

DPHEP: From Study Group to Collaboration.

Data Preservation in High Energy Physics



David South (DESY)
on behalf of the DPHEP Collaboration

arXiv:1205.4667



XXI INTERNATIONAL WORKSHOP ON
DEEP-INELASTIC SCATTERING AND
RELATED SUBJECTS
Marseille Congress Centre April 22-26 2013



Experimental particle physics in the collider era

➤ A wide variety of physics results from many, often very different experiments

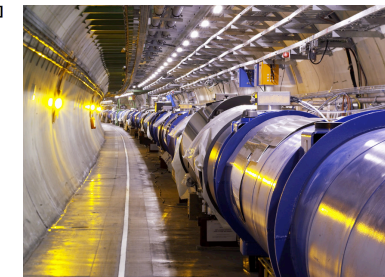
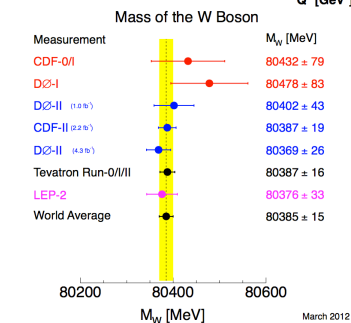
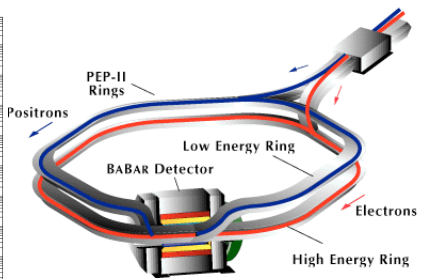
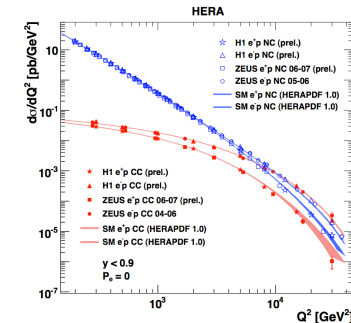
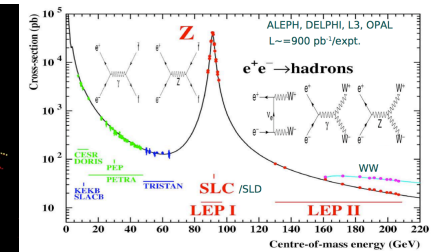
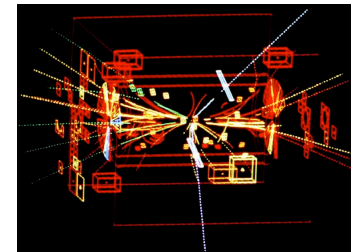
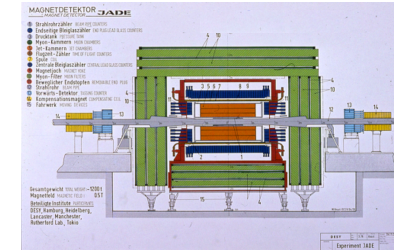
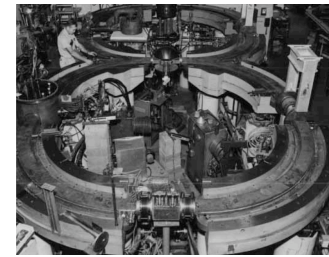
➤ Energy frontier probed with increasingly complex accelerator installations

- New experiments typically supersede previous, similar ones - but not always

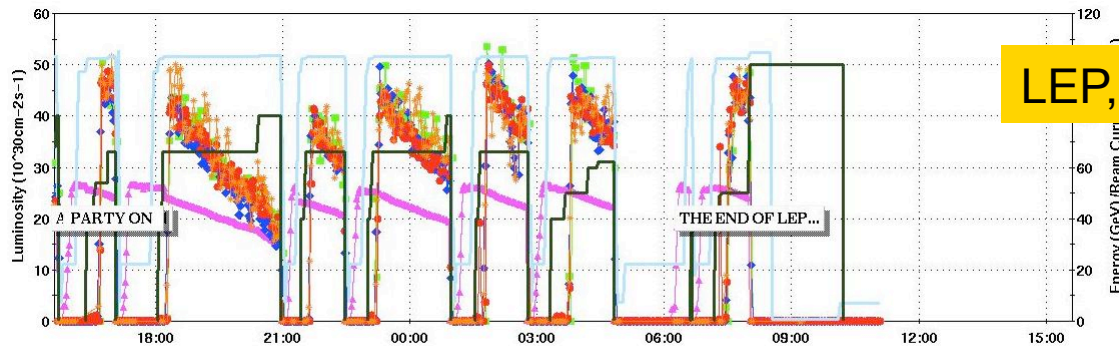
➤ Growth in size of the necessary international collaborations, as well as the diversity of the data management

➤ The age of the LHC has truly arrived

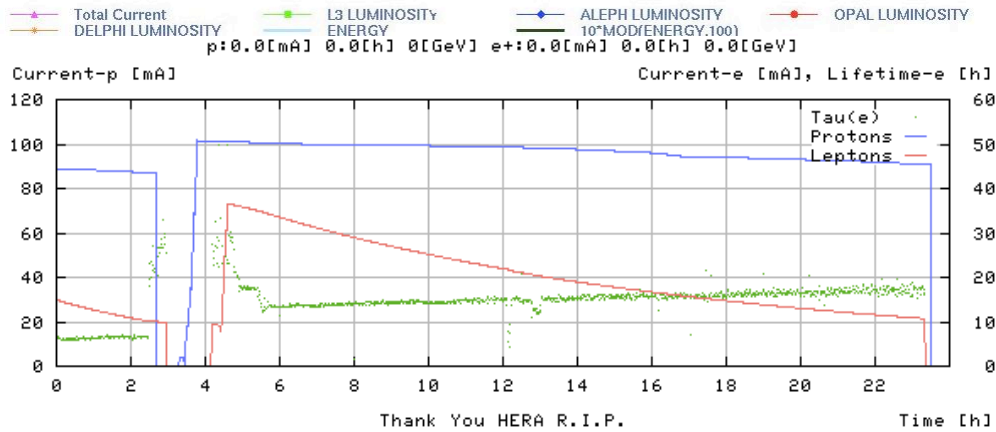
- Belle 2, HL-LHC, and other projects such as the ILC or next e-p/A collider are to come



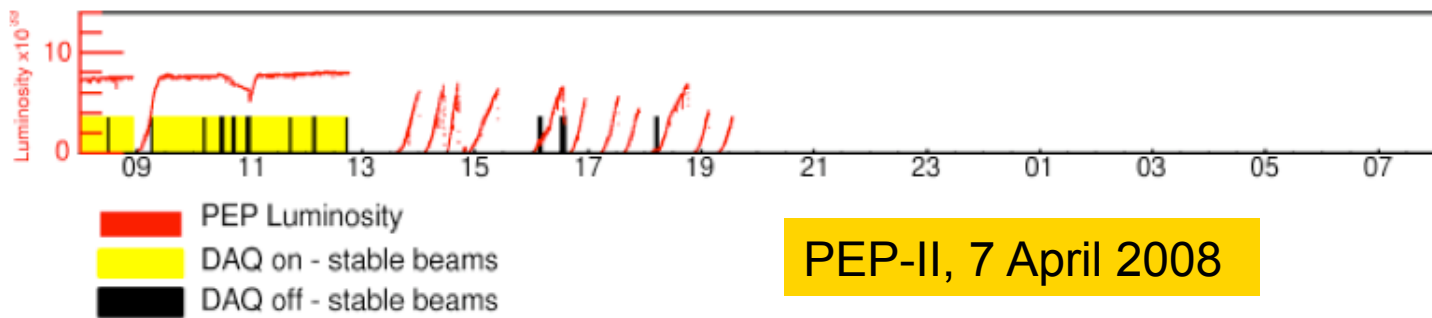
The last years have seen the end of several experiments



LEP, 2 November 2000



HERA, 30 June 2007



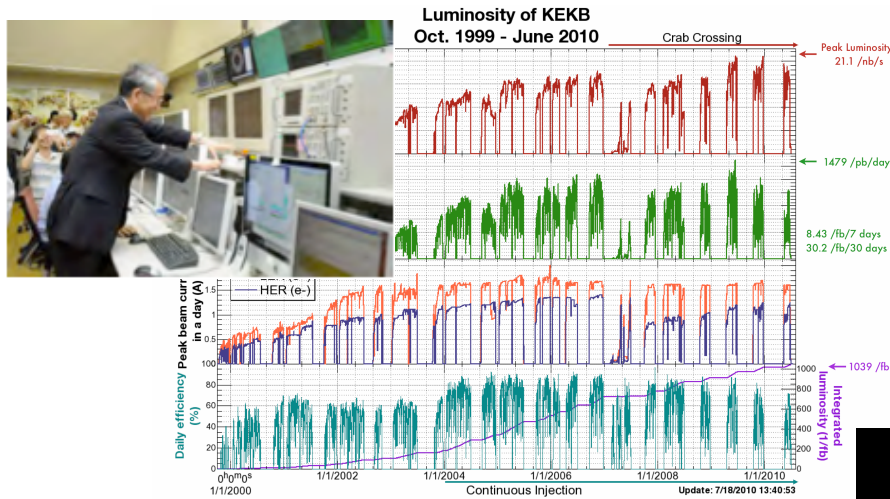
Mon Apr 7
 04:34h (19.0%) Stable Beams
 04:15h (93.1%) DAQ on
 84.6/pb Recorded Lumi
 1.3% Downtime
 18s DCH paused (# 4)

PEP-II, 7 April 2008

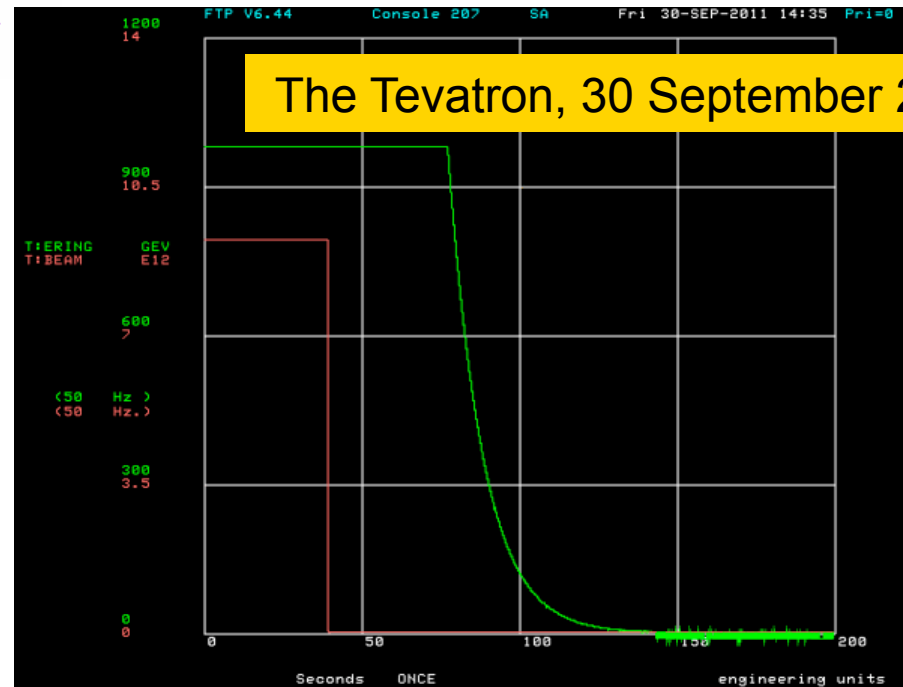


The last years have seen the end of several experiments

KEKB, 30 June 2010



The Tevatron, 30 September 2011



What do you do when the collisions have stopped ?

- > Finish the analyses! But then what do you do with the data?
 - Until recently, there was no clear policy on this in the HEP community
 - It's possible that older HEP experiments have in fact simply lost the data
- > Data preservation, including long term access, is generally not part of the planning, software design or budget of an experiment
 - So far, HEP data preservation initiatives have been in the main not planned by the original collaborations, but rather the effort a few knowledgeable people



> The conservation of tapes is not data preservation!

- *“We cannot ensure data is stored in file formats appropriate for long term preservation”*
- *“The software for exploiting the data is under the control of the experiments”*
- *“We are sure most of the data are not easily accessible!”*



The difficulties of data preservation in HEP

- > Handling HEP data involves large scale traffic, storage and migration
 - The increasing scale of the distribution of HEP data can complicate the task
- > Who is responsible? The experiments? The computing centres?
 - Problem of older, unreliable hardware: unreadable tapes after 2-3 years
 - The software for accessing the data is usually under the control of the experiments
- > Key resources, both funding and person-power expertise, tend to decrease once the data taking stops
- > And a rather key ingredient to all this is: *why do it?*
 - Can the relevant physics cases be made?
 - Who says we want to do this anyway?
 - Is the benefit of all this really worth the cost and effort?



DPHEP: An international study group on data preservation

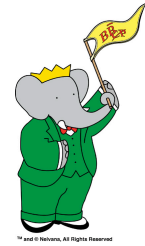
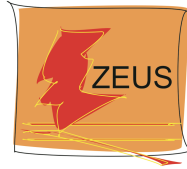


- > First contacts established in September 2008
 - Group since grown to over 100 contact persons
 - Endorsed as an ICFA panel since summer 2009
 - Initial make up of the group was driven by the coincidence of the end of data taking at several large colliders – SLAC, HERA, Tevatron - but now has grown to include others including the LHC experiments

- > Steering Committee with representatives from all members in addition to an International Advisory Committee



DPHEP: An international study group on data preservation



Institute of High Energy Physics
Chinese Academy of Sciences



Jefferson Lab



Science & Technology
Facilities Council



INSPIRE

BROOKHAVEN
NATIONAL LABORATORY





DPHEP

ICFA Study Group on Data Preservation and Long Term Analysis in High Energy Physics

Joint DASPOS / DPHEP-7 Workshop at CERN, March 21-22 2013

Publication from May 2012, available here:

Status Report of the DPHEP Study Group: Towards a Global Effort for Sustainable Data Preservation in High Energy Physics

6th DPHEP Workshop at CPPM, Marseille, November 19-21 2012

DPHEP@CHEP2012: DPHEP Session at CHEP 2012, New York, May 24 2012

High Energy Physics experiments initiate with this **Study Group** a common reflection on **data persistency and long term analysis** in order to get a common vision on these issues and create a multi-experiment dynamics for further reference.

The objectives of the Study Group are:

- Review and document the physics objectives of the data persistency in HEP.
- Exchange information concerning the analysis model: abstraction, software, documentation etc. and identify coherence points.
- Address the hardware and software persistency status.
- Review possible fundings programs and other related international initiatives.
- Converge to a common set of specifications in a document that will constitute the basis for future collaborations.

Since August 2009, the Study Group is endorsed by **ICFA (International Committee for Future Accelerators)**.

A series of workshops have been held by the Study Group, access to which can be found using the links below.

The 3rd workshop was preceded by a public **symposium**.

The **first DPHEP publication**, containing the initial recommendations of the Study Group was released in December 2009.

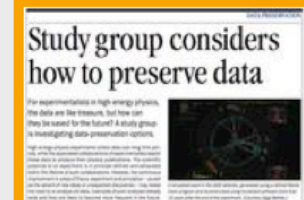
This was followed by a more comprehensive, **second publication** in May 2012.



Latest DPHEP Status Report released May 2012, available on arxiv:1205.4667

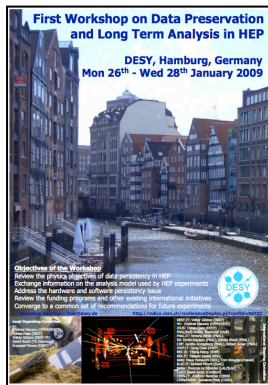


CERN Courier, April 2009



DPHEP: An international study group on data preservation

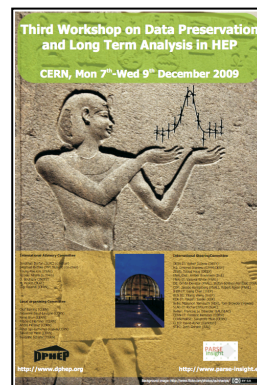
> Series of DPHEP workshops held since 2009



Jan 2009: DESY



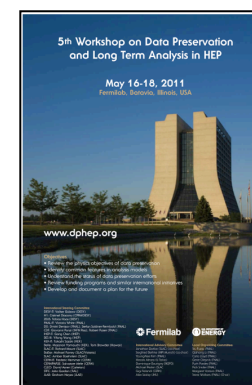
May 2009: SLAC



Dec 2009: CERN



Jul 2010: KEK



May 2011: Fermilab



Nov 2012: CPPM

> The first task of the group was to establish the working directions

- “To confront data models, clarify the concepts, set a common language, investigate technical aspects, compare with other fields handling large data.”

> Initial findings published in an interim report December 2009

- Focus on four key areas of the study group: **Physics Case for Data Preservation, Preservation Models, Technologies, Governance**

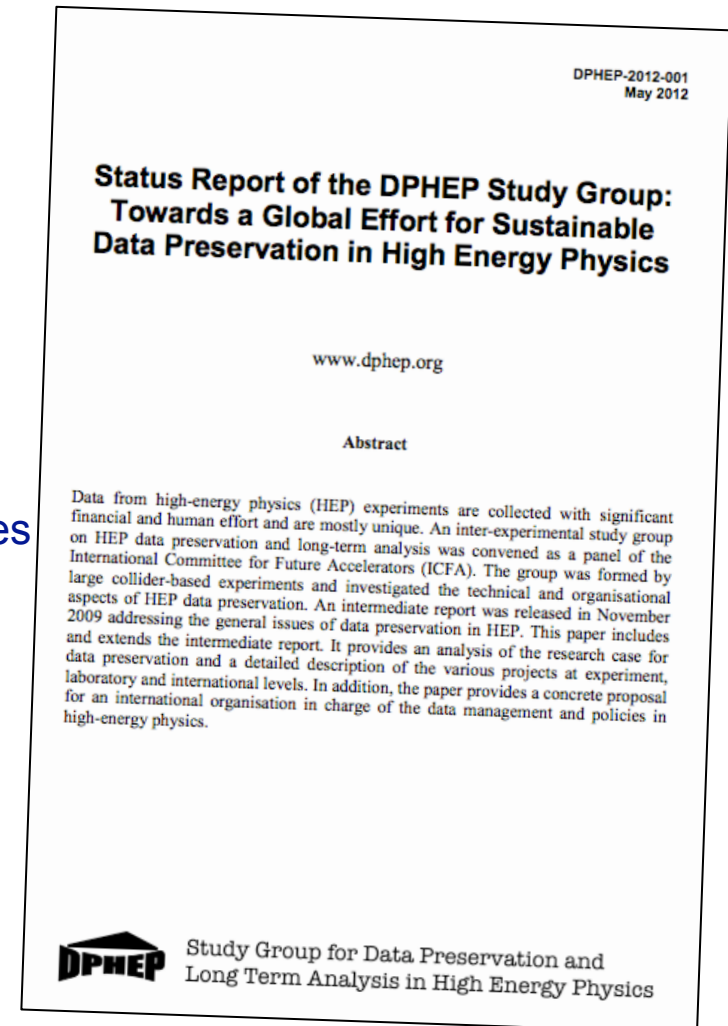
arXiv:0912.0255



Large scale DPHEP publication

- > Released May 21st 2012, 93 pages
- > Full status report of the activities of the DPHEP study group, including:
 - Tour of data preservation activities in other fields
 - An expanded description of the physics case
 - Defining and establishing data preservation principles
 - Updates from the experiments and joint projects
 - FTE estimates for these and future projects
 - Next steps to establish fully DPHEP in the field

arXiv:1205.4667



DASPOS

- New initiative in the funded in the US by the NSF, “Data and Software Preservation for Open Science” (DASPOS)
 - Focussed on high energy physics data from the LHC and the Tevatron
 - Complimentary effort to DPHEP, series of workshops focussing on different issues
 - Seventh DPHEP workshop held last month at CERN jointly with DASPOS

DASPOS Workshops

- Use Cases for Archived Data and Software in HEP
- Commonality with other Disciplines
- Data Model and Query Semantics
- Software Sustainability
- Preservation Policy
- Technical Storage Architectures

Data Preservation in High Energy Physics

The DPHEP effort, a Study Group for Data Preservation in High Energy Physics under the auspices of the International Committee on Future Accelerator (ICFA), has produced a study outlining the current state of data preservation within HEP, including an extensive overview of other disciplines. They suggest a series of guidelines for HEP data preservation efforts, as well as a framework for global coordination. Their conclusions include a recognition of the scientific potential for data re-use, especially the desirability to preserve full analysis capability. They also emphasize the urgency required to begin and sustain global, coordinated data preservation efforts.

For the purposes of discussing preservation efforts, the DPHEP studies have identified different types of HEP data that span the full set of possibilities, ranging from publications, metadata, associated documentation of all types, software, digital information (the data themselves) and finally expertise and human resources. Digital information can be categorized by four tiers covering the scope from publications to the raw data and the software used to process it. As outlined by DPHEP, the four tiers are:

1. Published results, along with additional analysis-related information, leading to more complete documentation of a given analysis
2. Processed data available in a simplified format (i.e., particle four vectors) that can be used for outreach and simplified additional analyses
3. The full processed experimental data and simulated data and the associated software for accessing and analyzing the data
4. The full raw data of the experiment and all of the reconstruction software necessary for processing the data into a form where it can be useful for analysis, as well as the simulation software needed for modeling

DASPOS Data and Software Preservation for Open Science

ABOUT

PEOPLE

WORKSHOPS

RESEARCH

REPORTS

The massive data sets accumulated by High Energy Physics (HEP) experiments represent the most direct result of the often decades-long process of construction, commissioning and data acquisition that characterize this science. Many of these data are unique and represent an irreplaceable resource for potential future studies. Forward-thinking efforts for preservation are necessary now in order to achieve the relevant parameters, analysis paths and software to preserve the usefulness of these rich and varied data sets.

"Ten or 20 years ago we might have been able to repeat an experiment. They were simpler, cheaper and on a smaller scale. Today that is not the case. So if we need to re-evaluate the data we collect to test a new theory, or adjust it to a new development, we are going to have to be able to reuse it. That means we are going to need to save it as open data..."

Rolf-Dieter Heur 2008
Director General, CERN

Second Workshop Scheduled

The second DASPOS Workshop has been scheduled for Thursday, July 25, 2013, at ACM/IEEE JCDL in Indianapolis, Indiana. [More information](#)



Data and Software Preservation for Open Science, DASPOS, represents an initial exploration of the key technical problems that must be solved to provide appropriate data, software and algorithmic preservation for HEP, including the contexts necessary to understand, trust and reuse the data. While the archiving of HEP data may require some HEP-specific technical solutions, DASPOS will create a template for preservation that will be useful across many different disciplines, leading to a broad, coordinated effort.

Discovery and Coordination >

Series of highly-structured public workshops to define, discuss and document the

Prototyping and Experimentation >

Key areas of research: data and query models and software sustainability

The DASPOS Team >

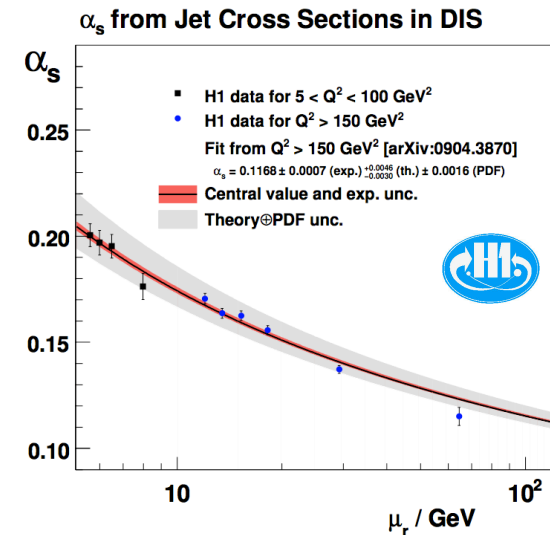
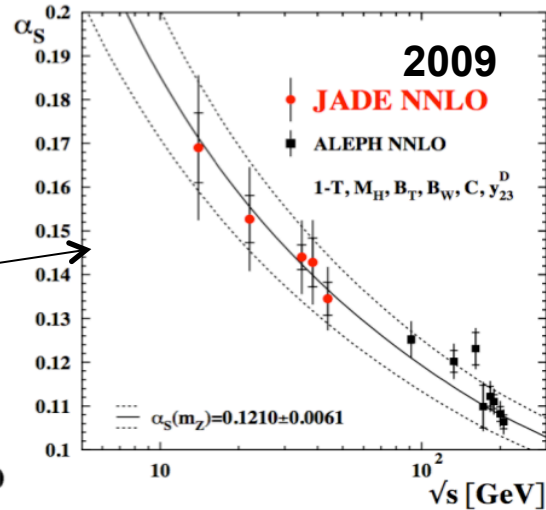
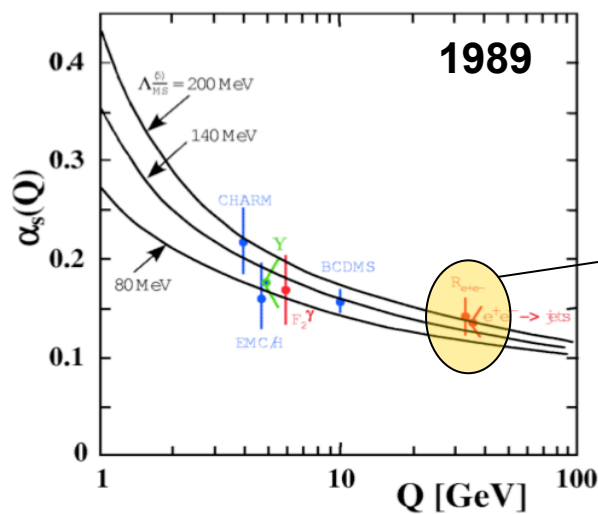
Computer science experts, experienced digital librarians, and experts in data-intensive

Building the physics case: Reasons to preserve HEP data

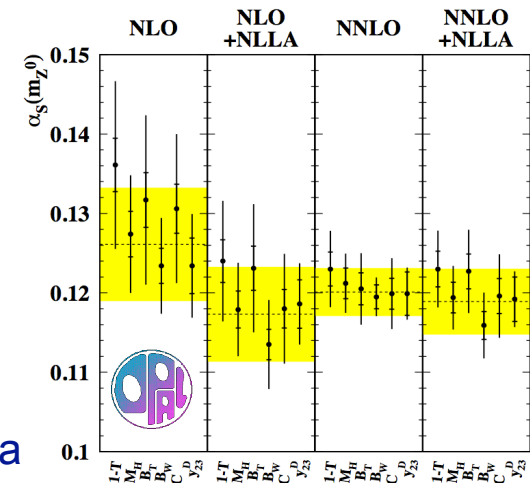
- > Long term completion and extension of an existing physics program
 - Up to 10% of papers are finalised in the “archival mode”
 - Gain in scientific output of the experiments
- > Cross-collaboration and combinations of physics results
 - During the active lifetime of similar experiments at one facility: LEP, HERA, TeVatron
 - And later across larger boundaries: Belle/BaBar, TeVatron/LHC
- > Revisit old measurements or perform new ones
 - Access to newly developed techniques, comparisons to new theoretical models
 - Unique data sets available in terms of energy, initial states
- > Use in scientific training, education, outreach
 - Simplified formats: associated exercises to perform e.g. composite-particle reconstruction, finding signals in the background, ...



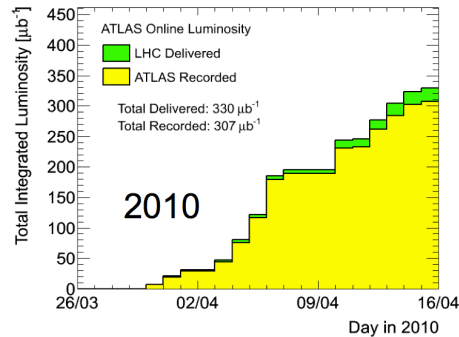
Example: Revisit old measurements or perform new ones



- Access to newly developed techniques, comparisons to new theoretical models
 - History may be repeated with the HERA α_s measurements
- Unique data sets are available in terms of initial state particles and energy
 - If no LHeC or alternative, all we have are the HERA e^+p data
 - Tevatron $p\bar{p}$ are also unique, A_{FB} , high-x jets, ...
 - Fixed target experiments, ... others, ...

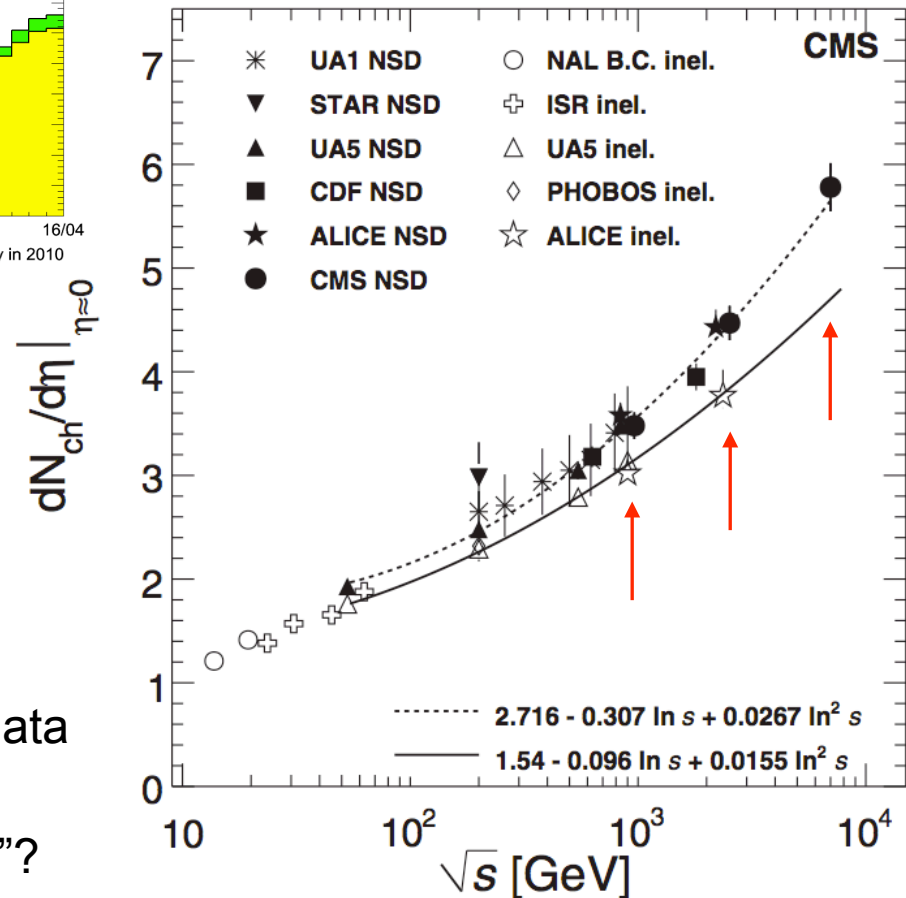


And what about LHC 900 GeV and 2.32 TeV data? 7 TeV data?

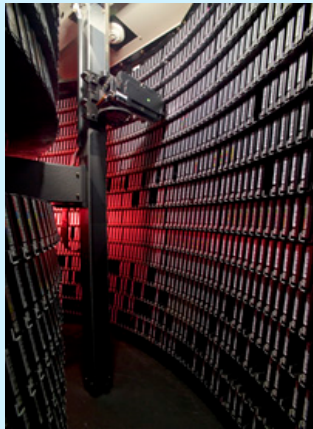


Centre-of-mass Energy	0.9 TeV	2.36 TeV
Selection	Number of Events	
BPTX Coincidence + one BSC Signal	72 637	18 074
One Pixel Track	51 308	13 029
HF Coincidence	40 781	10 948
Beam Halo Rejection	40 741	10 939
Beam Background Rejection	40 647	10 905
Valid Event Vertex	40 320	10 837

- > Early LHC measurements made using data at a unique centre of masses
- > Is the 2010 low pile up 7 TeV data “safe”?
- > What happens to Run 1 data when the 14 TeV collisions come? Something like what happened at the TeVatron?

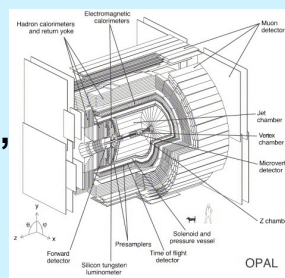


An important question: What is HEP data?



Digital information
The data themselves, volume estimates for preservation data of the order of **a few to 10 PB**
Other digital sources such as databases to also be considered

Software
Simulation, reconstruction, analysis, user, in addition to any external dependencies

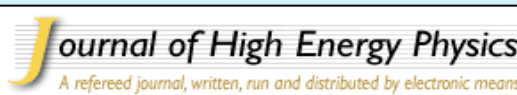
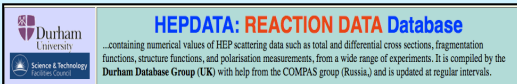


CERNLIB Access
• Access to the CERN Program Library is free of charge to all HEP users worldwide.
• Non-HEP academic and not-for-profit organizations: 1KSF/year

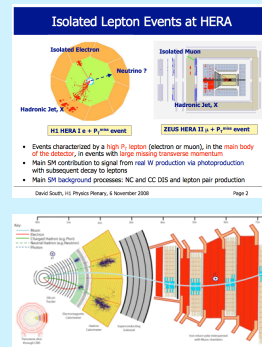
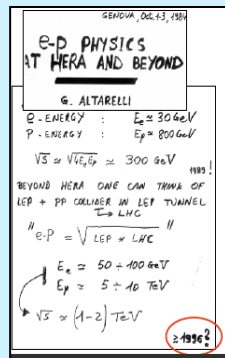
Meta information
Hyper-news, messages, wikis, user forums..



Publications arXiv.org



Documentation
Internal publications, notes, manuals, slides



Expertise and people



DPHEP models of HEP data preservation

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

- These are the original definitions of DPHEP preservation levels from the 2009 publication
 - Still valid now, although interaction between the levels now better understood
- Originally idea was a progression, an inclusive level structure, but now seen as complementary initiatives
- Three levels representing three areas:
 - Documentation, Outreach and Technical Preservation Projects



Level 1: Documentation

- > Dedicated documentation task forces set up by many experiments
 - Much material from pre-web days, or using all kinds of web applications

- > **Non-digital:** Cataloguing, scanning, photographing older material

- Papers, notes, drawings, pre-web talks, detector schematics, blueprints, logbooks...
- New *Virtual Archives* established



- > **Digital:** Securing and consolidating the existing content

- Online shift tools, detector configuration files, electronic logbooks, detailed run info, web content from out-dated servers with dead links, wikis, meetings, talks, ...
- Replacement of old web servers by VMs, hosted by the computer centres
- Replacement of old pages to newer technologies such as wikis
- Use of external services like INSPIRE for hosting collaboration material

Sort by: year | asc. | - or rank by - | Display results: 100 results | single list | Output format: Brief format | ZEUS

ZEUS Internal Notes 1,807 records found 1 - 100 ►► jump to record: 1 Search took 0.02 seconds.

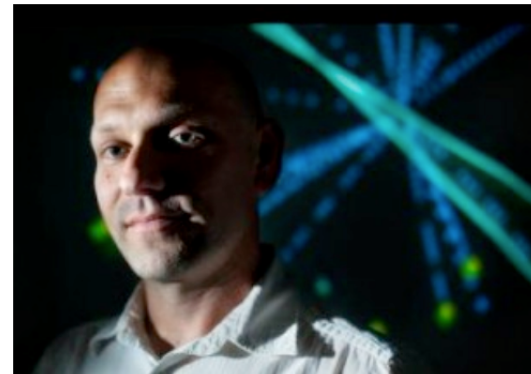
1. Inelastic J/psi helicity analysis: a comparison between the ZEUS and the H1 results.
A. Bertolin. ZEUS-IN-10-001.
[References](#) | [BibTeX](#) | [LaTeX\(US\)](#) | [LaTeX\(EU\)](#) | [Harvmac](#) | [EndNote](#)
[Detailed record](#) - [Attribute this paper](#)
2. Inclusive-jet production in NC DIS with HERA II.
J. Terron C. Glasman. ZEUS-IN-10-002.
[References](#) | [BibTeX](#) | [LaTeX\(US\)](#) | [LaTeX\(EU\)](#) | [Harvmac](#) | [EndNote](#)

INSPIRE



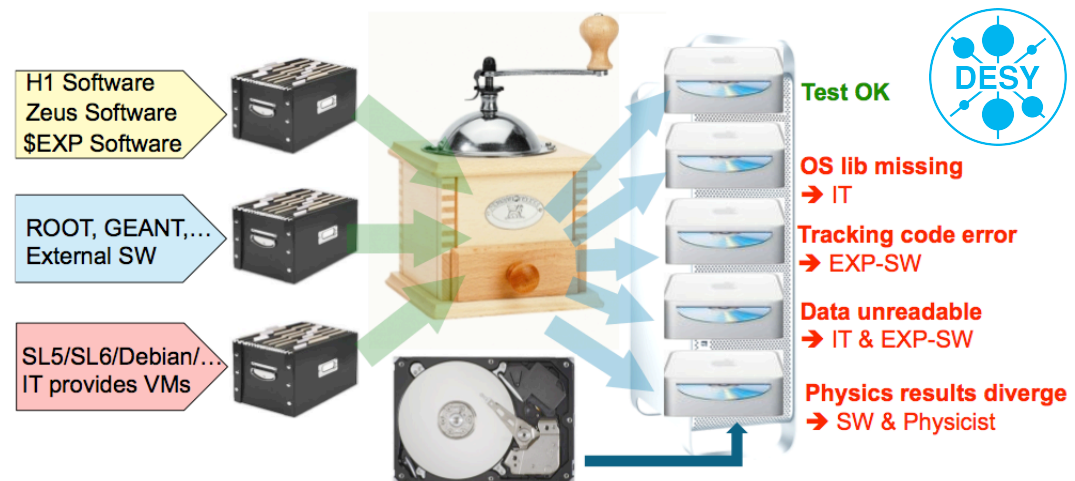
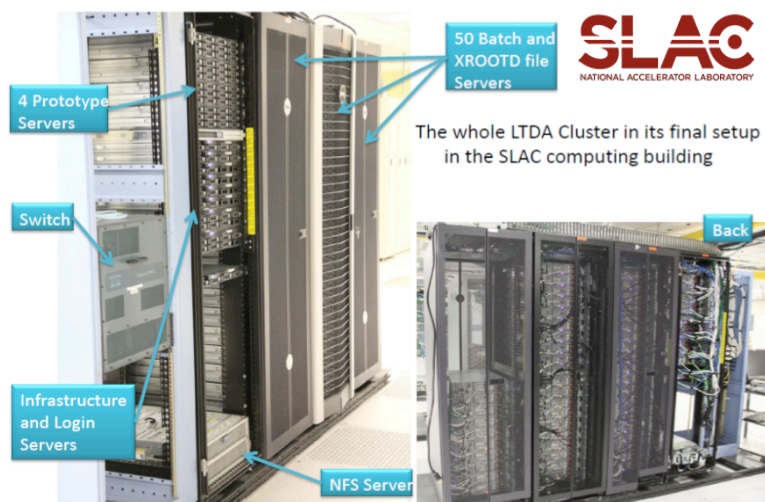
Level 2: Simplified formats for outreach

- Within DPHEP and the member collaborations there are generic ideas, such as common formats and user interfaces
 - In terms formats, much can be learned from other fields such as astrophysics or life sciences
- Such outreach formats in HEP are typically based on ROOT, containing particle 4-vectors and simple event information
 - Composite-particle reconstruction, finding signals
 - Initiatives in place at BaBar, Belle and LHC experiments
- Simplified formats also provide an ideal way of transferring data between experiments and theory
 - Allowing new models to be tested on HEP data



Technical projects: Levels 3 and 4

- > This is really the main focus of the data preservation effort
 - Access to analysis level data, MC and the analysis level software
 - In addition for level 4 this includes the reconstruction and simulation software
- > It's not about the data, but about still being able analyse it
 - Either keep your current environment alive as long as possible
 - Or adapt and validate your code to future changes as they happen
 - Two complimentary approaches taken by BaBar at SLAC and the HERA experiments at DESY, both employing virtualisation techniques, but in different ways



Summary and future working directions of DPHEP

- > The DPHEP Study Group has established itself in the HEP community reached a milestone in the publication of the latest report, which contains a comprehensive appraisal of data preservation in HEP
 - Increased participation in the last years, especially from the LHC community

- > DPHEP will continue to investigate and take action in areas of coordination, preservation standards and technologies
 - Also in cooperation with other initiatives such as DASPOS, as well as expanding the experimental reach and inter-disciplinary cooperation

- > In 2013 DPHEP will make the transition from Study Group to Collaboration
 - Final draft of Collaboration Agreement prepared contributors will be asked to sign during 2013
 - Target upcoming FP8/Horizon2020 EU funding to realise and fully deploy a variety of projects

Status Report of the DPHEP Study Group: <http://arxiv.org/abs/1205.46677>

Joint DASPOS/DPHEP-7 Workshop, March 2013: <https://indico.cern.ch/conferenceDisplay.py?confId=233119>

CHEP 2012 talk: <http://indico.cern.ch/contributionDisplay.py?sessionId=0&contribId=607&confId=149557>

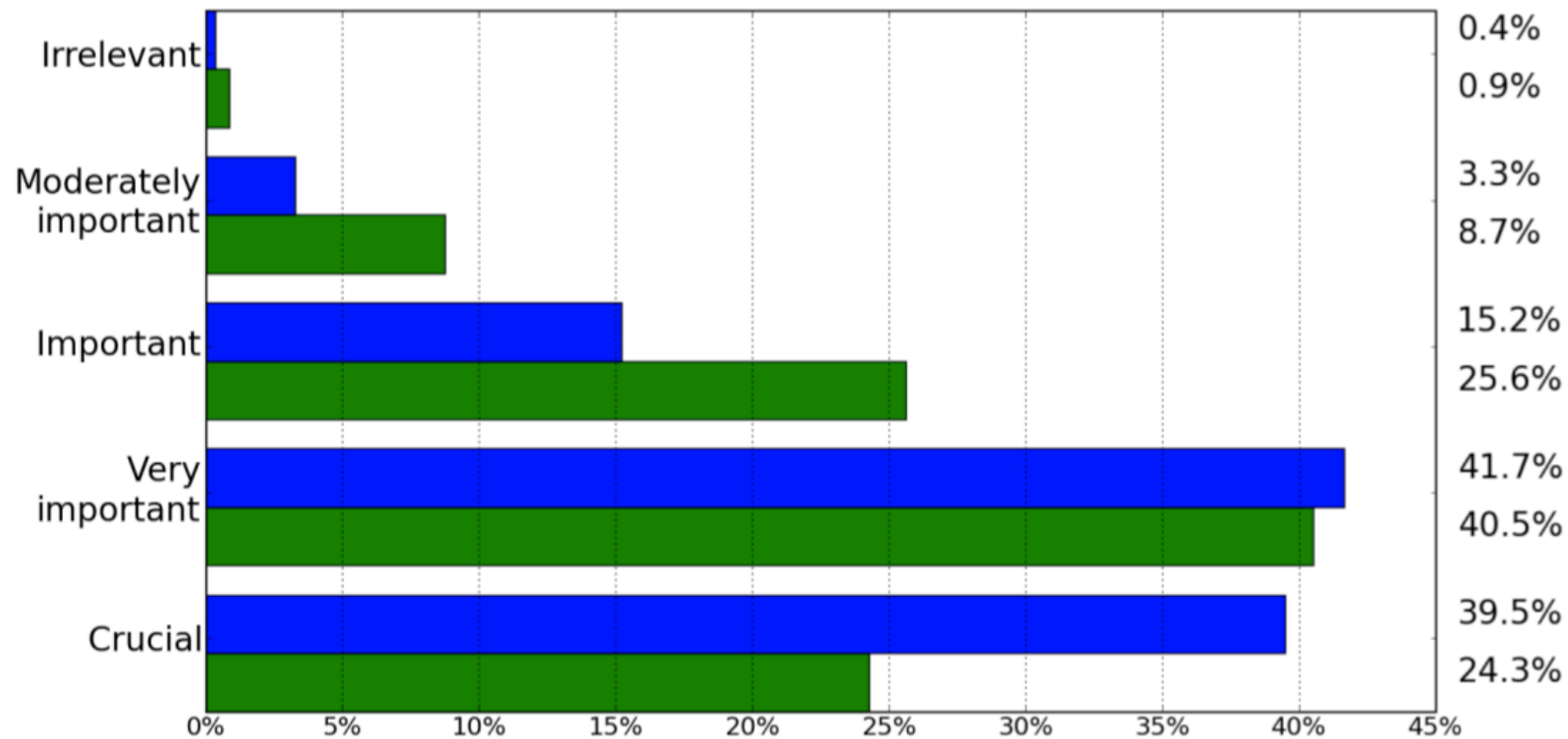
Seminar from November 2011: http://www.desy.de/dvsem/WS1112/south_talk.pdf

EXTRAS



Support for data preservation in the HEP community

In your opinion, how important is the issue of data preservation ?
(top/blue: theorists, bottom/green: experimentalists)

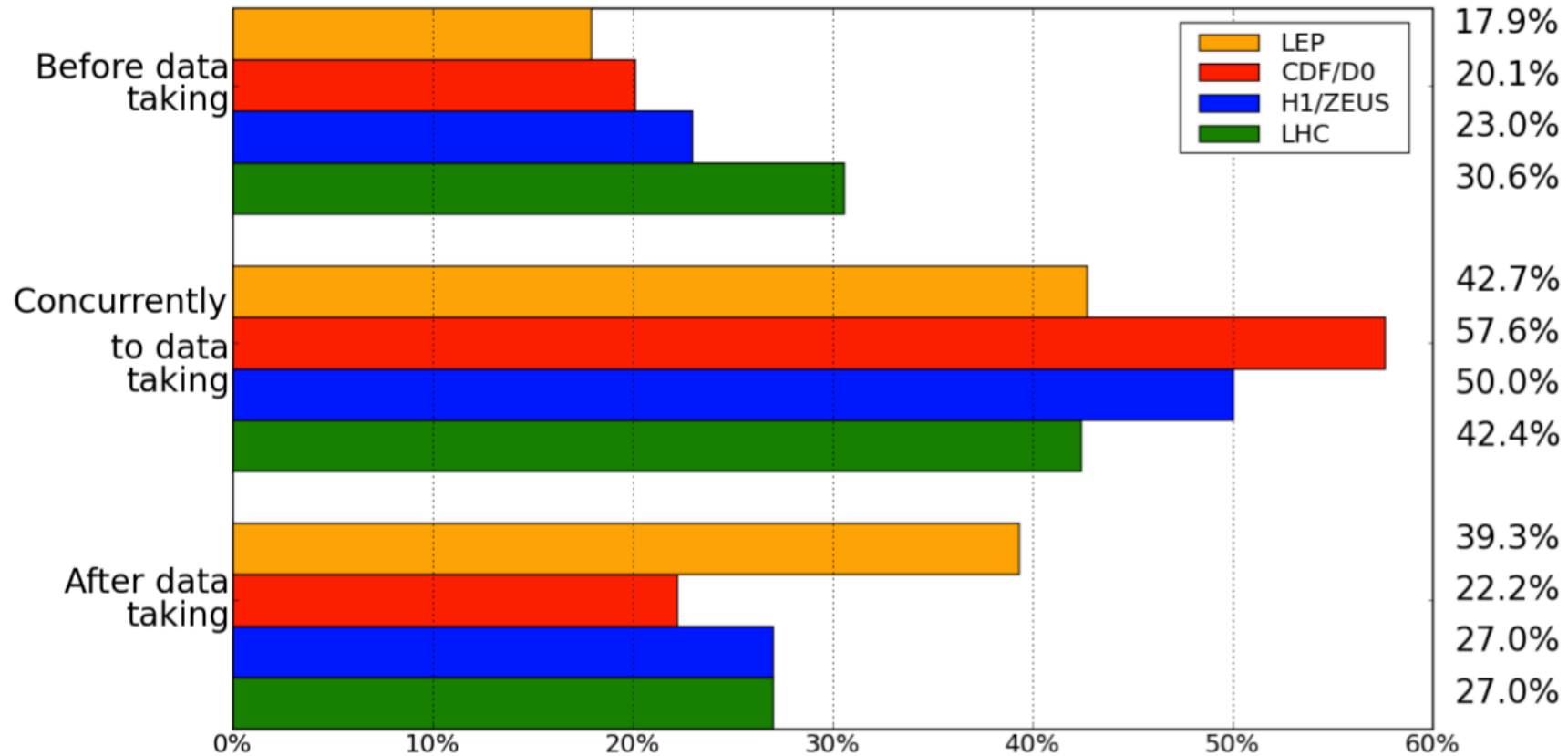


arXiv:0906.0485



Support for data preservation in the HEP community

In your opinion, when should this effort start in order to be the most effective ?



arXiv:0906.0485



Transition scenario and resources at the experimental level

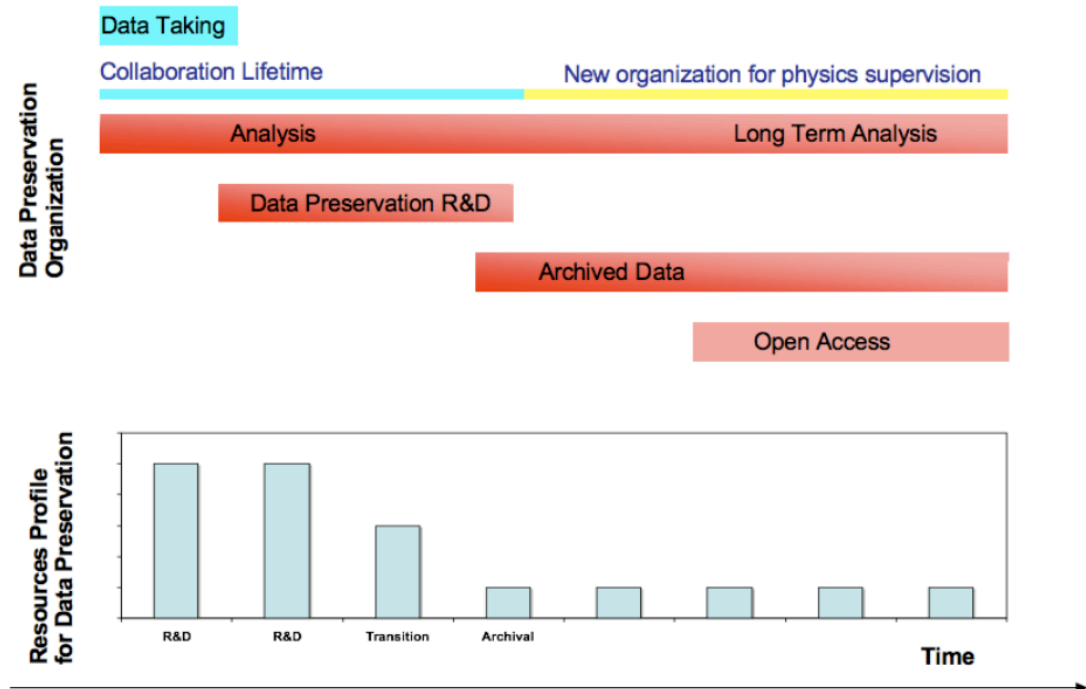
➤ Planning the transition to a long term analysis model

➤ R&D phase needed to develop the projects for the transition

➤ Long term custodianship of the physics data

➤ Resources / experiment

- Typically a surge of 2-3 FTEs for 2-3 years, followed by steady 0.5-1.0 FTE per experiment/lab
- This should be compared to 300-500 FTEs for many years per experiment!

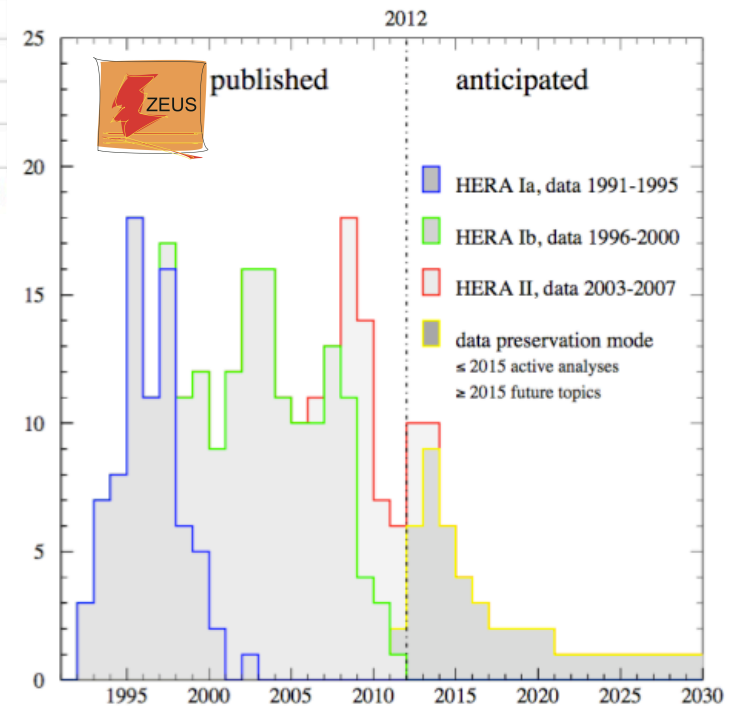
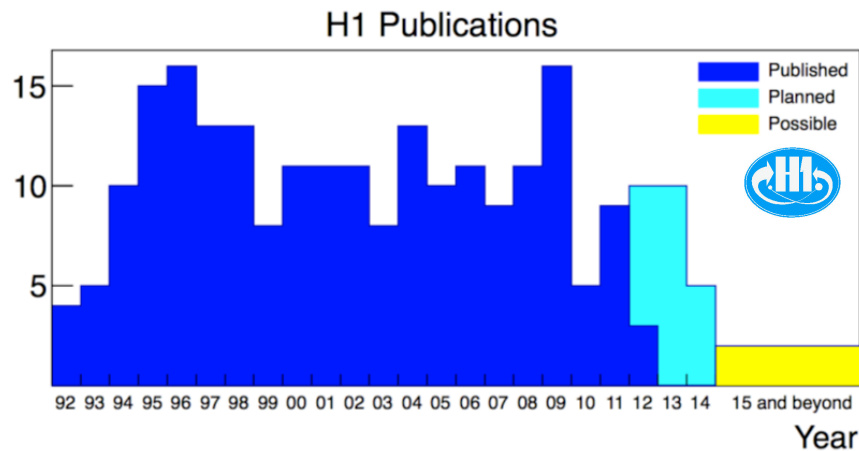
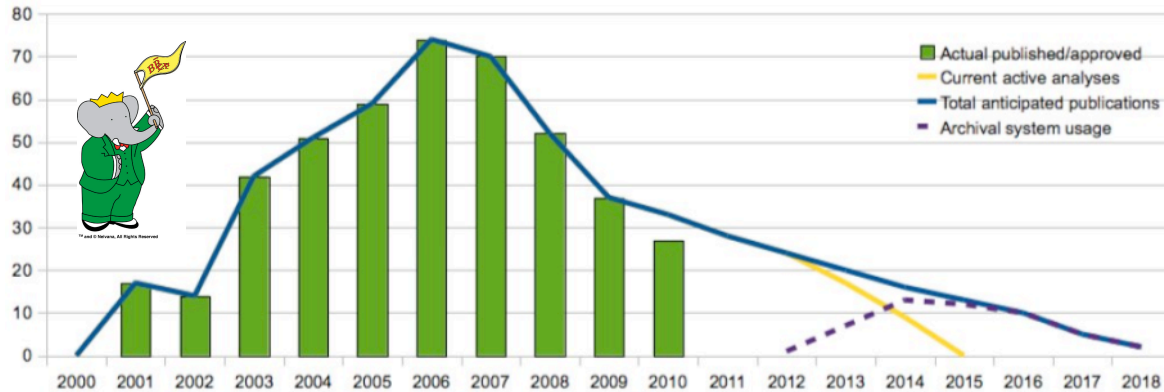


Cost estimates represent typically **much less than 1%** of the original investment

Scientific return: **O(10%)** in number of publications



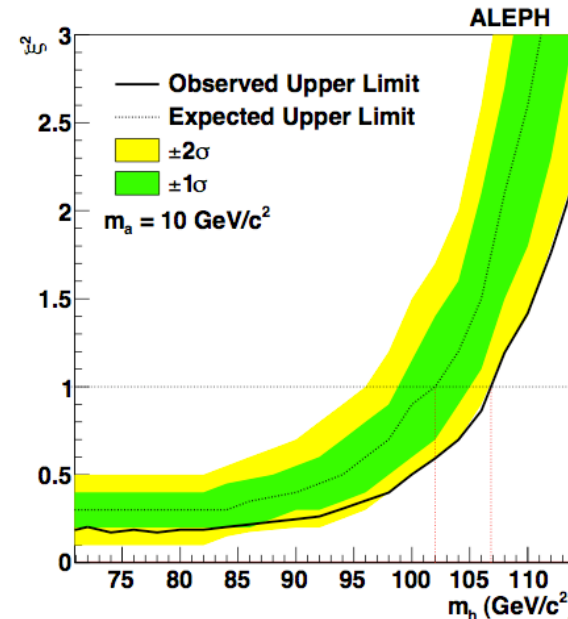
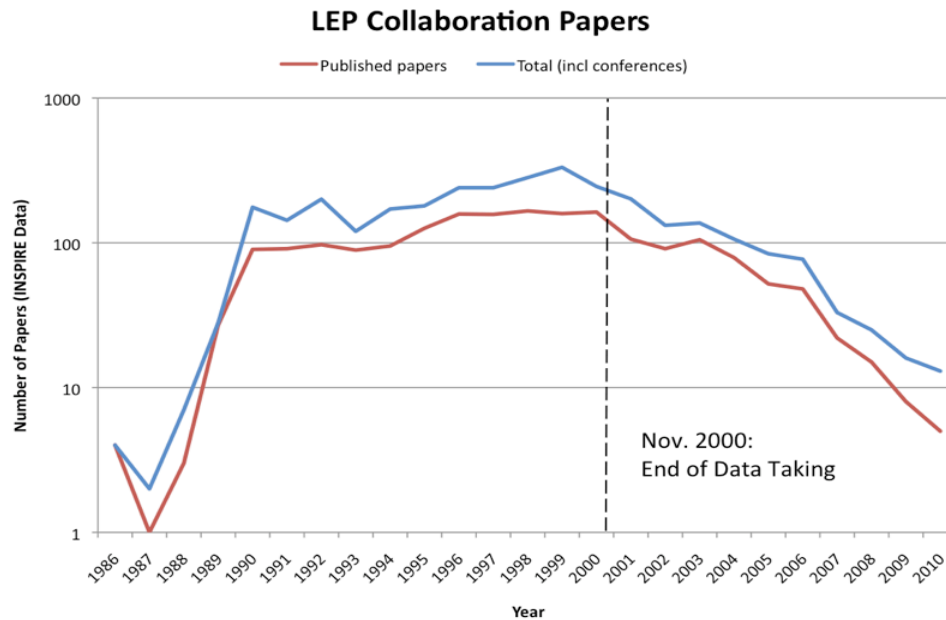
Long term completion of the physics programme



- Similar publication tails predicted by the BaBar, H1 and ZEUS experiments, taking into consideration the plans for data preservation



Long term completion of the physics programme



- The publication tail of LEP is long, with new papers still appearing
- Well over 300 papers produced since the end of collisions in 2000
- Recent analysis of LEP data gave unique limits on a novel Higgs model
- Similar, if not longer publication tails predicted by the BaBar, H1 and ZEUS experiments, after taking into consideration the plans for data preservation



Cross-collaboration combinations of physics results

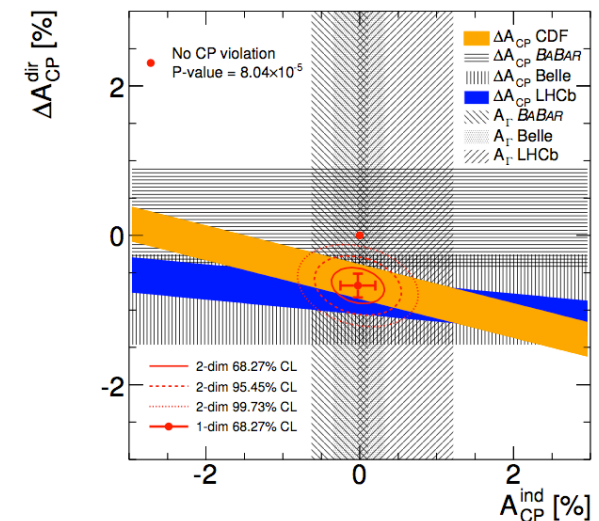
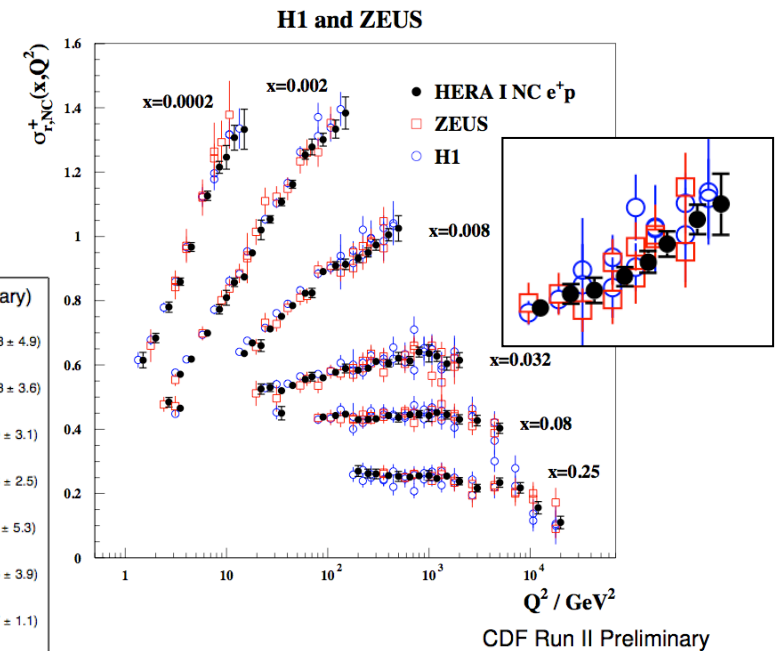
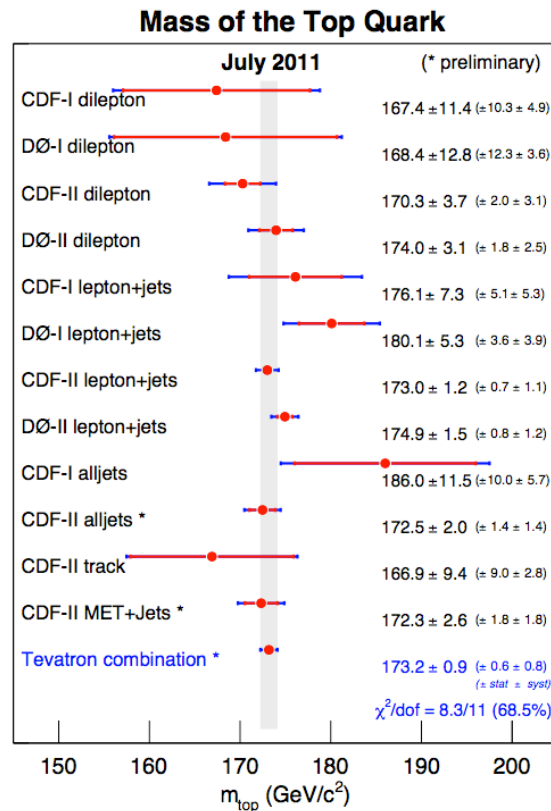
> Combination of data from multiple experiments to produce new scientific results

- Improved precision and increased sensitivity

> Comparison of experimental results

- Complimentary information from different physics
- Verification of experimental observations

> Both objectives facilitated by data preservation



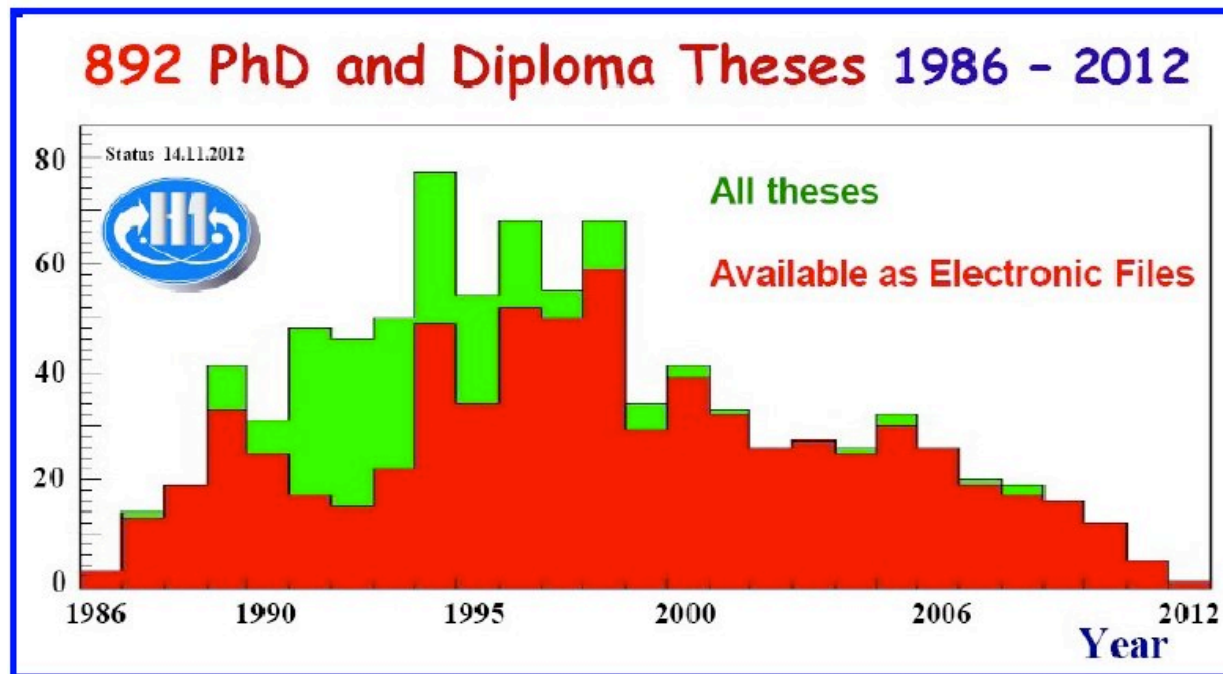
Level 1: Documentation

- > The organisation of documentation turns out to be quite a task
 - Dedicated task forces set up by many of the experiments
 - Much material from pre-web days, or using all kinds of web applications
- > **Non-digital:** Cataloguing, organisation, scanning or photographing of appropriate of papers, notes, drawings, talks from pre-web days, detector schematics, blueprints, logbooks, ...
 - *New Virtual Archives established by the experiments*
- > **Digital:** Old online shift tools, detector configuration files, electronic logbooks, detailed run information, web content from out-dated servers with dead links, various wikis, meetings, talks, ...
 - Replacement of old web servers by VMs, hosted by the computer centres
 - Replacement of old pages to newer technologies such as wikis (use of (T)wikis much more prevalent in the LHC era)
 - Use of external services for hosting collaboration material



H1 Theses

- > Since October 2010, **106** H1 theses discovered not previously known to the collaboration; **18** since this summer, latest ones only last week
- > Scanning and linking these to the official H1 pages is given high priority



- > **Currently, of the 892 known H1 theses 197 are not available in electronic form: ~ 22% not available to the H1 community!**



Documentation projects with INSPIRE

- Internal notes from all HERA experiments now available on INSPIRE
 - Experiments no longer need to provide dedicated hardware for such things
 - Password protected now, simple to make publicly available in the future



The screenshot shows the INSPIRE website interface. At the top left is the INSPIRE logo with 'HEP' underneath. To the right, a message states: 'Welcome to INSPIRE! INSPIRE is out of beta and ready to replace SP please email us at feedback@inspirehep.net'. Below this is a navigation bar with links for HEP, INST, HELP, SPIRES, and HEPNAMES. The main content area is titled 'ZEUS Internal Notes'. It contains a search instruction: 'Use "find *" for SPIRES-style search ([other tips](#))'. There is a search input field with a 'Search' button and links for 'Easy Search' and 'Advanced Search'. A link '[find in ZEUS-IN-10004](#)' is visible. A message reads: 'This collection is restricted. If you are authorized to access it, please click on the Search button.' At the bottom, there is a footer with links for 'HEP', 'Search', and 'Help', and text: 'Powered by [Invenio](#) v1.0.0-rc0+', 'Problems/Questions to feedback@inspirehep.net', and 'Last updated: 19 Oct 2011, 03:15'.



Documentation projects with INSPIRE

- Internal notes from all HERA experiments now available on INSPIRE
 - Experiments no longer need to provide dedicated hardware for such things
 - Password protected now, simple to make publicly available in the future

The screenshot shows the INSPIRE login interface. At the top, there is a navigation bar with the INSPIRE logo and a welcome message: "Welcome to INSPIRE! INSPIRE is out of beta and ready to replace SPIRES. Please email us at feedback@inspirehep.net". Below this is a secondary navigation bar with links for HEP, INST, HELP, SPIRES, and HEPNAMES. The main content area is titled "Login" and includes a message: "This collection is restricted. If you think you have right to access it, please authenticate yourself." There are input fields for "Username:" (containing "zeus") and "Password:". A checkbox labeled "Remember login on this computer." is present, along with a "login" button and a link for "(Lost your password?)". A note at the bottom states: "Note: You can use your nickname or your email address to login." The footer contains the text: "HEP: Search: Help Powered by Invenio v1.0.0-rc0+ Problems/Questions to feedback@inspirehep.net".



Documentation projects with INSPIRE

- Internal notes from all HERA experiments now available on INSPIRE
 - Experiments no longer need to provide dedicated hardware for such things
 - Password protected now, simple to make publicly available in the future

The screenshot displays the INSPIRE website interface. At the top, there is a navigation bar with the INSPIRE logo and a welcome message: "Welcome to INSPIRE! INSPIRE is out of beta and ready to replace SPI". Below this, a search bar is visible with the text "ZEUS Internal Notes" and "10 records found". The search results are listed as follows:

- 1. Inclusive-jet production in NC DIS with HERA II.**
J. Terron C. Glasman, ZEUS-IN-09-004.
[References](#) | [BibTeX](#) | [LaTeX\(US\)](#) | [LaTeX\(EU\)](#) | [Harvmac](#) | [EndNote](#)
[Detailed record](#) - [Similar records](#)
- 2. Three-subjet distributions in neutral current deep inelastic scattering.**
E. Ron C. Glasman, J. Terron. ZEUS-IN-09-003.
[References](#) | [BibTeX](#) | [LaTeX\(US\)](#) | [LaTeX\(EU\)](#) | [Harvmac](#) | [EndNote](#)
[Detailed record](#) - [Similar records](#)
- 3. 2009 Guide to Funnel: The ZEUS Monte Carlo Production Facility.**
A. Parenti. ZEUS-IN-09-002.
[References](#) | [BibTeX](#) | [LaTeX\(US\)](#) | [LaTeX\(EU\)](#) | [Harvmac](#) | [EndNote](#)
[Detailed record](#) - [Similar records](#)
- 4. Automated calculation of radiative correction to electron-proton charged current DIS at HERA.**
I. Marfin. ZEUS-IN-09-001.
[References](#) | [BibTeX](#) | [LaTeX\(US\)](#) | [LaTeX\(EU\)](#) | [Harvmac](#) | [EndNote](#)
[Detailed record](#) - [Similar records](#)

On the left side of the screenshot, there is a sidebar with a search bar and a "Log" button. At the bottom of the sidebar, there is a "Note:" section. The main content area also has a "Note:" section.



Documentation projects with INSPIRE

- Internal notes from all HERA experiments now available on INSPIRE
 - Experiments no longer need to provide dedicated hardware for such things
 - Password protected now, simple to make publicly available in the future

The screenshot displays the INSPIRE website interface. At the top, there is a navigation bar with links for HEP, INST, HELP, SPIRES, and HEPNAMES. A welcome message states: "Welcome to INSPIRE! INSPIRE is out of beta and ready to replace SPIRE please email us at feedback@inspirehep.net".

The main content area shows a list of "ZEUS Internal Notes" with the following entries:

- 1. Inclusive-jet production in**
J. Terron C. Glasman, ZEUS-IN-09-004.
[References](#) | [BibTeX](#) | [Detailed record](#) - [Similar records](#)
- 2. Three-subjet distributions**
E. Ron C. Glasman, J. Terron C. Glasman, ZEUS-IN-09-003.
[References](#) | [BibTeX](#) | [Detailed record](#) - [Similar records](#)
- 3. 2009 Guide to Funnel: The**
A. Parenti, ZEUS-IN-09-002.
[References](#) | [BibTeX](#) | [Detailed record](#) - [Similar records](#)
- 4. Automated calculation of**
I. Marfin, ZEUS-IN-09-001.
[References](#) | [BibTeX](#) | [Detailed record](#) - [Similar records](#)

A detailed view of the first note is shown, titled "Inclusive-jet production in NC DIS with HERA II - C. Glasman, J. Terron . ZEUS-IN-09-004". It includes a file download section for "ZEUS-09-004" with a version 1 file named "ZEUS-09-004.ps.gz" (130.74 KB) dated 21 Sep 2011, 18:13.

- The ingestion of other documents is under discussion, including theses, preliminary results, conference talks and proceedings, paper drafts, ...
 - More experiments working with INSPIRE, including CDF, D0 as well as BaBar



INSPIRE: Paper histories



Welcome to INSPIRE β . Please go to SPIRES if you are here by mistake.
Please send feedback on INSPIRE to feedback@inspire-hep.net

HEP :: HELP ... SPIRES HEPNAMES :: INST :: CONF :: EXP :: JOBS

[Home](#) > Events with Isolated Leptons and Missing Transverse Momentum and Measurement of W Production at HERA

Information | References (52) | Citations (8) | **H1 internal**

Events with Isolated Leptons and Missing Transverse Momentum and Measurement of W Production at HERA.

H1 Collaboration (F.D. Aaron (Bucharest, IFIN-HH & Bucharest U.) *et al.*) [Show all 256 authors.](#)
2009

Eur.Phys.J. C64 (2009) 251-271
e-Print: [arXiv:0901.0488 \[hep-ex\]](https://arxiv.org/abs/0901.0488)

Abstract: Events with high energy isolated electrons, muons or tau leptons and missing transverse momentum are studied using the full e^+p data sample collected by the H1 experiment at HERA, corresponding to an integrated luminosity of 474 pb^{-1} . Within the Standard Model, events with isolated leptons and missing transverse momentum mainly originate from the production of single W bosons. The total single W boson production cross section is measured as 1.14 pb (stat.) $\pm 0.14 \text{ (sys.) pb}$, in agreement with the Standard Model expectation. The data are also used to establish limits on the $WW\gamma$ gauge couplings and for a measurement of the W boson polarisation.

Keyword(s): INSPIRE: [W: production](#) | [transverse momentum: missing-energy](#) | [DESY HERA Stor](#) | [H1](#)

Record created 2009-01-05, last modified 2010-04-11 [Similar records](#)

[Abstract](#) and [Postscript](#) and [PDE](#) from arXiv.org
[Journal Server](#)
[Reaction Data \(Durham\)](#)

Export
[BibTeX](#), [EndNote](#), [LaTeX\(US\)](#), [LaTeX\(EU\)](#), [NLM](#), [DC](#)

- > Envisage an additional link for H1 members only
- > Provides additional information such as preliminary results, earlier draft versions and documentation from the publication procedure



INSPIRE: Paper histories



Welcome to INSPIRE ?. Please go to SPIRES if you are here by mistake. Please send feedback on INSPIRE to feedback@inspire-hep.net

HEP :: HELP SPIRES HEPNAMES :: INST :: CONF :: EXP :: JOBS

[Home](#) > [Events with Isolated Leptons and Missing Transverse M](#)

[Home](#) >> [Search Results](#)

Information **References (52)** Citation

Events with Isolated Leptons and Missing Transverse Momentum and Measurement of W Production at HERA

Events with Isolated Leptons and Missing Transverse Momentum and Measurement of W Production at HERA

PUBLICATION HISTORY

Preliminary Results

[HEP-EPS 2007 conference paper](#) | July 2007
[Prepared for Deep Inelastic Scattering 2007](#) | April 2007
[Prepared for 42nd Rencontres de Monod \(Electroweak\)](#) | January 2007
[Prepared for the 62nd DESY PRC](#) | October 2006
[ICHEP 2006 conference paper](#) | July 2006
[Prepared for the 60th DESY PRC](#) | November 2005
[HEP-EPS 2005 conference paper](#) | July 2005
[Lepton Photon 2005 conference paper](#) | June 2005
[Prepared for Deep Inelastic Scattering 2005](#) | April 2005
[Prepared for the 58th DESY PRC](#) | October 2004
[Analysis of High Pt HERA II Data](#) | ICHEP 2004 conference paper | August 2004
[High Pt Analysis of the HERA II Data](#) | Prepared for Deep Inelastic Scattering 2004 | April 2004

T0 talks

[Pre-T0 Talk](#) | 08.02.2008
[T0 Talk](#) | 24.07.2008
[T0 Addendum](#) | 14.08.2008

Paper Drafts

[First Draft](#) | [Answers to Draft](#) | 15.08.2008
[Second Draft](#) | [Answers to Draft](#) | 19.11.2008
[Referee Report](#) | 20.11.2008
[Final Version](#) | 06.01.2009

Abs data with prod also

Key

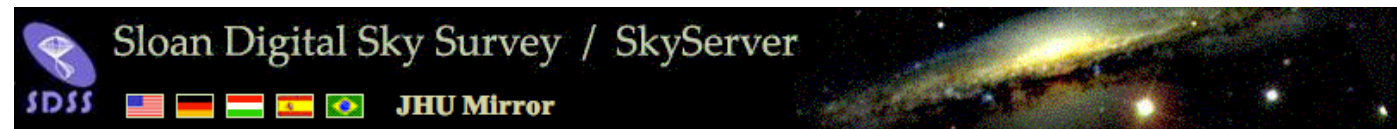
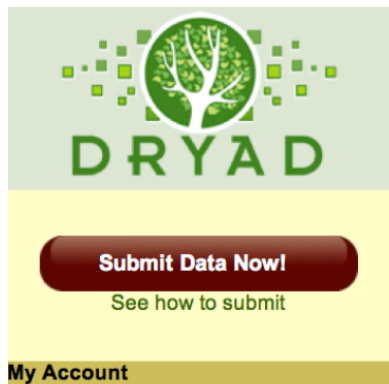
Record created 2009-01-05, last mod

[Abstract and Postscript](#)
[Journal S](#)
[Reaction Data](#)



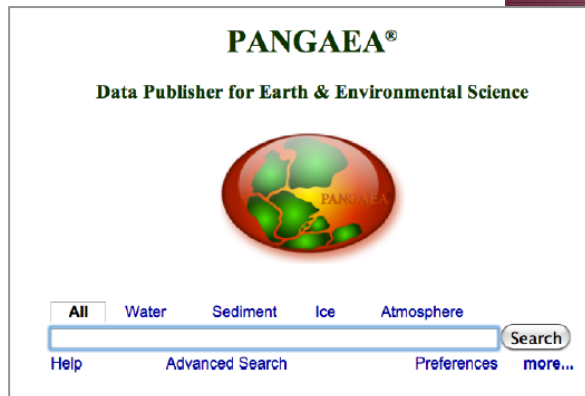
Initiatives in other fields

- Data preservation and in particular open access and data sharing are present in other fields such as:
 - Astrophysics, molecular biology, earth sciences, humanities and social sciences



Blue Ribbon Task Force
on Sustainable Digital Preservation and Access

[About Us](#) | [Members](#) | [Publications](#) | [Bibliography](#) | [News Center](#) | [Intra](#)



[Home](#) | [News](#) | [Docs](#) | [WCS](#) | [Samples](#) | [Libraries](#) | [Viewers](#) | [Utilities](#) | [Keywords](#) | [Conventions](#) | [Resources](#)

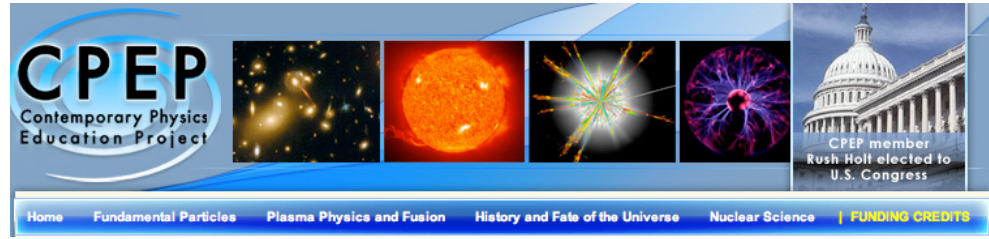
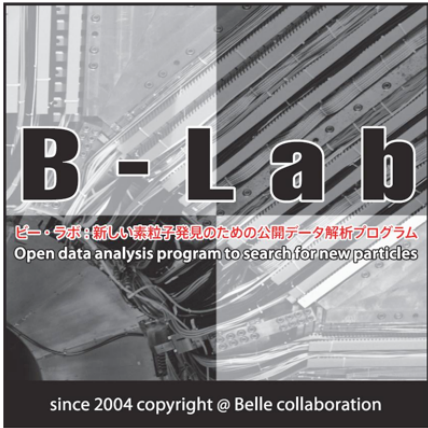
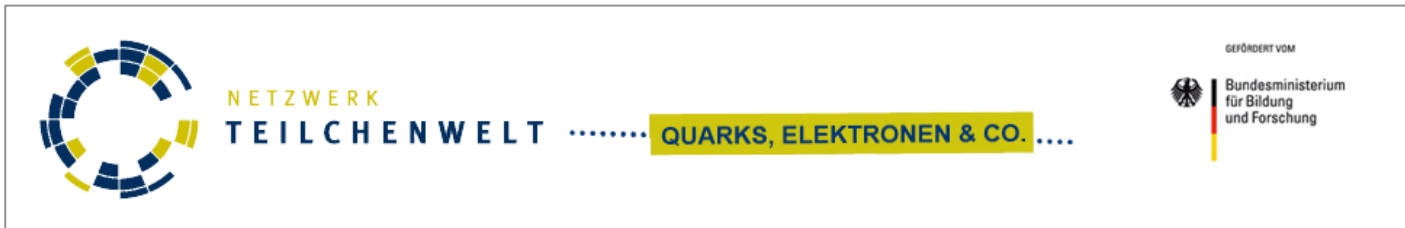
The FITS Support Office

at NASA/GSFC



HEP outreach initiatives

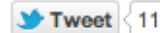
- Many initiatives promoting outreach efforts and to improve the public understanding of science in general



Science Hack Day: Increasing the access to LHC data

<http://cms.web.cern.ch/news/cms-public-data-activity-scoops-prize-nairobi>

CMS public data activity scoops prize in Nairobi



An application using real event data from CMS has won "Best Science" prize in a public "Science Hack Day" held in Nairobi between 13th and 15th April 2012. Science Hack days bring together a wide range of enthusiastic members of the public to create something completely new using existing scientific systems or data.

The winning application visualized real CMS di-muon events from the 2011 LHC run, which are made public for use in various educational programmes, such as the [IPPOG Masterclasses](#), [Quarknet](#) and [I2U2](#). The application showed an animation of muons produced in CMS superimposed on a map of the world, showing where they would go if they were to continue without stopping (which they don't in reality).

Other prizes were awarded to Leah Atieno, a 15-year-old high-school student, for a voice-controlled walking robot and Denis Munene for a crowd-mapping platform to help promote the fight against malaria.

The Nairobi event, involving 240 developers, is part of broader series of Science Hack Day events. CMS data previously featured in another very successful event in San Francisco.

News article by Gythan Munga, [HumanIp](#)

[See photos of the event](#)

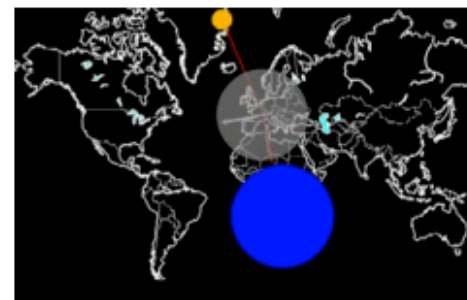
[Youtube film](#)

[Link to more Science hack events](#)

2012-04-20, by Lucas Taylor



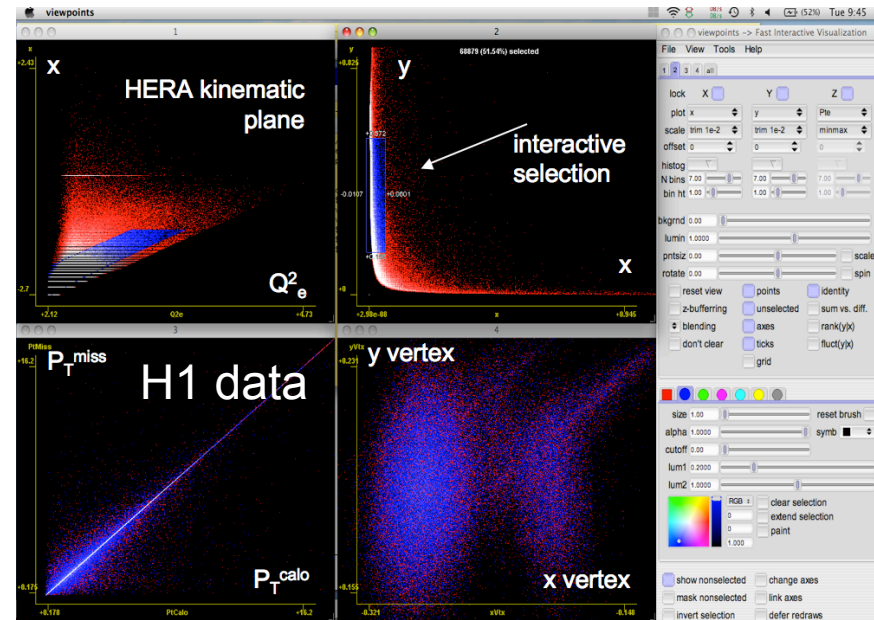
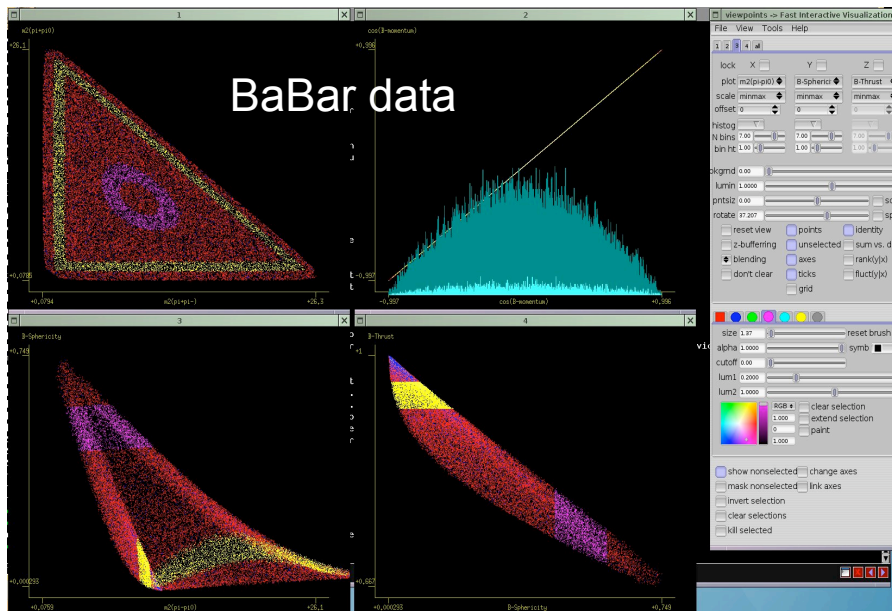
CMS use of public data in a "Science Hack" event in Nairobi. Photo credit: [Matt Biddulph](#), via Flickr



Application developed to visualise where muons from CMS would go if they continued forever



Some Outreach activity within DPHEP



> One nice collaboration was between BaBar and H1, using the NASA Viewpoints application

> Common format:
simple text

#	Q2e	y	x	PtCalo	PtMiss	Ex	Ptx	Phix	Thx	Ee	Pte	Phie	The	Empz	xVtx	yVtx	zVtx
377.673	0.174	0.021	2.769	2.769	189.685	15.153	-11.780	8.231	26.226	17.665	164.138	137.656	50.542	-0.237	0.207	2.582	
185.111	0.399	0.005	2.133	2.133	41.933	12.652	87.669	36.327	18.252	10.544	-93.948	144.713	57.878	-0.248	0.192	-12.829	
187.320	0.211	0.009	2.584	2.584	51.742	9.773	78.869	13.682	23.482	12.160	-106.349	148.813	55.164	-0.266	0.208	2.968	
264.266	0.508	0.005	0.238	0.360	35.343	11.738	-138.270		64.975	15.984	11.407	41.034	134.465	57.043	-0.225	0.210	5.925
229.056	0.043	0.052	4.204	5.067	65.601	19.196	72.870	17.842	28.485	14.805	-98.351	148.685	58.941	-0.237	0.199	-7.082	
275.596	0.121	0.022	4.277	4.282	78.331	18.413	51.596	14.380	26.750	15.562	-139.235	144.425	55.018	-0.234	0.200	-5.038	
240.102	0.183	0.013	3.513	3.434	67.134	17.402	85.049	17.201	24.719	14.004	-92.840	145.491	56.060	-0.266	0.202	-1.606	
451.996	0.209	0.021	1.723	1.723	49.126	17.196	66.018	24.927	25.936	18.913	-114.452	133.180	55.810	-0.259	0.190	-13.705	
524.251	0.572	0.009	2.170	2.170	43.738	17.555	171.073	61.182	16.573	14.987	-11.341	115.274	58.543	-0.249	0.208	4.410	
391.944	0.000	0.000	2.107	2.107	183.513	21.270	75.875	6.693	31.602	19.959	-100.713	140.834	58.375	-0.245	0.203	-0.630	
201.600	0.212	0.009	4.441	4.441	44.890	17.098	-92.989	27.261	23.578	12.605	86.968	147.683	55.361	-0.243	0.212	7.653	
335.881	0.052	0.064	16.769	16.769	29.256	1.142	-90.021	2.250	29.219	17.848	83.461	142.349	52.723	-0.242	0.214	10.137	
286.039	0.009	0.315	2.514	2.514	194.560	18.922	-83.365	5.616	29.944	16.837	92.126	145.787	56.826	-0.254	0.193	-11.169	
207.703	0.137	0.015	8.095	8.095	84.993	21.487	82.549	15.129	25.701	13.389	-95.886	148.605	53.237	-0.258	0.205	0.895	
387.371	0.358	0.011	1.272	1.272	78.266	15.460	-91.456	17.173	21.232	15.772	93.071	132.027	55.684	-0.236	0.198	-5.242	
855.333	0.509	0.017	2.588	2.588	88.511	23.066	70.622	21.191	21.306	20.499	-110.759	105.828	56.196	-0.237	0.178	-24.198	
154.527	0.667	0.002	3.509	3.509	72.273	8.810	174.450	101.604	10.598	7.176	-28.067	137.379	92.478	-0.240	0.207	4.578	
304.756	0.025	0.121	1.622	1.622	120.020	17.765	-145.756		8.522	29.678	17.240	39.272	144.486	55.298	-0.228	0.193	-5.659
278.950	0.627	0.004	3.813	0.726	37.163	9.588	124.342	60.056	12.831	10.205	-53.435	127.311	52.247	-0.243	0.190	-10.069	



HERA data for preservation



Final data reprocessing to mDST completed in 2009

- Basic preserved data format: ROOT based “Common Ntuples”
- Ultimately RAW, MDST data and MC removed from robots, keep only cNuptles
- Final production of data/MC cNuptles started, to be completed early 2013



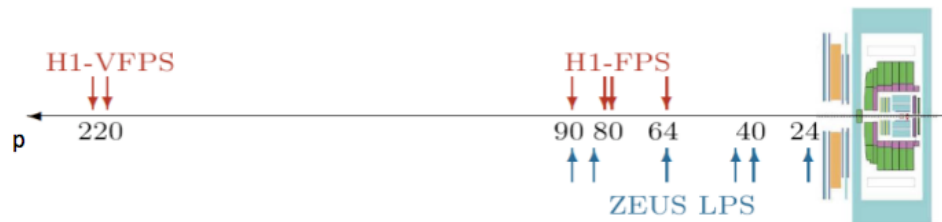
Final reprocessing (DST-7) of HERA II data in 2009, HERA I done in 2012

- Final version of *common analysis software environment + files*, H1OO also done
- Preserve RAW data, as well as DST-7 and H1OO 4.0 versions
- Large MC production of up $2 \cdot 10^9$ events / year, preserved MC sets to be decided



Final data and MC production completed in 2012

- Main format for analysis is the mDST, this is the one to be preserved
- Importantly for HERMES, all data/MC productions now moved to dCache



Dialogue with DESY machine group concerning their HERA data



HERA data for preservation



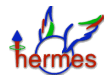
Final data reprocessing to mDST completed in 2009

- Basic preserved data format: ROOT based “Common”
- Ultimately RAW, MDST data and MC removed from ... nuptles
- Final production of data/MC cNuptles started ... 2013



Final reprocessing (DST-7) of HERA ... done in 2012

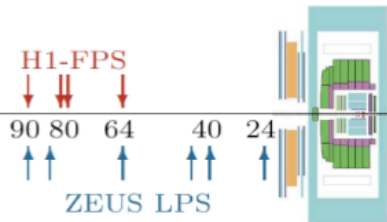
- Final version of common analysis ... + files, H100 also done
- Preserve RAW data, as ... 4.0 versions
- Large MC production ... year, preserved MC sets to be decided



Final data and ... ed in 2012

- Main for ... mDST, this is the one to be preserved
- In ... all data/MC productions now moved to dCache

Total for HERA experiments: ~ 1 PB
Data preservation is not about the data!



Dialogue with DESY machine group concerning their HERA data



Long term archival storage for the HERA data

- > Dedicated system too costly in both hardware and person-power
- > All collaborations now using dCache for mass storage, and this system will continue at DESY-IT for the LHC, photon-physics, others..
 - By using dCache this is completely transparent for user, relying on IT admin work
 - Active check for the data consistency on disk level, tape copy only for the case if disk copy is corrupted; corresponding checks also on the tape (checksum)
 - For data which have no copy on disk: two copies on tapes (different technology old vs new, but both still readable), regular migration to new media
 - Data which should be archived, but not online all the time – re-pack into larger files
- > System properties also defined by which data needed “always online”

- Initial estimates:
 - ~ 1 PB in total
- Different strategies visible

	Online	Not online	Total
H1	~ 250 TB	~ 100 TB	~ 350 TB
ZEUS	~ 250 TB		~ 250 TB
HERMES	~ 40 TB	~ 350 TB	~ 390 TB
Total	~ 540 TB	~ 450 TB	~ 990 TB



Isn't it obvious, virtualisation will solve everything?

My first and very naïve ansatz

- OK, why don't we just put everything in a virtual machine?
 - Data archival is done elsewhere, just need "to plug that into the VM"
 - Your VM contains everything you need to develop and run code and analysis
- The problem would then be reduced to maintain virtual images, and maintain their ability to run. In the Cloud era, seems like a trivial task
- Problems: Everything in IT is a moving target:
 - Will your network always be the same?
 - Will your access protocol always be the same?
 - Are you sure you do not need new software (e.g. MC generators) that require a new OS?
 - Are you sure your i386/SL4 VM will produce the same results when emulated on a quantum computer in NN years?
 - What about service you need, like CondDB,...
- Naïve virtualization will not work... but still, virtualization can help



Freezing vs rolling

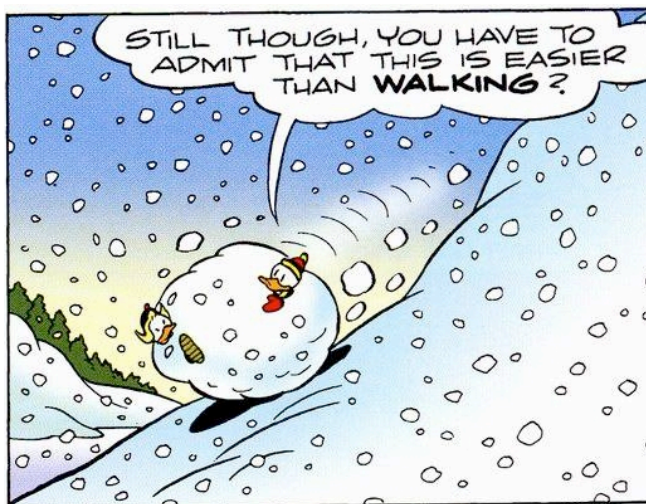


> Pro Freezing

- One-time effort, very small maintenance outside of analysis phase
- Also allows software w/o code (but might fail with DRM / licensing issues)

> Cons Freezing

- Rely on certain standards and protocols that may evolve
- Potential performance problems



> Pro Test-driven migration

- Usability and correctness of code is guaranteed at every moment
- Data accessibility and integrity can be checked as well
- Fast reaction to standard/protocol changes
- General code quality can improve, as designed for portability and migration

> Cons Test-driven migration

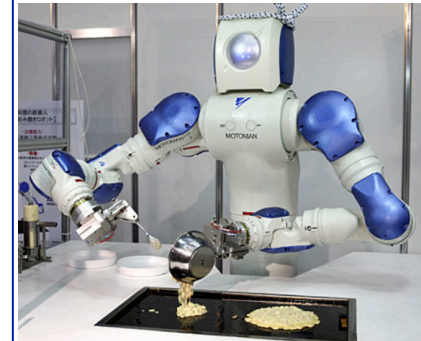
- Needs long-time intervention, more man-power and resources needed
- Some knowledge of the frameworks must be passed to maintainers

Pizza Preservation



- > Couple of days
 - Fridge
- > Couple of month
 - Deep freezer
- > Couple of years???
 - Preserve the recipe
 - Practice it often: You will not forget the recipe and you can detect variations in external dependencies

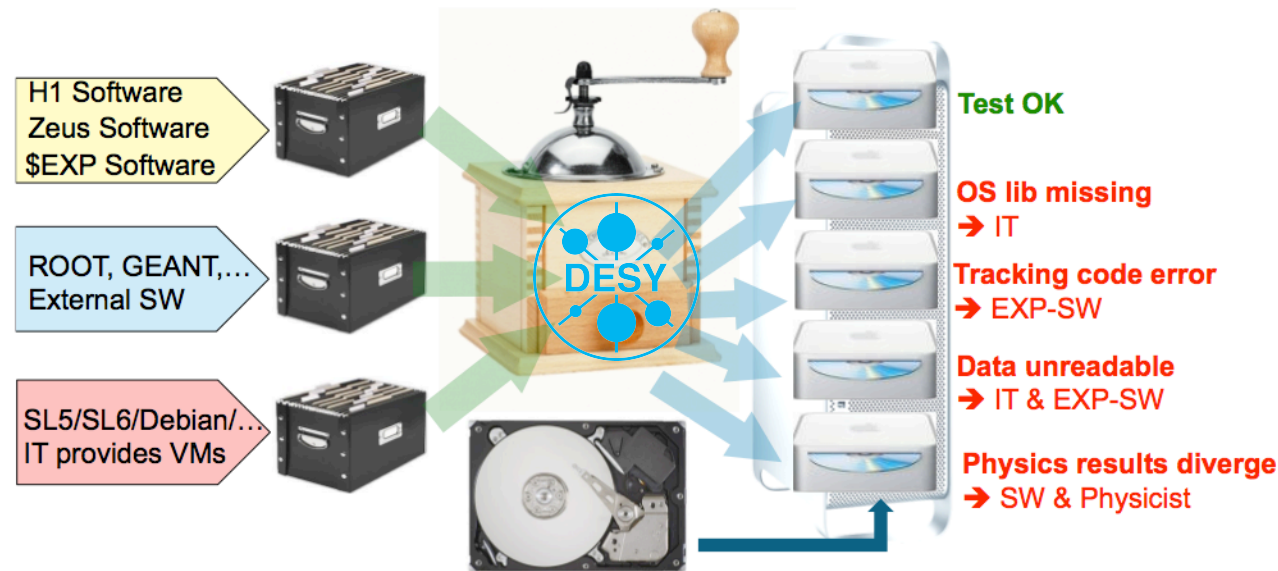
Y. Kemp, D. Ozerov,
CHEP 2012



- > Whilst freezing the software and environment is easy to do, long term use and correctness of the results not guaranteed
 - Naïve assumption virtualisation solves everything breaks down at the first security hole
- > Freezing software is *OK* if the timeline and scope are reduced, but if changes are needed this is more difficult the longer software is frozen
- > Better to cook the same recipe again and again (and maybe even allow it to be improved), validating the output *automatically*
 - Virtualisation can help!



The Software Preservation System @ DESY



- > Automated validation system to facilitate future software and OS transitions
 - **Uses virtualisation techniques to repeatedly run well defined tests**
 - **Perform checks of different and evolving environments (OS, s/ware configuration)**
 - **Stand alone system: No hidden dependencies or /afs access etc: rigorous testing**
 - **Automatically check these results against predefined, default values**
 - **Notify when test results differ from these values**
 - **Separate responsibilities of IT and the experiments**



The Software Preservation System @ DESY



- > Automated validation system to facilitate future software and OS transitions
 - **Uses virtualisation techniques to repeatedly run well defined tests**
 - **Perform checks of different and evolving environments (OS, s/ware configuration)**
 - **Stand alone system: No hidden dependencies or /afs access etc: rigorous testing**
 - **Automatically check these results against predefined, default values**
 - **Notify when test results differ from these values**
 - **Separate responsibilities of IT and the experiments**

The Software Preservation System @ DESY



- > Automated validation system to facilitate future software and OS transitions
 - **Uses virtualisation techniques to repeatedly run well defined tests**
 - **Perform checks of different and evolving environments (OS, s/ware configuration)**
 - **Stand alone system: No hidden dependencies or /afs access etc: rigorous testing**
 - **Automatically check these results against predefined, default values**
 - **Notify when test results differ from these values**
 - **Separate responsibilities of IT and the experiments**

The Software Preservation System @ DESY



- > Automated validation system to facilitate future software and OS transitions
 - **Uses virtualisation techniques to repeatedly run well defined tests**
 - **Perform checks of different and evolving environments (OS, s/ware configuration)**
 - **Stand alone system: No hidden dependencies or /afs access etc: rigorous testing**
 - **Automatically check these results against predefined, default values**
 - **Notify when test results differ from these values**
 - **Separate responsibilities of IT and the experiments**

The Software Preservation System @ DESY



- > Automated validation system to facilitate future software and OS transitions
 - Uses virtualisation techniques to repeatedly run well defined tests
 - Perform checks of different and evolving environments (OS, s/ware configuration)
 - Stand alone system: No hidden dependencies or /afs access etc: rigorous testing
 - Automatically check these results against predefined, default values
 - Notify when test results differ from these values
 - Separate responsibilities of IT and the experiments

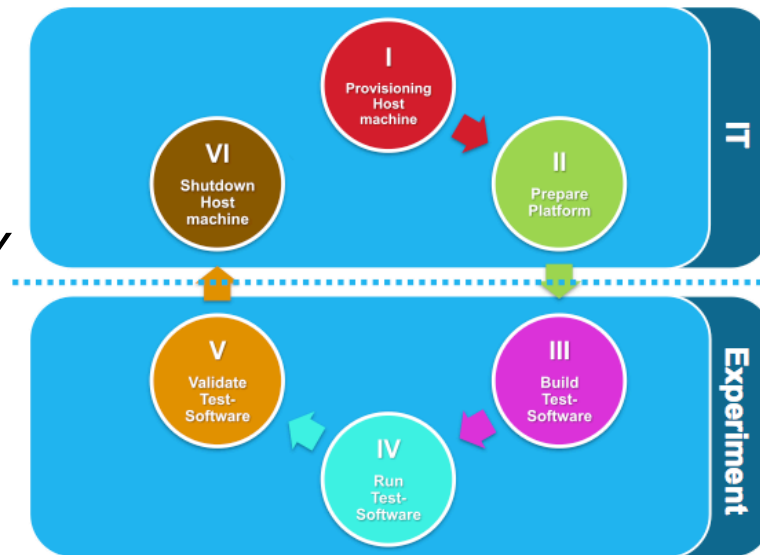
First test runs in pilot project at CHEP 2010

	SL4	SL5	Fedora 13	
ROOT V5.26	-no F77 compiler gfortran found -libX11 MUST be installed	Estimated ROOTMARKS: 1534.29	Estimated ROOTMARKS: 1512.76	Compilation
H1Data analysis	Processed 47243 events with J/Psi candidates Histogram written to jpsi_mods.root	Processed 47243 events with J/Psi candidates Histogram written to jpsi_mods.root	Processed 47243 events with J/Psi candidates Histogram written to jpsi_mods.root	Run pre- compiled tgz using compat libs
ZEUS MC prod	> ls -lh ZEUSMC.HFSZ627.E89 54.GRAPE.Z01 4.2 MByte	> ls -lh ZEUSMC.HFSZ627.E89 54.GRAPE.Z01 4.2 MByte	> ls -lh ZEUSMC.HFSZ627.E89 54.GRAPE.Z01 4.2 MByte	Run pre- compiled tgz using compat libs
HERA-B	Compilation OK DB connect fails	Compilation OK DB connect fails	Compilation failed – needs code change	Compilation



The sp-system: Towards the full implementation

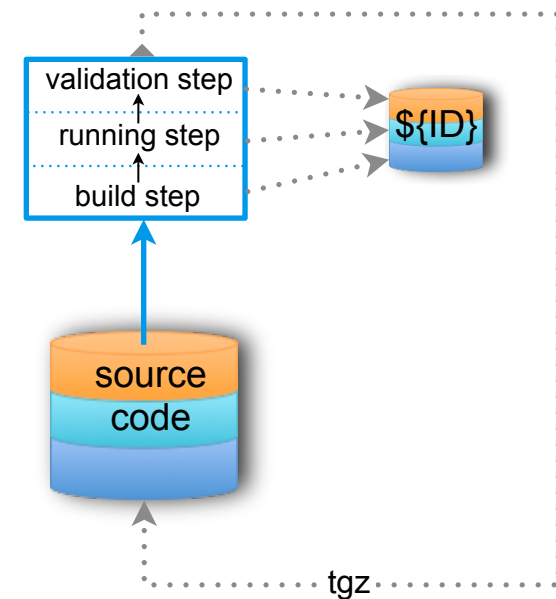
- > Pilot project in 2010
 - Single configuration, simple tests
- > Full implementation now installed at DESY
- > Common baseline of SLD5 / 32-bit achieved in 2011 by all experiments
 - Sound starting point for validation
- > Following OS configurations now available in sp-system:
 - sl5.6/64(gcc4.4), sl5.7/32(gcc4.4), sl5.7/64(gcc4.1), sl5.7/32(gcc4.1), sl6.2/64(gcc4.4)
- > In addition, to multiple ROOT versions
 - 5.26.00d, 5.28.00c, 5.30.05, 5.32.00, 5.34.01
- > 64-bit systems a major step toward migrations to future OS and hardware
 - SL6 will only be supported in 64 bit variant at DESY
 - NFS4.1 technology, to be used in dCache, native only in SL6.2/64 or higher



Running jobs in the sp-system

> Initial step

- Compilation of analysis (level 3) and sim/rec (level 4) software
- **Or:** use tar-balls with pre-compiled software
- Provide access to software
 - Copy tar-balls to persistent storage
- All output kept in directory with unique name



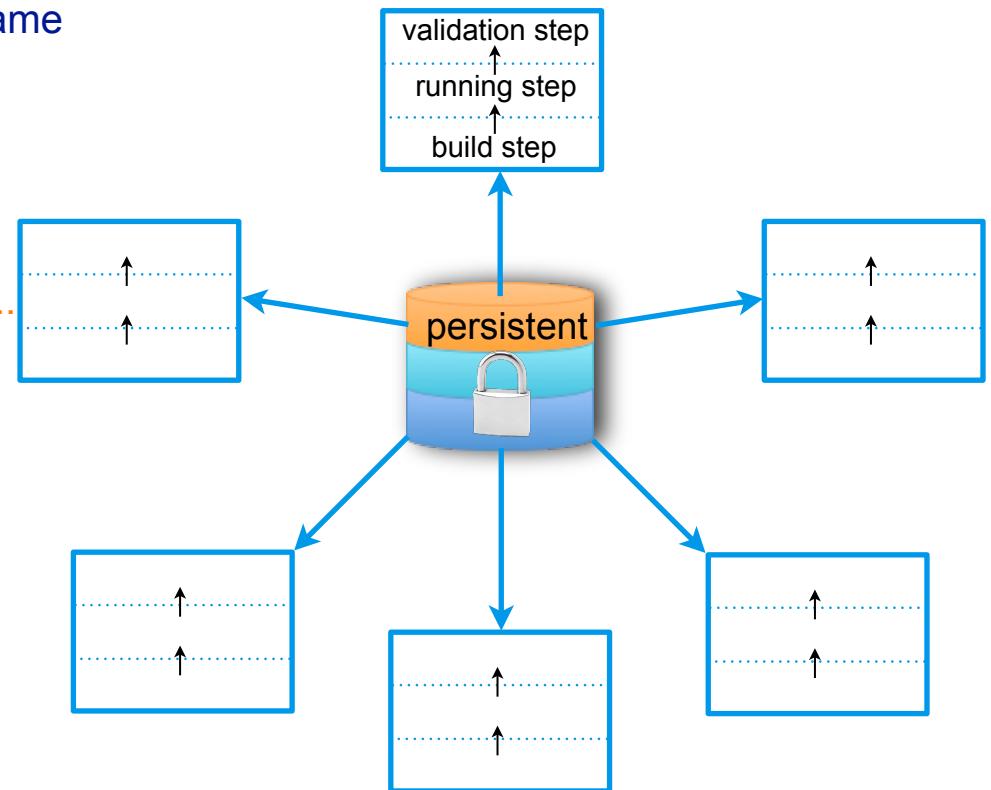
Running jobs in the sp-system

> Initial step

- Compilation of analysis (level 3) and sim/rec (level 4) software
- **Or:** use tar-balls with pre-compiled software
- Provide access to software
 - Copy tar-balls to persistent storage
- All output kept in directory with unique name

> Run parallel tests

- Set up software environment
- Validate binaries with persistent input
 - e.g. event display, database access, ...



Running jobs in the sp-system

> Initial step

- Compilation of analysis (level 3) and sim/rec (level 4) software
- **Or:** use tar-balls with pre-compiled software
- Provide access to software
 - Copy tar-balls to persistent storage
- All output kept in directory with unique name

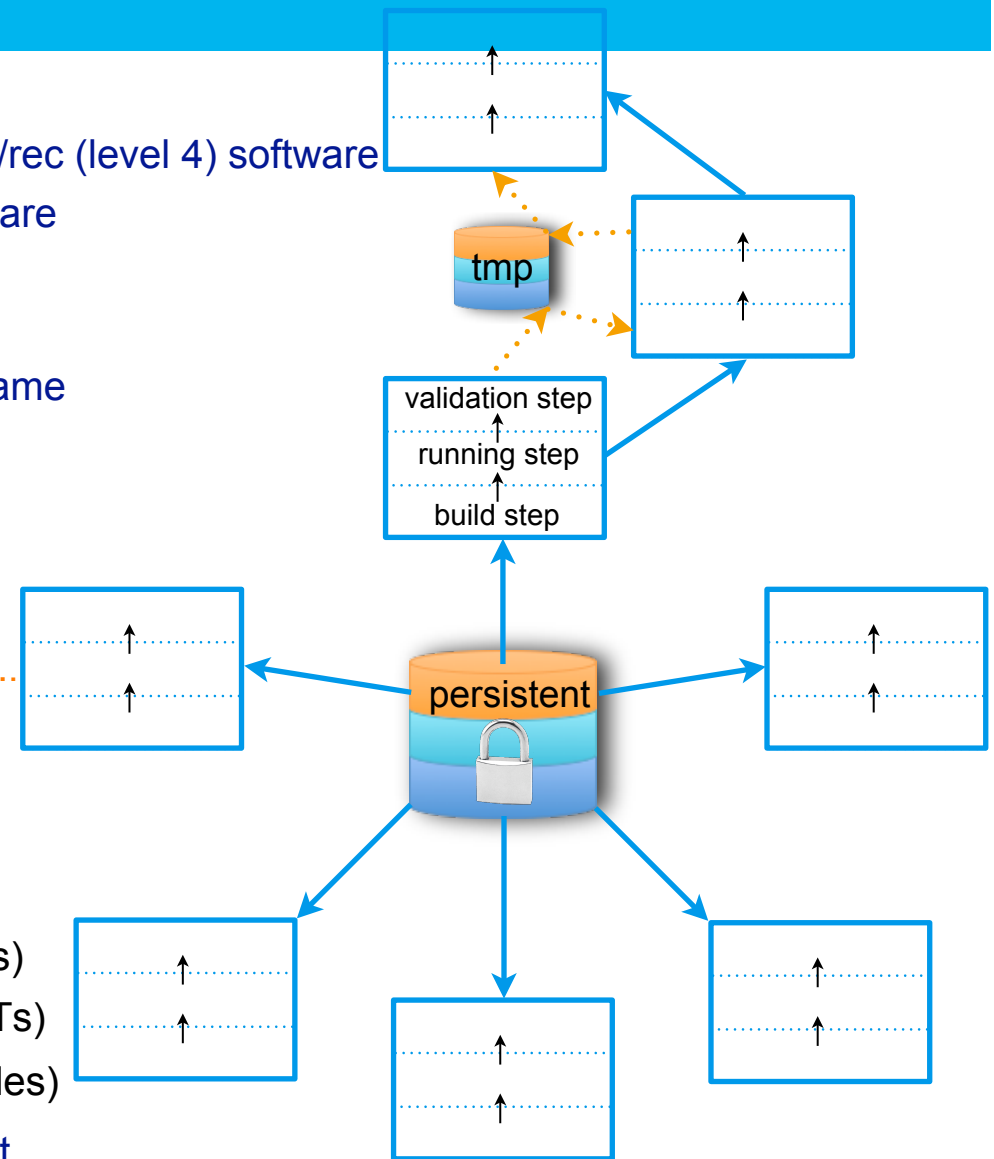
> Run parallel tests

- Set up software environment
- Validate binaries with persistent input
 - e.g. event display, database access, ...

> Run sequential tests

- Set up software environment
- Validate file production
 1. **MC generation** (produce gen files)
 2. **Reconstruction** (gen. files → DSTs)
 3. **Analysis level** (DSTs → ROOT files)
- Tests use output of previous test as input

> Results remain accessible or can be reproduced with identical results



Running jobs in the sp-system

> Initial step

- Compilation of analysis (level 3) and sim/rec (level 4) software
- **Or:** use tar-balls with pre-compiled software
- Provide access to software
 - Copy tar-balls to persistent storage
- All output kept in directory with unique name

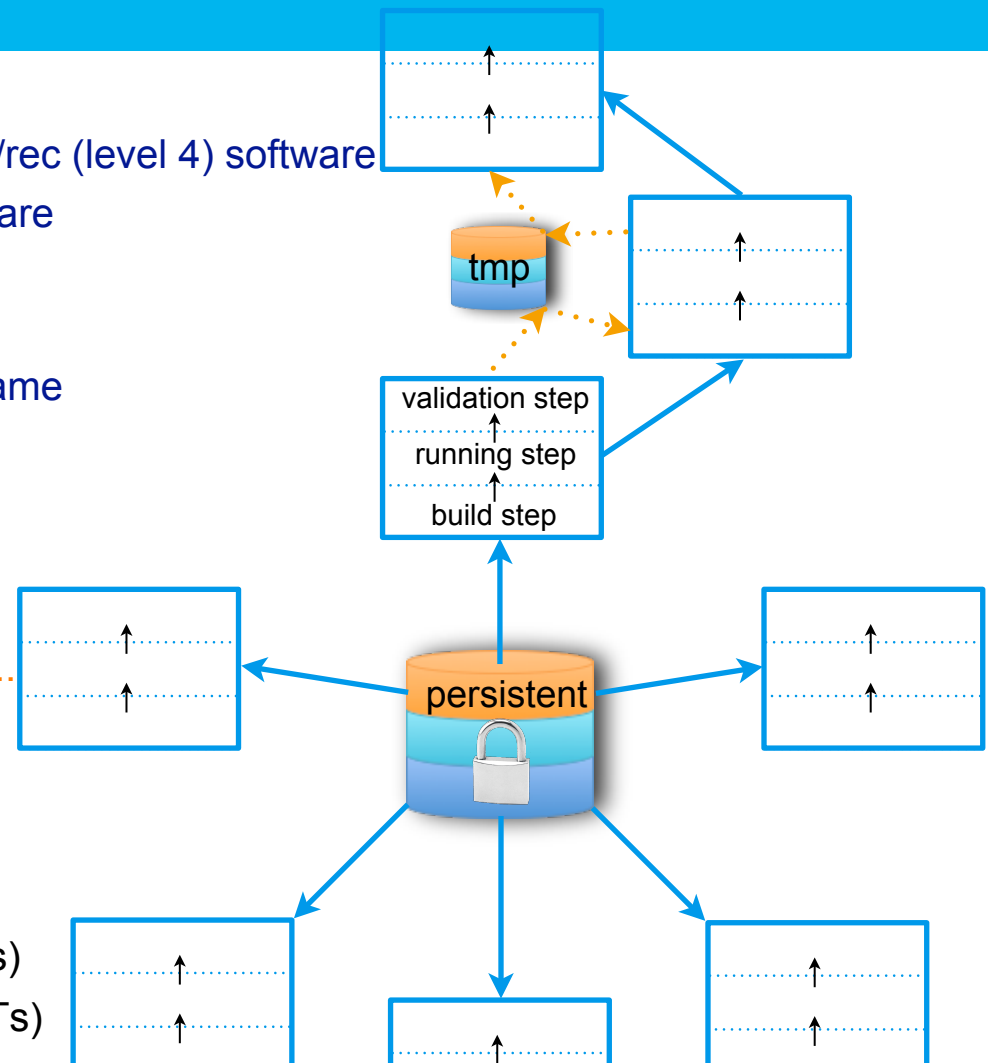
> Run parallel tests

- Set up software environment
- Validate binaries with persistent input
 - e.g. event display, database access, ...

> Run sequential tests

- Set up software environment
- Validate file production
 1. MC generation (produce gen files)
 2. Reconstruction (gen. files → DSTs)
 3. Analysis level (DSTs → ROC)
- Tests use output of previous test as

> Results remain accessible or can



→ It is essential to have robust definition of complete set of experimental tests
 The nature and number dependent on desired preservation level

First sketch of H1 tests

```

*****
*****
++h1 executables
*****
*****
antlr
batch_kinit
carli
chk_tree
dlg
fpack
fplist
fpmmerge
fsubset
h1ftemu
h1geanonly
h1ieeefp.o / h1ieeefp.cpp
h1rec
h1sim
h1simcheck
h1simrec
hostr
H4shis
H4his
H4m
H4s
look
ltab
ndbint
nqs2pbs
pbs_tclsh
pbs_wish
pbsdsh
pbsnodes
printjob
printtracking
qaller
qdel
qdisable
qenable
qhold
qmgr
qmove
qmsg
qorder
qreun
qris
qrun
qselect
qsig
qstart
qstat
qstop
qsub
qterm
refresh
refresh_kinit
tracejob
xpbs
xpbsmon
    
```

55

```

*****
*****
++h1 libraries
*****
*****
#cemlib-gcc44
libLHAPDF.so
libriadne412.a
libbases.a
libbos.a
libcascade2.a
libdatman.a
libdiffm.a
libpack.a
libpack.so
libgksdummy.a
libh1bstrec.a
libh1eclass.a
libh1ftemu.a
libh1geang.a
libh1geanh.a
libh1geant.a
libh1i4.a
libh1look.a
libh1mcutit.a
libh1ndb.a
libh1phan.a
libh1qt.a
libh1rec.a
libh1sim.a
libh1trig.a
libh1util.a
libheracles*.a
libheracles*.so
libhztool.a
libjset74.a
liblook.a
libpythia62.a
libpythia64.a
librappap31.a
libshift.a
    
```

20
??

36

```

*****
*****
++h1oo packages
*****
*****
H1Analysis
H1AnalysisExample
H1Arrays
H1Banks
H1Benchmarks
H1Binning
H1Bos2oop
H1CalcPointers
H1CalcWeights
H1Calculator
H1CalibTrigger
H1CaloTrigger
H1Clusters
H1Cuts
H1ElecCalibration
H1Examples
H1Filter
H1Finder
H1Geom
H1HadronicCalibration
H1Hat
H1HatFilter
H1HfsFinder
H1JetFinder
H1Macros
H1Mods
H1MuonFinder
H1NonepBgFinder
H1OOBanks
H1Ods
H1PartEmFinder
H1PhysUtils
H1Pointers
H1QCDFunc
H1Red
H1SVFit
H1Selection
H1Skeleton
H1SoftLeptonId
H1Steering
H1SubDetInfo
H1Tools
H1Tracks
H1TrkFinder
H1UserCim
H1UserDstar
H1UserFit
H1UserLifetime
H1Wrappers
oo_tools
#share
    
```

51

"only" UseTiming(v2)
x2
x3
x2
x2
x2
+ Mayfield
lots

```

*****
*****
++h1oo binaries
*****
*****
AnalysisExample
AnalysisExampleExtraction
AnalysisExamplePlots
H1Bos2oop
H1Makeptr
L12Root
MakeInputTable
TestQCDFunc
batchAnalysis
boosted Jets
checkcim
cintsteering
clusters_ods
copyMyEvents
create_eventlist
dbaccess
deleteJobs
dst2all
dst2ods
dstar_mods
empz_hat
h1red
h1root
jpsi_mods
kaonfind_ods
l1te_hat
lumicalc
mergeAnalysis
mynkicim
ods2modshat
oolist
columi
oomclumi
oomove
oosubset
read_dstartree
read_eventlist
read_ods
read_usertree
rerun_finder
rerun_rec
resubChains
snapshot
steermanager
test_binning
write_eventlist
    
```

46

```

*****
*****
++h1oo libraries
*****
*****
libH1Analysis.so
libH1AnalysisExample.so
libH1Arrays.so
libH1Benchmarks.so
libH1Binning.so
libH1CalcPointers.so
libH1CalcWeights.so
libH1Calculator.so
libH1CaloTrigger.so
libH1Clusters.so
libH1Cuts.so
libH1ElecCalibration.so
libH1Filter.so
libH1Filter_odsonly.so
libH1Finder.so
libH1Geom.so
libH1HadronicCalibration.so
libH1Hat.so
libH1HatFilter.so
libH1HfsFinder.so
libH1JetFinder.so
libH1MagFieldOO.so
libH1Mods.so
libH1MuonFinder.so
libH1NonepBgFinder.so
libH1OOBanks.so
libH1Ods.so
libH1PartEmFinder.so
libH1PhysUtils.so
libH1Pointers.so
libH1QCDFunc.so
libH1Red.so
libH1RedLook.so
libH1Red_bos.so
libH1SVFit.so
libH1Selection.so
libH1Skeleton.so
libH1SoftLeptonId.so
libH1SoftLeptonId_Impl...so
libH1Steering.so
libH1SubDetInfo.so
libH1Tools.so
libH1Tracks.so
libH1TrkFinder.so
libH1UserCim.so
libH1UserDstar.so
libH1UserDstar_fill.so
libH1UserFit.so
libH1UserFit_Filter.so
libH1UserLifetime.so
libH1UserTiming.so
libH1UserTiming_fill.so
libH1Wrappers_bos.so
libH1Wrappers_fastjet.so
libH1Wrappers_geom.so
    
```

74

```

libH1Wrappers_Jumi.so
libH1Wrappers_ndb.so
libH1Wrappers_neurobays.so
libSISconePluginOO.so
libUser.so
libbosutil.so
libcemlibOO.so
libfastjetOO.so
libfortran.so
libfortranpatchOO.so
libfortranshared.so
libfortranstat.a
libpackOO.so
libh1ndbOO.so
libh1recOO.so
libmbddummy.so
libneurobaysOO.so
libsisconeOO.so
libutildummy.so
    
```

~100...so compiler
~60 core files
25+12 H1so + 115mb + 115mb
H1so "tests" ~37
+ Suptec → 10, includes MC, DR?
+ dst2all
+ fpack 3
+ ndbint, including only ~60
S>10 analysis, including:
+ all hat → for each
+ all calc 15020
event dply → haly env
10



First sketch of H1 tests

> Validate compilation of

- ~100 (shared) library objects
- ~60 executables
- MC generators not yet included
- most important:

- simrec - reconstruction / dst production
- dst2all - h100 file production

> Validate correct running of

- 37 x h100
- 4 x fpack, ndb
- ≥10 x h1simrec → dst2all → analysis

One test for every run period + MC / (DQHat)

- ... Let's say about 60 executables

> Run and validate physics analyses

- (At least) one test for every run period
- Inclusive & all HAT/H1 Calculator variables
- 5-10 'real' physics analyses

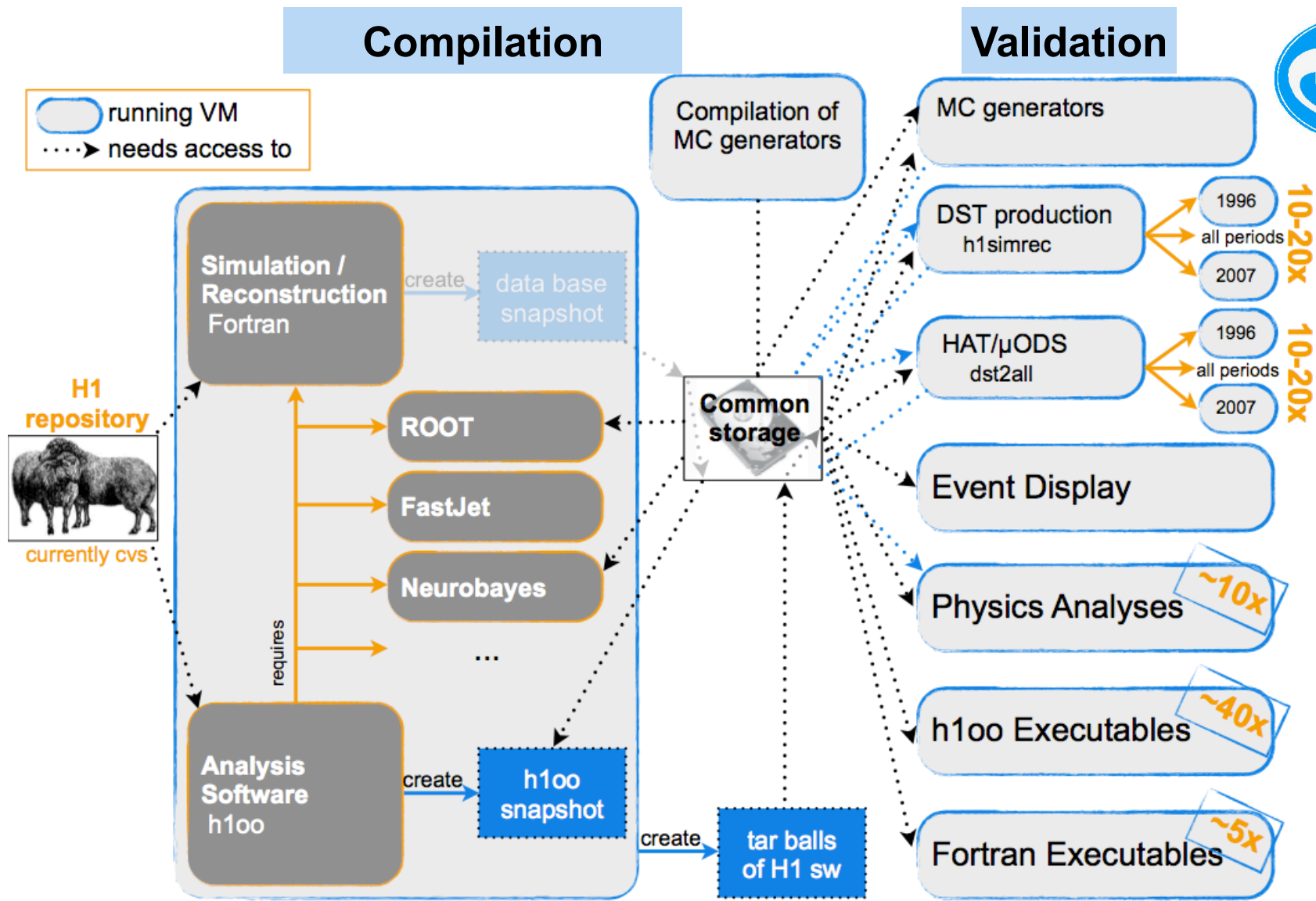
> Check event display

Handwritten notes on a grid background, enclosed in an orange box. The notes are organized into sections:

- comp**: ~100 .so compile hrs.
- exec**: ~60 executables
- exec**: 25 + 12 H1Exmde + H1moro
- H100 "tests"**: (37)
- + **simrec** → (10), including MC, DQ?
- + **dst2all**
- + **fpack** (3)
- + **ndbint**, including exit (~60)
- 5 > 10 analyses, including:**
 - + all hat → for each
 - + all calc (15 > 20)
- event display** → longy exes
 - menu
 - r.hly desten.
 - tree (multi etc.)
 - dmp.
- Σ ≈ 250 tests**



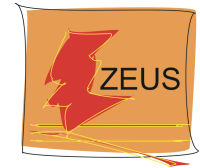
Example structure of experimental tests: H1 (Level 4)



Including compilation of individual packages: about 250 tests planned by H1

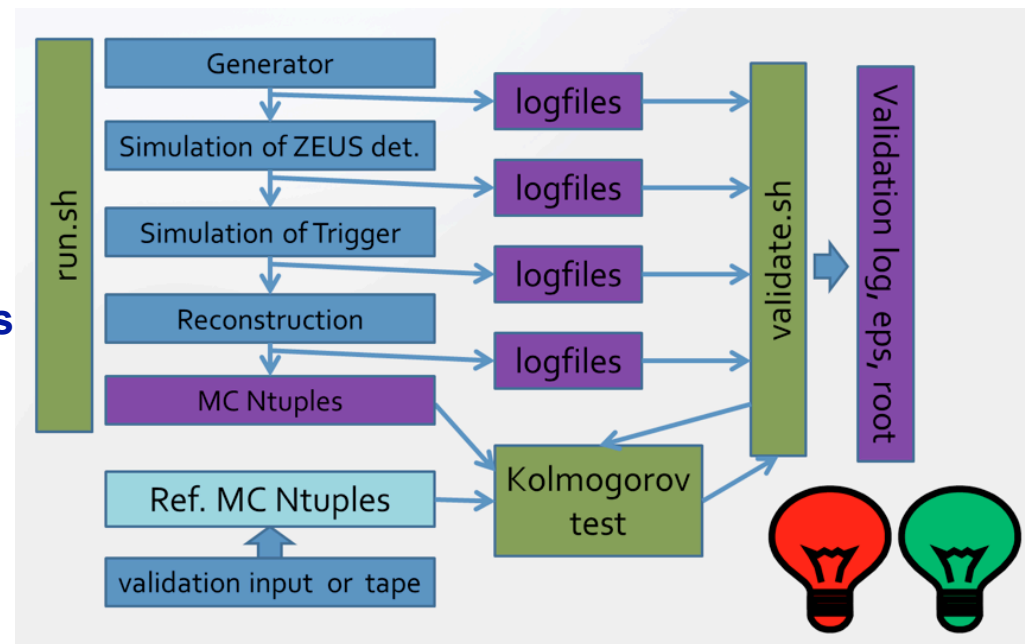


Example structure of experiment tests: ZEUS (Level 3 + MC chain)






- > ZEUS strategy: use ROOT based analysis level Common Ntuples as data format for preservation – DPHEP level 3
- > Only external dependence is ROOT
 - Validation of new ROOT versions included as analysis level tests in the **sp-system**

- > However, the MC production chain pre-compiled executables will also be preserved as a standalone package
 - **Remaining ZEUS SL3 executables continue to work on the SL6/64 OS**
- > In addition, an interface for new generators is developed, which is also included in the validation system

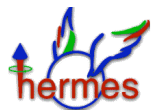


Putting it all together

Operating System		SL5 32bit				SL5 64bit					SL6 64bit
		External Dependencies				Cernlib		Fastjet	Neuro-bayes		
Process		5.26	5.28	5.30	5.32	2005	2006	2.3.3	2008 0312	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 uDST (precompiled s/w)	█	█	█	█						█
	Reconstruction (precompiled s/w)	█									█
	Producing uDST (precompiled s/w)	█									█
	Validation	█	█	█	█	█					

Full chain, including compilation of all H1 software, from MC generation, through to validation of analysis level (e.g. high Q^2 neutral current) histograms now in place within the sp-system

Putting it all together



Process	Operating System	SL5 32bit				SL5 64bit					SL6 64bit	
		External Dependencies	ROOT 5.26	ROOT 5.28	ROOT 5.30	ROOT 5.32	Cernlib 2005	Cernlib 2006	Fastjet 2.3.3	Neuro-bayes 2008 0312		Neuro-bayes 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 uDST (precompiled s/w)												
Reconstruction (precompiled s/w)								No dependence				
Producing uDST (precompiled s/w)												
Validation												

Here a much finer granularity needed for displaying the results !

Digesting the validation results


- Display the results of the validation in a comprehensible way: web based interface
- The test determines the nature of the results
 - Could be simple yes/no, plots, ROOT files, text-files with keywords or length, ...

H1 Validation Results

List of available validation runs: success error(s) work to be done not in list

- [H1_64bit_VT79_4.0.21](#)

Description of used software version:
H1_64bit_VT79_4.0.21



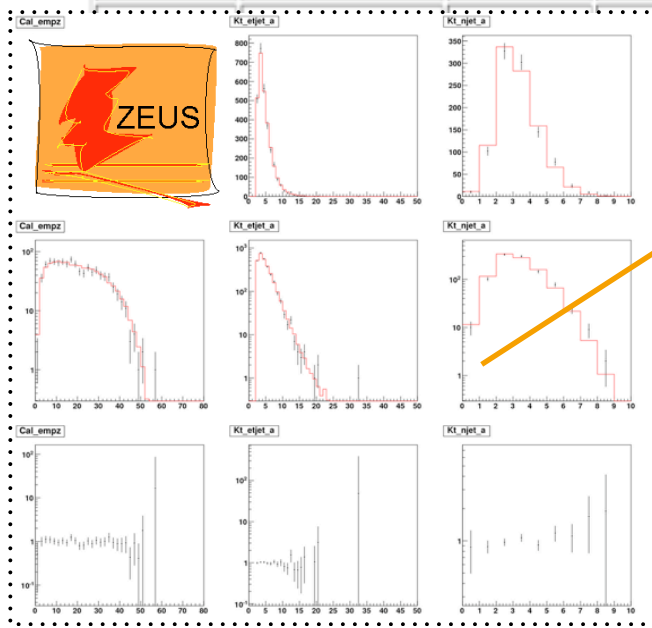
	12:23 17.01. 2012	17:38 30.01. 2012	08:06 04.02. 2012
cernlibs			
fastjet			
neurobayes			
h1unix			
h1icefp			
bos			
...			

test number	operating system	root version	staus	std output file	error file	plots	root file
58	s15.6_64	5.28.00c	OK	out	err	plots	root
						plots	root
						plots	root
						plots	root
						plots	root
						plots	root
						plots	root
						plots	root

Results of Tests

Tests run with software version:
H1_64bit_VT79_4.0.21
 created on: 04.02.2012 (08:06)

	HERA I					HERA II				
no specific year	1996	1997	1998	1999	2000	2003	2004	2005	2006	2007
dst2all										
dumpHATvariables										
jpsi_mods										
ndbint										
...										



Opening ZEUSMC.HFIX627.F15419.4B70.TEEST.Z01.root

You have chosen to open

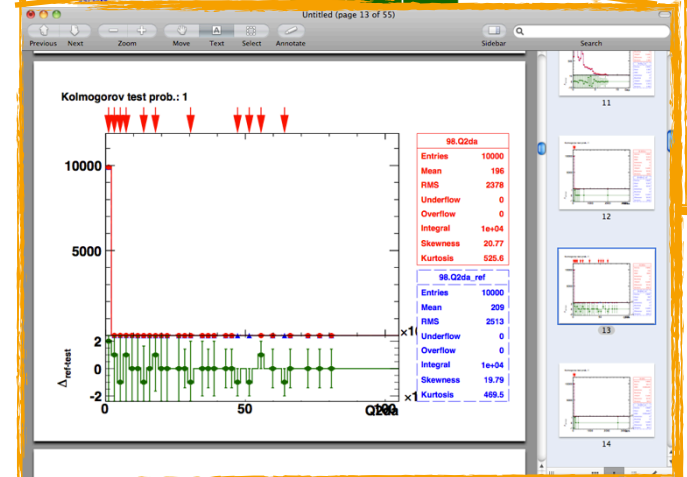
ZEUSMC.HFIX627.F15419.4B70.TEEST.Z01.root
 which is a: root File (40.7 MB)
 from: <http://www-zeus.desy.de>

What should Firefox do with this file?

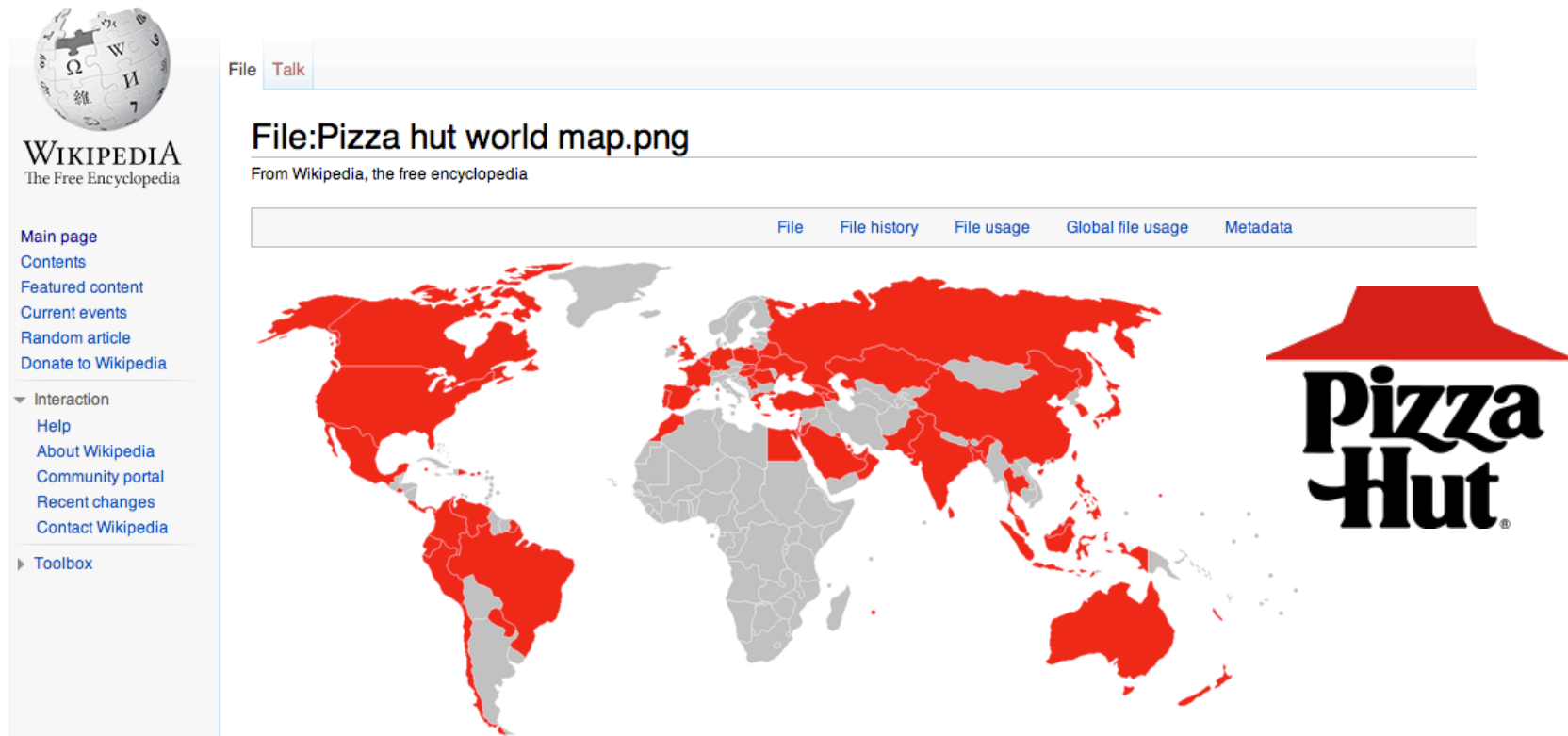
Open with

Save File

Do this automatically for files like this from now on.



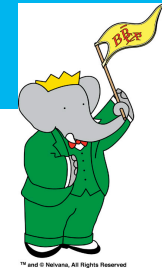
Deployment



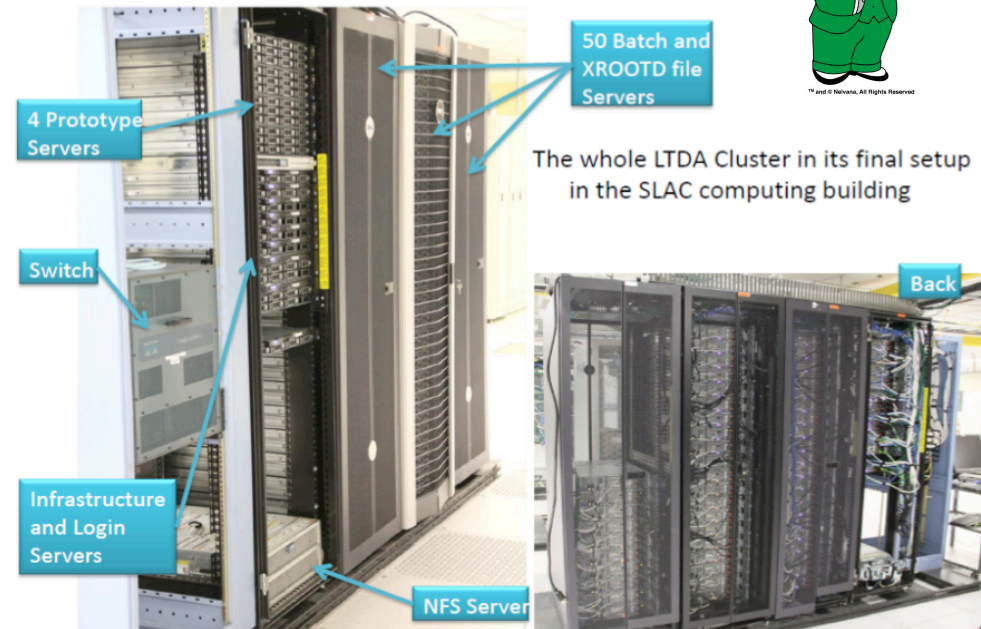
- The whole point of the `sp-system` is **not** to provide a future resource for the experiments, but rather to provide a recipe which can be deployed
 - At DESY, this means for example exploring alternative resources such as the local BIRD cluster, the National Analysis Facility (dedicated to LHC, unlikely) or the Grid



The BaBar Long Term Data Access archival system



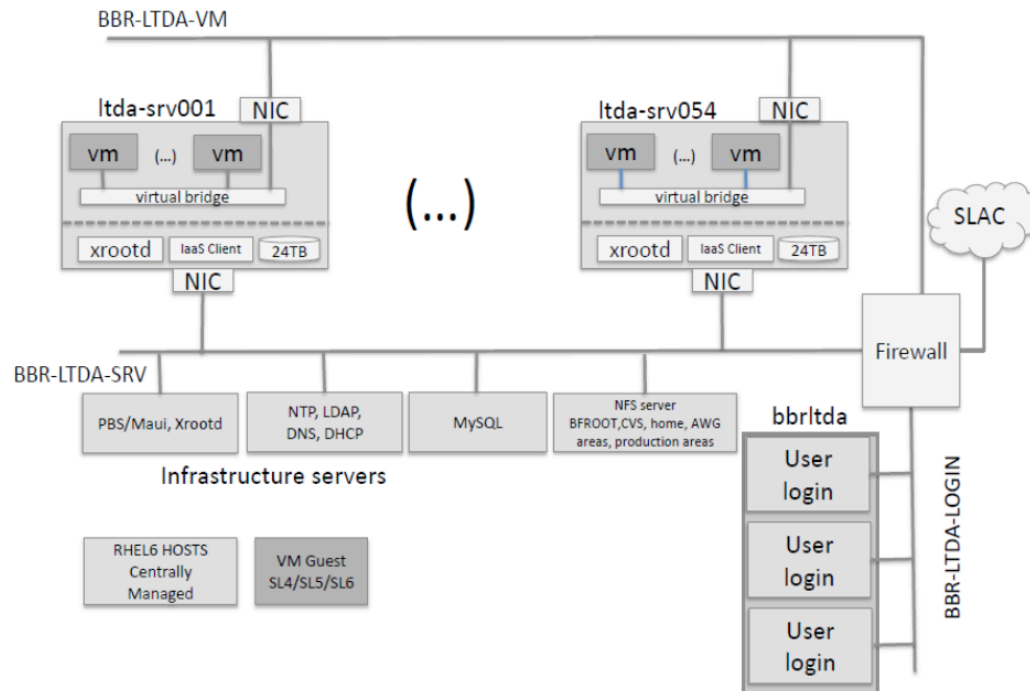
- > New BaBar system installed for analysis until at least 2018
- > Isolated from SLAC, and uses virtualisation techniques to preserve an existing, stable and validated platform
- > Complete data storage and user environment in one system



- > Required large scale investment: 54 R510 machines, primarily for data storage, as well as 18 other dedicated servers
 - Resources taken into account in experiment's funding model during analysis phase!
- > From the user's perspective, very similar to existing BaBar infrastructure



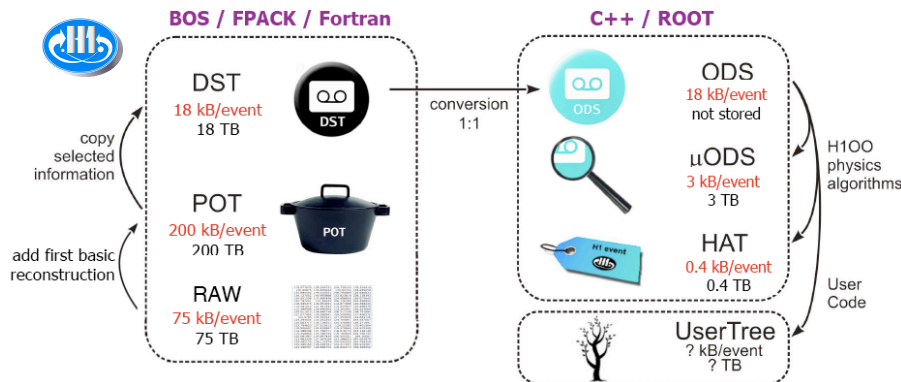
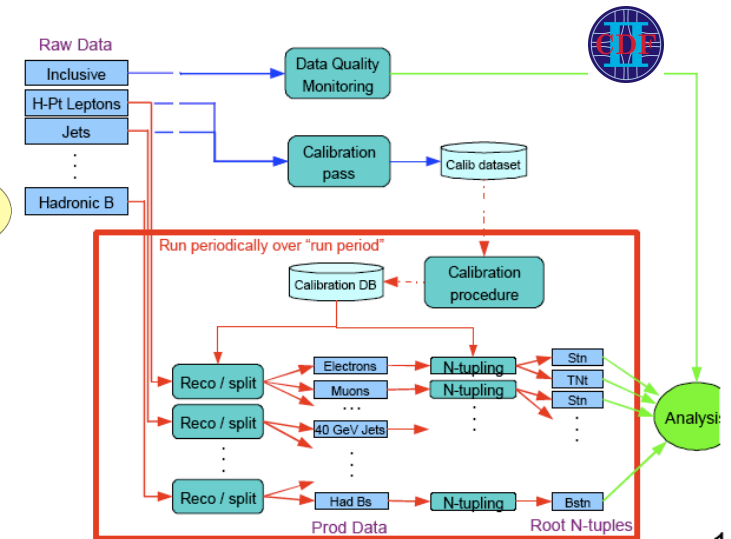
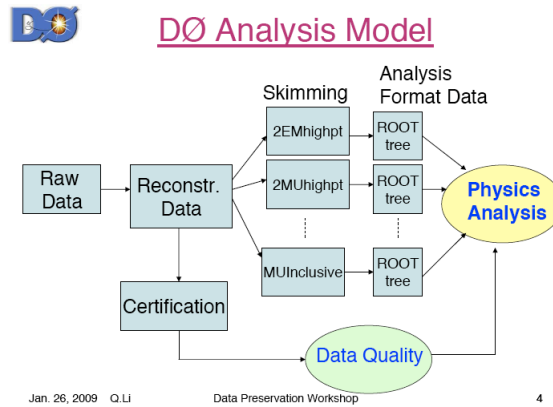
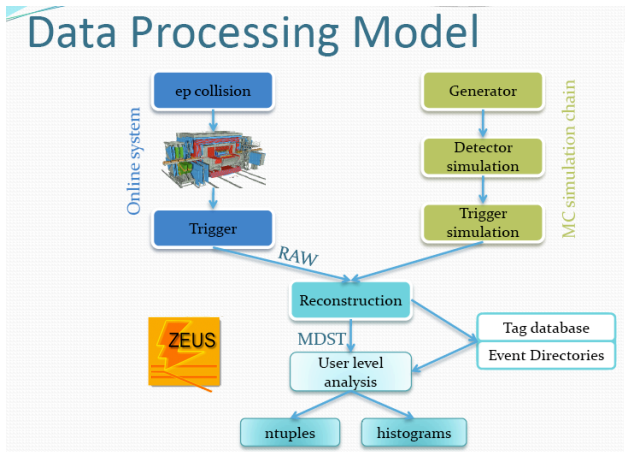
The BaBar Long Term Data Access archival system



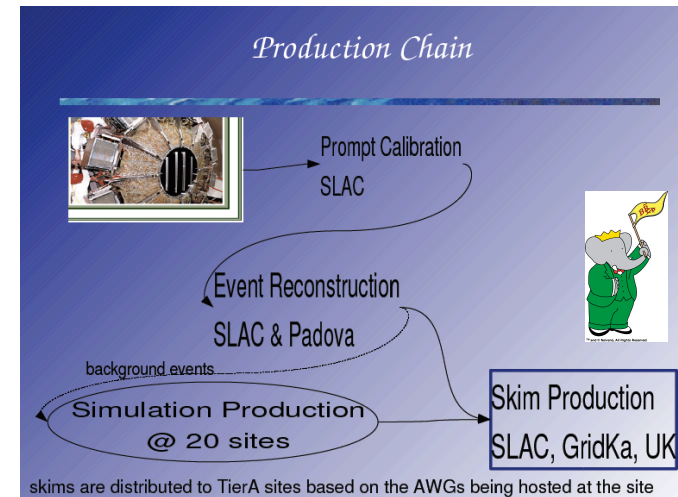
- > Crucial part of design is to allow frozen, older platforms to run in a secure computing environment
- > *Naïve* virtualisation strategy, not enough
 - Cannot support an OS *forever*
 - Security of system under threat using old versions

- > Achieved by clear network separation via firewalls of part storing the data (more modern OS) and part running analysis (the desired older OS)
- > Other BaBar infrastructure not included in VMs is taken from common NFS
- > More than 20 analyses now using the LTDA system as well as simulation

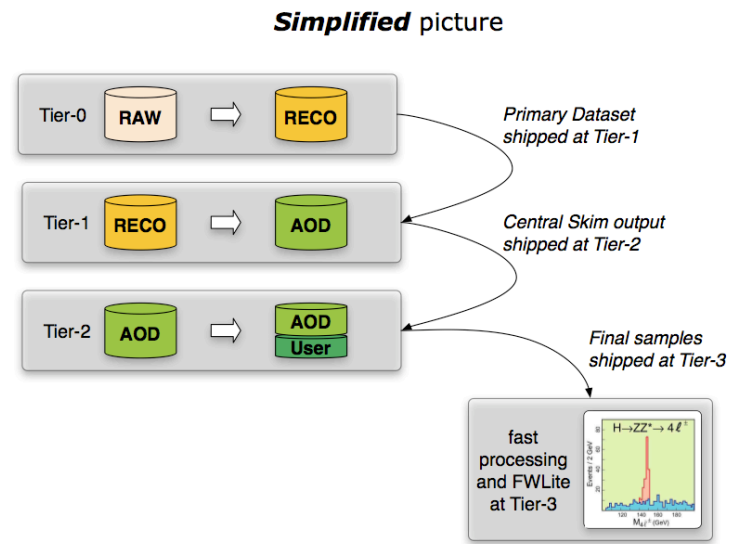
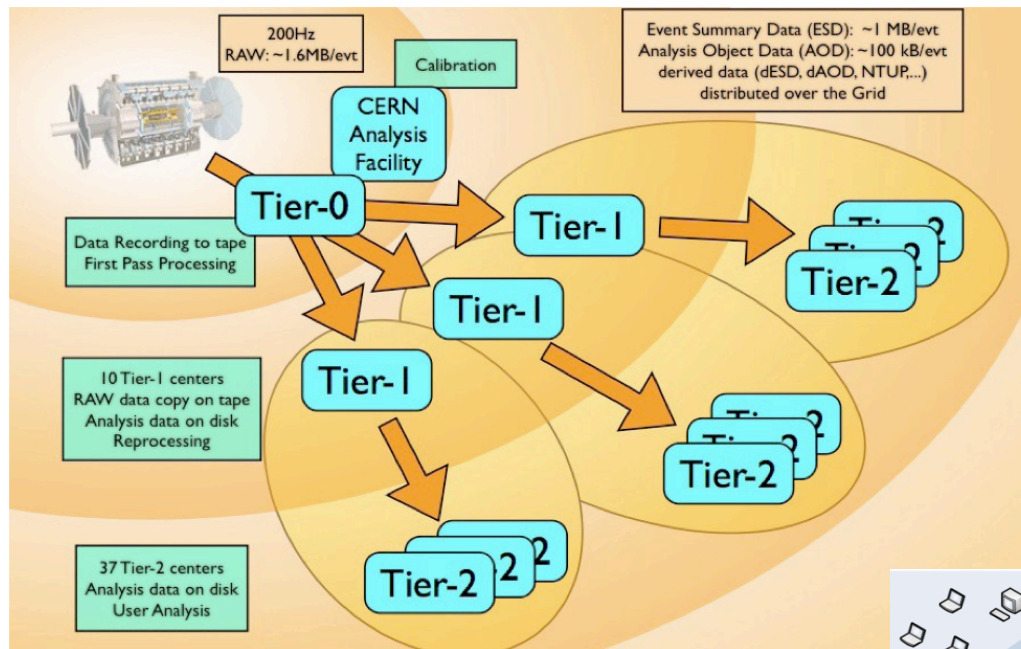
Data analysis models in HEP



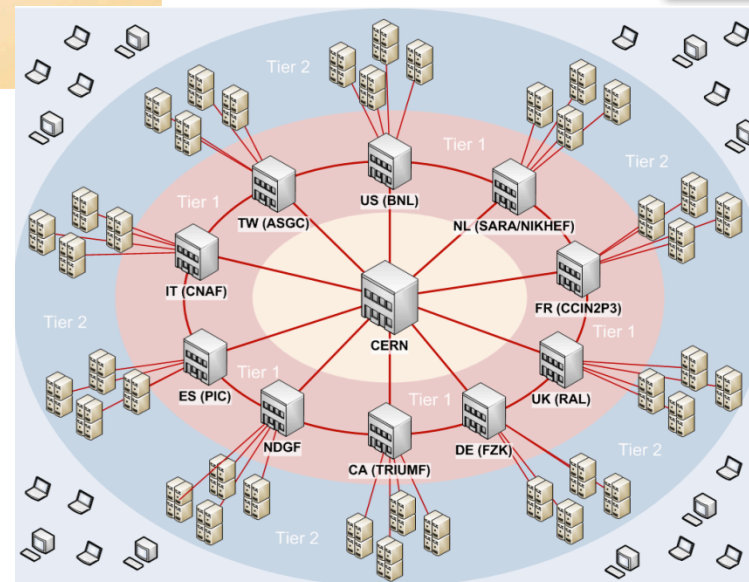
- > Complicated, at first glance different
- > Familiar descriptions of data analysis chain, from reconstruction to analysis level
 - RAW (→ POT) → DST → ntuple → analysis



Data analysis models in HEP in the LHC era



- > More skims - yes
- > More distribution - *certainly*
- > More complexity - *perhaps..*
- > Data placement is key, but analysis-wise it's still very similar to what we had before



Summary of information from the (pre-LHC) experiments

	BaBar	H1	ZEUS	HERMES	Belle	BESIII	CDF	DØ
End of data taking	07.04.08	30.06.07	30.06.07	30.06.07	30.06.10	2017	30.09.11	30.09.11
Type of data to be preserved	RAW data Sim/rec level Data skims in ROOT	RAW data Sim/rec level Analysis level ROOT data	Flat ROOT based ntuples	RAW data Sim/rec level Analysis level ROOT data	RAW data Sim/rec level	RAW data Sim/rec level ROOT data	RAW data Rec. level ROOT files (data+MC)	Raw data Rec. level ROOT files (data+MC)
Data Volume	2 PB	0.5 PB	0.2 PB	0.5 PB	4 PB	6 PB	9 PB	8.5 PB
Desired longevity of long term analysis	Unlimited	At least 10 years	At least 20 years	5-10 years	5 years	15 years	Unlimited	10 years
Current operating system	SL/RHEL3 SL/RHEL 5	SL5	SL5	SL3 SL5	SL5/RHEL5	SL5	SL5 SL6	SL5
Languages	C++ Java Python	C C++ Fortran Python	C++	C C++ Fortran Python	C C++ Fortran	C++	C C++ Python	C++
Simulation	GEANT 4	GEANT 3	GEANT 3	GEANT 3	GEANT 3	GEANT 4	GEANT 3	GEANT 3
External dependencies	ACE CERNLIB CLHEP CMLOG Flex GNU Bison MySQL Oracle ROOT TCL XRootD	CERNLIB FastJet NeuroBayes Oracle ROOT	ROOT	ADAMO CERNLIB ROOT	Boost CERNLIB NeuroBayes PostgresQL ROOT	CASTPR CERNLIB CLHEP HepMC ROOT	CERNLIB NeuroBayes Oracle ROOT	Oracle ROOT



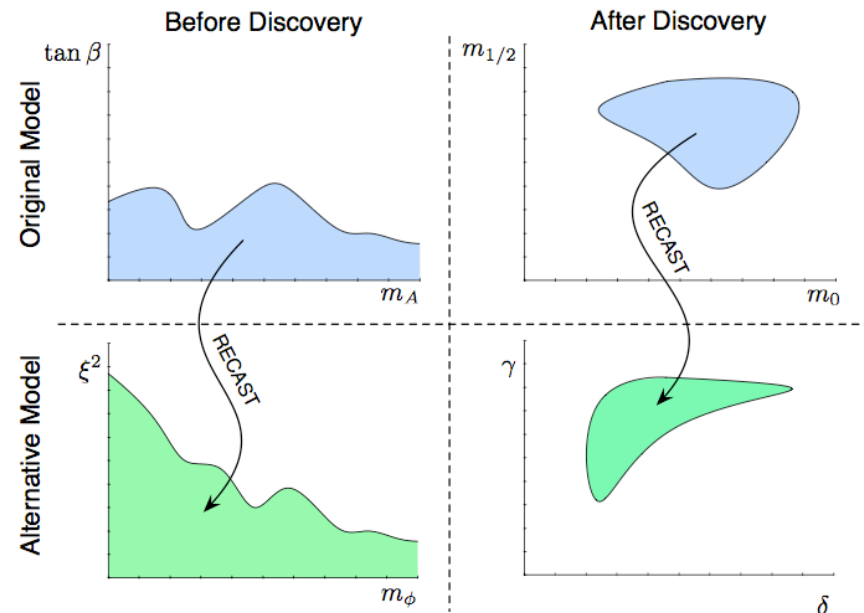
A multi-preservation level tool: RECAST

arXiv:1010.2506

- > Framework developed to extend impact of existing analyses
- > Complementary approach of analysis archival, encapsulating the full event selection, data, backgrounds, systematics

- > Idea is to **recast** existing physics search results to constrain alternate model scenarios

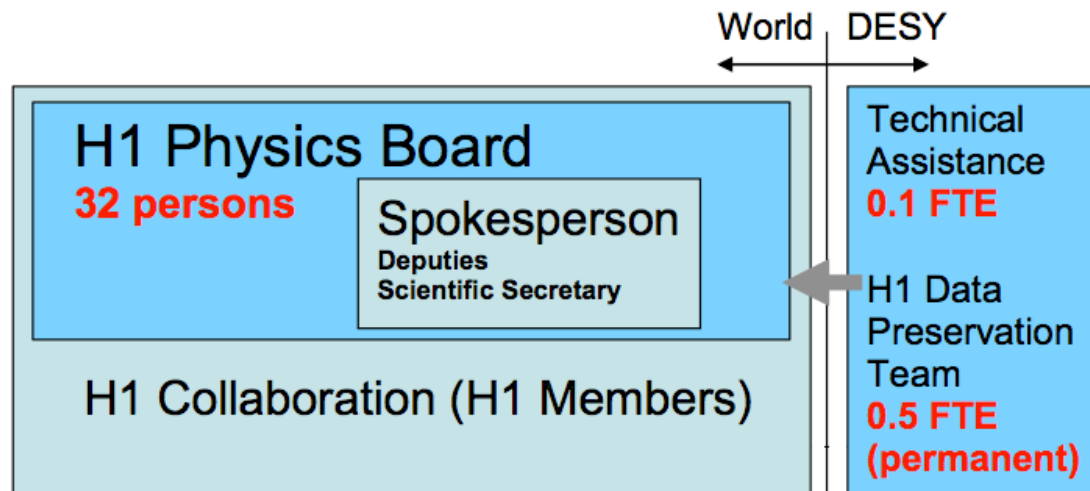
- Complete information from original analysis contained in the data
- Already performed on ALEPH data, LHC experiments investigating



- > RECAST does not fit directly into the DPHEP preservation levels
 - Levels 3 and 4 are in the back-end, containing the complete archived analyses
 - However, only the selection in the publication is preserved, it could also be described as additional information, more like level 1



Changing face of the HERA collaborations



- > H1 moved to a new collaboration management model in July 2012
 - Formation of **H1 Physics Board**, to replace Collaboration Board (institute based)
 - Future author list policies also set down in new constitution approved by collaboration
- > ZEUS (and HERMES) management teams retain same model as before, but similarly to H1 the collaborating institute layer is now removed
 - Remaining physics ZEUS working groups are now consolidated to a single physics group



DPHEP person power requirements

	Project	Goals and deliverables	Resources and timelines	Location, possible funding source, DPHEP allocation
Experiment and laboratory Priority: 1	Experimental Data Preservation Task Force	Install an experiment data preservation task force to define and implement data preservation goals.	1 FTE installed as soon as possible, and included in upgrade projects	Located within each computing team. Experiment funding agencies or host laboratories. DPHEP contact ensured, not necessarily as a displayed FTE.
	Facility or Laboratory Data Preservation Projects	Data archivist for facility, part of the R&D team or in charge with the running preservation system and designed as contact person for DPHEP.	1-2 FTE per laboratory, installed as a common resource.	Experiment common person power, support by the host labs or by the funding agencies as a part of the on-going experimental program. A fraction 0.2 FTE allocated to DPHEP for technical support and overall organisation.
Multi-experiment Priority: 3	General validation framework	Provide a common framework for HEP software validation, leading to a common repository for experiments software. Deployment on grid and contingency with LHC computing also part of the goals.	1 FTE	Installed in DESY, as present host of the corresponding initiative. Funding from common projects. Cooperation with upgrades at LHC can be envisaged. Part of DPHEP.
	Archival systems	Install secured data storage units able to maintain complex data in a functional form over long period of time without intensive usage.	0.5 FTE	Multi-lab project, cooperation with industry possible. Included in DPHEP person power.
	Virtual dedicated analysis farms	Provide a design for exporting regular analysis on farms to closed virtual farm able to ingest frozen analysis systems for a 5-10 years lifetime.	1 FTE	The host of this working group should be SLAC. Funding could come from central projects and can be considered as part of DPHEP.
	RECAST contact	Ensure contact with projects aiming at defining interfaces between high-level data and theory.	0.5 FTE	Installed with proximity to the LHC, the main consumer of this initiative, with strong connections to the data preservation initiatives that may adopt the paradigms.
	High level objects and INSPIRE	Extend INSPIRE service to documentation and high-level data object.	0.5-1.5 FTE	Installed at one of the INSPIRE partner laboratories.
	Outreach	Install a multi-experiment project on outreach using preserved data, define common formats for outreach and connect to the existing events.	1 FTE central + 0.2 FTE per experiment	A coordinating role can be played by DPHEP in connection with a large outreach project existing at CERN, DESY or FNAL. The outreach contributions from experiments and laboratories can be partially allocated to the common HEP data outreach project and steered by DPHEP.
Global Priority: 2	DPHEP Organisation	DPHEP Project Manager	1 FTE	A position jointly funded by a combination of laboratories and agencies.

LEP Paper Tables

	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total	2004-2009
ALEPH	46	42	24	34	12	9	4	4	2	607	65
DELPHI	64	30	31	58	21	19	7	7	2	678	114
L3	51	40	23	52	16	11	5	2	0	578	86
OPAL	61	38	32	55	9	11	4	3	2	675	84
All	222	150	110	199	58	50	20	16	6	2538	349

Table 1: Statistics of peer-reviewed publications of the LEP collaborations.

Papers 2004-2009	ALEPH	DELPHI	L3	OPAL	All
Electroweak	17	26	22	24	89
QCD	19	25	19	22	85
Higgs Searches	6	14	8	9	37
SUSY Searches	4	7	5	9	25
Exotica Searches	5	12	10	7	34
Flavour Physics	6	15	4	5	30
Exclusive Channels	3	8	8	2	21
Cosmo-LEP	3	3	6	0	12
Other	2	4	4	6	16
Total	65	114	86	84	349

Table 2: Distribution of physics topics in LEP publications in the years 2004-2009.

