

Long Term Data Preservation in High Energy Physics

The summary of the summary

Patrick Fuhrmann

Motivation and first consequences

- HERA experiments need/want to allow for long term analysis of experiment data.
- This requires :
 - Determining the value of the data depending on time, which mainly depends on future experiments.
 - Safely storing data for some decades.
 - Documentation of the used format or storing self-describing formats on media.
 - Maintenance/porting of reconstruction and analysis software.
 - Public accessibility, Licenses, copyright and ownership of future publications.
 - Some FTE (to be determined) for a long time. (permanent position, money)
 - Determining the price per year of doing all this.
- Consequence : Initiating a working group of HEP groups to profit from synergy.



The Group

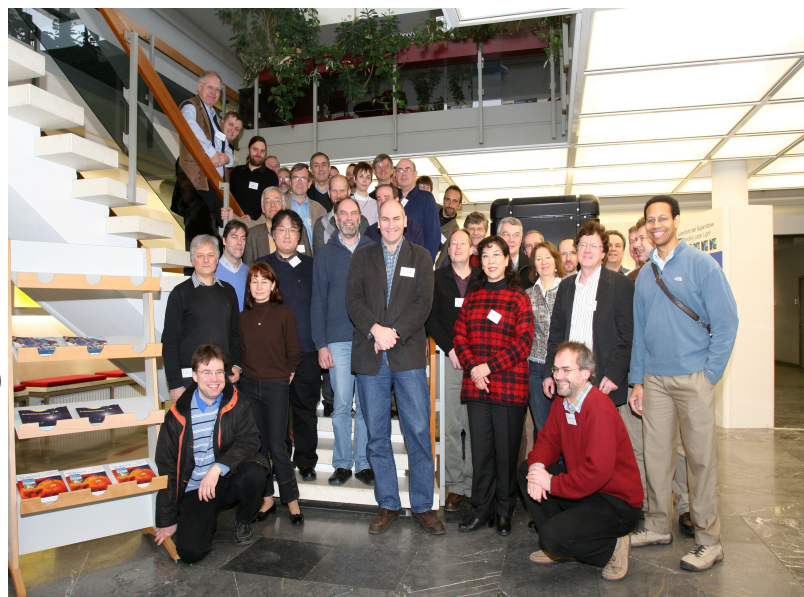
Study Group for Data Preservation and
Long Term Analysis in High Energy Physics.



www.dpHep.ORG

With prominent members of (from) :

- › SLAC (BaBar, SPIRES)
- › FERMILab (D0, CDF, Computing, DOE)
- › DESY (IT, H1, Zeus, Hermes)
- › KEK, Belle
- › IN2P3 (IT)
- › gridKa
- › CERN (ROOT, IT, LEP, L3, PARSE, ALEPH)
- › IHEP
- › BES III
- › LEO
- › RAL



- WP1 : Physics case for data preservation in HEP
 - Survey of possible benefits from data preservation
 - Including business models
 - Including links with other research fields
- WP2 : Preservation Models.
 - ee, ep, pp input
 - Priorities, costs and benefits, link to technologies
- Collaborations, Governance and Data access policies
 - Including contacts with general initiatives
- Technologies and facilities
 - Survey and assesement of existing infrastructures in HEP and their adaptability to data preservation requirement
 - Reflection on the impact of the new technologies on the data preservation methods

Agenda :

- Report from experiments.
- Computing centers and technologies.
- Workshop discussions : ee, pp, ep
- Reports from past experiments on data preservation.
- Open access and long term collaborative governance.
- Options for long term data analysis.

Results :

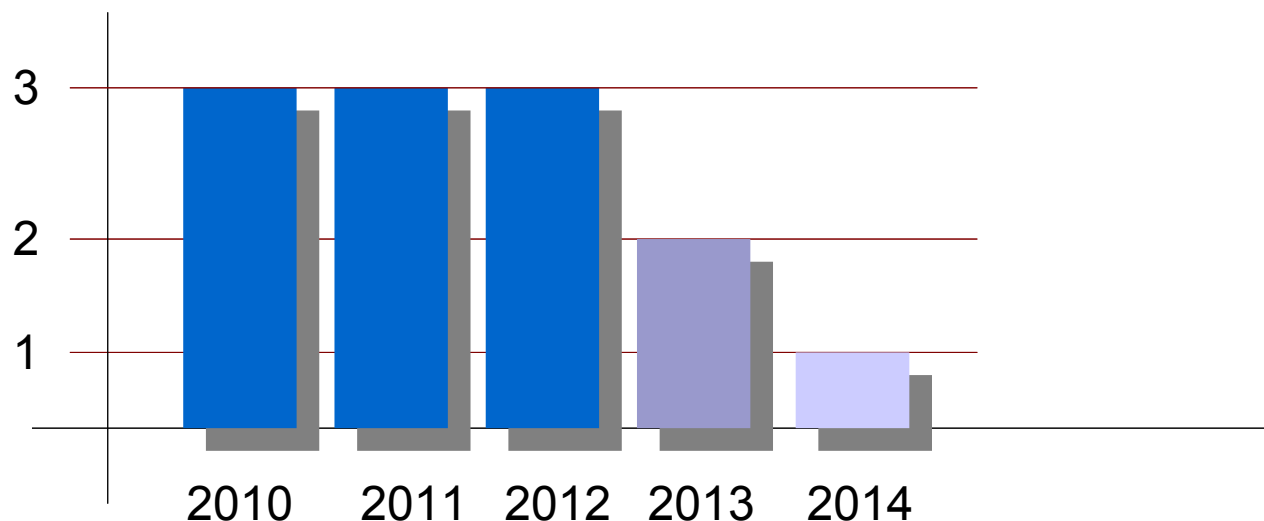
- Executive Summary (see www.dphep.org)
- CERN courier article (draft)
- Proceedings (Due march 3)
- Next meeting June/July at SLAC
- **First concrete plans for the HERA experiments.**

Discussions between ep experiments

Source : Presentation by David South

- Possible future use cases.
- Models for preservation.
- H1 versus Zeus : a common repository
- Certification, Validation models.
- Risk analysis, unpredictable events.
- Infrastructure planing.

FTE's for solution development (mainly for HERA experiments)



Make one position permanent at DESY IT for data preservation and for enabling long term analysis.

DESY IT - Data Custodian:

Someone permanently taking care of the HERA data, making sure that it still fulfills the requirements and taking action in case the "validation system" (a deliverable of the development project) signals a potential problem with the preserved data or software.

- HERA experiments need to plan for long term data analysis.
- International working group, including many more HEP experiments should coordinate efforts and make use of synergy effects.
 - DPHEP.org
- Within the 'dphep' group, the HERA experiments already presented a concrete plan on how to investigate in the various difficult areas of data preservation and long term analysis.
- For about 5 years, investigation and development should dominate the effort, followed by the actual preservation work.
- Sustained funding including permanent position(s).
- DESY IT already involved in the project.