

Preservation of the Direct Photon and Neutral Meson Analysis in the PHENIX Experiment at RHIC

Gabor David¹, Maxim Potekhin² and Dmitri Smirnov²

¹Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA

²Department of Physics, Brookhaven National Laboratory, Upton, NY, USA

E-mail: david@bnl.gov

Abstract. The PHENIX Collaboration has actively pursued a Data and Analysis Preservation program since 2019, the first such dedicated effort at RHIC. A particularly challenging aspect of this endeavor is preservation of complex physics analyses, selected for their scientific importance and the value of the specific techniques developed as a part of the research. For this, we have chosen one of the most impactful PHENIX results, the joint study of direct photons and neutral pions in high-energy d+Au collisions. To ensure reproducibility of this analysis going forward, we partitioned it into self-contained tasks and used a combination of containerization techniques, code management, and robust documentation. We then leveraged REANA (the platform for reproducible analysis developed at CERN) to run the required software. We present our experience based on this example, and outline our future plans for analysis preservation.

1 Introduction

The PHENIX experiment [1] was one of the two major experiments at the Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory (BNL). It was built in the 1990s and started taking data in 2000, playing the leading role in the discovery of the Quark-Gluon Plasma (QGP) and studying its properties [2, 3, 4]. The experiment continued to operate through 2016 and recorded a total of 24PB of raw data, covering a large variety of ions and collision energies, made possible by the versatility of RHIC.

While PHENIX has already published more than 220 physics papers, a few important analyses are still ongoing even now, and there is a possibility of new analyses that may start in the future, triggered by new physics insights or emergence of new analysis methods and tools. This is not uncommon – see for instance the recent re-analysis of HERA H1 data taken 2006-2007 [5]. However, in the era of rapidly changing hardware and software environments and fading institutional memory, ensuring that the data can be meaningfully analyzed decades from now is a very complex task. This work area is usually referred to as Data and Analysis Preservation (DAP). Ideally, DAP should allow a completely new analysis of the data (new observable or method), but reaching a reduced goal, namely making an existing and published analysis fully reproducible at any later time is already an important achievement.

In this paper we report on the successful preservation of the high transverse momentum direct photon and π^0 analysis [6] of the 200 GeV d+Au collision data that was taken in 2016. As seen in the iconic Fig.1 measurement of the nuclear modification factor, particularly for neutral mesons and direct photons is one of the supreme capabilities of PHENIX. Also, observation of the large hadron suppression and non-suppression of photons was one of the groundbreaking results at RHIC, decisive in establishing the formation of the QGP in large ion collisions. Apart of being highly interesting in its own right, we have chosen the d+Au analysis as our first DAP target, because this is the first time π^0 and γ have been analyzed and published in the same workflow, intertwined, and with a method that substantially reduces previous systematic uncertainties. Also, the analysis has important implications to the determination of collision centrality in small system collisions.

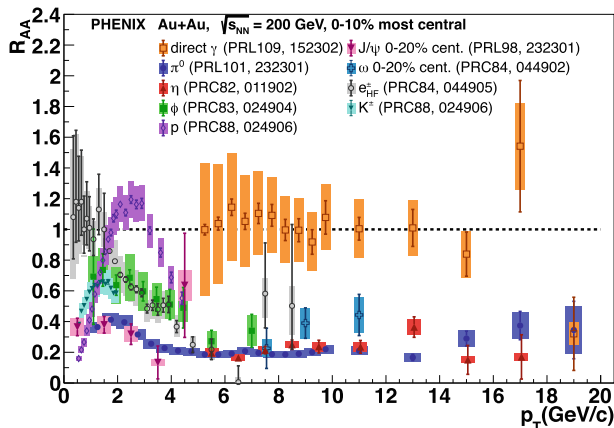


Figure 1: Nuclear modification factor in 200 GeV Au+Au collisions for various hadrons and direct photons.

2 Summary of the Data and Analysis Preservation effort in the PHENIX experiment

Data and Analysis Preservation (DAP) became a focus area for the Collaboration in 2019. At its core it has three principal components [7]

- Data preservation, including raw/processed data, databases, metadata
- Management and preservation of necessary core and analysis-specific software and its configuration
- Knowledge management: capturing and making available all relevant information even after the original analyzers are not available and "institutional memory" fades

2.1 Data preservation

PHENIX [8] relies on its host (Brookhaven National Laboratory) for archiving of its raw data and a subset of the derived data products [9]. A special area within the disk mass storage has been allocated for keeping intermediate data crucial for Analysis Preservation, including the analysis of direct photons in d+Au collisions. Conditions and calibrations data are stored in a database, and the issue of preserving these data will be addressed at a later time. The publication-level data such as data points used in tables in plots is reliably preserved by PHENIX Collaboration, by leveraging the HEPData Portal at CERN. A vast majority of all published PHENIX results are now preserved in this manner.

2.2 Software preservation

The PHENIX software stack, the result of over 25 years of continuous development, is inherently complex and includes a number of legacy components. Preserving this software presents significant challenges, particularly in maintaining compatibility amid ongoing component upgrades and evolving operating systems. To address these challenges, the PHENIX Collaboration has adopted containerization technology, specifically Docker, which provides an efficient solution for preserving the software environment. Docker enables the encapsulation of the entire software stack, ensuring both its continued operation and portability across different systems.

In addition to containerization, we have prioritized reproducibility in our computational workflows for physics analysis. To achieve this, we integrated the REANA platform [10], which facilitates the orchestration of complex workflows within a containerized environment. REANA not only ensures that our analyses are reproducible but also that they can be reliably executed in the future, irrespective of changes in the underlying infrastructure.

This combined approach of Docker-based containerization and the REANA platform has proven effective in overcoming the challenges of software preservation, ensuring that the PHENIX software stack remains functional and accessible for future analyses.

2.3 Knowledge management

Even a perfectly preserved software can't be useful without accompanying "know-how" for specific applications and related configuration details. This is being addressed via the PHENIX Knowledge Management

effort. PHENIX has successfully migrated a large number – more than 620 – of its research documents (including PhD theses, technical write-ups, conference presentations etc) to a modern digital repository — the Zenodo platform at CERN [11]. The previously fragmented Web materials have been curated and migrated to a new website, created with the goals of simplicity, security, performance and ease of long-term management [8]. All these measures combined allowed us to create a functioning foundation for Analysis Preservation.

3 The Analysis of π^0 and direct γ in d+Au collisions

3.1 The physics issues addressed in the analysis

The complete description of the context, significance, issues and physics consequences drawn from the analysis of d+Au data are beyond the scope of this paper and are described in [6], as well as in a dedicated write-up for this analysis [12]. Some of the key results in heavy-ion experiments such as PHENIX are the nuclear modification factors (essentially the qualitative difference between particle production in heavy ion vs. proton-proton collisions) which are studied as a function of centrality (the amount of overlap between the colliding ions). Conclusions are based in a large part on the measured invariant yields (spectra) of π^0 and direct γ as a function of centrality.

The steps of the analysis leading to those invariant π^0 and direct γ yields are shown in Fig. 2 and preserved in our REANA implementation. This workflow is designed such that can be run on a batch system and has been demonstrated to reproduce the fundamental quantities being studied, such the spectra in [6], thus satisfying the first and foremost goal of DAP: reproducibility of a published result. This is illustrated in Fig. 3. While the analysis as a whole is preserved, some of its parameters, cuts and conditions in the individual macros can be changed, making somewhat different analyses of the same data possible, so it can be viewed as a meaningful step toward the more ambitious, extended DAP goal of new analyses of preserved old data. The entire workflow and all relevant code, along with the necessary configuration files is preserved in a version control system accessible to users with BNL SDCC credentials here [13], with the necessary input files stored in [14]. The foundational software stack used in PHENIX has been preserved in a Docker image, ensuring portability.

3.2 The REANA implementation

The REANA implementation of our analysis workflow, as illustrated in Fig. 2, is designed to encapsulate and automate the complex sequence of steps required for data analysis in PHENIX.

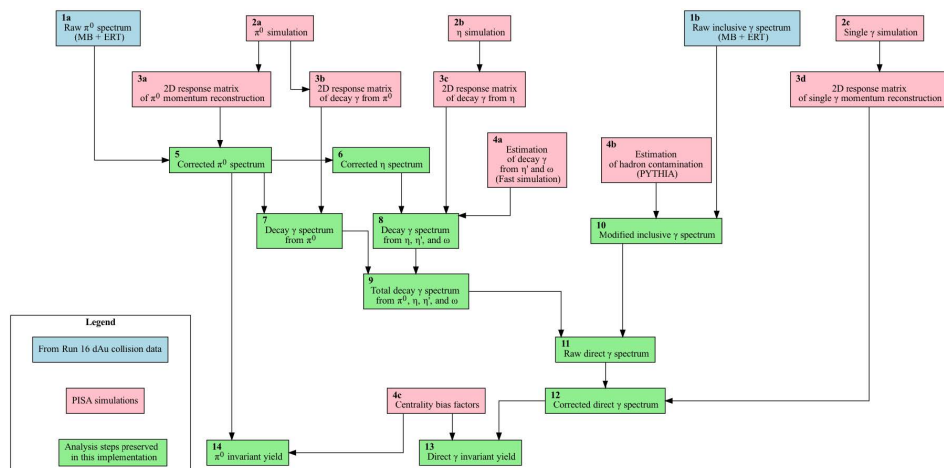


Figure 2: Workflow of the π^0 and direct γ analyses in PHENIX, as preserved. See the text for details.

In this workflow, blue boxes represent operations performed on a condensed output derived from a comprehensive scan of all available data, stored as a series of multidimensional histograms. These condensed outputs, essential for various stages of the analysis, are stored within the containerized environment. Pink boxes indicate operations involving simulated particles embedded in real events, generating what are known as “response matrices,” which are 2D histograms that capture the detector’s response to different physical scenarios. These matrices are also preserved within the container, ensuring that

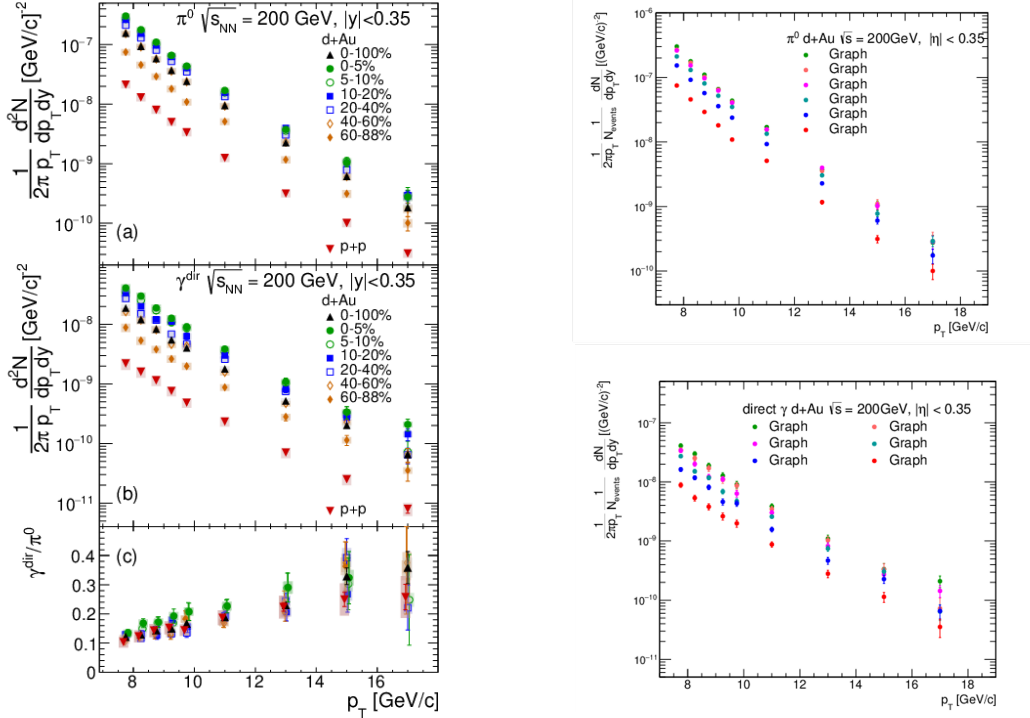


Figure 3: Left: final results of π^0 and direct γ production as a function of centrality in 200 GeV $d+Au$ collisions [6]. also shown are the corresponding spectra for 200 GeV $p+p$ collisions, taken from a different analysis, and a derived quantity, the point-by-point ratio of the π^0 and γ spectra (bottom panel). Right: π^0 (top) and γ spectra as reconstructed in the REANA implementation shown in Fig. 2.

all necessary inputs for analysis are readily accessible. Green boxes correspond to the actual analysis steps, which involve executing the preserved ROOT macros. Analysts can edit these macros to modify parameters, apply different cuts, or adjust conditions according to their specific analysis objectives. Each step in the workflow is accompanied by a verbal description to guide users through the process.

To validate the robustness of our preservation approach, the containerized analysis was tested by a computer-literate individual who is not a PHENIX collaborator. This individual successfully executed the entire analysis chain, with the results shown in Fig. 3. The left panel displays the π^0 and direct γ spectra (and their ratio) as originally published, while the right panel shows the same spectra generated from the preserved analysis chain. This successful replication by an external user underscores the effectiveness of the REANA implementation in preserving not just the data, but also the full analytical capability for future research.

3.3 Lessons learned

The PHENIX Collaboration has been largely successful in its Data and Analysis Preservation program (DAP), however this required a substantial amount of effort, against the background of limited personnel available for work in this area. Adding to the challenge, the PHENIX hardware and software design started in early 1990s, and for obvious reasons the software environment underwent major changes since then. Over the years, the software stack became very large, with complex code dependencies and in some cases redundancies. In view of this, we conclude that the single most important lesson from our efforts is that a well-designed, long-term DAP component of the experiment needs to be created by professionals from the very beginning of the experiment and the required standards meticulously followed during its entire lifetime, even if it appears costly or counterproductive at any particular moment. We recognize that a robust analysis preservation effort may be considered a hurdle on the path to quickly obtaining new

and exciting publishable results. For that reason, another important lesson is that clear, stable guidance is needed from the very beginning, on how to balance these contradictory requirements. One other dimension of the work we are doing is balancing the legitimate requirements of openness (since the data, the code and the analysis techniques need to be eventually made public due to the policies of the funding agencies), against the also legitimate requirements of data protection and cybersecurity. Reconciling all of this adds to the already significant challenge, and once again, clear guidance and sufficient funding are instrumental in achieving the worthy goals of the Data and Analysis Preservation.

4 Summary

The PHENIX experiment stopped taking data in 2016, having accumulated a large amount of data, much of it with a unique physics reach. Analyses of these data are ongoing and even new analyses can be expected to take place in the future, requiring a long term Data and Analysis Preservation effort. Such effort started in PHENIX in 2019 with limited manpower. Substantial progress has been achieved in the area of Knowledge Management, leveraging a web technology suited for long-term maintenance and a modern digital repository. Data tables for almost all PHENIX publications are now available on the HEPData portal [15]. The PHENIX software environment has been preserved with the use of containerization technology. Recently the analysis chain of a high-impact PHENIX paper has been implemented in REANA in the containerized format. The next project is REANA implementation of another signature PHENIX analysis, charmonium production via the $J/\Psi \rightarrow \mu\mu$ decay channel.

References

- [1] K. Adcox et al., “PHENIX detector overview,” *Nucl. Instrum. Meth. A*, vol. 499, pp. 469–479, 2003.
- [2] —, “Suppression of hadrons with large transverse momentum in central Au+Au collisions at $\sqrt{s_{NN}} = 130\text{-GeV}$,” *Phys. Rev. Lett.*, vol. 88, p. 022301, 2002.
- [3] S. S. Adler et al., “Elliptic flow of identified hadrons in Au+Au collisions at $s(NN)^{1/2} = 200\text{-GeV}$,” *Phys. Rev. Lett.*, vol. 91, p. 182301, 2003.
- [4] K. Adcox et al., “Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration,” *Nucl. Phys. A*, vol. 757, pp. 184–283, 2005.
- [5] V. M. Mikuni, “Multi-differential Jet Substructure Measurement in High Q^2 ep collisions with HERA-II Data,” *PoS*, vol. EPS-HEP2023, p. 231, 2024.
- [6] N. J. Abdulameer et al., “Disentangling centrality bias and final-state effects in the production of high- p_T π^0 using direct γ in d+Au collisions at 200 GeV,” 2023, arXiv:2303.12899.
- [7] M. Potekhin, “Data and Analysis Preservation in the PHENIX Experiment at RHIC,” *J. Phys. Conf. Ser.*, vol. 2438, no. 1, p. 012020, 2023.
- [8] <https://www.phenix.bnl.gov/>.
- [9] C. Hollowell, J. Barnett, C. Caramarcu, W. Strecker-Kellogg, A. Wong, and A. Zaytsev, “Mixing HTC and HPC Workloads with HTCondor and Slurm,” *J. Phys. Conf. Ser.*, vol. 898, no. 8, p. 082014, 2017.
- [10] T. Šimko, L. Heinrich, H. Hirvonsalo, D. Kousidis, and D. Rodríguez, “REANA: A System for Reusable Research Data Analyses,” *EPJ Web Conf.*, vol. 214, p. 06034, 2019.
- [11] <https://zenodo.org/communities/phenixcollaboration/>.
- [12] <https://zenodo.org/records/8169171>.
- [13] <https://git.racf.bnl.gov/gitea/PHENIX/ana-dAuPi0Photon>.
- [14] <https://zenodo.org/records/8117401>.
- [15] <https://www.hepdata.net/>.