PROCEEDINGS OF SCIENCE

The ZEUS long term data preservation project

Andrii Verbytskyi[*†]
Max-Planck Institut für Physik (Werner Heisenberg Institut), Föringer Ring 6, München 80605, Germany
E-mail: andrii.verbytskyi@mpp.mpg.de

The ZEUS data preservation (ZEUS DP) project assures continued access to the data and documentation related to the experiment. It aims to provide the ability to continue the generation of valuable scientific results from these data in the future. This talk presents analysis opportunities within the ZEUS DP project and its benefits for the physics community. The implications of the preserved data are discussed in the context of current data analyses, verification of contemporary and future theoretical models and the planning of future experiments and analyses. The detailed documentation on the data preservation effort can serve as an example for the data preservation efforts in current and future experiments.

XXIV International Workshop on Deep-Inelastic Scattering and Related Subjects
11-15 April, 2016
DESY Hamburg, Germany

[*]Speaker.
[†]On behalf of the ZEUS collaboration

PoS(DIS2016)264

1. Introduction

The data collected by high energy physics experiments are crucial for the understanding of particle physics. Among the many high energy physics projects the experiments on the HERA collider hold a special place, since it was (and still is) the world's only *ep* collider. Therefore the HERA data are unique and should be preserved.

The data preservation effort aims to ensure the long-term availability of the data after the end of the data taking and funding periods of the experimental collaborations. In this way it can increase the physics potential of the experiments, allowing long-term data analysis, re-using and re-analysing the data, combining results between experiments and using the data for education, training and outreach purposes. The data preservation efforts for the HERA experiments are part of a larger data preservation effort [1].

2. HERA and ZEUS

The Hadron Elektron Ring Anlage (HERA) was a ring accelerator at DESY in Hamburg [2], so far the only electron(positron)-proton collider in the world. It was running between 1991 and 2007 with a break in 2001-2002. During the break the HERA ring was upgraded: the luminosity was increased and electron polarisation systems were installed. The data taking period before the break is referred to as HERA-I, while the one after the break is referred to as HERA-II. During almost all the time of running the electron(positron) beam energy was close to 27.5GeV, while the proton beam energies varied, being 820GeV and 920GeV for HERA-I, 920GeV, 575GeV and 460GeV for HERA-II.

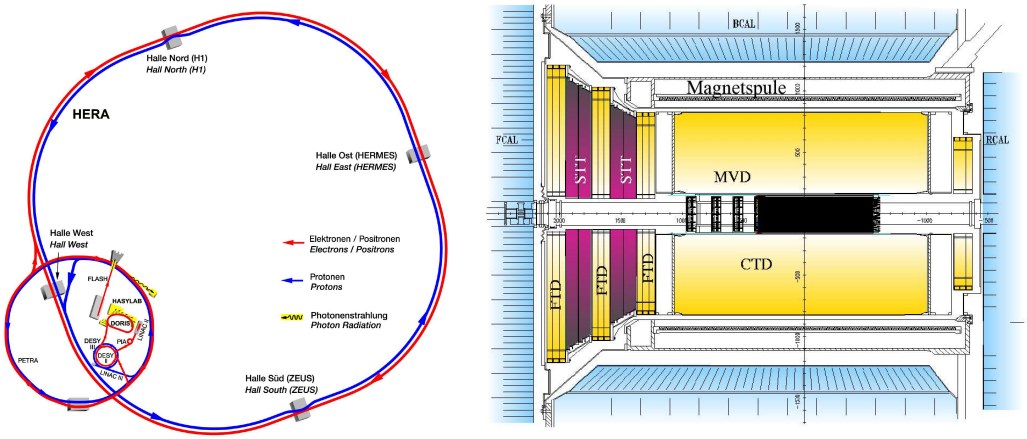

Figure 1: Schematic view of HERA accelerating ring, preacceleration facilities and experimental halls (left) and general schematic view of ZEUS tracking system (right) [3].

ZEUS was an experiment taking data in *ep* collisions at HERA. A full description of the original ZEUS detector design can be found elsewhere [2]. Briefly, ZEUS was a general purpose detector developed for a broad range of physics studies. It had a size of $12\,\mathrm{m} \times 11\,\mathrm{m} \times 20\,\mathrm{m}$ and a weight of 3600 tons. The major components of ZEUS included

a system for charged particle tracking for polar angles $7^\circ < \theta < 164^\circ$ within a high-field superconducting solenoid, a high resolution depleted-uranium calorimeter and a tracking system for muons. The detector was completed by dedicated forward detectors and a luminosity monitor.

In total ZEUS took more than 360 billion events of ep collisions which corresponds to a total luminosity close to $0.5\,\mathrm{fb}^{-1}$.

3. Physics case

After the end of funding period in 2014, the ZEUS collaboration has produced numerous publications using the preserved data (see Fig. 2). As of 2016 the collaboration actively produces valuable physics results [4–7]. It is foreseen that many valuable results can be obtained with the preserved ZEUS data [8, 9]. Briefly, some topics of interest are: the proton structure (e.g. F_2 and F_L and the strangeness content of the proton), combination of various diffraction measurements, studies of jets and event shapes with N(N)LO predictions, studies of photon structure, prompt photons, measurements of electroweak couplings, searches of instantons etc.

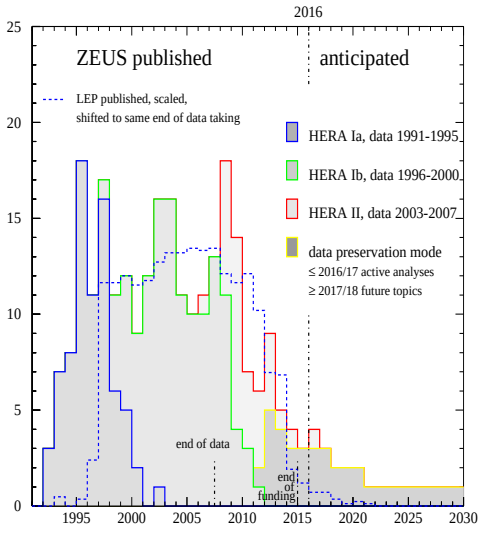

Figure 2: The number of ZEUS published and anticipated publications per year. The colours and grey shades indicate the usage of data from the different data taking periods, and the contribution from results produced from the preserved data. The corresponding numbers from the four LEP experiments combined, shifted forward by 7 years to account for the difference in end of data taking and normalised to the period 1997-2009 for ZEUS. The excess of the ZEUS distribution over the normalised LEP distribution beyond the end of funding in 2014 can be considered as a result of dedicated data preservation.

4. Common Ntuples

Before the end of 2012, the ZEUS analysis model used the data stored in Mini Data Summary Tapes (MDST) in ADAMO format. To perform an analysis a large set of ZEUS-specific software together with its customisation was needed. It was not intended to be maintained when the available manpower was reduced. Instead, the ZEUS analysis software was used to create common-usage ntuples for real and Monte Carlo(MC) data. These are referred to as Common Ntuples(CN) and, by their structure, are simple ROOT [10][1] files.

[1] PAW files with an identical information were created as well

They contain the necessary information to perform ongoing and planned future analyses; all ongoing and most of the recently published analyses actively use them.

CN are backward compatible as these are based on the existing documented ntuple scheme. This makes use of the expertise and knowledge accumulated in ZEUS analyses over the last decades. As ROOT is the main analysis tool for LHC it is expected to survive in a backwards compatible way for a very long time.

5. Bit preservation

One of the most important tasks of the ZEUS data preservation is to save the data bits. The ZEUS data selected for the preservation consist of the CN (ROOT/some PAW) and corresponding log files, in total ≈ 250Tb in 1.1 million files. The data is stored in DESY[2] and MPCDF[3].

In the DESY IT centre the data is stored in the DESY DPHEP dCache [11] system. The system has two instances with identical content: the disk pool system and backup archive tape system. The disk pools can be accessed with a very low latency from the DESY NAF[4] computing facility and two dedicated machines inside DESY. A DESY/ZEUS account is required for it.

In MPCDF centre the data is stored in the tape system and in the disk pool. The disk pool system contains only a fraction of the data stored in the archive tape system, namely the data and most important Monte Carlo simulated samples. On request any specific part of the archived data can be transferred to the disk pool. The disk pools can be accessed with a very low latency from the MPCDF/MPP[5] machines or with multiple protocols worldwide. For the former a MPCDF/MPP account with appropriate permissions is needed. For the access to the data worldwide a grid certificate with the ZEUS Grid Virtual Organisation [12] membership is needed.

6. Software preservation

For an analysis of ZEUS data an appropriate software is needed. In this section the most important software packages for the ZEUS analysis with the preserved data are described: the analysis framework, event display, file catalogue and MC production packages.

An **analysis framework** is a core analysis software. It is made to read, analyse the data and perform the calculations. Before switching to the data preservation mode of analysis, considerable efforts were made by the ZEUS collaboration to minimise the amount of software required for the analysis framework. As a result, the standard ROOT package is the framework for the ZEUS analysis. It is documented in detail elsewhere [10].

[2]http://www.desy.de, Deutsches Elektronen Synchrotron, Notkestr. 85, Hamburg DE-22607

[3]http://www.mpcdf.mpg.de, Max-Planck Computing and Data Facility, Gießenbachstr. 2, Garching DE-85748

[4]http://naf.desy.de/, National Analysis Facility

[5]http://www.mpp.mpg.de, Max-Planck Institut für Physik, Föhringer Ring 6, München DE-80805

An **event display** is a program for the visualisation of particle collision event data in a detector. It serves to adjust and check complicated reconstruction techniques by looking at single typical events and to produce illustrations of sophisticated analysis techniques for talks and papers (see Fig. 3). During the HERA-II period an event display called ZeVis was developed using ROOT libraries. Originally it was a client-server application where the dedicated server performed the reconstruction from MDSTs creating a ROOT file with the required information. The file was transmitted to the client via the network for the visualisation. In the context of data preservation the client was modified to read the information from the CN. The detailed description of the package is given in Ref. [14].

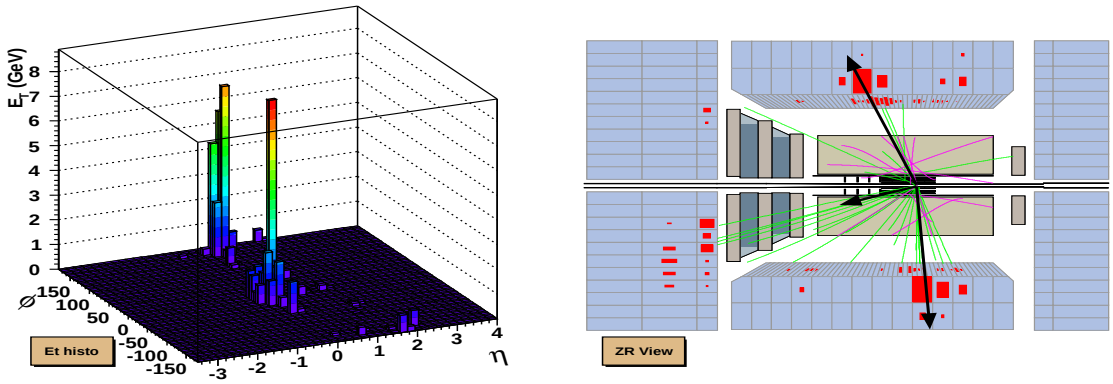

Figure 3: One of the selected events from the Z^0-boson analysis [13]. The transverse energy, $E_T(\eta,\phi)$, distribution in the CAL sells (left) and ZR event view (right) are shown.

A **filecatalog** is a database with an information on the available data samples. The ZEUS filecatalog, called CNINFO, provides information on the available data and Monte Carlo samples in CN format: the number of events, the data taking period, CN version, run numbers etc. The information can be extracted from the database with corresponding utilities. The database itself is stored in SQLite3 [15] and plain text formats. An option to edit (e.g. add new MC samples) is foreseen.

Some of the future analyses might require the generation of new MC samples. The **MC generators** used during active data taking are deeply integrated into ZEUS software and produced the events in the ZEUS-specific event record format. As of 2016 it is possible to use these for MC event generation, but an integration of a new generator to this scheme requires significant efforts or is not possible at all. To solve this problem, an utility for a conversion of the output of MC generators to ZEUS-specific format was developed. The conversion is possible for multiple formats, including HEPMC2 [16], HEPMC3 [17] and HEPEVT [18]. In the data preservation mode of analysis this is the favourable way to produce MC events and the only way to do it with the new MC generators.

For the purpose of ZEUS DP the full software chain for the detector simulation and reconstruction of MC simulated events, completely independent from the original analysis environment, has been prepared. The software chain is called ZEUS MC Standalone Package (**ZMCSP**) and is based on the software used for the ZEUS MC simulations on the

Grid [12]. Briefly, the package is a tarball that contains all the libraries and executable needed for the production of the MC simulated samples. The input for the package is the MC generated events and the output is the CN files (ROOT and/or PAW) described above. The software chain has been designed to be as much as possible independent from the host system. This makes it possible to use ZMCSP outside the DESY environment and benefit from other computing resources available in member institutes and elsewhere worldwide. Hereby, for the MC event generation, the Grid technology is especially attractive as it has been successfully exploited for many years for the same tasks in ZEUS [12].

As of 2016, the described software for ZEUS analysis can be used on many contemporary systems. The discussed long-term solutions for a system that can run the software are the NAF facility and virtual machines. Briefly, the first option is to use the standard analysis farm (or a part of it) that will be maintained at DESY with an option of access to CN on a long term. The second option is a customised CentOS [19] installation image suitable for various (virtual) machines with the relevant software. The image and the relevant instructions are provided on request (see Sec. 8).

7. Documentation preservation

The main efforts to preserve the digital and paper documentation of the ZEUS experiment were conducted in DESY. A brief description of the activities can be found in Ref. [20]. Concepts for long-term preservation of the ZEUS digital documentation were worked out in collaboration with the DESY IT division and DESY library. These are based on the InSpire[6] system and centrally maintained web servers[7]. As of 2016 all of the ZEUS internal notes are available (with password) on InSpire. The archive of non-digital information is hosted by the DESY library. It includes all ZEUS notes written before 1995, transparencies presented at meetings before 2000, technical drawings etc. It is also foreseen that in some cases it might be useful to add new or update the existing documentation.

8. Access rules

In the context of the data preservation it is important to clarify the rules for the data and software access. The access rules for the ZEUS data as of 2016 in the data preservation mode are similar to these during the active phase of the collaboration. In addition, interested analysers from institutes not in ZEUS can get access to the data under conditions to be negotiated with the collaboration spokesperson.

9. Conclusions

The ZEUS data preservation project has multiple levels and many participants. The presented work describes its implementation and potential benefits. The ultimate goal is to encourage potential collaborators to perform new analyses and continue to deliver valuable physics results.

[6] http://inspirehep.net

[7] http://www-zeus.desy.de, http://wwwzeus.mpp.mpg.de

References

[1] DPHEP Study Group, Z. Akopov et al., Status report of the DPHEP study group: Towards a global effort for sustainable data preservation in high energy physics. (2012). arXiv:1205.4667.

[2] ZEUS Collaboration, The ZEUS Detector. Status Report (unpublished), DESY (1993). http://www-zeus.desy.de/bluebook/bluebook.html.

[3] V.E. Schonberg, Measurement of beauty and charm photoproduction using inclusive secondary vertexing with the ZEUS detector at HERA, Ph.D. thesis, Bonn U., BONN-IR-2010-05, 2010.

[4] ZEUS Collaboration, H. Abramowicz et al., Search for a narrow baryonic state decaying to pK^0_S and $\bar{p}K^0_S$ in deep inelastic scattering at HERA. Phys. Lett. **B759**, 446 (2016). arXiv:1604.02220.

[5] ZEUS Collaboration, H. Abramowicz et al., Combined QCD and electroweak analysis of HERA data. Phys. Rev. **D93**, 092002 (2016). arXiv:1603.09628.

[6] ZEUS Collaboration, H. Abramowicz et al., Limits on the effective quark radius from inclusive *ep* scattering at HERA. Phys. Lett. **B757**, 468 (2016). arXiv:1604.01280.

[7] ZEUS Collaboration, H. Abramowicz et al., Measurement of the cross-section ratio $\sigma_{\psi(2s)}/\sigma_{J/\psi(1S)}$ in deep inelastic exclusive *ep* scattering at HERA. Nucl. Phys. **B909**, 934 (2016). arXiv:1606.08652.

[8] A. Bacchetta et al., Summary of workshop on future physics with HERA data. (2016). arXiv:1601.01499.

[9] A. Geiser, Possible future HERA analyses. (2015). arXiv:1512.03624.

[10] ROOT team, I. Antcheva et al., ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization. Comput.Phys.Commun. **182**, 1384 (2011).

[11] dCache team, dCache Documentation, 2016, http://www.dcache.org.

[12] H. Stadie et al., Monte Carlo mass production for the ZEUS experiment on the Grid. Nucl. Instrum. Meth. **A559**, 43 (2006).

[13] ZEUS Collaboration, H. Abramowicz et al., Production of Z^0 bosons in elastic and quasi-elastic *ep* collisions at HERA. Phys. Lett. **B718**, 915 (2013). arXiv:1210.5511.

[14] O. Zenaiev, Charm production and QCD analysis at HERA and LHC, Ph.D. thesis, U. Hamburg, Dept. Phys., DESY-THESIS-2015-012, 2015.

[15] SQLite team, SQLite Documentation, 2016, https://www.sqlite.org.

[16] M. Dobbs and J.B. Hansen, The HepMC C++ Monte Carlo event record for high energy physics. Comput. Phys. Commun. **134**, 41 (2001).

[17] HepMC3 team, HepMC3 Documentation, 2016, https://www.cern.ch/HepMC.

[18] T. Sjostrand et al., A proposed standard event record, LEP Physics Workshop Geneva, Switzerland, February 20, 1989. (1989).

[19] The CentOS project, CentOS home page, 2016, http://www.centos.org/.

[20] ZEUS Collaboration, J. Malka and K. Wichmann, The ZEUS data preservation project. J. Phys. Conf. Ser. **396**, 022033 (2012).