The ZEUS long term data preservation project: Monte Carlo generation and DP outside of ZEUS

Andrii Verbytskyi

ZEUS Analysis Forum
Hamburg,
February 2, 2016

Content

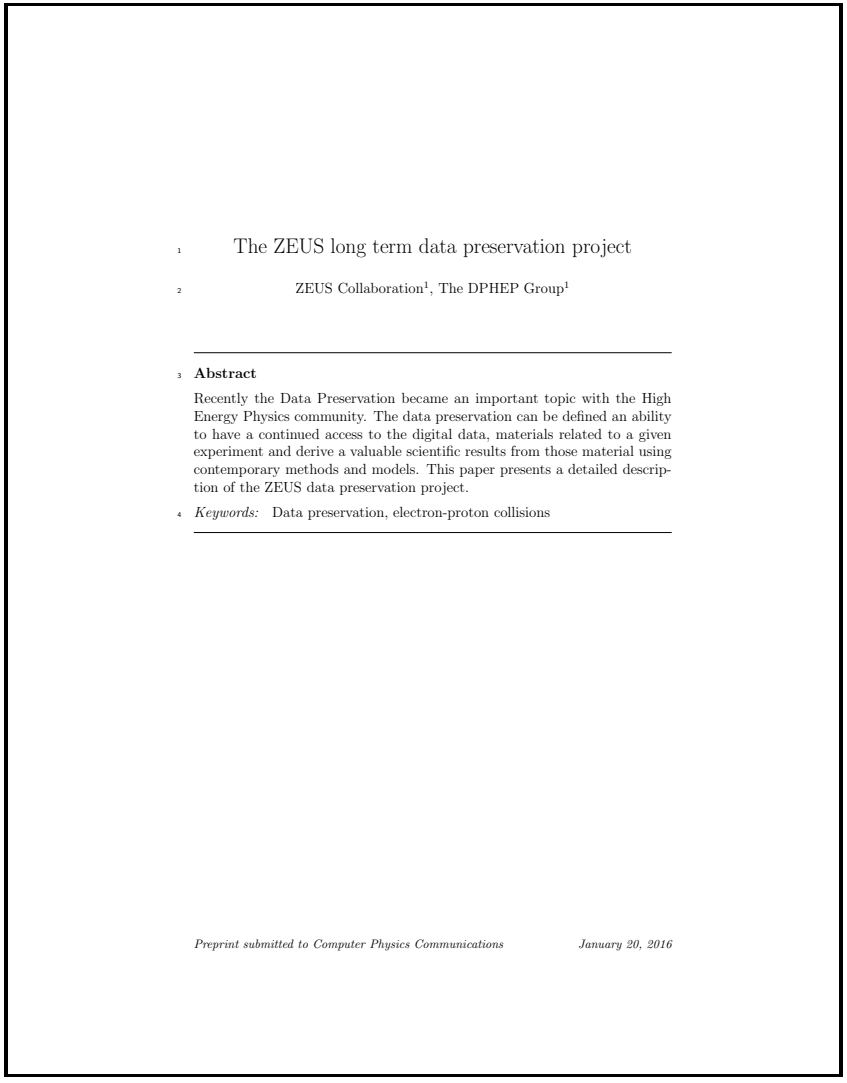
- Brief reminder about the paper, for details see Achim's talk.
- Monte Carlo generation capabilities (the interesting part).
- Data preservation outside of ZEUS (the boring part).

Data preservation paper for ZEUS

- Most data preservation papers are either conference proceedings or contain plans only $\rightarrow$ ZEUS paper will provide the **first** full description of the data preservation;
- Coverage:
  - Documentation on the data;
  - Experiment policy on data access and usage;
  - Manual for a possible analysis;
  - Manual for the MC generation, including new MC generators;
  - Statements on dedicated resources;
  - ...
- From the point of view of physics: **The document should contain enough information for an estimation of particular analysis opportunity with the preserved data.**

See an extensive description in Achim's talk!

Data preservation paper for ZEUS



The ZEUS long term data preservation project

ZEUS Collaboration[1], The DPHEP Group[1]

Abstract

Recently the Data Preservation became an important topic with the High Energy Physics community. The data preservation can be defined an ability to have a continued access to the digital data, materials related to a given experiment and derive a valuable scientific results from those material using contemporary methods and models. This paper presents a detailed description of the ZEUS data preservation project.

Keywords: Data preservation, electron-proton collisions

Preprint submitted to Computer Physics Communications *January 20, 2016*

So far 69 pages, 13 tables, 7 figures, 87 references, 19 listings.

Motivation for the MC generation

- Some analyses can be improved with new MC.
- Tests of new MC generators.
- Tests of analyses in new experiments.

What do we have

- Full MC simulation and reconstruction chain is preserved in ZMCSP, **ZEUS Monte Carlo Standalone package**. The input comes in ADAMO format, the output are the Common Ntuples. By J.M.&A.V.
- formoza, an utility that converts HEPEVT-like event records (ZEUS HEPEVT) to ADAMO format. By Y.G.
- HepMC3 library that can convert HepMC event records to other event records, including these readable by formoza. By HepMC developers and A.V.

Theory of MC generation: how does it work

- With an old generator:
  - Run it using the old ZEUS steering cards.
- With a new generator:
  - Set-up the generator, to produce events in HepMC format.
  - Convert the output to HEPEVT-like event record with HepMC3 library.
  - Convert the HEPEVT-like event records to ADAMO using formoza.
- Rename the resulting files according to the desired trigger periods.
- Use the renamed files as an input for ZMCSP package.

The detailed instructions for every step are in the Data Preservation paper and ZEUS Data Preservation page in MPP.
https://wwwzeus.mpp.mpg.de/dphep.html

Practice of MC generation: How does it work

- The first two samples were generated this summer. See Iurii's talk. It was found that the key to success is to **read and follow the instructions**.
- There is a lot of computing resources around, where it is possible to generate even large samples. 20-50-100M events is NOT a problem.
- In practice the generation with new MC generators is not much complicated, however only tiny samples were produced with SHERA 2.2 and blackhat 0.9.9.

Example: SHERPA2.2+Pythia8+blackhat0.9.9

```
(run){
  EVENTS 5000;
  # technical parameters
  NJET:=4; QCUT:=5; SDIS:=1.0;
  LJET:=2,3; LGEN:=BlackHat;
  ME_SIGNAL_GENERATOR Comix Amegic LGEN;
  EVENT_GENERATION_MODE Weighted;
  RESPECT_MASSIVE_FLAG 1;
  CSS_KIN_SCHEME 1;
  BEAM_1 -11 27.5; BEAM_2 2212 920;
  PDF_SET_1 None;
  # hadronization tune
  PARJ(21) 0.432; PARJ(41) 1.05; PARJ(42) 1.0; PARJ(47) 0.65; MSTJ(11) 5;
  FRAGMENTATION Lund; DECAYMODEL Lund;
}(run);
(processes){
  Process -11 93 -> -11 93 93{NJET};
  CKKW sqr(QCUT/E_CMS)/(1.0+sqr(QCUT/SDIS)/Abs2(p[2]-p[0]));
  NLO_QCD_Mode MC@NLO {LJET};
  Order (*,2); Max_N_Quarks 6;
  ME_Generator Amegic {LJET};
  RS_ME_Generator Comix {LJET};
  Loop_Generator LGEN;
  PSI_ItMin 25000 {3};
  Integration_Error 0.03 {3};
  End process;
}(processes);
(selector){
  Q2 -11 -11 4 1e12;
}(selector)
```

The output, 5000 events

Problems of MC generation

- The validation of output relays on the tests that were performed by J.M. It is recommended to do them for every particular MC generation setup. The MC input was deleted, but some samples survived $\rightarrow$ should be tested very soon.
- It was not discussed how to store the generated samples and documentation. Technically everything prepared for that. The plan is to produce a small CNINFO database for newly generated samples and store them in MPP (MPCDF). Very small amount of work.

Conclusions&TODOs

- From the technical point of view the data preservation for ZEUS is completed;
- TODO: discuss details, finish the documentation and do promotion.
- MC generation is an important and interesting feature in ZEUS DP.
- If you want a new MC – contact me and read the docs.

The boring part: What is happening outside of ZEUS?

The experiments

Known data preservation efforts:

- ALICE, ATLAS, CMS, LHCb, LHCf, MOeDAL: CERN
- ZEUS,H1, HERMES, HERA-B: DESY and MPP (ZEUS and H1 only)
- OPAL, ALEPH, DELPHI: CERN, MPP (OPAL), INFN (ALEPH)
- JADE: MPP
- CDF and D0: FERMILAB, INFN, Uni Mainz
- BELLE: KEK, MPP + more
- BABAR: SLAC
- BES: IHEP
- More experiments claimed at http://hep-project-dphep-portal.web.cern.ch The information on every project is very limited. In many cases these are **private communications**.
- Astroparticle and other experiments: MAGIC and CRESST at MPP. NOT A PART OF DPHEP!

DPHEP, the collaboration

- DPHEP study group paper – 2009
- DPHEP first paper – 2012
- DPHEP collaboration – 2014
- 8 dedicated workshops since 2009
- DPHEP sessions on CHEP, WLCG meetings, ACAT etc. The most recent is this week: DPHEP/WLCG workshop in Lisbon https://indico.cern.ch/event/444264/.
- Contributions to multiple workshops and conferences in physics and data science with a rate of 3-4 workshops per year.

DPHEP reading: looking closer

On the meetings on data preservation, quite often the following resources are referenced:

- http://opendata.cern.ch/
- http://hep-project-dphep-portal.web.cern.ch
- http://www.dphep.org/
- https://www.opensciencedatacloud.org/
- https://www.eudat.eu/
- +EU commission directives, US Department of Energy statements, WLCG guidelines, policies of funding agencies, management plans, etc.

Outcome

However, a search on keyword:"data preservation" in InSpire gives **27** entries.

- **21** are proceedings, the remaining 5 (should be papers) are
  - S. Amerio *et al.* [DPHEP Collaboration], "Status Report of the DPHEP Collaboration: A Global Effort for Sustainable Data Preservation in High Energy Physics," arXiv:1512.02019 [hep-ex].
  - A. Pace, "Technologies for large data management in scientific computing," Int. J. Mod. Phys. C **25** (2014) 1430001.
  - Z. Akopov *et al.* [DPHEP Study Group Collaboration], "Status Report of the DPHEP Study Group: Towards a Global Effort for Sustainable Data Preservation in High Energy Physics," arXiv:1205.4667 [hep-ex].
  - R. Mount *et al.* [DPHEP Study Group Collaboration], "Data Preservation in High Energy Physics," arXiv:0912.0255 [hep-ex].
  - C. Diaconu and D. South, "Study group considers how to preserve data," CERN Cour. **49N4** (2009) 21.
- only **5** items out of 27 have more than 10 pages, most contributions are really short.
- **There is no even one complete description of DP. Can ZEUS be the first one?**

Outcome: ZEUS items

- J. Malka [ZEUS and DESY DPHEP Group Collaborations], "The ZEUS data preservation project,"
- J. Malka *et al.* [ZEUS Collaboration], "The ZEUS data preservation project," J. Phys. Conf. Ser. **396** (2012) 022033.

Note: these are quite informative.

Best examples and featured examples on the market

- B. Aubert *et al.* [BaBar Collaboration], "The BABAR Detector: Upgrades, Operation and Performance," Nucl. Instrum. Meth. A **729** (2013) 615
- P.A. Movilla Fernandez [Jade Collaboration], "A guide to the resurrected JADE data and software," JADE Computer Note 103, unpublished (2003), http://wwwjade.mppmu.mpg.de/JCN103.ps.gz
- S.L. Lloyd *et al.* [OPAL Collaboration], "The OPAL Primer," unpublished (2004), https://opal.web.cern.ch/opal/manuals/oprimer/oprimer/oprimer.html

What about detector article on ZEUS?

Physics with preserved data?

It is hard to distinguish DP mode versus normal mode, but you know when you see it. Not a complete list, but it gives a hint:

- ZEUS and H1 publications can be considered.
- BABAR, BELLEI are active.
- LHC Run-I?
- OPAL and JADE at MPP:
  - N. Fischer *et al.* [OPAL Collaboration], "Measurement of observables sensitive to coherence effects in hadronic Z decays with the OPAL detector at LEP," Eur. Phys. J. C **75** (2015) 12, 571
  - J. Schieck *et al.* [JADE Collaboration], "Measurement of the strong coupling α_S from the three-jet rate in $e + e$- - annihilation using JADE data," Eur. Phys. J. C **73** (2013) 3, 2332
  - and more
- suspected DP articles/analyses: J. Cuth, K. Merkotan, M. Schott and S. Webb, "Determination of the Transverse Momentum of W Bosons in Hadronic Collisions via Forward Folding Techniques," arXiv:1512.03276 [hep-ex].
- The information on data usage in DP mode is limited.

Conclusions

- DPHEP is very active in planning, but the number of implemented plans is limited.
- The succesfull cases were outside of DPHEP.
- ZEUS can be presented as a good example of *actually doing* preservation being a part of DPHEP.