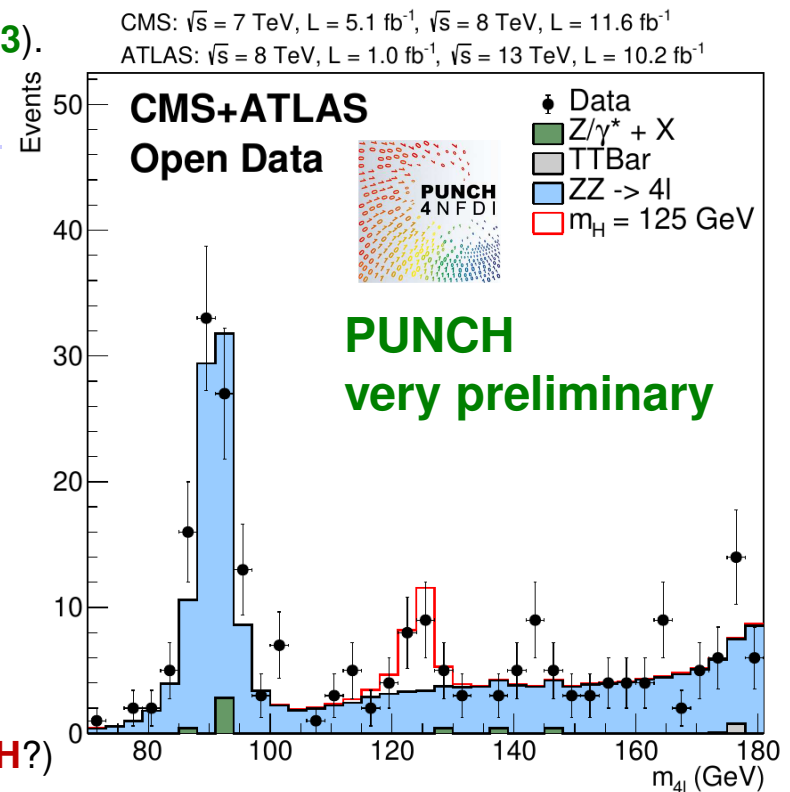- **Relevant scripts to be stored on AIP gitlab**

Setup of corresponding Research Products, TA4/WP1 (Ivan)
for PUNCH derived data sets and entire example
(on the platform, **TA4/WP4,** AIP team).
- **Integrate transformed data files into the WP1 research product description setup (for data files).**
- **Integrate scripts into the WP1 research product description setup (for software).**
- **Implement the workflow into the WP4 ReAna setup.**
- **Integrate this ReAna setup into the WP1 research product setup (for workflows). -> Summer student project starting July 18**

Any overlap with TA3? (e.g. accessibility of **ROOT versions in PUNCH**?)

Make the example publicly available on the PUNCH platform

CMS: $\sqrt{s}$ = 7 TeV, L = 5.1 fb^{-1}, $\sqrt{s}$ = 8 TeV, L = 11.6 fb^{-1}
ATLAS: $\sqrt{s}$ = 8 TeV, L = 1.0 fb^{-1}, $\sqrt{s}$ = 13 TeV, L = 10.2 fb^{-1}

CMS+ATLAS Open Data

- Data
- Z/γ^* + X
- TTBar
- ZZ -> 4l
- m_H = 125 GeV

PUNCH very preliminary

Events

m_{4l} (GeV)

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

A.Geiser, T. Bonnevier-Wallstedt, with support by TA2 and TA4/WP4, DESY, 30.8.23, Inter-TA use case meeting

After some initial period of being stalled (thanks to **all those who tried to help**), made **significant progress over the last two weeks:**

Thanks to **setting up oidc agent** on one of the DESY CMS workgroup servers, **recovered access to compute4punch**, and were able to **access the CERN store (eospublic) from there.** Thanks a lot to **Christoph (TA2) and DESY IT**!

recovered access to storage4punch (Bonn store) directly from root with xrootd, from CMS workgroup servers. Thanks a lot to **Oliver (TA2)!**
small issue with the interference between grid and token access still being sorted out.
Also thanks to **Oliver** for **copying part of the datasets** from **hifis** (formerly dcache test) to the **Bonn xrootd store.**

established **access to the hifis** or **desy stores locally** or **from compute4punch** for e.g. **download with curl** (thanks**, Christoph!**)**,** but interactive xrootd access does not work yet.
web interface to hifis (for listing its structure) **seems to work** (on some browsers).

documentation of some purely CMS examples was successfully **translated to md** (thanks, **Tilde!**)**,** stored on **internal AIP gitlab** (thanks, **Harry!**)**,** and **extended by examples using the Bonn xrootd store** (thanks**, Tilde!**)**.** See
https://gitlab-p4n.aip.de/punch/intra-docs-content/-/blob/master/docs/TA4/WP3/NanoAODRun1Examples/NanoAODRun1Examples.md
(the documentation of the ntuple content was already stored there last year,
https://gitlab-p4n.aip.de/punch/intra-docs-content/-/blob/master/docs/TA4/WP3/Workbook.md)
Some of the **related pictures and scripts** are also **stored there,** in subdirectories (thanks, **Tilde!**).
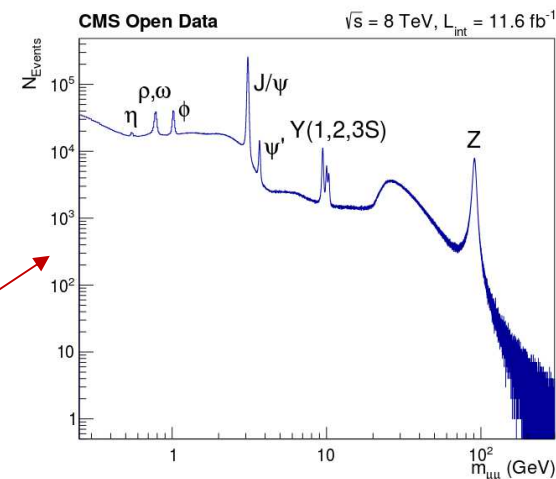
To run the version of the script that takes data from the Bonn store:

- `source /cvmfs/sft.cern.ch/lcg/views/LCG_104/x86_64-centos7-gcc12-opt/setup.sh`
- `eval `oidc-agent``
- `oidc-add punch-aai`
- `export BEARER_TOKEN=$(oidc-token punch-aai)`
- `root -l MuHistos_punch_Bonn.cxx++`

The modified 2012 DoubleMuParked outreach example (C++/python with RDataFrame)

A.Geiser, T. Bonnevier-Wallstedt, DESY, 30.8.23, Inter-TA use case meeting

One of these CMS examples has been **successfully installed and run on the CERN eospublic store on the AIP Reana instance** (thanks to Harry and Tilde!)

A first draft of the **documentation of the PUNCH implementation of the H->4L example** (local version, still under construction, thanks Tilde!) can be found in
https://gitlab-p4n.aip.de/punch/intra-docs-content/-
/blob/master/docs/TA4/WP3/Higgs%20to%204l/README_Higgsto4l_2.md

implementation of parts of this example on AIP ReAna in four levels is ongoing (thanks **Tilde!**):
level 1: just access and appreciate picture and its description (immediate, done)
level 2: read in intermediate histogram files (so far locally, will be moved to public AIP gitlab) **and run script to produce picture** (~ 1 CPU-Minute, tested OK)
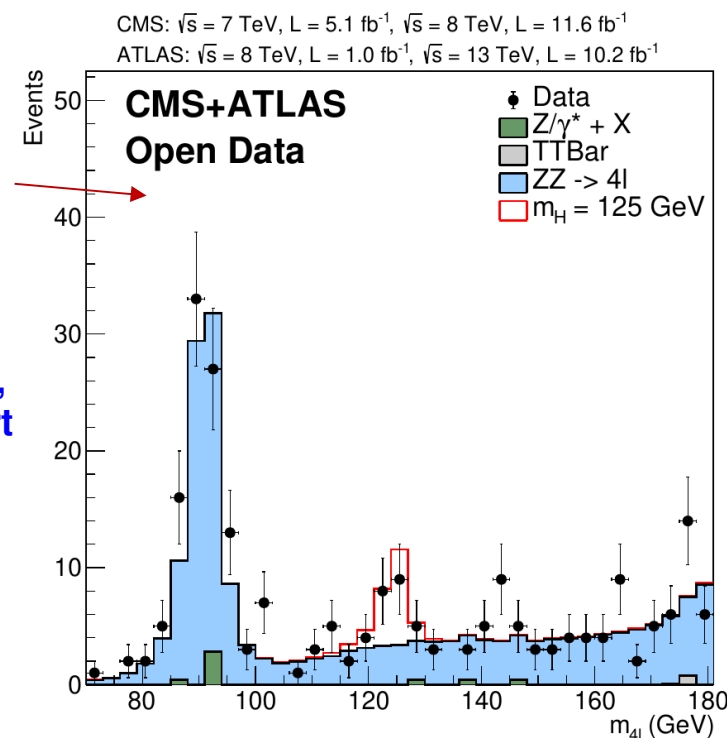**level 3: read part of the CMS ntuple input files from CERN eospublic store, and produce the corresponding histogram files which supersede part of the precompiled histogram files of level 2.** (~ 1 CPU-hour)
Run the script to produce the plot. Being tested.
Plan: in addition, store the transformed ATLAS files (small enough) on Zenodo and read them from there (without access restrictions)
level 4: treat all input files and run the full example (~ 1 CPU-day)
This is not possible yet.

CMS: $\sqrt{s}$ = 7 TeV, L = 5.1 fb^{-1}, $\sqrt{s}$ = 8 TeV, L = 11.6 fb^{-1}
ATLAS: $\sqrt{s}$ = 8 TeV, L = 1.0 fb^{-1}, $\sqrt{s}$ = 13 TeV, L = 10.2 fb^{-1}

CMS+ATLAS Open Data

- Data
- $Z/\gamma^* + X$
- TTBar
- ZZ -> 4l
- m_H = 125 GeV

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

A.Geiser, T. Bonnevier-Wallstedt, DESY, 30.8.23, Inter-TA use case meeting

Request to TA2:

Create a **subset of one of the PUNCH stores** with **no access restrictions and store all the transformed CMS and ATLAS files there** (to make Zenodo superflous). **H->4L should then work for almost anybody from almost anywhere** (after registration with ReAna)

and/or

Request to TA4/WP4 & TA2:

Make the **Bonn store accessible from within AIP Reana with PUNCH credentials.**
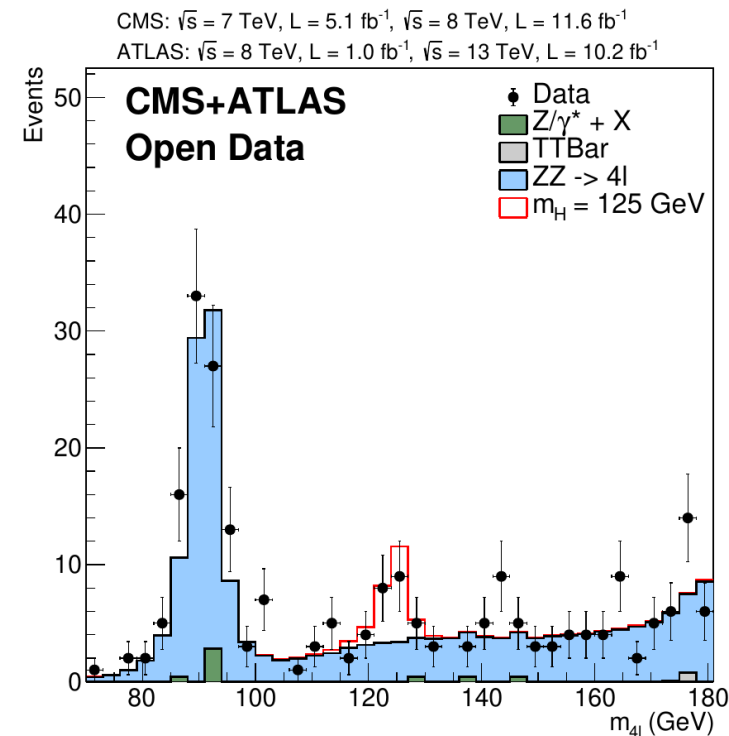**H->4L should then work for almost anybody with PUNCH credentials.**

in addition:

Request to TA3 (?):

Maintain documentation of and access to Root/xrootd versions suitable for PUNCH (e.g. via cvmfs)
E.g. version suitable for Bonn store (thanks **Oliver!**)**:**
source /cvmfs/sft.cern.ch/lcg/views/LCG_104/x86_64-centos7-gcc12-opt/setup.sh

Access to Bonn store with xrootd from compute4punch might be possible with this version, not tested yet.

CMS: $\sqrt{s}$ = 7 TeV, L = 5.1 fb^{-1}, $\sqrt{s}$ = 8 TeV, L = 11.6 fb^{-1}
ATLAS: $\sqrt{s}$ = 8 TeV, L = 1.0 fb^{-1}, $\sqrt{s}$ = 13 TeV, L = 10.2 fb^{-1}

CMS+ATLAS Open Data

- Data
- Z/γ^* + X
- TTBar
- ZZ -> 4l
- m_H = 125 GeV

m_{4l} (GeV)

Backup

A. Geiser, inter-TA use case meeting

original CMS legacy research data
(2 PB on CERN /eospublic via
CERN Open Data portal)
2010 data (100%, legacy format 1)
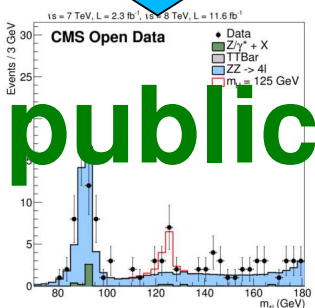and 2011/12 (70%, legacy format 2)

public

original ATLAS legacy research data (not public
by ATLAS collaboration

nonpublic

simplified educational
ATLAS Open Data 2012
(on CERN /eospublic via
CERN Open Data portal)
(10%, simplified format 3)

simplified educational
ATLAS Open Data 2016
(on separate ATLAS
Open Data portal)
(25%, simplified format 4)

public

original CMS legacy software
(from public github via
CERN Open Data portal)
(2 different versions, run on two
different legacy VMs or containers)

public

VM with dedicated
software package 1

VM with dedicated
software package 2
or Jupyter notebook

public

download

TA2 compute4punch project

produce
histograms

many CPU
months

public

apply data transformation
interface (versions 1 & 2)

now public

apply data transformation
interface (versions 3 & 4)

PUNCH TA4

public

76 different samples with common
unified & simplified research level
data format

50% public (CERN eos)
50% storage4punch

"single" script, < 1 CPU day
documentation+metadata

PUNCH
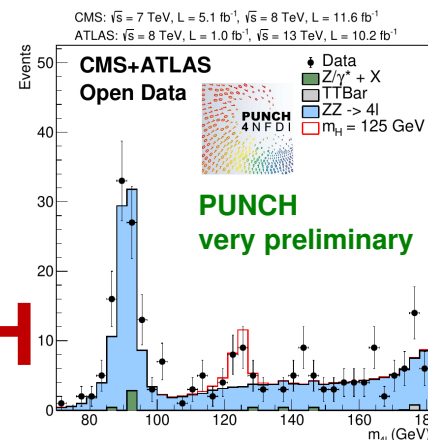
public

PUNCH
very preliminary

76 different samples with common unified & simplified research level data format, **via the PUNCH platform**

"single" script, < 1 CPU day, documentation + metadata

CMS: √s = 7 TeV, L = 5.1 fb⁻¹, √s = 8 TeV, L = 11.6 fb⁻¹
ATLAS: √s = 8 TeV, L = 1.0 fb⁻¹, √s = 13 TeV, L = 10.2 fb⁻¹

CMS+ATLAS Open Data

- Data
- $Z/\gamma^* + X$
- TTBar
- ZZ -> 4l
- m_H = 125 GeV

PUNCH very preliminary

Events

m_{4l} (GeV)
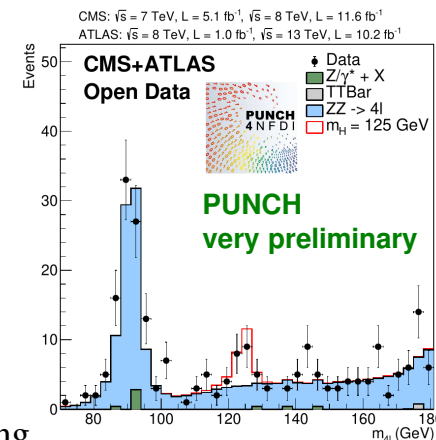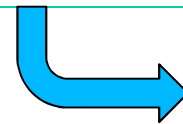
details being finalized