



Data Preservation at Fermilab (Tevatron and MINERvA)

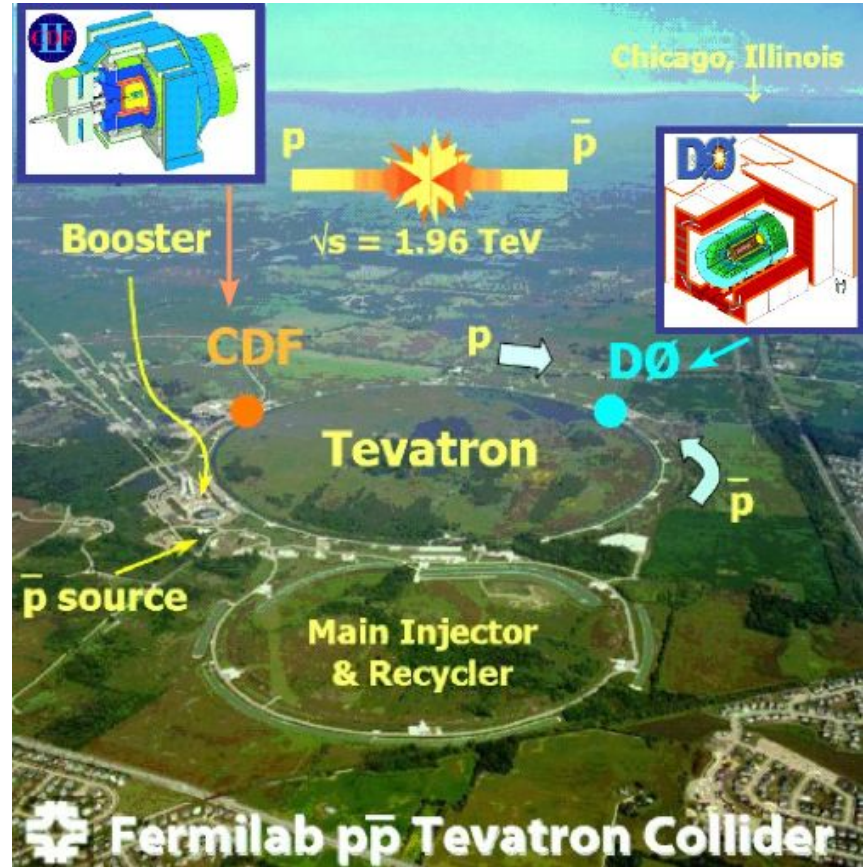
K. Herner, B. Jayatilaka, A. Delgado, L. Fields, R. Fine, D. Harris, K. McFarland, B. Messerly,
A. Olivier, D. Ruterbories, H. Schellman

DPHEP 2021

23 June 2021

Outline

- Tevatron
 - Recap of the R2DP project
 - Recent updates
- Non-collider experiments
 - MINERvA plans



Tevatron: The Run II Data Preservation Project

- Overall project goal was DPHEP Level 4 preservation of both [CDF](#) and [D0](#) through at least 2020 (end of SL6)
 - Efforts began in 2011 (final year of run) within each experiment
 - Fermilab SCD established R2DP project in 2012 to accomplish specific pieces of the experiments' programs; ran 2013-2015
 - Dedicated experts from CDF and D0, Fermilab SCD technical lead, and project manager
 - Several efforts from the experiments outside of project scope as well
- Technical work to change experiment code bases, transition data delivery and file catalog systems from Oracle to Postgres, move (some) releases to CVMFS, switch to running on IF computing infrastructure
- Project documented in a NIM paper:
 - S. Amerio et al., "Data Preservation at the Fermilab Tevatron", [Nucl. Instrum. Methods Phys. Res. Sect. A, 851, 1 \(2017\)](#). [arXiv:1701.07773](#)

Tevatron updates 2017-2021

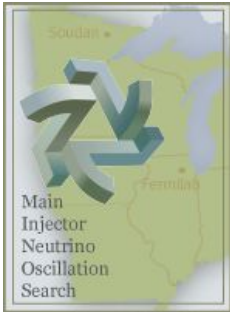
- Our last talk was in 2017 after the NIM paper came out. Since then:
 - D0
 - Didn't end up using much of the new infrastructure; kept using the legacy systems
 - Last of legacy batch computing retired in 2020 (SL6 EOL)
 - SL7 login node available
 - CDF
 - Continued using DP infrastructure including opportunistic use of FNAL grid
 - Established SL7 cloud node to continue after SL6 EOL
 - SL6-compiled code validated to run on SL7
- Both experiments now over 500 publications; two or three more papers expected from each experiment
 - Publications since 2017: 5 CDF, 9 D0, and 2 CDF+D0
- O(few dozen) people still active

Tevatron Future Plans

- FNAL is migrating Tevatron data to LTO8 media (~10 PB/experiment)
 - May take longer than last migration (took 2.5 yrs) since demand is lower and many other experiments want to use the tape drives
- Should a future analysis arise:
 - CDF SL6-built releases verified to run correctly in SL7; can simply run from CVMFS.
 - D0 will likely run everything in SL6 containers (likely Singularity); no plans to natively compile in SL7. Running the SL6 build in SL7 should also work, but detailed testing not done
- **Personnel likely not available to launch a major new initiative** (“major” = requires reprocessing data or generating significant new background MC)

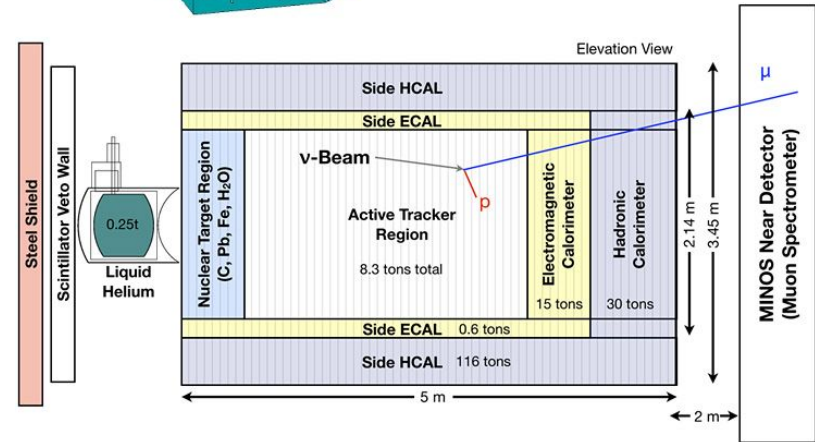
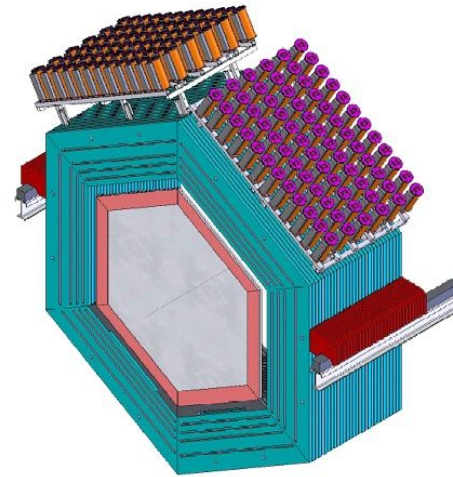
Non-collider experiments

- All DOE experiments must have some kind of [data management plan](#)
- A number of experiments (see below for a non-exhaustive list) already have entered, or will enter in the 2020s, their post-datataking period
- One of the first to hit this stage is MINERvA...



MINERvA Overview

- [MINERvA](#) measures neutrino- nucleus interactions in a variety of elements
- Approx. 65 collaborators
- Datataking completed in 2019
- Numerous [publications](#); more expected
- Publications often have an accompanying [data release](#) consisting of files needed to reproduce plots, etc.
- Uses GAUDI framework (now SL7)
- Already using most of the common Fermilab tools for job submission, file delivery, cataloguing



MINERvA Data Preservation Project

- The MINERvA data preservation project aims to ensure MINERvA data remains usable into the DUNE era.
- Three primary components:
 - **Unified ROOT tree** incorporating low- and high-level reconstructed objects needed to reproduce most MINERvA measurements (~500 TB total)
 - **The MINERvA Analysis Toolkit (MAT)** – a broadly applicable HEP software toolkit with the functionality necessary to turn ROOT trees into cross section measurements, including propagation of systematic uncertainties (see [VCHEP paper](#) for more info)
 - **A software package built on the MAT that reproduces MINERvA published analyses** and provides a template for new analyses
- Current status: beta versions of all three components exist and are being used to produce upcoming MINERvA results; also being tested by SBN collaborators
- The collaboration is investigating options for long-term storage that will allow use of data and analysis toolkit by the entire neutrino community (not just MINERvA collaborators)

Summary

- The CDF and D0 Collaborations have retired their legacy systems
 - Data being migrated (~10 PB per experiment); analysis now must occur with infrastructure created during the R2DP Project (2013-2015)
- Since the end of the R2DP Project, Several other Fermilab experiments are no longer taking data; others will end in the next few years
 - MINERvA has created a detailed DP plan and is executing it
 - Others will follow using lessons learned so far