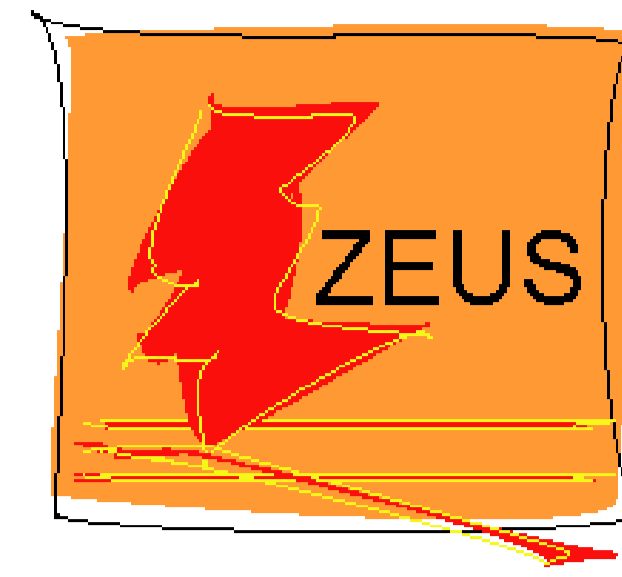


# Data Preservation in ZEUS Collaboration.



J. Malka\*, K. Wichmann\* of behalf of the ZEUS Collaboration

\* Deutsches Elektronen-Synchrotron, Hamburg, Germany



## DPHEP HEP Data Preservation Project

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

Data preservation increases the physics potential of experiments

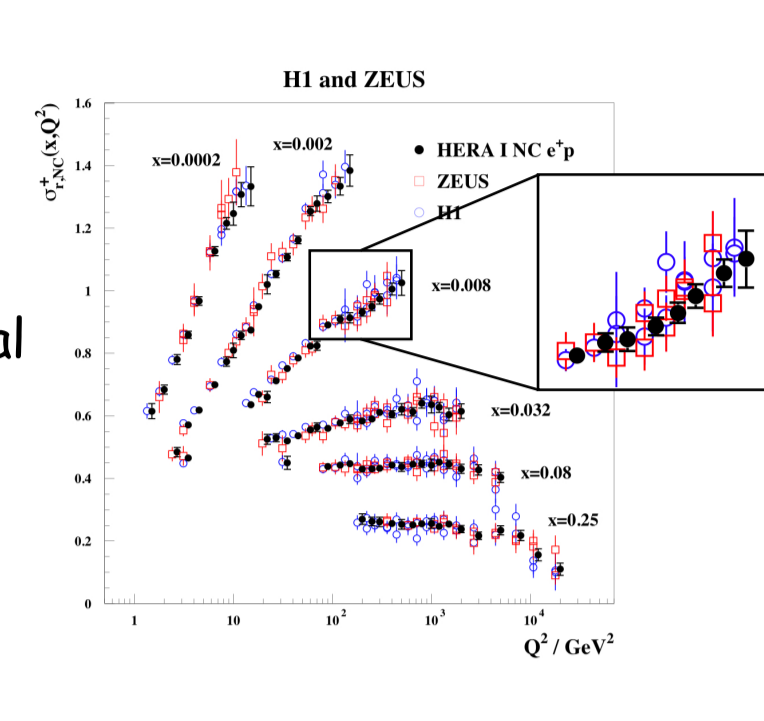
- Long-term data analysis
- Re-using and re-analyzing data
- Combining results between experiments
- Education, training and outreach

Preservation Model	Use Case
1. Additional information	Publication related information
2. Provide data in simplified format	Outreach, training
3. Preserve the analysis level software and data format	Full scientific analysis possible, based on existing reconstruction
4. Preserve the full simulation and reconstruction software as well as the basic level data	Retain the full potential of the experimental data

Increasing complexity, benefit and cost

## 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
  - combination of H1 and ZEUS results
- New observables may be used in the future
- Some measurements dominated by uncertainties related to the theory



## Future ZEUS Analysis Model – Beyond 2012

**ZEUS Collaboration aims for Preservation Model 3 & 4:**

- Preserve Data and MC samples in Common Ntuple (CN):
  - CN project started in 2006
- In addition maintain the ability of simulation of new MC after the end of the current analysis model and the current MC mass production system

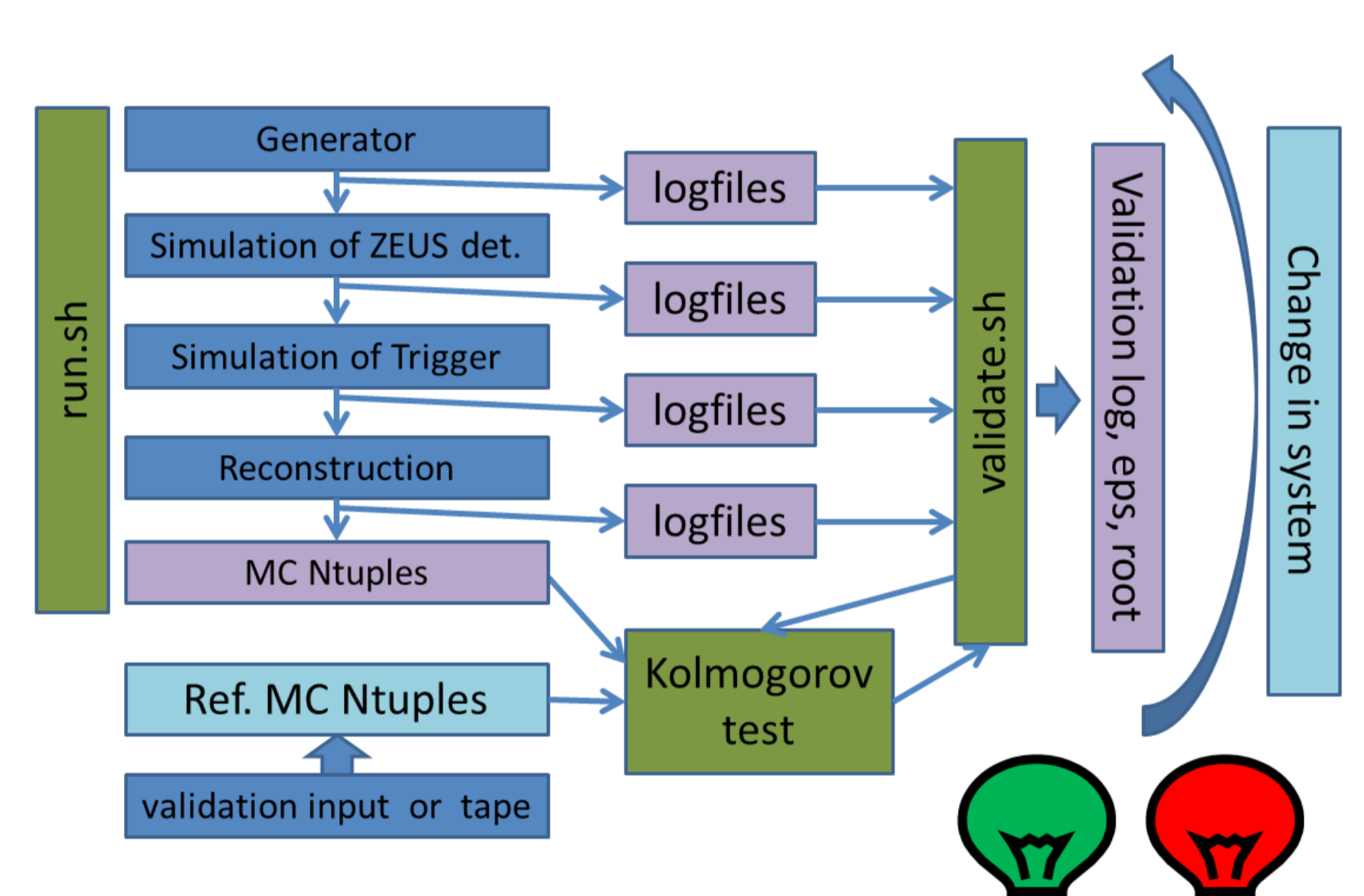
## ZEUS Common Ntuple Project

Old ZEUS analysis model based on MDST (Mini Data Summary Tape) contains a lot of external dependencies and cannot be maintained after 2012 when most of the manpower is going to be reduced. Therefore 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 (no object, no histograms)
- The resulting total ntuples size is expected to be between 10-20% of the size of MDST data
- The storage and access is unchanged with respect to the current model (tapes and dCache)

## Standalone MC Simulation Package

The ZEUS Collaboration plans to maintain the ability of Simulation of new MC after the end of the current analysis model and the current MC mass production system. It will contain the full chain from existing precompiled MC generators or generic MC interface (like HepMC standard format) to simulation of the detector to Common Ntuple production. All dependencies are included: calibration, conditions, alignment, geometry, executables, steering cards; all unnecessary dependencies are removed. The core of the package is based on the current MC production scheme on GRID.



Virtualisation techniques will be involved - a crosscheck against changes in operating system and other software changes.

## Validation of MC Simulation

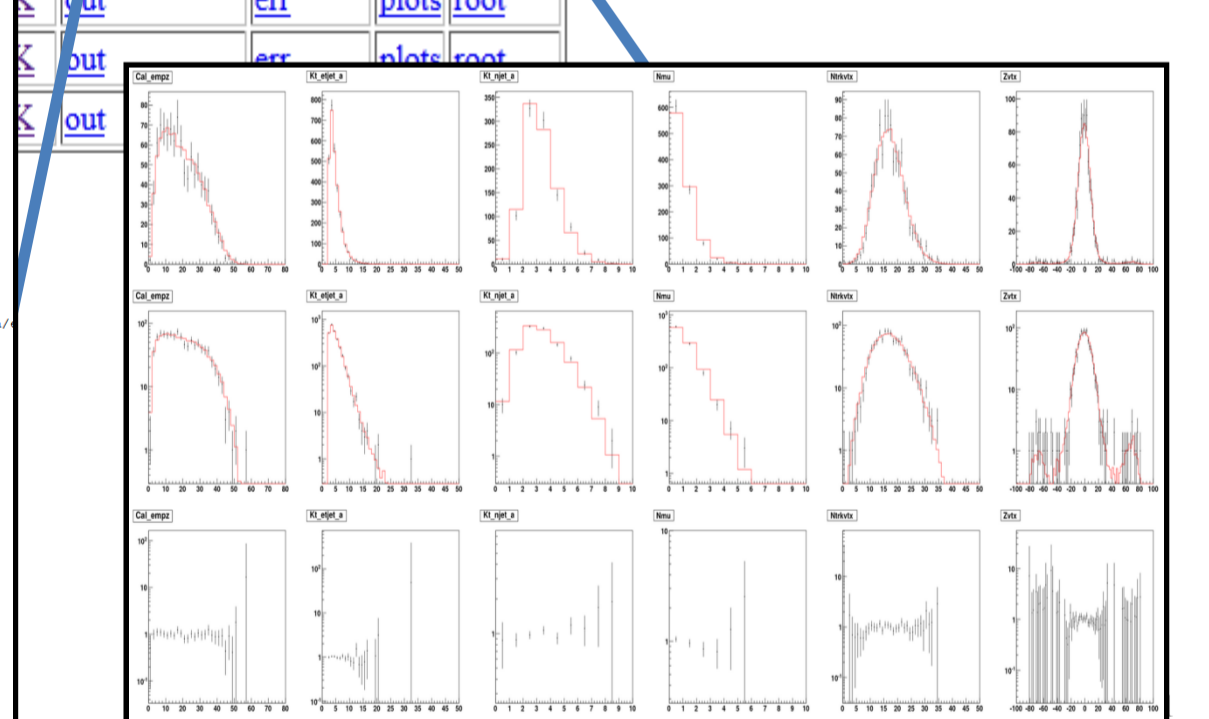
In order to make sure that changes of the operating system and/or ROOT version do not change the physics content of the simulation output, a dedicated validation test has to be established.

Validation is done with simple tools to compare different MC simulation releases based on current MC validation tools:

- There is a basic set of histograms for comparison
- Need to develop root based validation tools for MC and software packages on the level of physics analysis

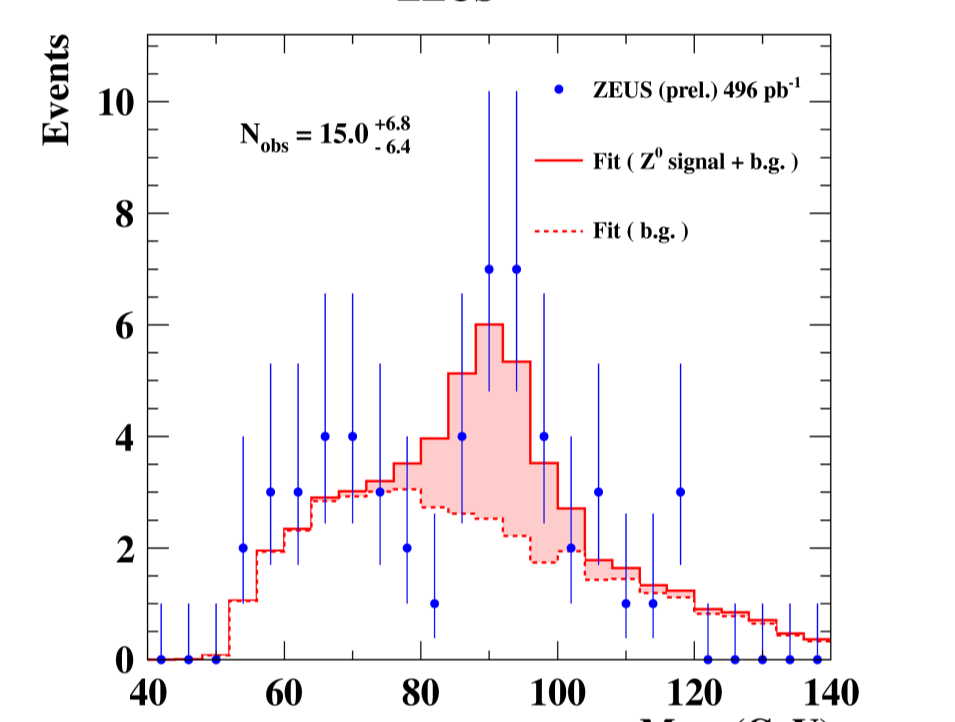
Information for zmcsp tests

test number	operating system	root version	status	std output file	error file	plots	root file
58	s45.6.64	5.28.00c	OK	out	err	plots	root
59	s45.7.32	5.28.00c	OK	out	err	plots	root
61	s45.6.64	5.30.05	OK	out	err	plots	root
60	s45.7.32	5.30.05	OK	out	err	plots	root



## Validation of Physics Analyses

It is very important to test the stability of physics analyses against future changes of the software environment. ZEUS uses an example physics analysis (Z<sup>0</sup> analysis) to perform essential checks using the same dedicated virtual system as for MC production.




- Validate access to data/MC in CN format
- Check results of the Z<sup>0</sup> analysis (event list, cross section, acceptance, invariant mass calculations) against various possible future changes
  - 32-64 bit machines
  - New ROOT versions
  - Speed
  - New operating systems
  - New data access schemes
  - ...

## 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 experiment.

Non-digital documentation is being collected, stored & cataloged

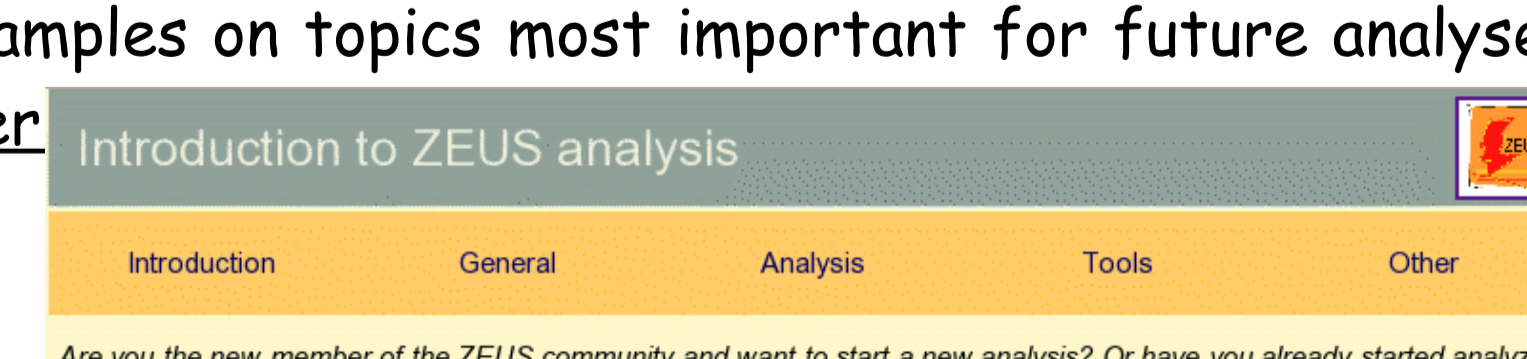
- ZEUS non-digital documentation sorted and safely stored
  - stored in DESY library archive, together with other HERA experiments
- some part of non-digital documentation digitized (ZEUS internal Notes)



Digital documentation:

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

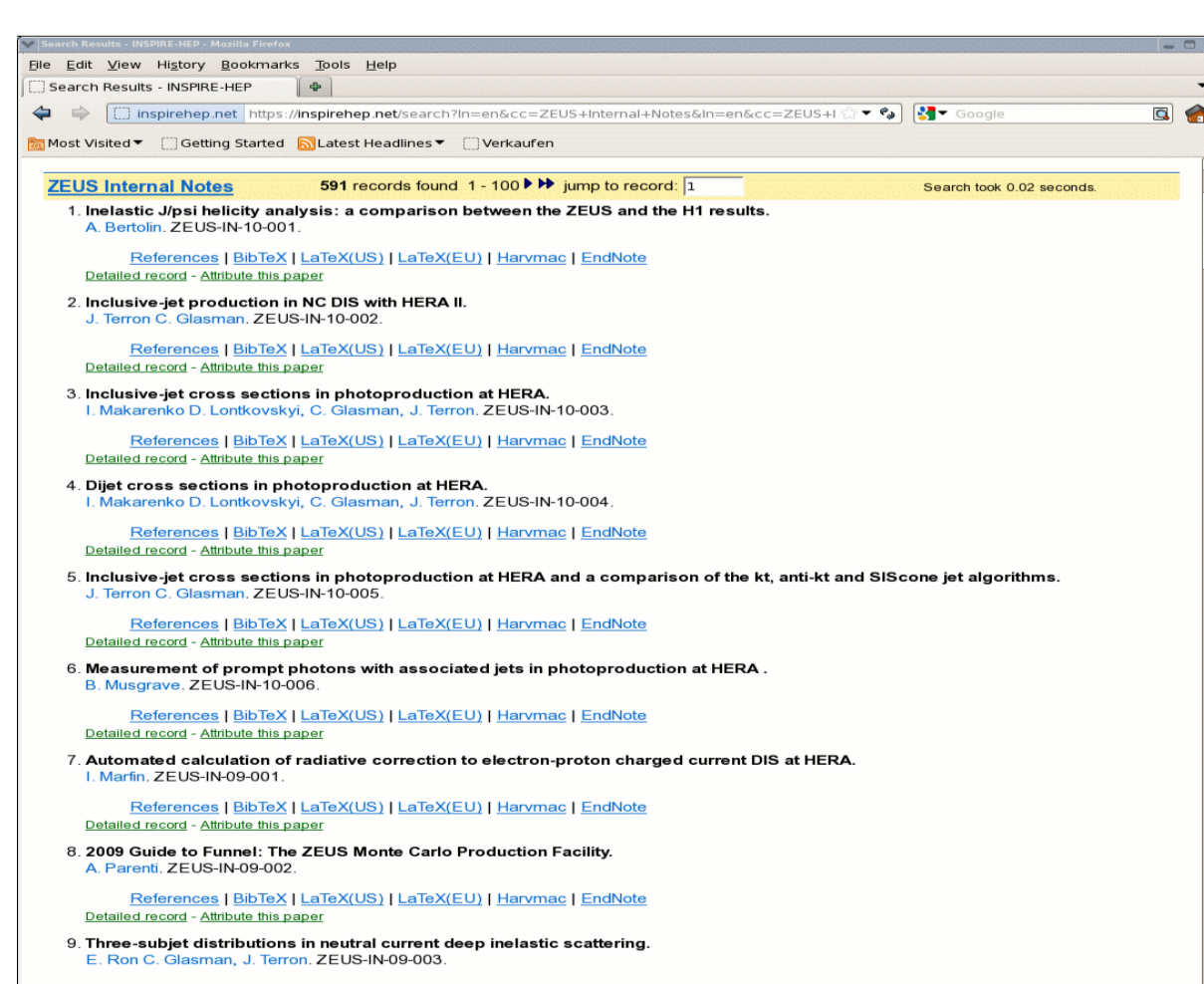
New condensed tutorials and examples on topics most important for future analyses are being prepared - ZEUS Primer



## INSPIRE

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: new dedicated effort to create an improved info storage like spires
- Inspire offers many convenient options for digitized documents archiving
- ZEUS Internal notes 1985-2011 submitted to INSPIRE (password protected)
- ZEUS thesis list in Inspire



## Long-Term Web Server Archiving

Most of essential digital information for analysis is stored on web servers - huge amount of various type of data to be safely stored

- DESY IT provides a virtual web server that can host static web pages and provides maintenance of this server in future
  - server configured and working
  - tools for migration being investigated & tested (wget, rsync)
  - need solution for ZEUS data bases - in touch with DESY WebOffice

