

Data Preservation Activities at DESY.

Katarzyna Wichmann and David South on behalf of the DESY-DPHEP Group



Software Validation and Status of Experiments Data Preservation Projects

In order to substantially extend the lifetime of the analysis capability (see blue box below) it is beneficial to migrate to the latest software versions and technologies for as long as possible. In collaboration with DESY-IT a framework has been developed to automatically test and validate the software and data against future changes and upgrades to the environment, as well as changes to the experiment software itself. (See a dedicated poster about the sp-system). The ZEUS Collaboration aims for a data preservation level between 3 and 4, preserving its data in a form of common ntuples and planning to maintain the ability of simulation of new Monte Carlo samples after the end of the current analysis model and the current MC production system. The H1 and HERMES Collaborations aim for the data preservation level 4, meaning keeping functional the full software chain from raw data to analysis level. The focus of is on the full migration to 64-bit starting with SLD5/64-bit. It is also very important to test the stability of physics analyses against future changes of the software environment. It is done with the same dedicated virtual sp-system.

Summary of present status of each collaboration DP project:

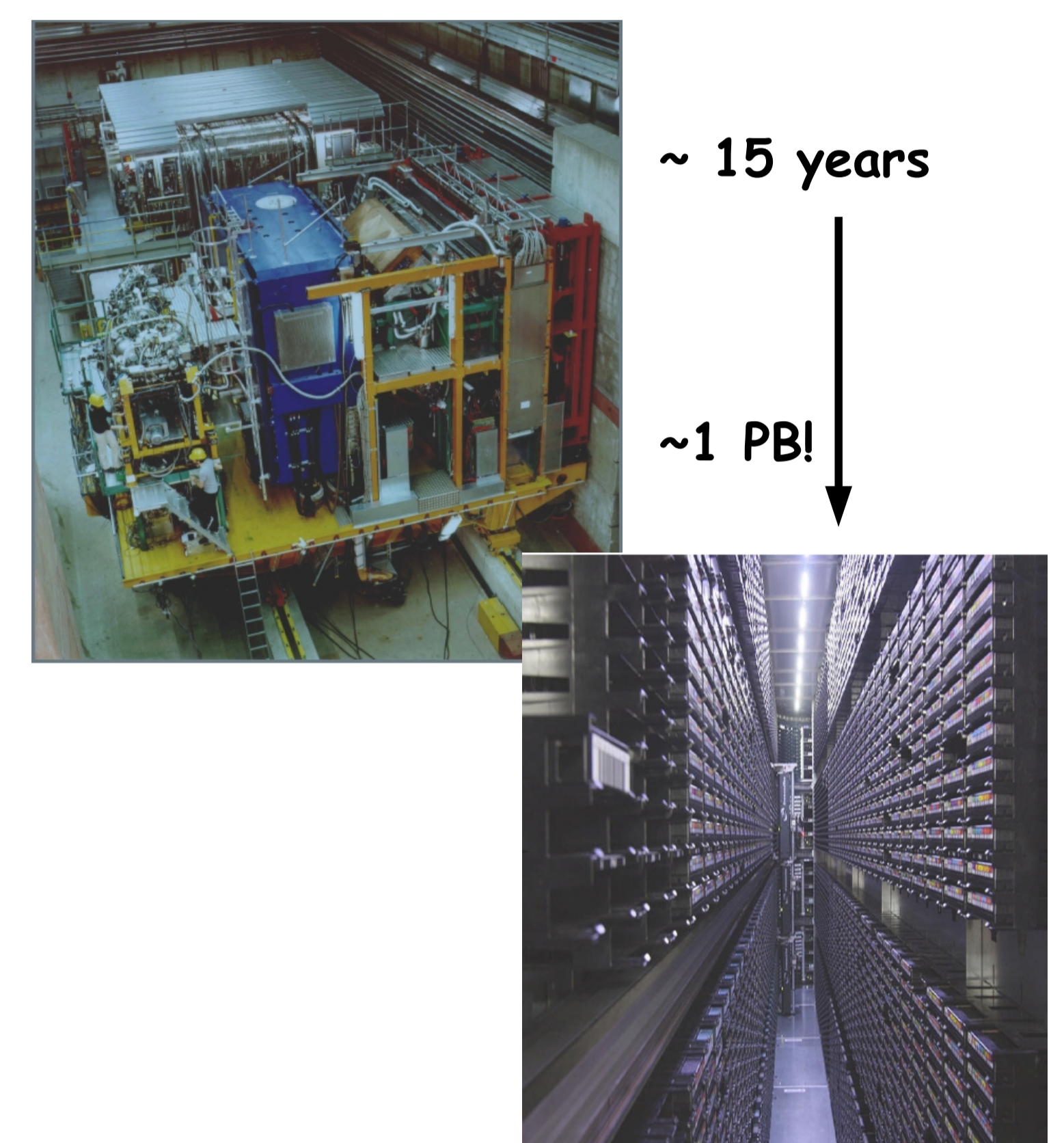
- Pre-compiled SL5/32bit software runs without problems
- Validation of stand-alone MC and common ntuple production ongoing
- First physics validation scripts implemented
- Final DST and analysis level data sets for DP produced
- Reconstruction & analysis software compiles in SL5/64bit
- Pre-compiled software successfully running on SL6/64bit
- Implementation of full chain of validation scripts chain ongoing
- Transition to central IT resources in 2nd half of 2014
- All experiment software successfully compiled
- Validation of results ongoing
- Production of Adamo based μ DST required for 64bit
- Archival mode in use already from 2013 on

Operating System	SL5 32bit				SL5 64bit				SL6 64bit	
	External Dependencies	ROOT	Cernlib	Fantjet	Neuro-0312	Neuro-3.3.0	Neuro-3.3.0	Neuro-3.3.0		
Process	5.26	5.28	5.30	5.32	2005	2006	2.3.3	2008	0312	Neuro-3.3.0
Accessing cNtuples (Data/MC)										
Creating cNtuples (Data/MC)										
ZMCSP (simulate/reconstruct MC)					No dependence					
Validation										
Compilation of s/w										
Generating MC files										
Producing DST files (Data/MC)										
Producing h1oo files (Data/MC)										
Accessing h1oo files (Data/MC)										
Accessing ndb snapshot										
Validation										
Compilation of s/w										
Accessing μ DST (precompiled s/w)										
Reconstruction (precompiled s/w)										
Producing μ DST (precompiled s/w)										
Validation										

ok ongoing to be done

Data Storage

Experimental data collected and simulated through years of operation have been stored in formats of various complexity. During active years data have been stored on local NFS filesystems as well as duplicated on tapes. In the Data Preservation phase all local resources are preferably transferred to lab-supported platforms. For each level a corresponding storage type is being developed by DESY-IT to assure prompt access and safe storage based upon the usage type.

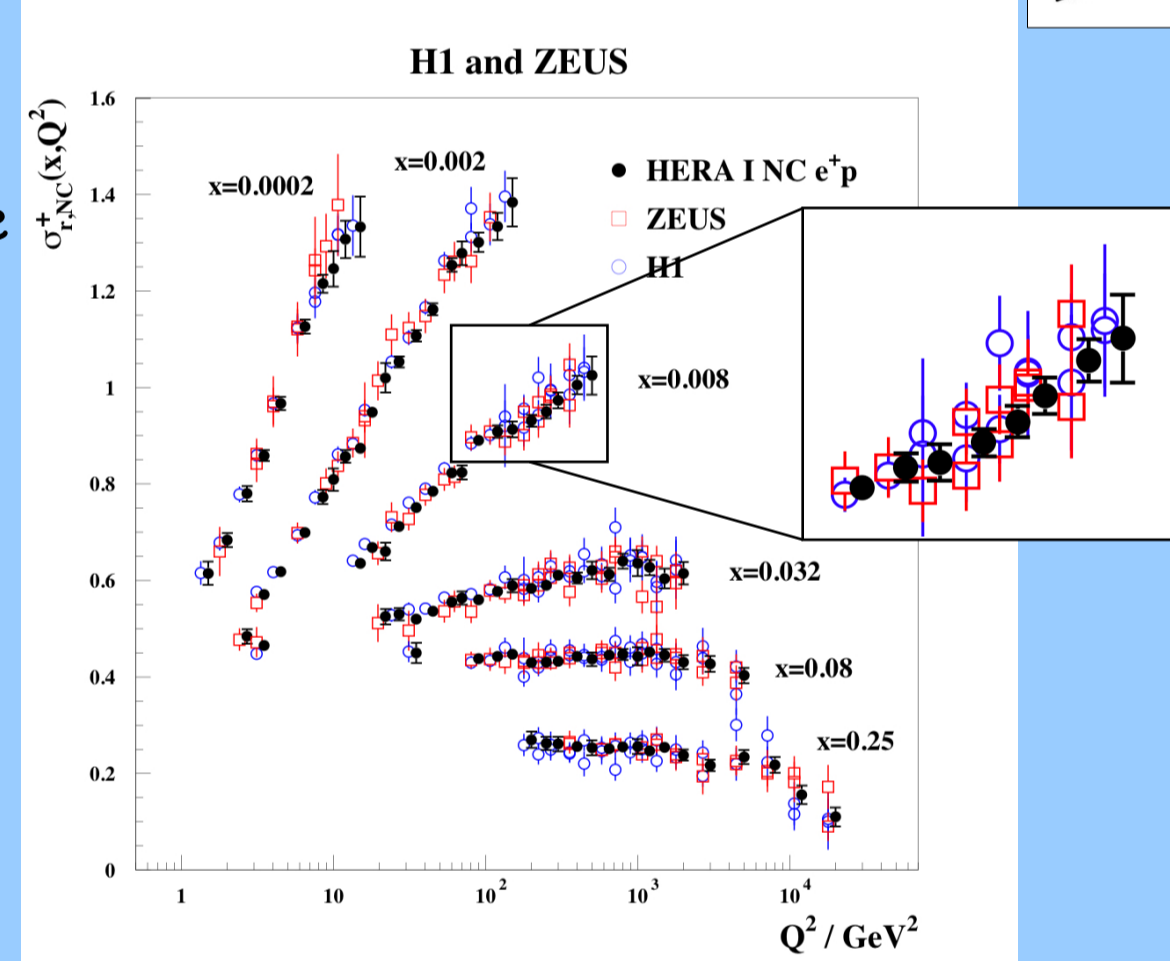


~ 15 years

~1 PBI

Why Preserve HERA Data?

- HERA data represent a unique achievement in HEP
- The potential of the data can be further exploited:
 - developments in theory and experimental methods
 - combined H1 & ZEUS results
- Possible use of new observables
- Some measurements dominated by uncertainties in theory



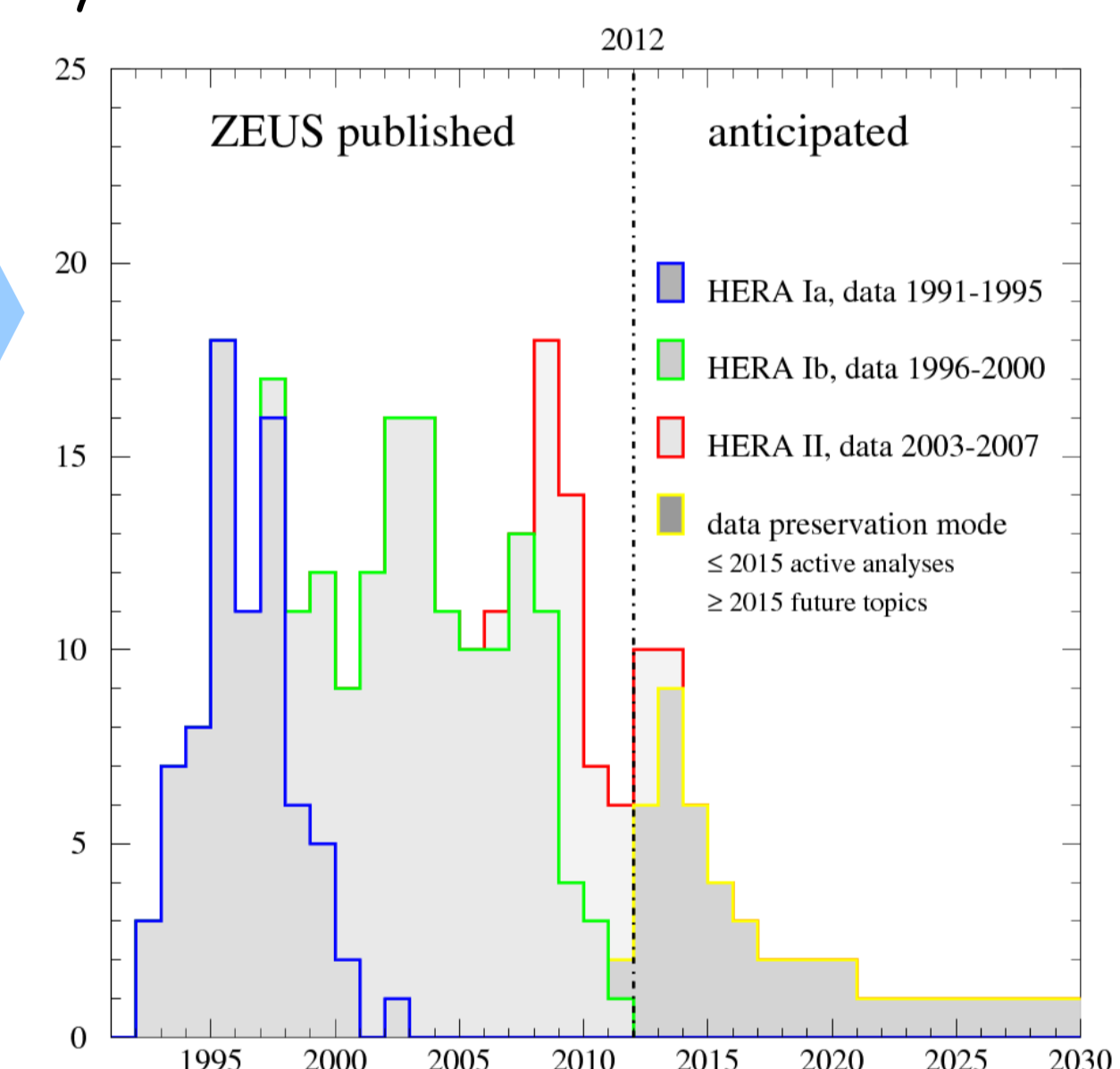
DPHEP HEP Data Preservation Project

High Energy Physics data collected by experiments are crucial to our understanding of particle physics. The DP effort aims to ensure long-term availability of these data after the end of the experimental collaborations. DP increases the physics potential of experiments

- Long-term data analysis
- Re-using and re-analyzing data
- Combining results between experiments
- Education, training and outreach

ZEUS Common Ntuple Project

The ZEUS current analysis software is used to create common usage ntuples (real and MC data) with content wide enough to incorporate all possible physics analyses.



- Superset of all potential ntuples for (almost) all possible physics analysis
- Simple ROOT ntuple format is used - flat ntuples

Preservation Model	Use Case
1 Provide additional documentation	Publication related info search
2 Preserve the data in a simplified format	Outreach, simple training analyses
3 Preserve the analysis level software and data format	Full scientific analysis, based on the existing reconstruction
4 Preserve the reconstruction and simulation software as well as the basic level data	Retain the full potential of the experimental data

All HERA Collaborations aim for data preservation level 3 or 4



Digital and Non-Digital Documentation

Great care is being taken to preserve as much as possible of the various documentation collected over the years of running of the experiments. Non-digital documentation is being collected, stored & cataloged

- non-digital documentation sorted and safely stored in DESY library archive
- part of the non-digital documentation digitized (eg. H1 Virtual Archive)



H1 Virtual Archive

- Physical Location of the H1 Archive
- Plenary Meetings
- Physics Meetings
- Thursday Weekly Meetings
- Collaboration Board (CB), Executive Committee (EC) Meetings
- Physics Working Groups
- Diploma and PhD Theses
- Trigger Meetings
- H1 Detector
 - General Documents (LoI, Proposal, Progress Reports...)
 - H1 Operation Meeting
 - H1 CDAQ Logbooks
 - H1 and HERA Startup (Various internal documents)
- Meetings of various DESY committees: Minutes, talks, documents

Original digital documentation:

- Former online monitoring and shift tools
- Web-based documentation, electronic logbooks, presentations, minutes...



Inspire gives the unique opportunity to conserve documentation, wikis, news forums and even data outside collaboration resources and keep it available and undisturbed "forever"

- Inspire offers many convenient options for digitized documents archiving
 - Internal notes and preliminary results (password protected), thesis list
 - Curator accounts - operated by experiments

ZEUS Preliminary Notes

212 records found 1 - 100

- Charm fragmentation fractions in PHP (In English)
 - ZEUS Collaboration, 2012.
 - ZEUS-pre-12-003
 - References | BibTeX | LaTeXJUS | LaTeXJEU | HarvMacs | EndNote
 - Link to website for CDS/2012: Link to Summary table for fractions
 - Detailed record - Attribute this paper
- Charm production in DIS using inclusive secondary vertices and extraction of F2cc (In English)
 - ZEUS Collaboration, 2012.
 - ZEUS-pre-12-002
 - References | BibTeX | LaTeXJUS | LaTeXJEU | HarvMacs | EndNote
 - Link to Public Info
 - Detailed record - Attribute this paper
- Search for elastic Z0 production in ep collisions at HERA (In English)
 - ZEUS Collaboration, 2012.
 - ZEUS-pre-12-001
 - References | BibTeX | LaTeXJUS | LaTeXJEU | HarvMacs | EndNote
 - Link to Full Text Link to ZEUS Internal Info

INSPIRE H1 Internal Notes

619 records found 1 - 25

- NLO calculations for diffractive dijets.
 - S. Schecterson, H1-09/07-428.
 - References | BibTeX | LaTeXJUS | LaTeXJEU | HarvMacs | EndNote
 - Link to Fulltext
 - Detailed record - Attribute this paper
- Improved Description of Electromagnetic Showers in H1FAST.
 - J. Marks M. Eberbrock, C. Rieker, C. Schwannberger (HDB1, MPIM, DESY), H1-01/05-617.
 - References | BibTeX | LaTeXJUS | LaTeXJEU | HarvMacs | EndNote
 - Link to Fulltext
 - Detailed record - Attribute this paper
- H1 Calorimeter DAQ Upgrade for HERA-II.
 - Pierre-Yves Duval Dirk Hoffmann, Claude Vallejo (CPPM), H1-07/02-599.
 - References | BibTeX | LaTeXJUS | LaTeXJEU | HarvMacs | EndNote
 - Link to Fulltext
 - Detailed record - Attribute this paper