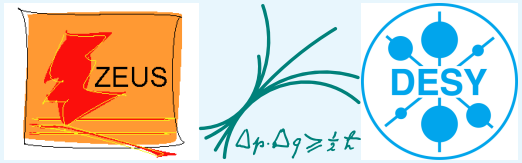


The ZEUS long term data preservation project

Andrii Verbytskyi on behalf of the ZEUScollaboration

24th International Workshop on Deep-Inelastic Scattering and Related Subjects
Hamburg,
April 13, 2016

HERA data preservation motivation

Future data (re-)analysis with new models and new approaches.

Modelling for the future experiments.

HERA reminder:

The only $e^{\pm}p$ collider, 1991-2007;

$27.5\,GeV$ $e^{\pm}$; $460, 575, 820, 920\,GeV$ p;

(Un)polarized $e^{\pm}$ collide with p;

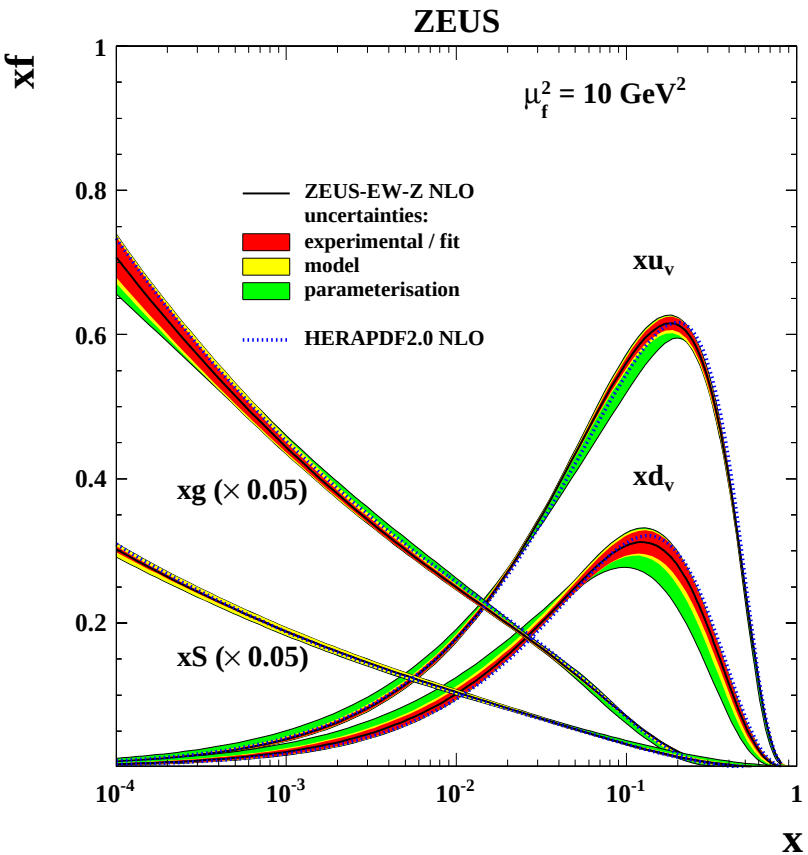
Polarized $e^{\pm}$ collide with $H/D/.../Xe$ targets;

p collide with nuclear targets.

When we say “data preservation”:

We mean:

- Data bits,
- Software,
- Experiment documentation,
- Data preservation policies and documentation



ZEUS DP:
10 papers since 2014
3 this year!

should be preserved in a way that allows **production of valuable physics results.**

Note: most of the results of recent yeas are produced in Data Preservation mode.

When we say “data preservation”:

We want results!

We are interested in: **Science** – physics.

We are not interested in: **Tools** – files, storages, clusters, systems etc.

The talk will describe some tools. . . but only with an intention to show how these **serve science.**

Use cases for HERA data

Something that now we are not aware about.

QCD:

- Proton structure, e.g. F_2 and F_L, strangeness in the proton;
- Diffraction, e.g. combination of measurements;
- Jets and event shapes with NNLO;
- Photon structure, instantons, pentaquarks, etc.

EW physics:

- Prompt photons;
- Electroweak couplings.

See arXiv:1601.01499 and arXiv:1512.03624 for details.

We discuss data preservation for ZEUS:

Experiment documentation;

Data bits;

Software and environment;

Data preservation policies and documentation.

ZEUS reminder:

1991-2007; p collide with (un)polarized $e^{\pm}$;

$27.5 GeV$ $e^{\pm}$; $460, 575, 820, 920 GeV$ p;

$0.5 fb^{-1}$/360M events of data.

Data preservation for ZEUS: Experiment documentation

Public:

251 ZEUS papers (1992-2016) are in the scientific databases (journals, libraries, arXiv, InSpire);
ZEUS related thesis are public and listed on InSpire;
Web-based documentation is in plain HTML on the `http://www-zeus.desy.de/`.
Paper documentation is stored in DESY library.

Internal:

ZEUS notes are in DESY library and password-protected on InSpire;
Web-based documentation is in plain HTML on the `http://www-zeus.desy.de/`. Internal presentations and Monte-Carlo documentation is included;
Paper documentation is stored in DESY library.

Updates for both:

The documentation is updated with recent papers, notes, etc.

Data preservation for ZEUS: Bits location

Data and MC is stored in DESY and MPCDF[1];

Data is preserved in a form of ROOT (and PAW) ntuples. No custom software is needed to read it;

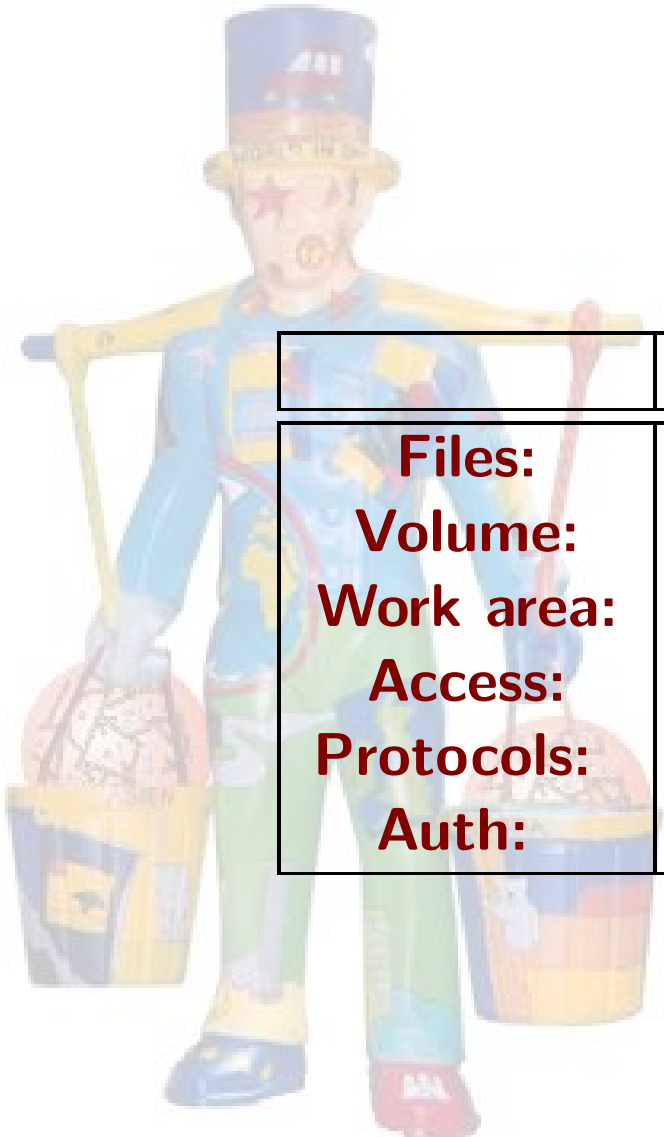
Data from MPCDF is accessible via multiple protocols worldwide.

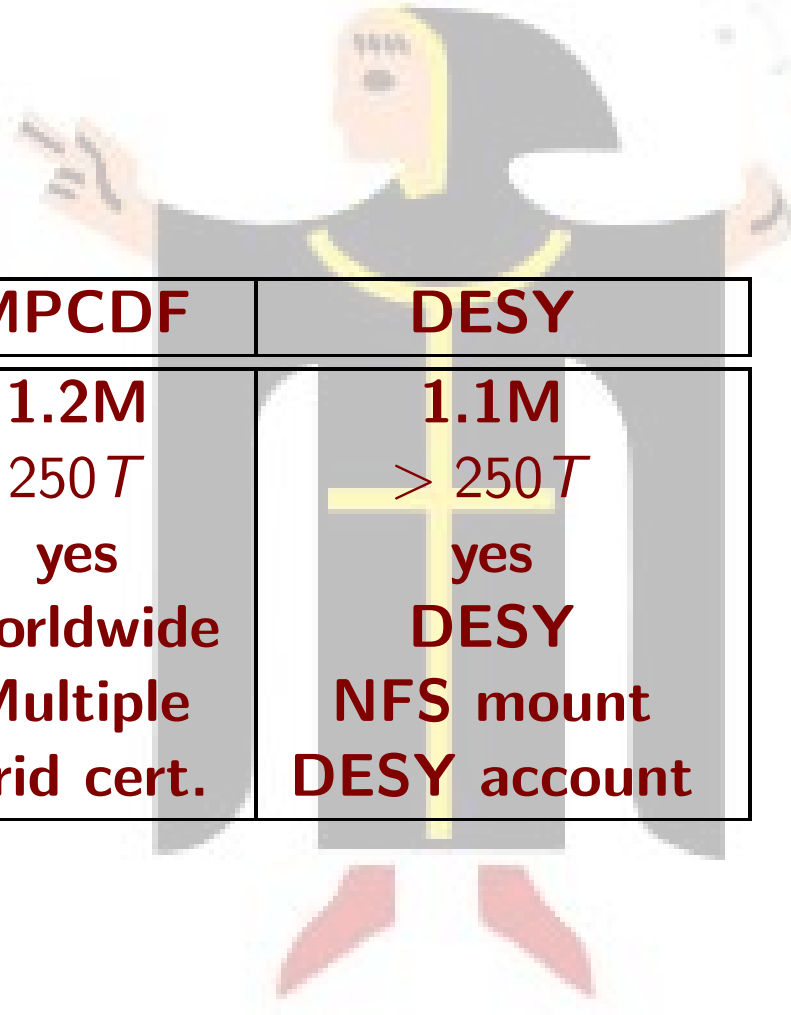
Data from DESY is accessible on DESY cluster/internal network.

Data is listed in text-based and SQLite3 databases;

[1] Max-Plank Computing and Data Facility, Garching bei München, D-85748

Data preservation for ZEUS: Bits statistics





	MPCDF	DESY
Files:	1.2M	1.1M
Volume:	$250\,T$	$> 250\,T$
Work area:	yes	yes
Access:	Worldwide	DESY
Protocols:	Multiple	NFS mount
Auth:	Grid cert.	DESY account

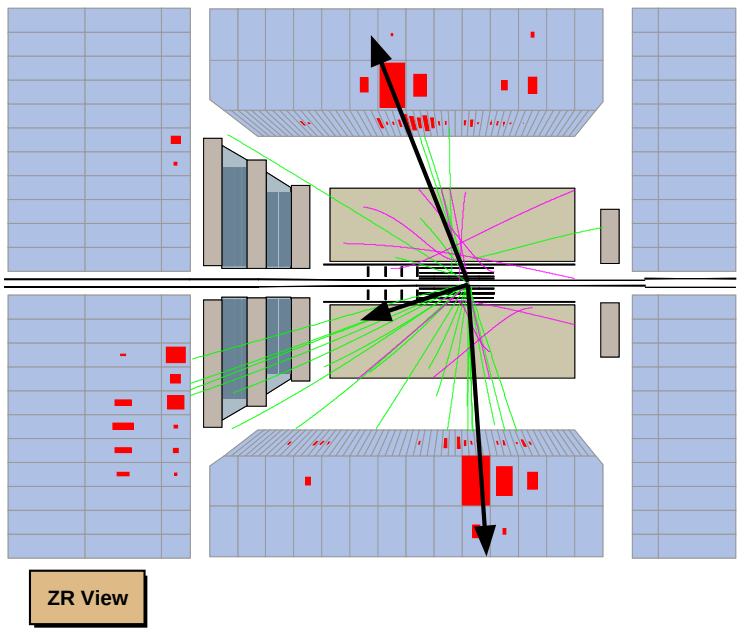
Data preservation for ZEUS: Software

Main software for the analysis is vanilla ROOT.

Additional software includes:

- ZEVIS, the event display based on ROOT;
- CNINFO, the event data base, based on ROOT and SQLite3;
- ZMCSP Monte-Carlo standalone generation packages – see next slides.

+any ROOT extension that will work for you...



ROOT
Data Analysis Framework

SQLite <Your favourite software>

Data preservation for ZEUS: Software environment

A certain environment is needed for the analysis. As of 2016 the demands are low and easy to fulfil:

DESY provides an access to a batch computing cluster.

In parallel:

Virtual machines(VM) looks like a very attractive **long-term** solution;

The way other experiments (LEP/LHC) are going.

Because of very generic requirements it is foreseen that both environments will remain functional for a long time.

Data preservation for ZEUS: Software environment/VM

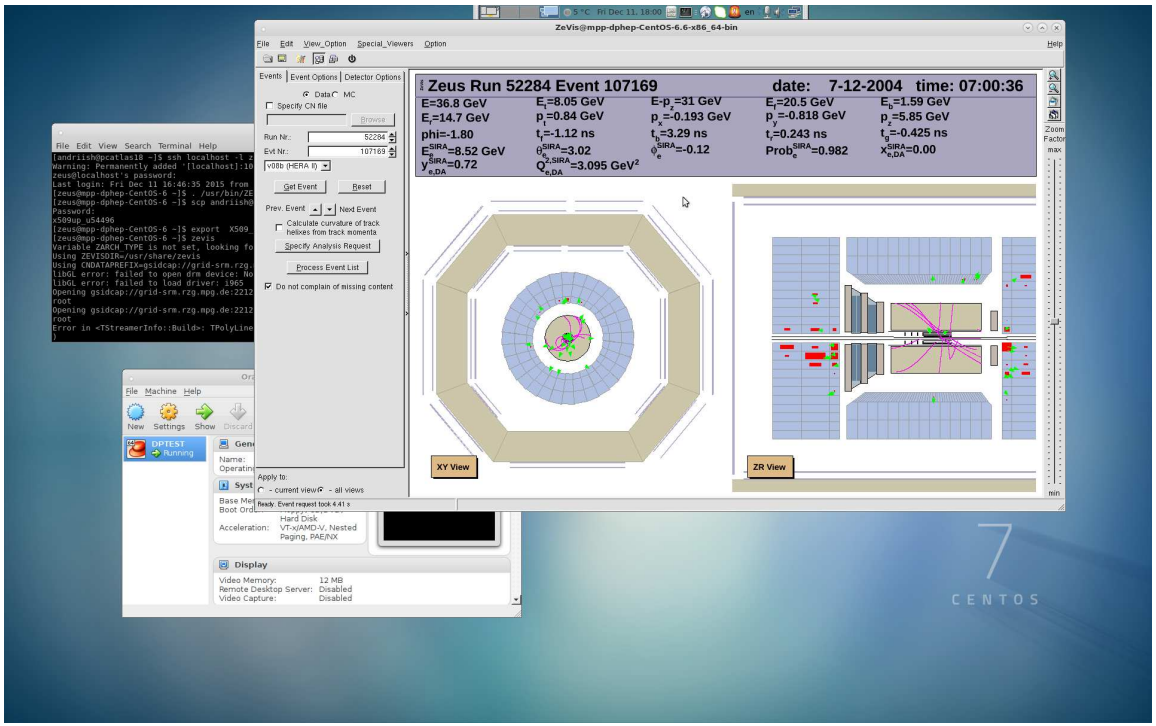
Virtualization for ZEUS has a two-fold purpose: it provides **benchmark** environment that suppose to work for a long time and it can be used, if desired, as a **super-portable production environment**.
It is based on DVD ISO image with SL6 and all software. It has options for:

- Automatic install on virtual or real hardware;
- Customisation, root privileges, etc.;
- Unlimited number of installations $\rightarrow$ potentially usable on clouds;
- Usage not restricted to any laboratory or virtualization software. Can run anywhere.

It is not necessary to use it in the production if something more productive like institute's cluster is working for you.

Data preservation for ZEUS: Software environment/VM

ZEUS software: ROOT, MC simulation, event display, file catalogue, setup scripts etc.
Modern MC generators, FastJet, cernlib, PAW, Rivet and other popular and “not really” packages.
Anything you will want to install...
Agree access and download it.



Data preservation for ZEUS: Policies

Should define how to:

- Access the data;
- Work with data in a proper way;
- Publish valid results;
- Set up a set of rules to solve possible tensions.

So far the details on policies of data usage can be clarified with the spokesperson.

Data preservation for ZEUS: The documentation

Coverage:

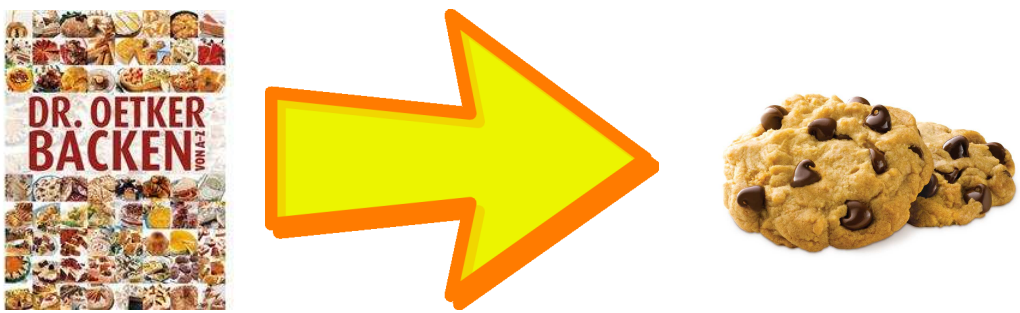
- Documentation on the data;
- Experiment policy on data access and usage;
- Manual for a possible analysis;
- Manual for the MC generation, including new MC generators;
- Statements on dedicated resources;
- ...

From the point of view of physics: **The documentation with enough information for an estimation of particular analysis opportunity with the preserved data.**

Data preservation for ZEUS: The MC recipe

Motivation:

- Some analyses can be significantly improved with new MC;
- New MC generators/models can be tuned;
- New experiments can use it for the technical studies.



Data preservation for ZEUS: The MC recipe ingredients

An instruction is prepared how to generate events with old ZEUS generators.

An interface that reads most common HEP event records and transforms them to input for ZEUS MC production chain has been created. This is an option for the modern generators. See an example card for SHERPA2.2+blackhat0.9.9 in the backup.

ZMCSP (ZEUS Monte Carlo Standalone Package) is a tarball with all the software needed for the reconstruction of MC simulated events. It has no external dependencies, runs on modern Grid clusters, virtual machine, a laptop[2]. On the Grid it can produce 50-100M events[3] per week. Supplemented with example of scripts and documentation.

[2] Not on a Mac, of course!

[3] ZEUS has 360M of data events

ZEUS Data Preservation summary

Data is accessible in DESY and MPCDF.

Documentation is stored in DESY library/Inspire/web-server;

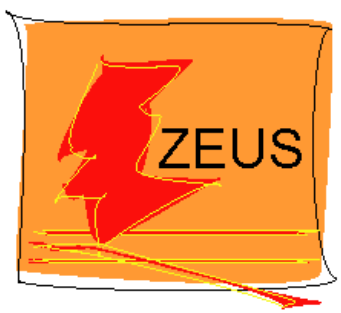
Analysis requires only standard software;

An option for MC production with new and old MC generators exists;

Virtualization+two dedicated DESY machines+ access to batch cluster.

Policy statements and documentation on Data Preservation in development.

New collaborators are welcome to analysis!



H1 data preservation

(Un)polarized $e^{\pm}$ are collided with protons.

- H1 data is accessible in DESY and MPCDF(most).
- Some documentation is on tapes only.
- Documentation is stored in DESY library/Inspire/web-server.
- Has an option for full re-reconstruction.
- Requires ROOT and corresponding custom classes.
- MC production is possible.
- Two dedicated DESY machines and access to DESY batch cluster.
- Grid VO resources.

HERMES data preservation

Polarized $e^{\pm}$ are fired at polarized H/D and unpolarized $H/.../Xe$ targets.

- HERMES data is accessible in DESY.
- Some data is on the tapes only and is not accessible directly.
- Documentation is stored in DESY library/Inspire/web-server.
- Has an option for full re-reconstruction.
- Custom C/FORTRAN or C++ analysis framework.
- MC production is possible.
- Two dedicated DESY machines and access to DESY batch cluster.
- Grid VO resources.
- Use cases are described in arXiv:1601.01499.

BACKUPS

Example: SHERPA2.2+Pythia6+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)
```

An example close to the one in SHERPA manual.