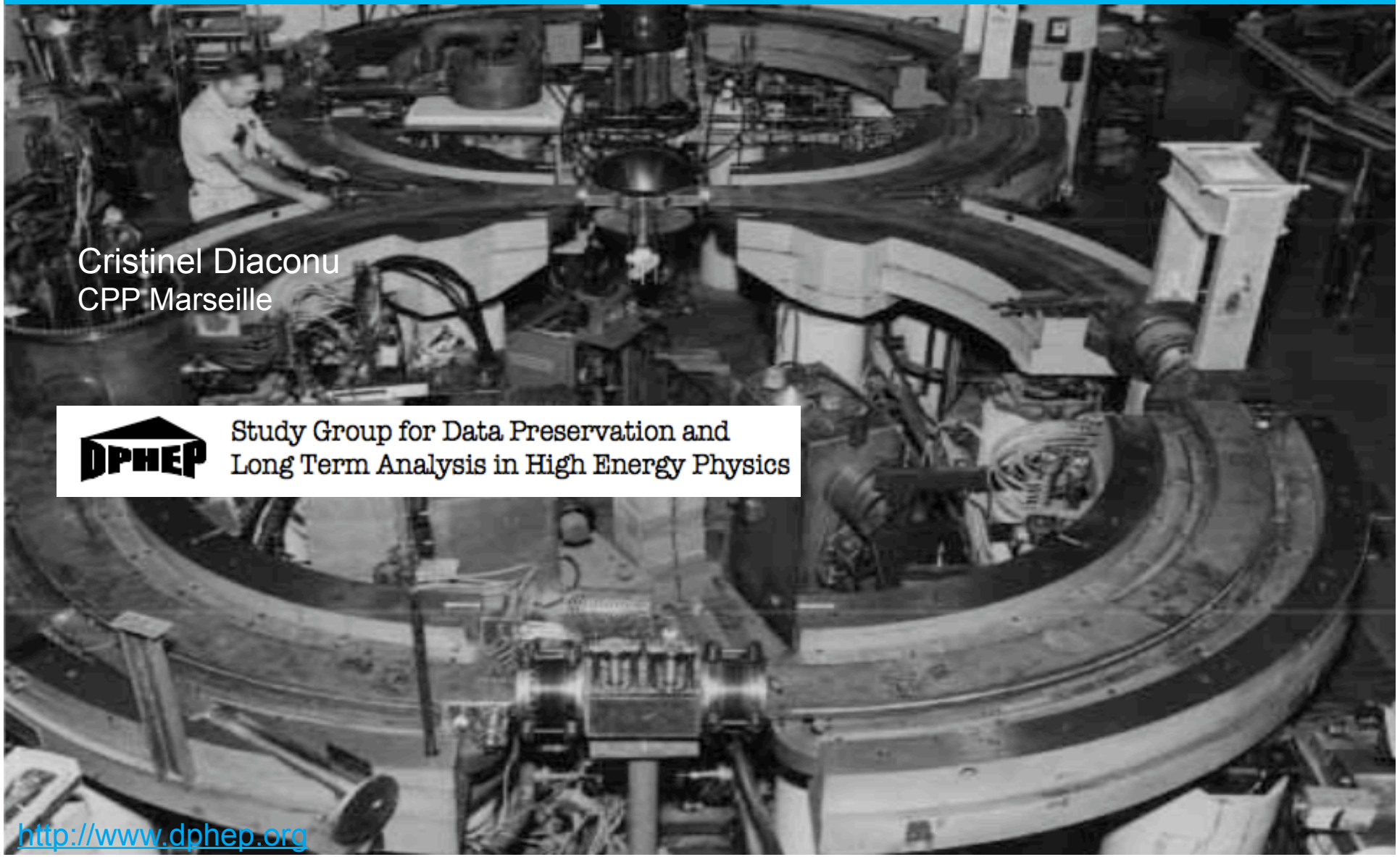
Scientific Data Preservation

The Case for High Energy Physics.



Cristinel Diaconu
CPP Marseille

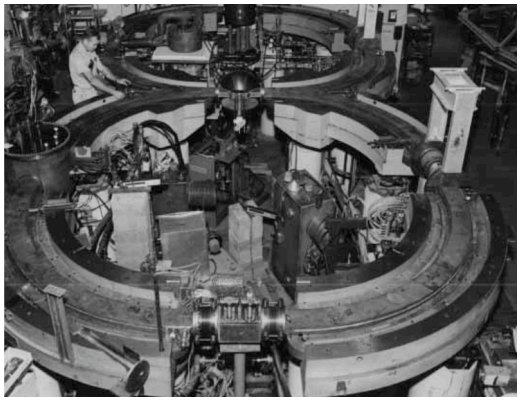
DPHEP Study Group for Data Preservation and Long Term Analysis in High Energy Physics

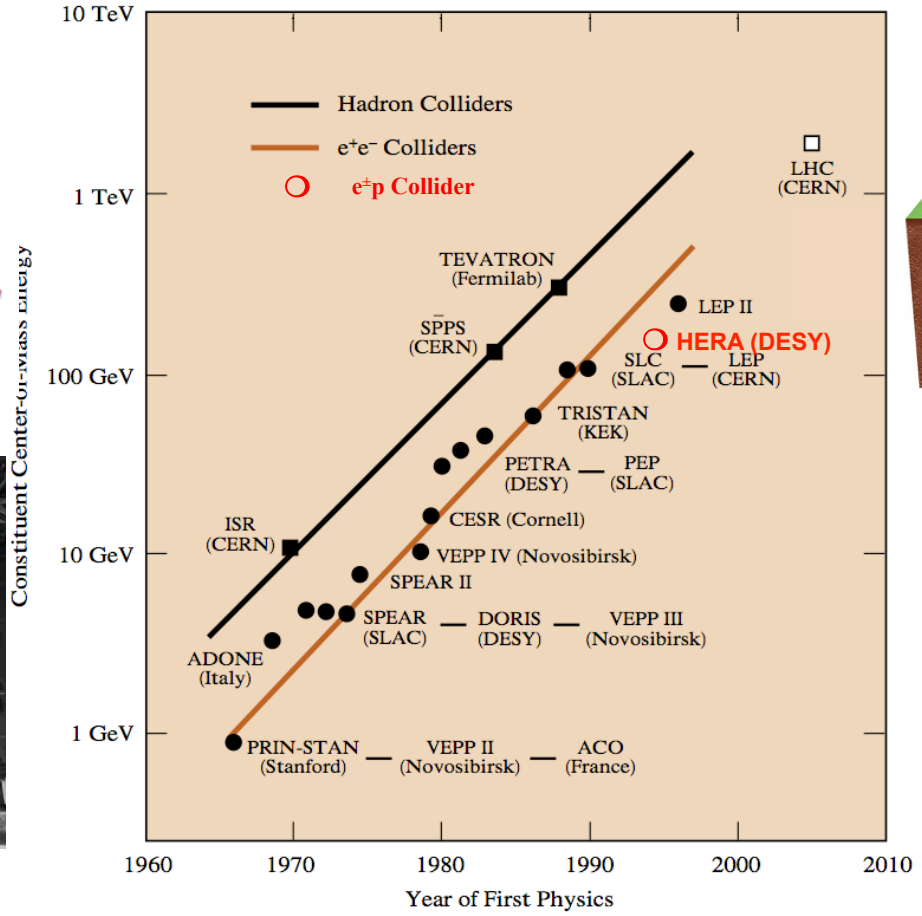
http://www.dphep.org

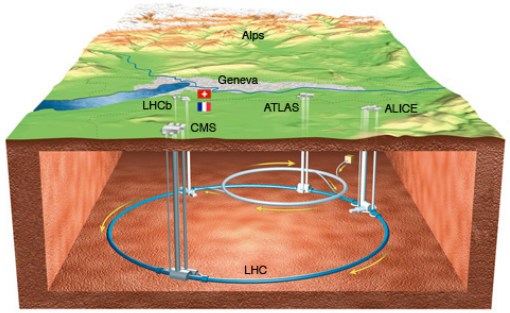
The Last 50 Years of High Energy Physics

PRIN-STAN, built late 1950's

The first colliding-beam machine, a double-ring electron-electron collider, built by a small group of Princeton and Stanford physicists. (Courtesy Stanford University)



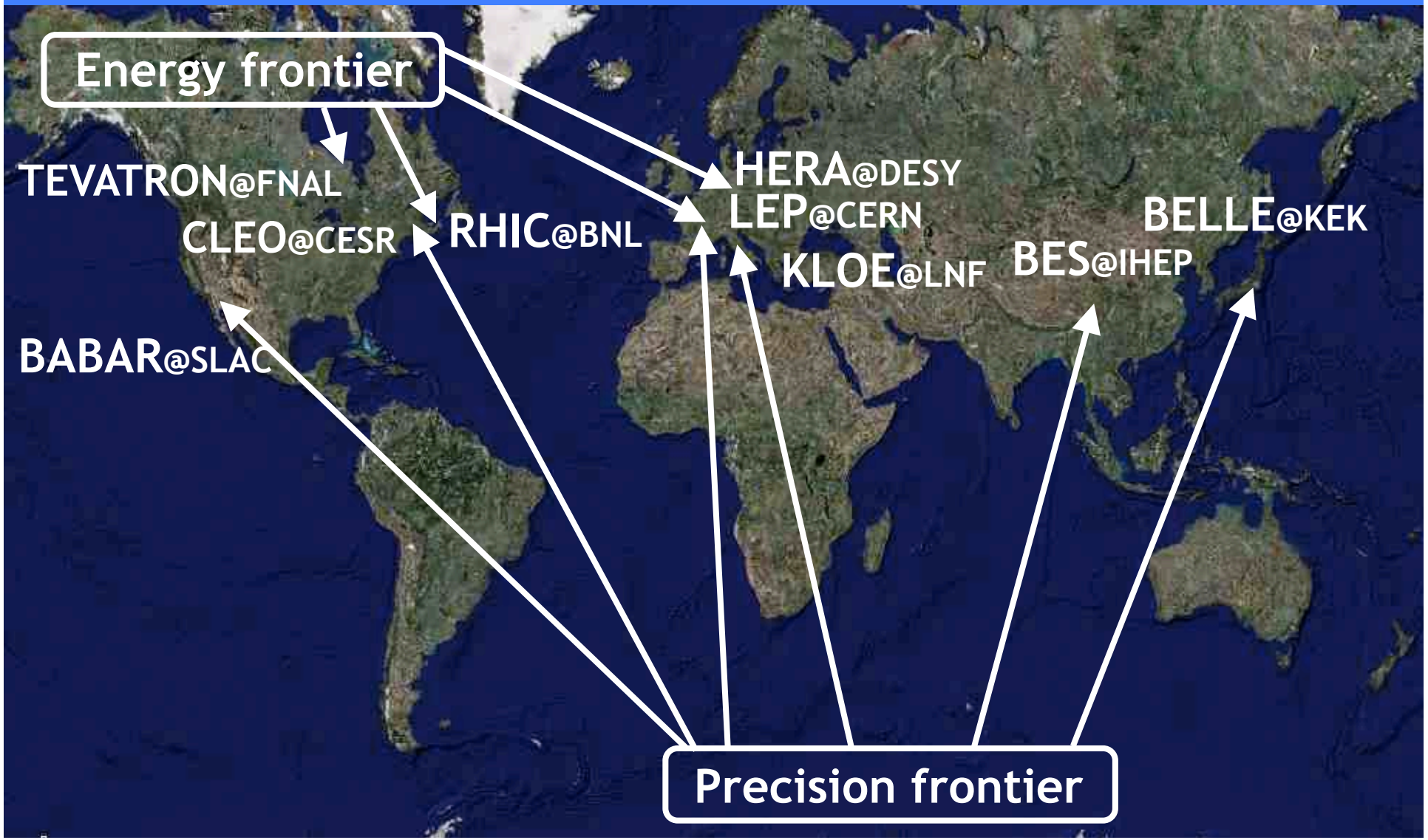




First collisions observed at the LHC in 2008; first data taking at 7 TeV now!

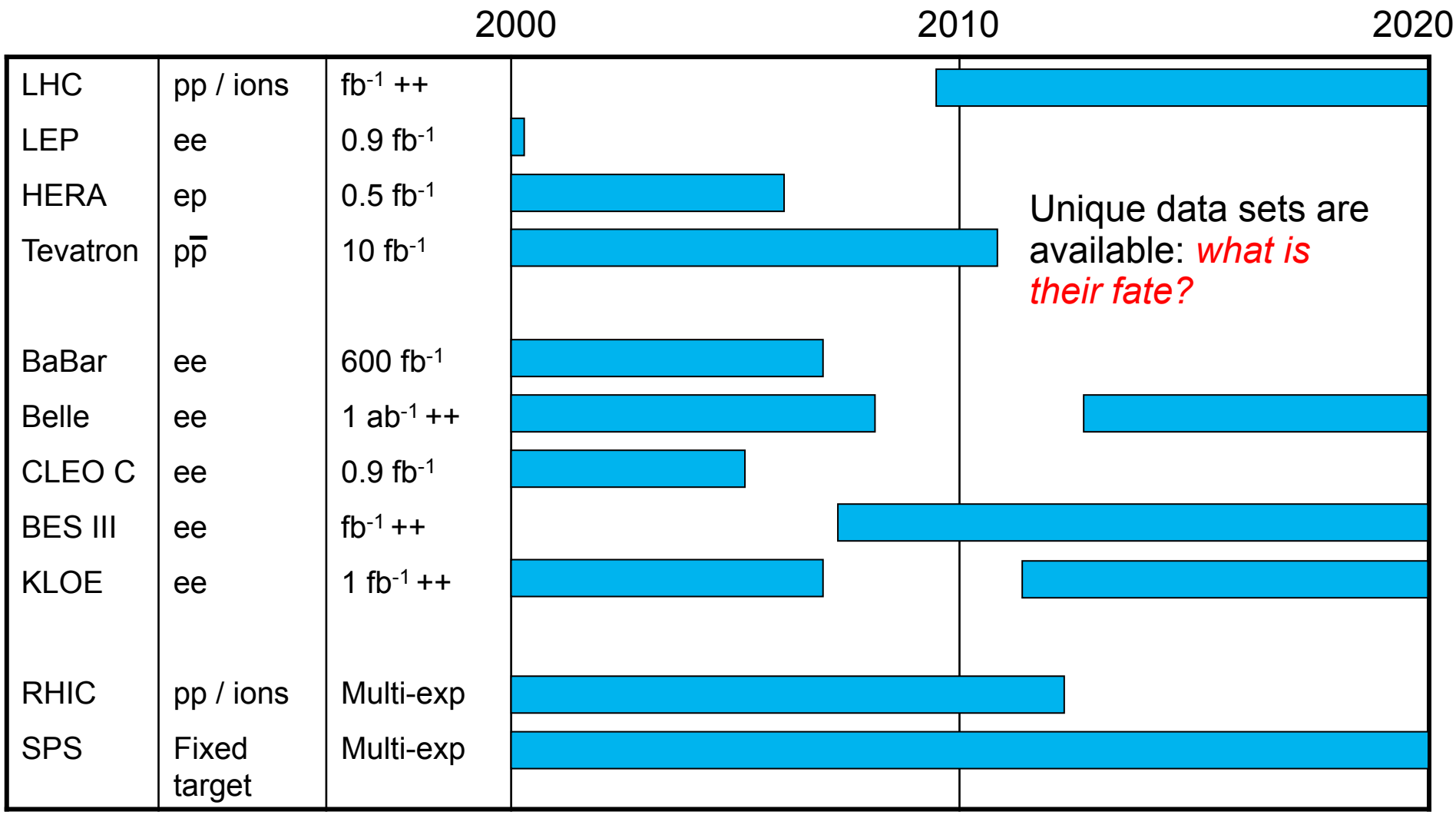
- Energy frontier probed with complex experimental installations
- New experiments normally supercede previous/similar ones - but not always..
- What is the present situation?

Active Experiments in the Pre-LHC Landscape



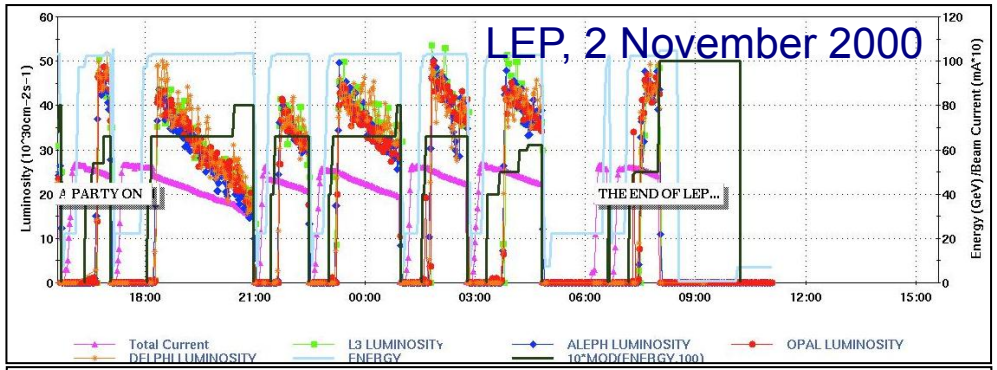
Important and unique experiments end after 10-20 years of data taking

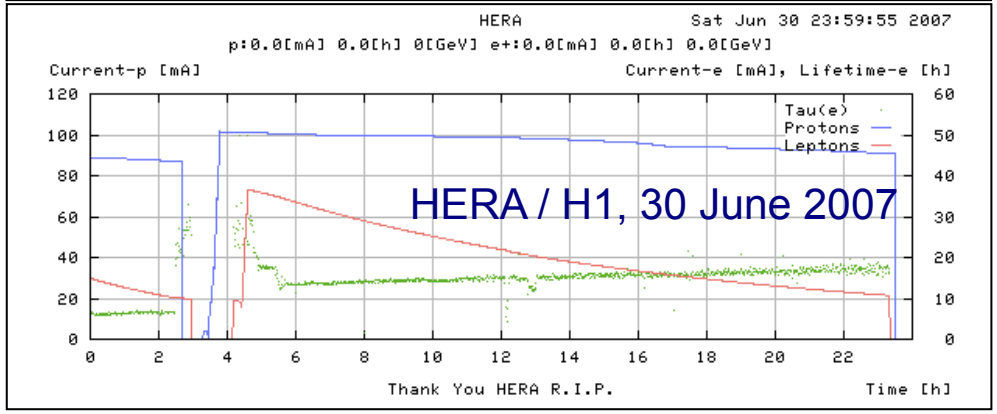
HEP Experimental Programmes in ± 10 Years

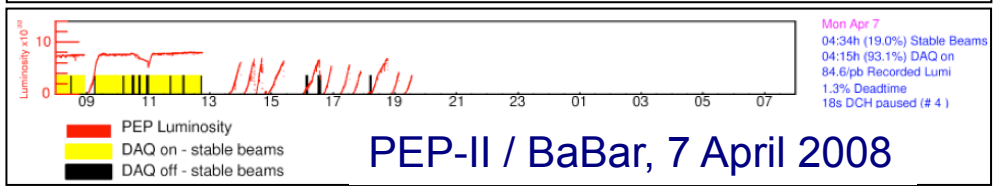


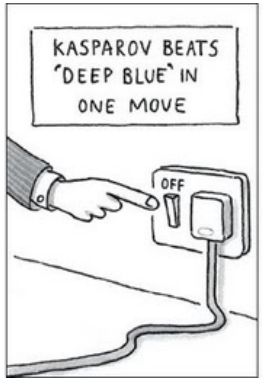
[not all programmes, dates are approximate, just to give the picture]

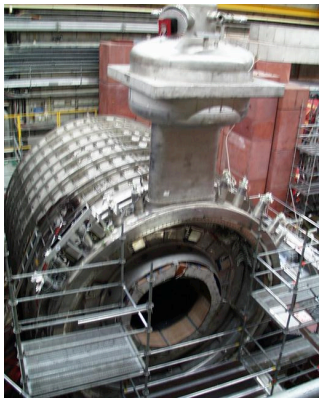
After the End of Data Taking

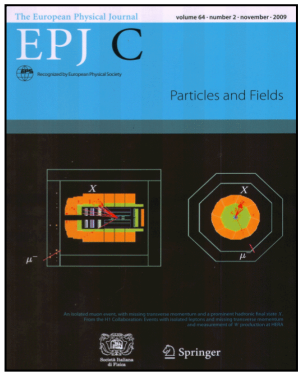










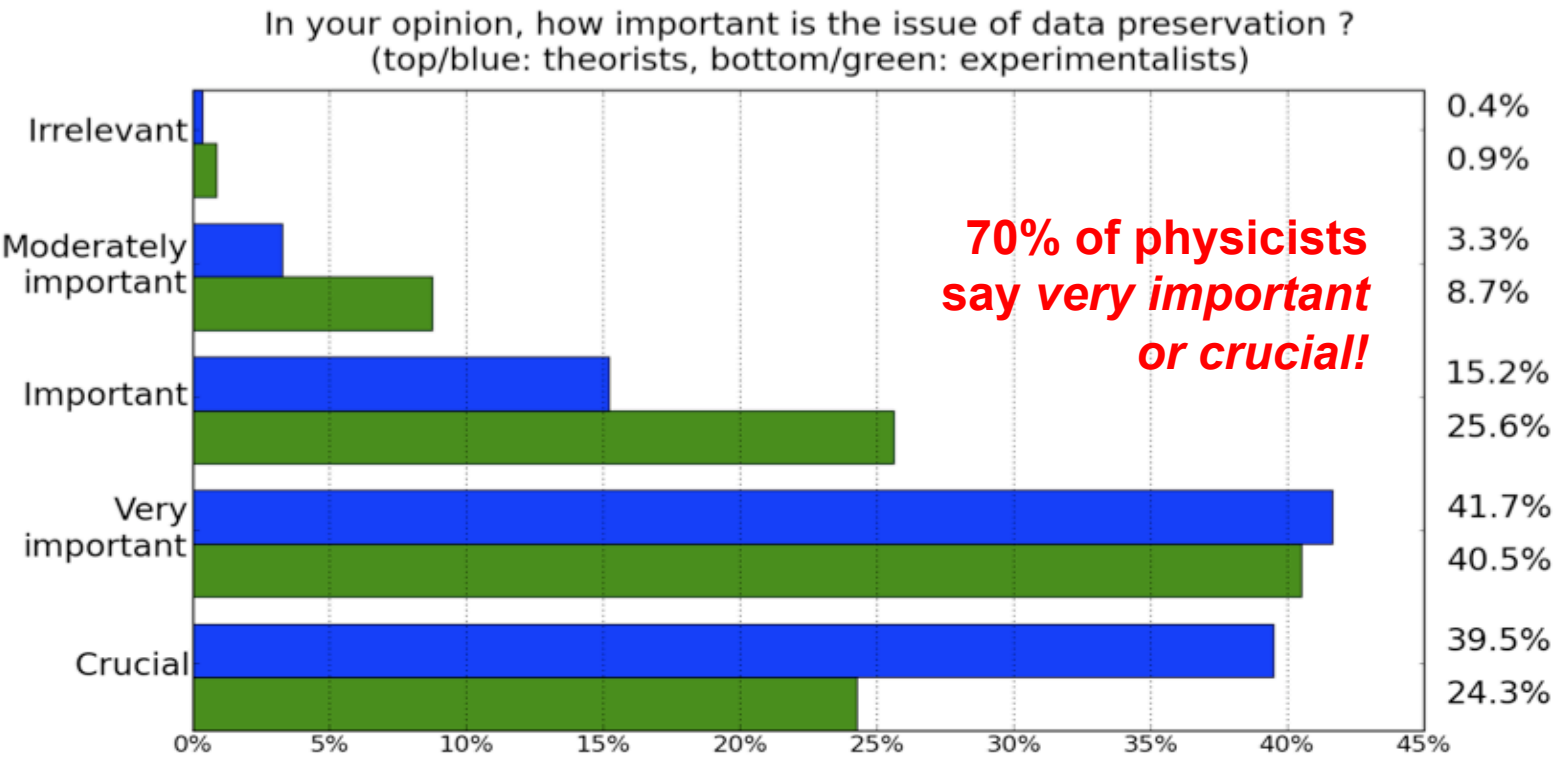


- Have an end of run party, dismantle the detector, finalize the analyses,.. *all in all about 5 years*
- *And then what do you do with the data?*

HEP Data Preservation: support in the community

PARSE.Insight is financed by the European Commission and run at CERN
arXiv:0906.0485

S. Mele

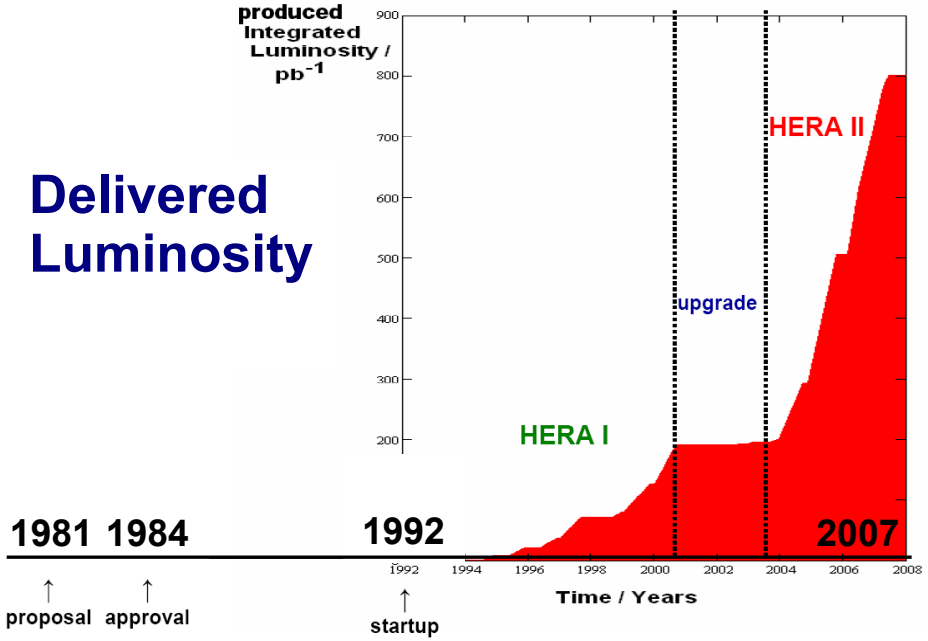


> However, no coherent strategy: in general, HEP data are lost

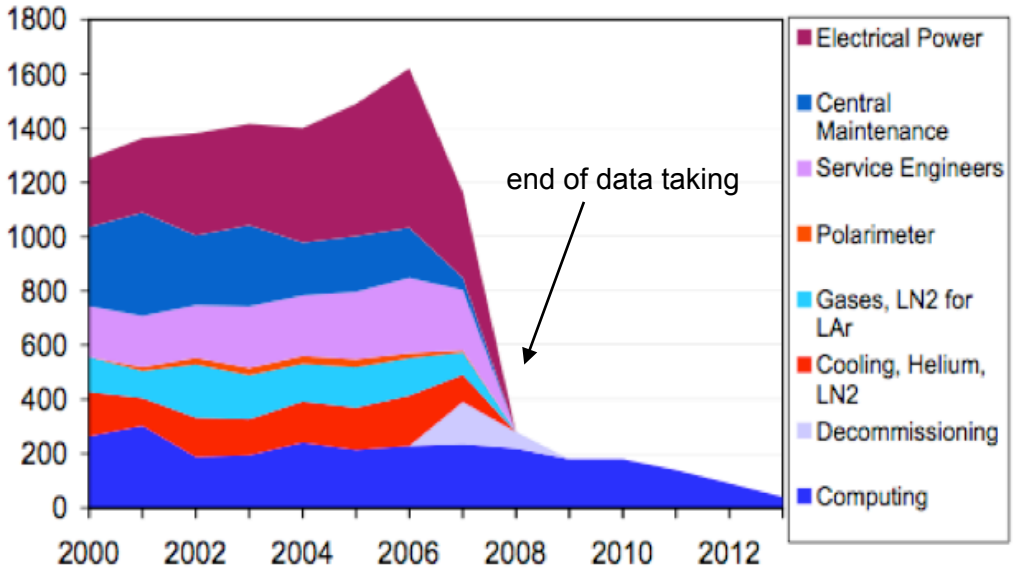
Why is it Difficult to Preserve HEP Data?

- Lots of data available to analyse at the end of collisions
- The existing resources (funding and expertise) then decrease when the data taking stops

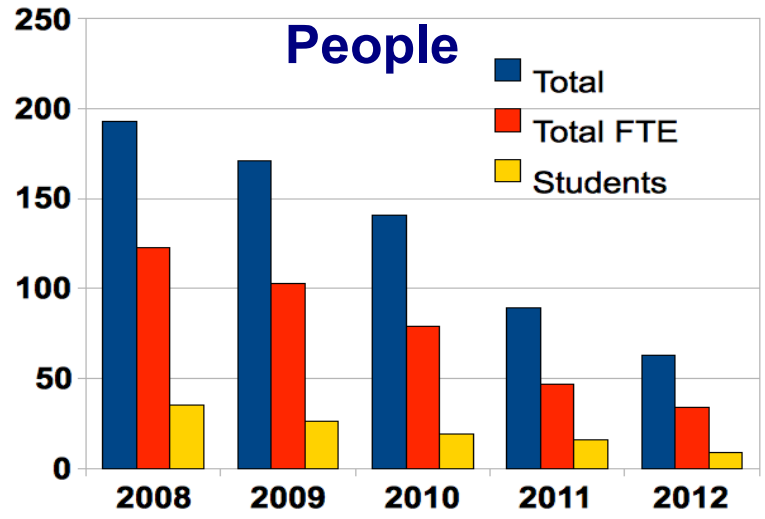
Delivered Luminosity



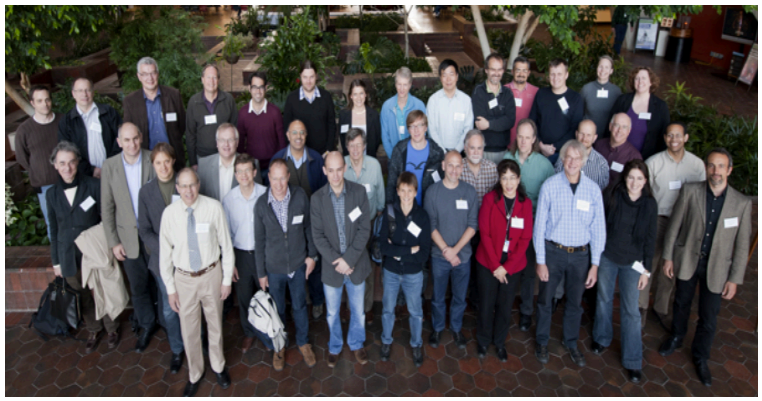
Funding



People



DPHEP: International Study Group on Data Preservation



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

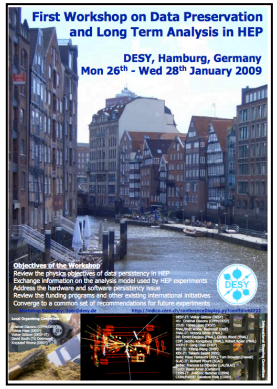
- Group has grown since 2008 to over 100 contact persons
- Endorsed by ICFA summer 2009
- **LHC** experiments joined in 2011

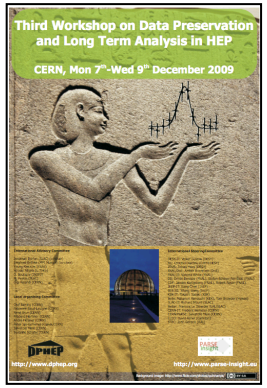
- Chair: Cristinel Diaconu (DESY/CPPM)
- Working Groups
  - Physics Cases: François Le Diberder (SLAC/LAL)
  - Preservation Models: D. South (DESY), Homer Neal (SLAC)
  - Technologies: Stephen Wolbers (FNAL), Yves Kemp (DESY)
  - Governance: Salvatore Mele (CERN)
- International Steering Committee
  - Participants from ee, ep and pp collider experiments
  - Associated computing centres at the labs
  - Some funding agencies
- International Advisory Committee
  - Chairs: Jonathan Dorfan (SLAC), Siegfried Bethke (MPIM)
  - Advisers: Gigi Rolandi (CERN), Michael Peskin (SLAC), Dominique Boutigny (IN2P3), Young-Kee Kim (FNAL), Hiroaki Aihara (IPMU/Tokyo), Alex Szalay (JHU)

DPHEP Activities

- First contacts established in September 2008, series of DPHEP workshops

  - Jan2009: DESY
  - May 2009: SLAC
  - Dec 2009: CERN
  - Jul 2010: KEK
  - May 2011: Fermilab





- Confront data models
- Clarify the concepts, set a common language
- Investigate technical aspects
- Compare and connect to other fields
  - astrophysics, life sciences, libraries …

DPHEP Visibility

CERN Courier, May 2009

DATA PRESERVATION

Study group considers how to preserve data

For experimentalists in high-energy physics, the data are like treasure, but how can they be saved for the future? A study group is investigating data-preservation options.

A simulated event in the JADE detector, generated using a refined Monte Carlo program and reconstructed using revitalized software more than 10 years after the end of the experiment. (Courtesy Siggi Bethke.)

High-energy-physics experiments collect data over long time periods, while the associated collaborations of experimentalists exploit these data to produce their physics publications. The scientific potential of an experiment is in principle defined and exhausted within the lifetime of such collaborations. However, the continuous improvement in areas of theory, experiment and simulation – as well as the advent of new ideas or unexpected discoveries – may reveal the need to re-analyse old data. Examples of such analyses already exist and they are likely to become more frequent in the future. As experimental complexity and the associated costs continue to increase, many present-day experiments, especially those based at colliders, will provide unique data sets that are unlikely to be improved upon in the short term. The close of the current decade will see the end of data-taking at several large experiments and scientists are now confronted with the question of how to preserve

the complexity of the hardware and a more dynamic part closer to the analysis level. Data analysis is in most cases done in C++ using the ROOT analysis environment and is mainly performed on local computing farms. Monte Carlo simulation also uses a farm-based approach but it is striking to see how popular the Grid is for the mass-

of data that should be PB and 10 PB for each dards but nonetheless long-term data varies preparation was fore- conservation initiatives data analysis. are widely recognized ble computing centres de Calcul de l'Institut, sique des Particules, uber of DESY and Erik Facility – showed that

21

methods climate

Science February 2011

information technologies many scientific sharing challenges visualization access example digital knowledge community project Dealing with Data

Rescue of Old Data Offers Lesson for Particle Physicists

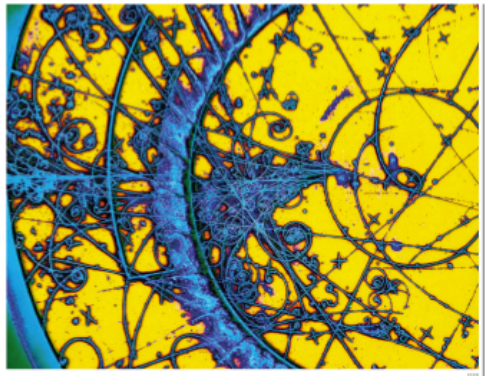
Old data tends to get forgotten as physicists move on to new and better machines.

nature news

May 2011

Symmetry, December 2009

Wissenschaft



Die Hieroglyphen von morgen

An Beschleunigern sind immense Datenmengen entstanden – die Archivierung beginnt erst jetzt

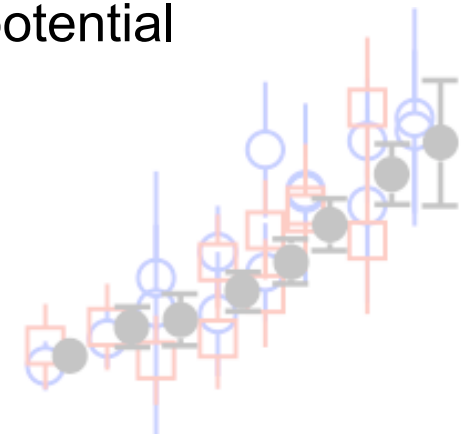
Der Teilchenzoo

Berliner Zeitung and Frankfurter Rundschau, February 2010

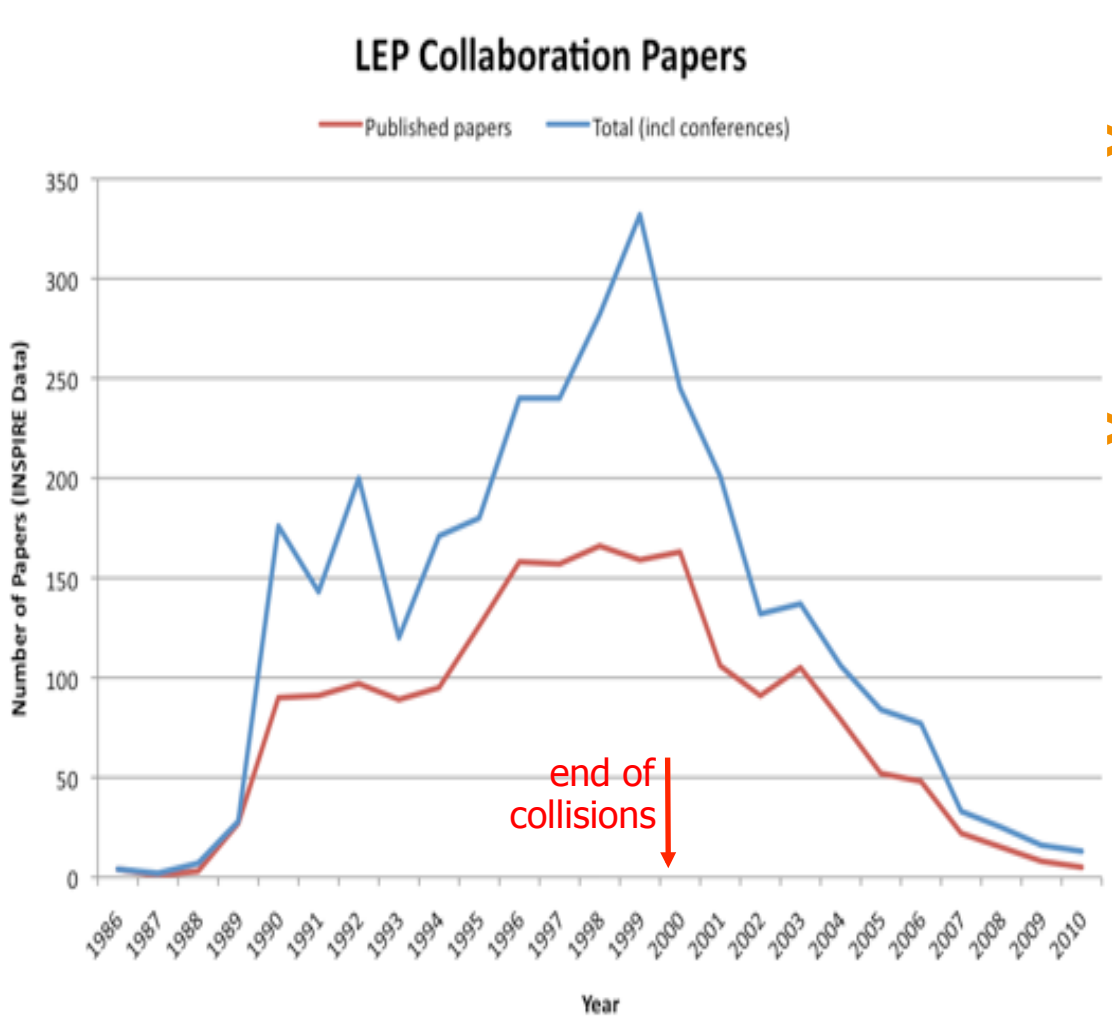
Why? Physics Case for Data Preservation

- Several physics cases can be presented for preservation
  - Long term completion and extension of existing physics program: safeguarding the data
  - Cross collaboration between experiments - usually done towards the end of the programs
  - Re-use of old data: go back and do something new
  - Use in scientific training, education, outreach

- HEP data are mostly unique and have true scientific potential



Physics case for data preservation: the tail of the physics program

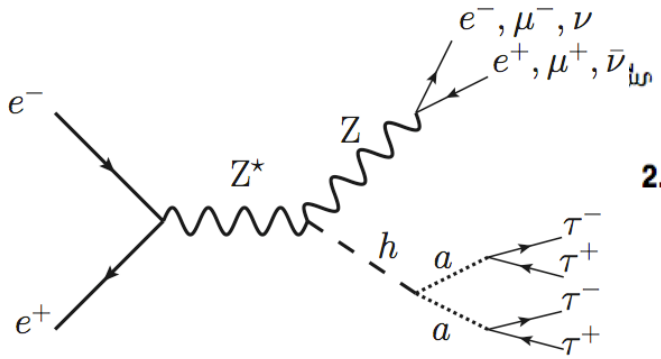


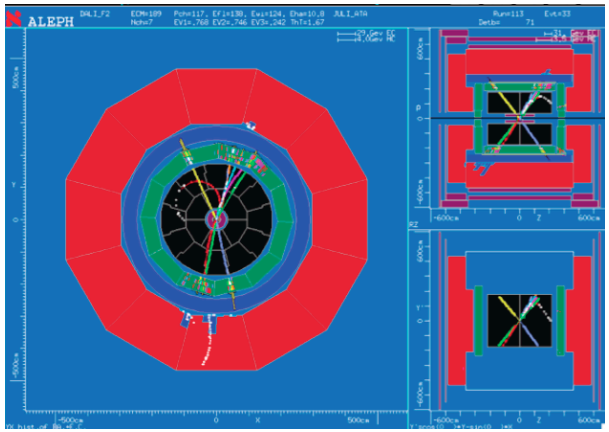
The tail of the physics program

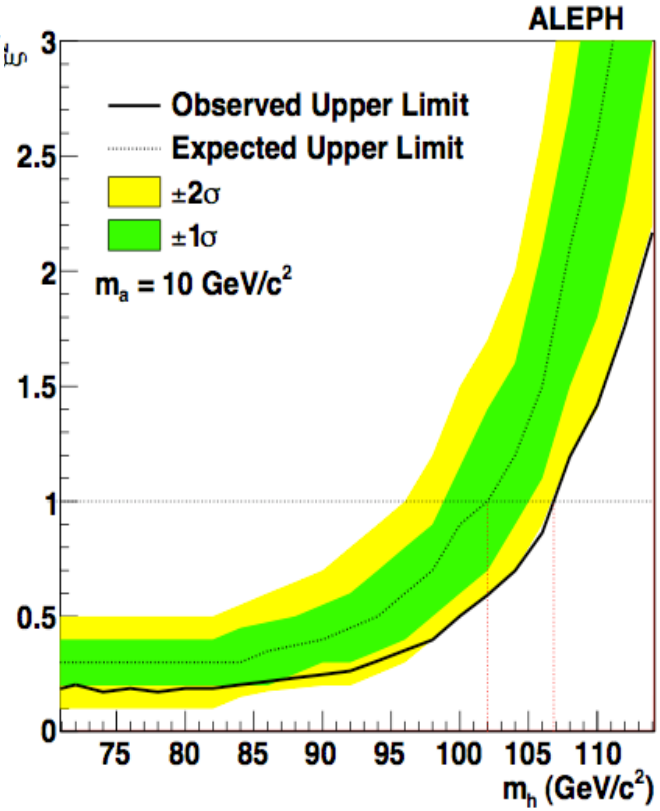
- Physics subjects are published after the end of collisions and/or collaborations
- 5-10% of the papers are finalized in the "archival mode"
  - Large number of publications well after data taking stopped
  - Large variety of topics
  - Legacy publications (full data, combined results) came later

Physics case: searches in previous data sets

- Theory and “common sense” evolve
- ALEPH: Unique physics case analysed 10 years after the end of collisions
  - and 5 years after the official end of the collaboration







EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)

Search for neutral Higgs bosons decaying into four taus at LEP2

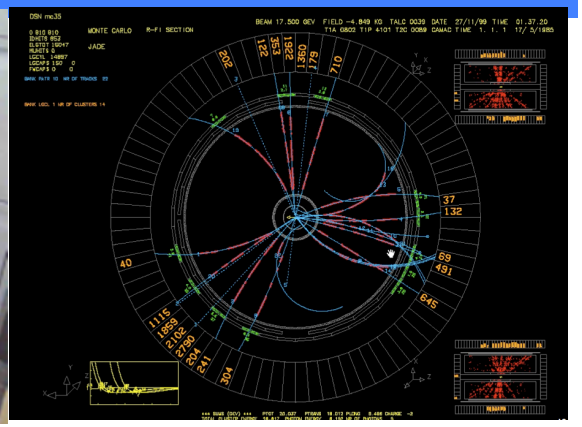
The ALEPH Collaboration*)

Abstract

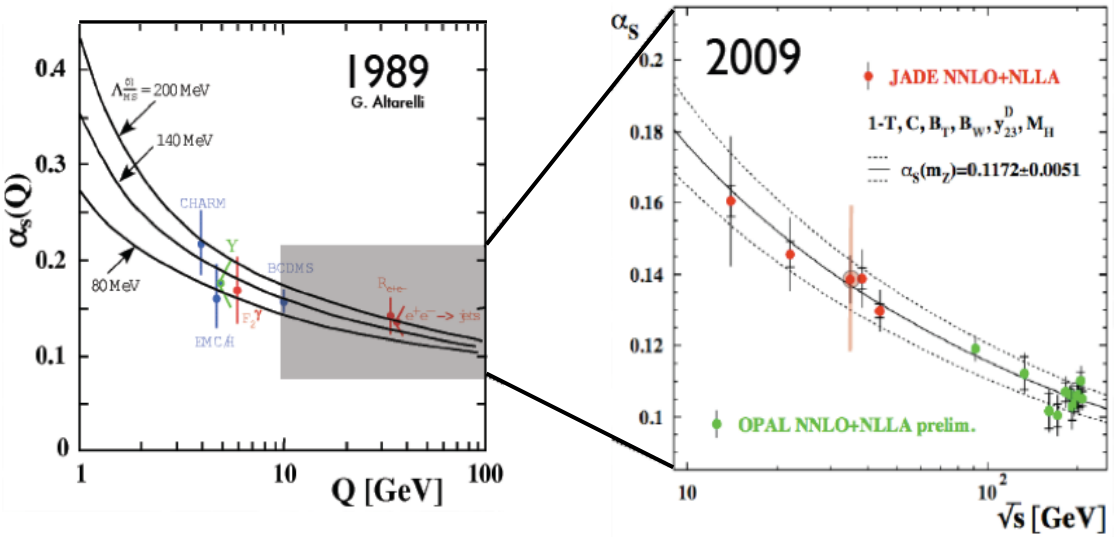
A search for the production and non-standard decay of a Higgs boson, h, into four taus through intermediate pseudoscalars, a, is conducted on 683 pb^{-1} of data collected by the ALEPH experiment at centre-of-mass energies from 183 to 209 GeV. No excess of events above background is observed, and exclusion limits are placed on the combined production cross section times branching ratio, $\xi^2 = \frac{\sigma(e^+e^- \to Zh)}{\sigma_{SM}(e^+e^- \to Zh)} \times B(h \to aa) \times B(a \to \tau^+\tau^-)^2$. For $m_h < 107$ GeV/c^2 and $4 < m_a < 10$ GeV/c^2, $\xi^2 > 1$ is excluded at the 95% confidence level.

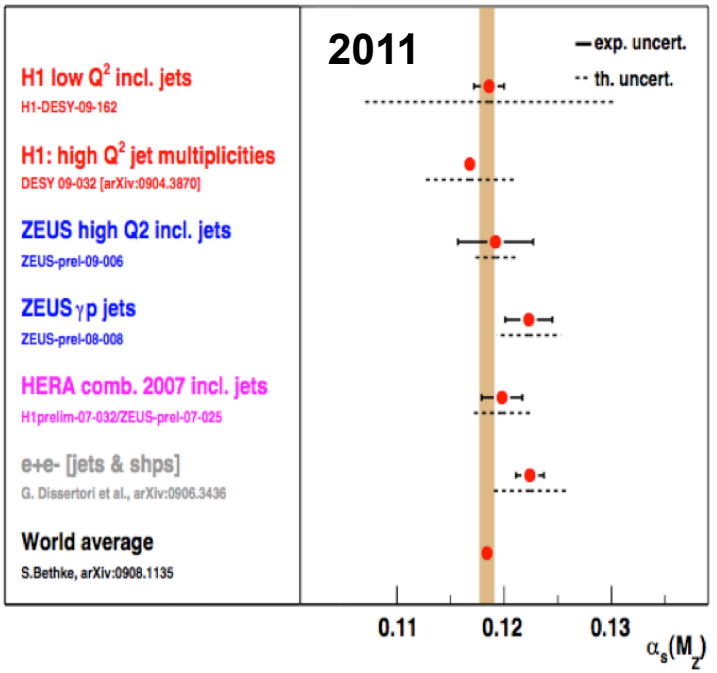
Physics case: Improvement in theory and simulation

- JADE: Required full raw data preservation, software revitalisation, individual initiatives…



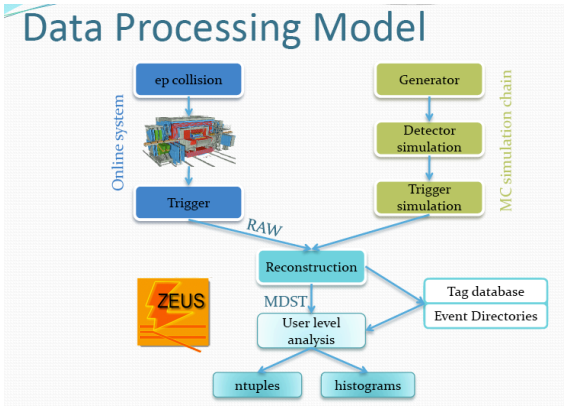
10 recent publications

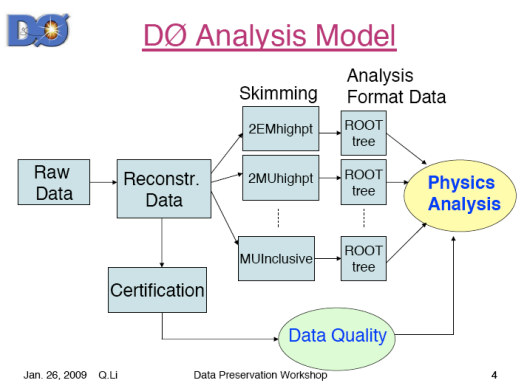


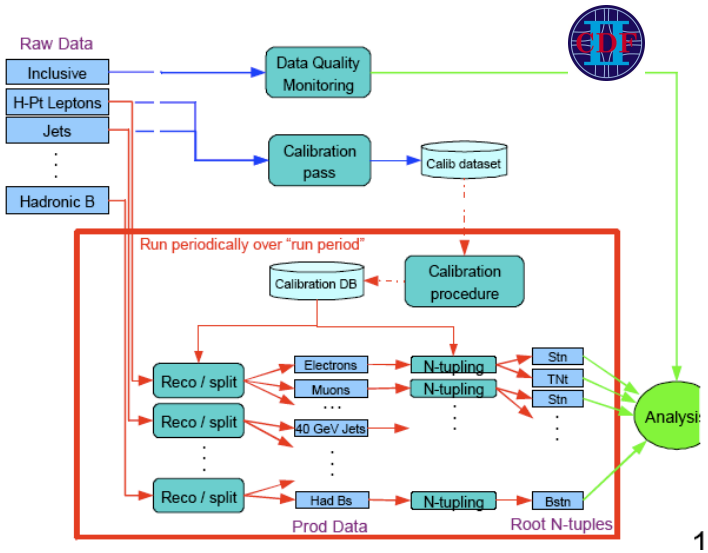


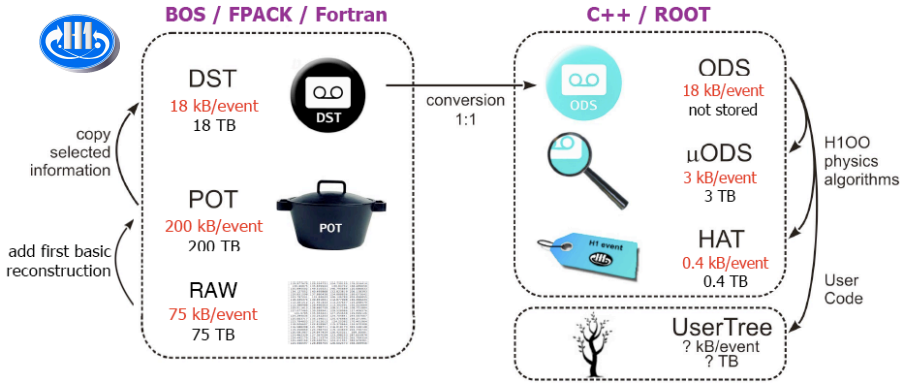
- **Around 10% of measurements are dominated by non-experimental errors: theory (N^nLO?) and simulation..**

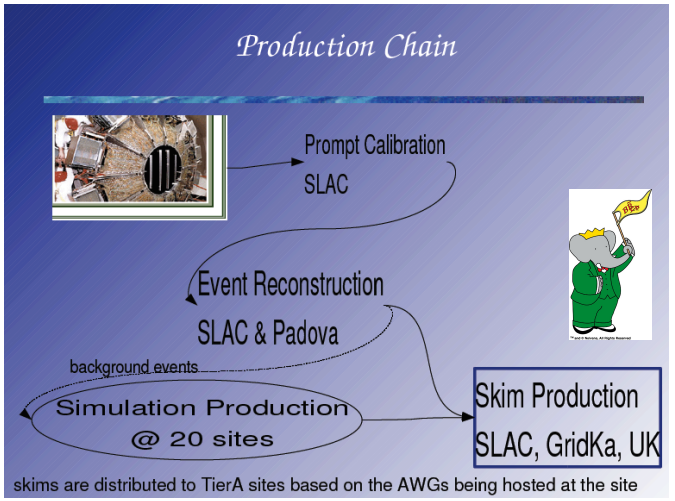
Data Analysis Models in HEP





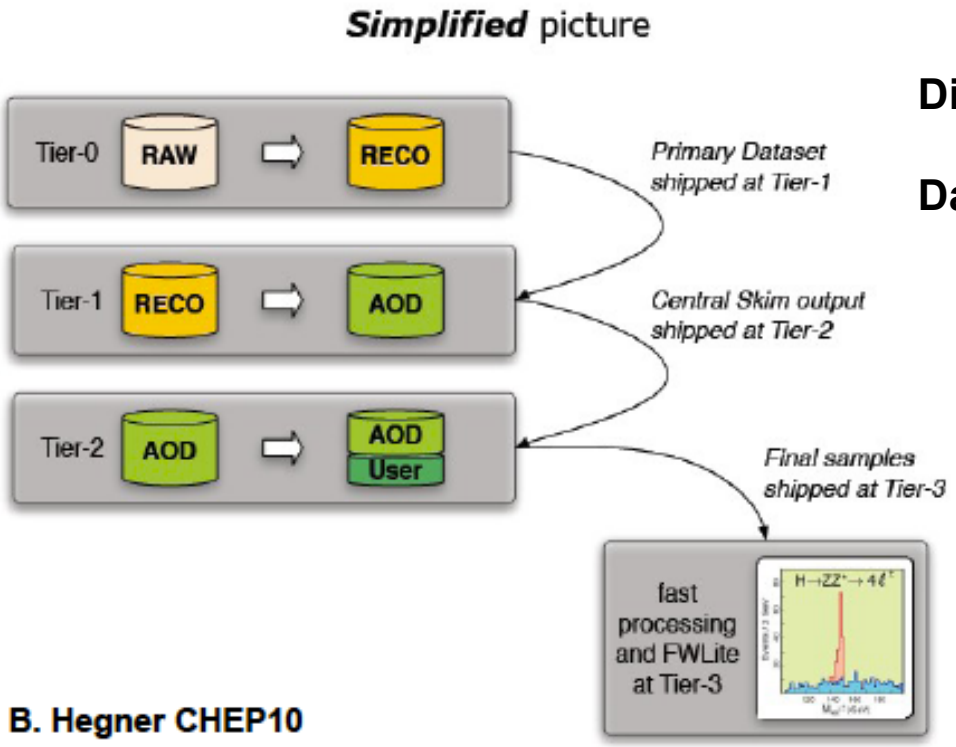






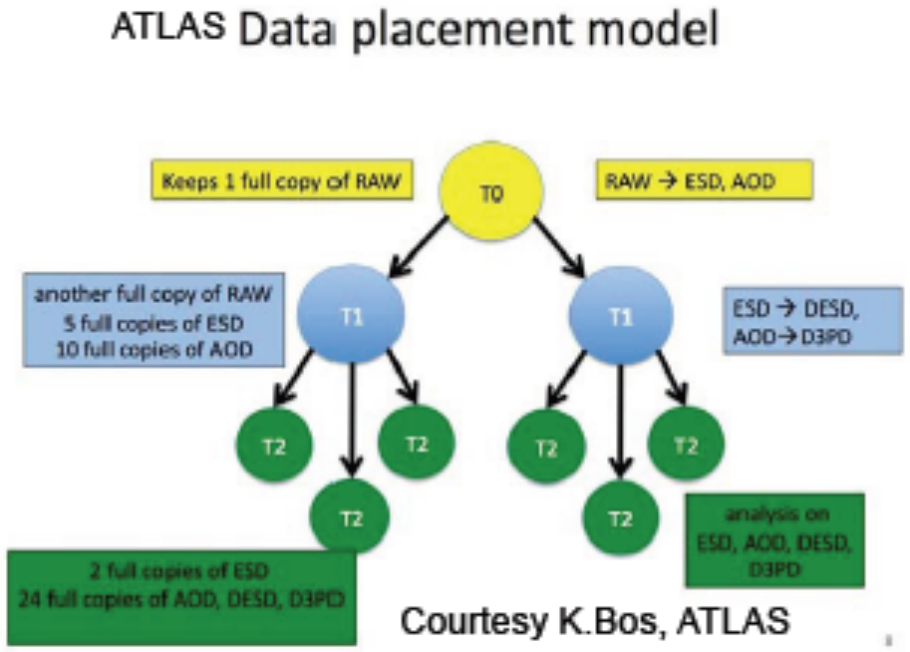
- Complicated, at first glance different
- Familiar descriptions of data analysis chain, from reconstruction to analysis level
  - RAW → POT → DST → *ntuple* → analysis

LHC: new aera, similar reduction models….



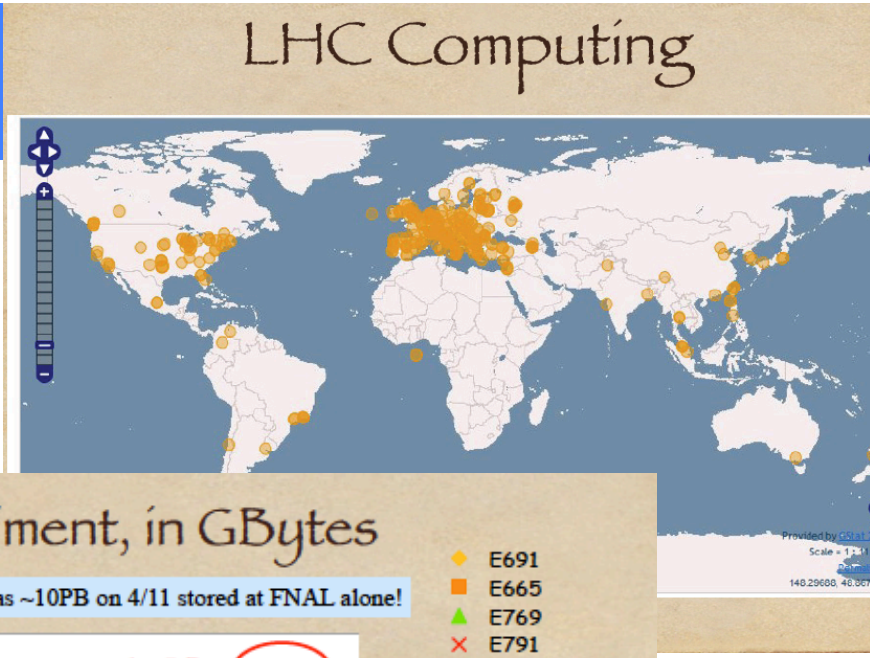
Distributed computing necessary

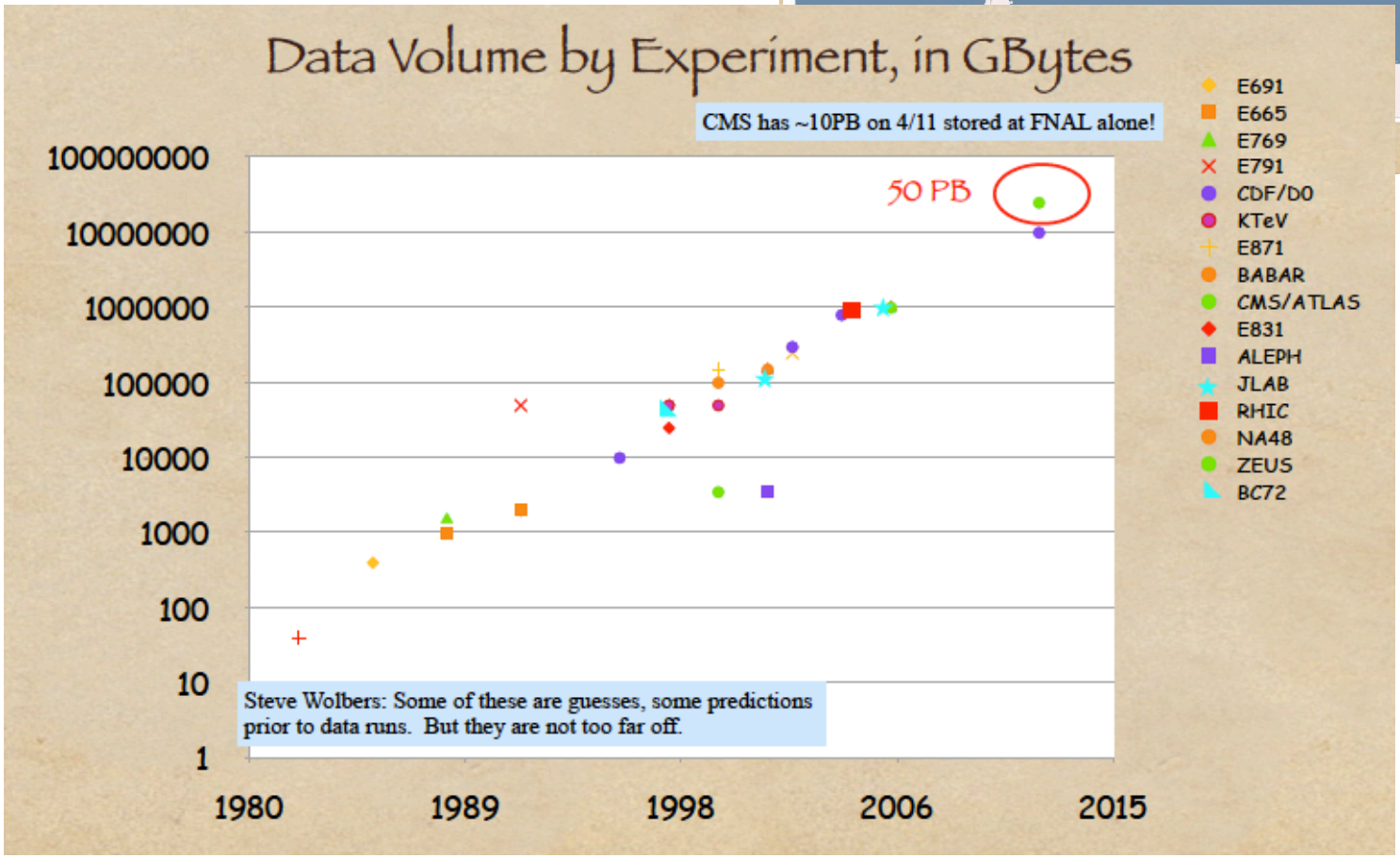
Data placement strategy is a key ingredient



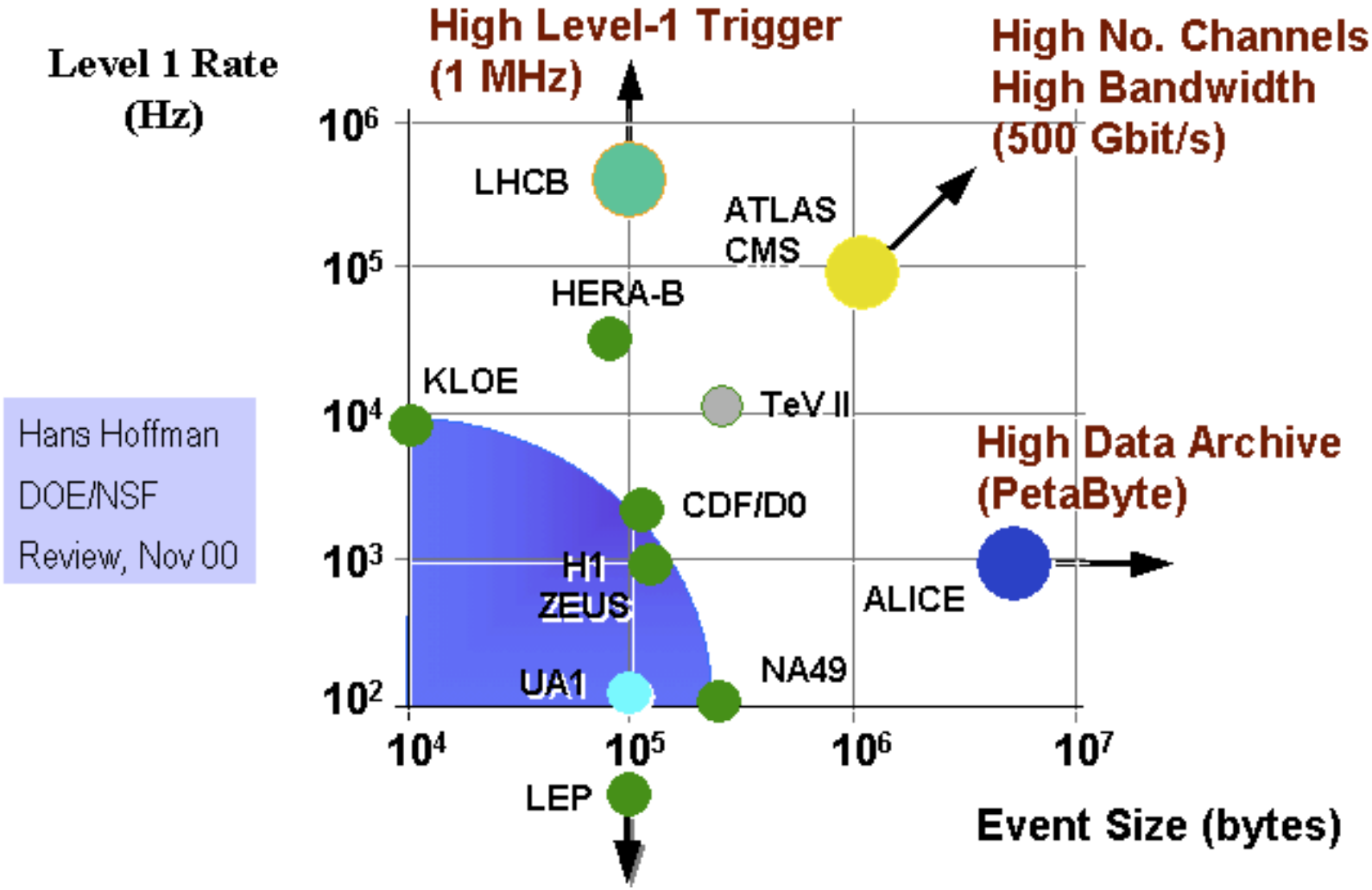
…but different scales

- The whole world does LHC computing
- Data Volume is very large (~100 Pb)





The scaling has worked so far

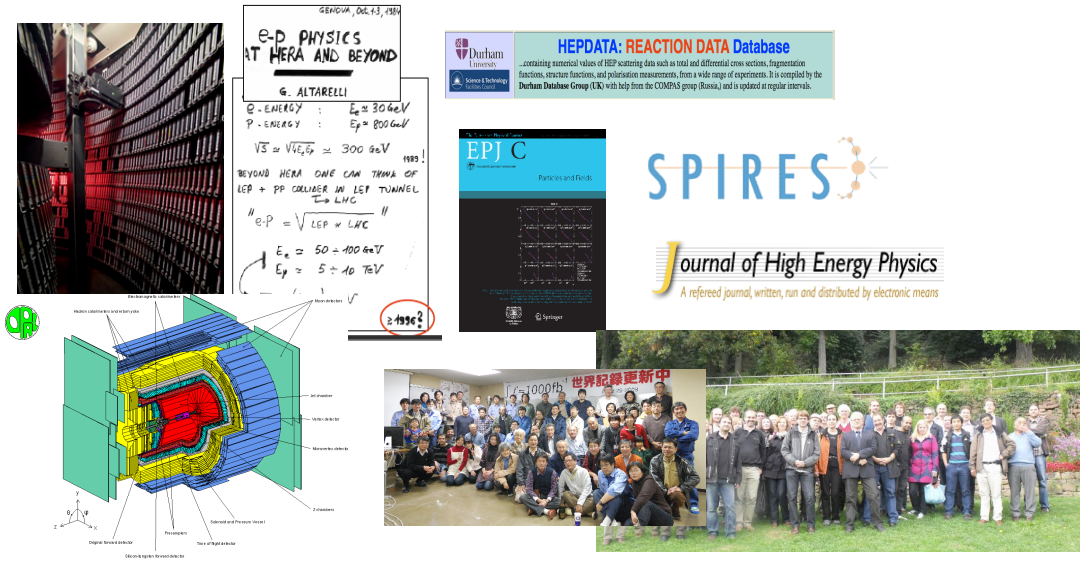


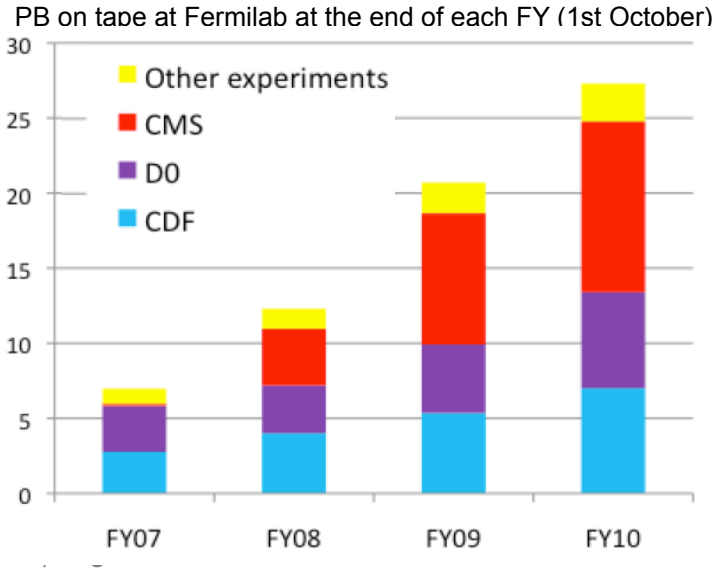
HEP Data: How much is it?

- Discussions in DPHEP lead to a number of around 0.5 to a few PB / exp. (LHC to ~100Pb)
  - Depending on preservation model
- HEP Computing centres are able to store the data, but:

  It have costs and has to be plannned

  Data preservation is not only about the digital data!
- HEP Data include much more than bits:
  - **Software**, meta-data, documentation, publications, expertise

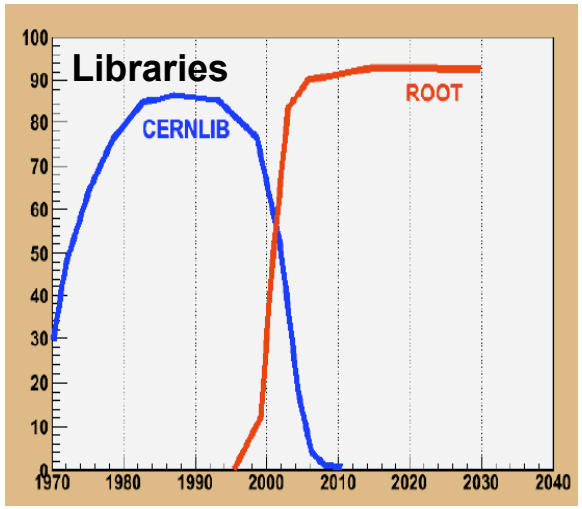


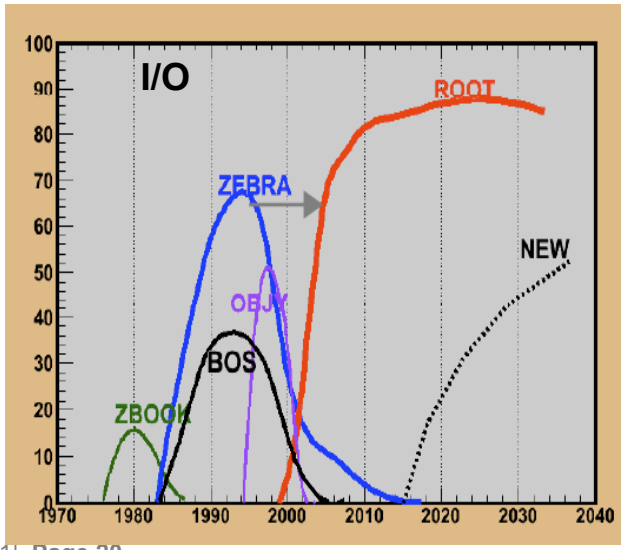


A serious issue: the software maintenance

- Freezing: Technology preservation
  - Virtualisation techniques provide the software environment, freeze the hardware
  - Preparation step is not saved, lifetime limited as well
- Better: Continuous migration
  - Follow technology changes, external software, new OS, redesign, recompile etc
  - Virtualisation can help here too
- Preparation is not trivial
  - New operational model
  - Dependencies etc.
- Supervision is needed for both data and software
  - Data archivist position

R. Brun





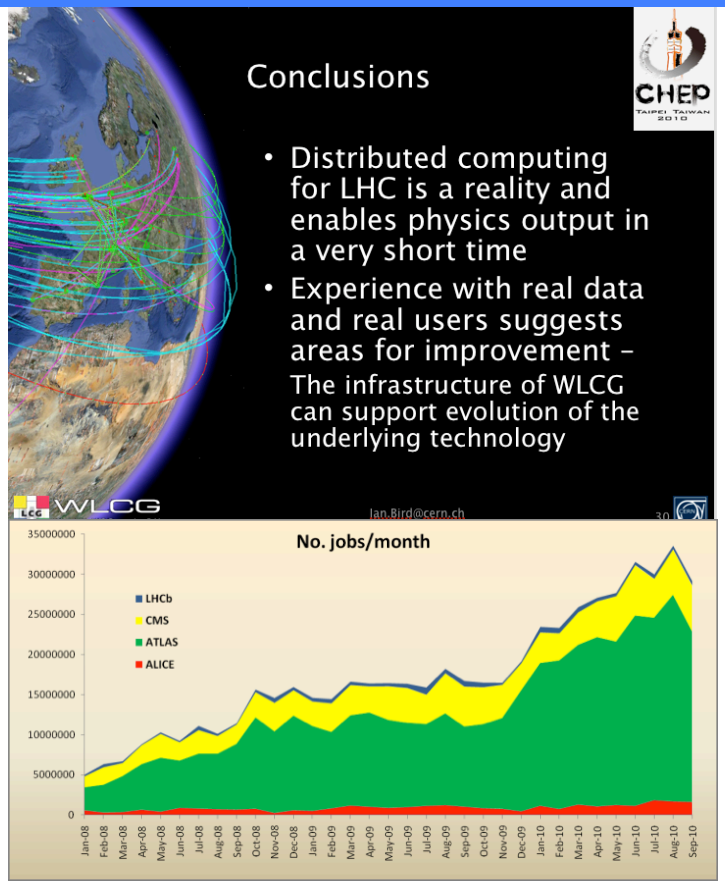
Data Preservation Models identified by DPHEP

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

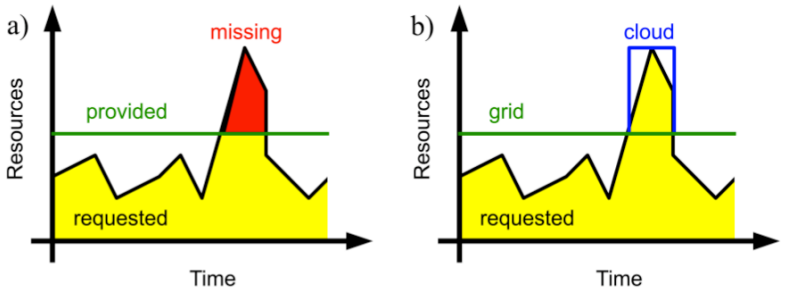
Cost, complexity, benefits

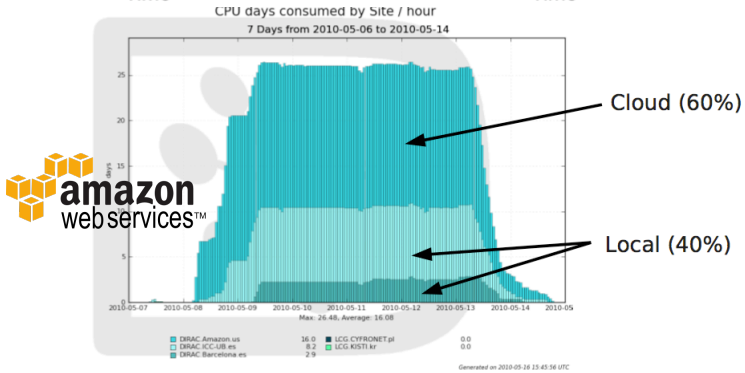
- Levels 1 and 2 still require some work! (Inspire project)
- Only with the full flexibility does the full potential of the data remain
  - Level 4 type programme was required by the JADE and ALEPH re-analyses

The Data Analysis Model is also Evolving



- Cloud computing allows to buy resources on demand
  - Well suited to absorb peaks in varying resource demand

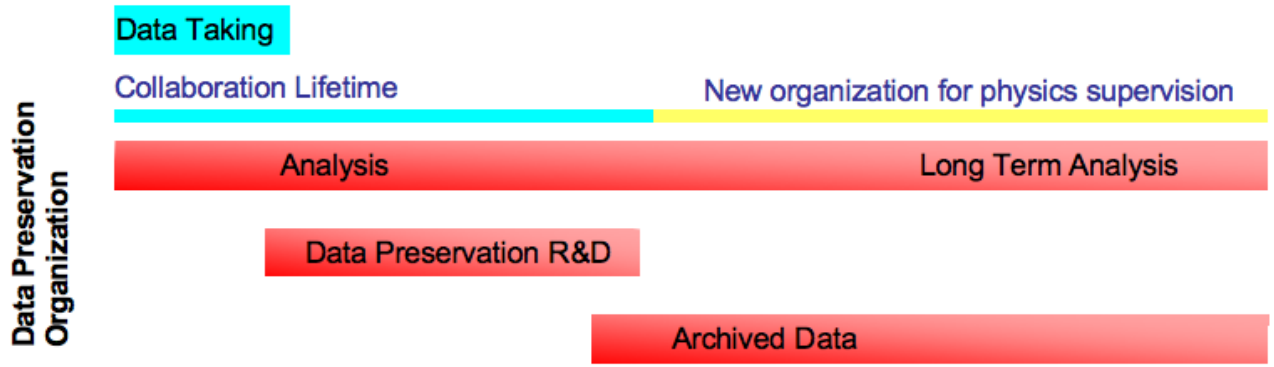


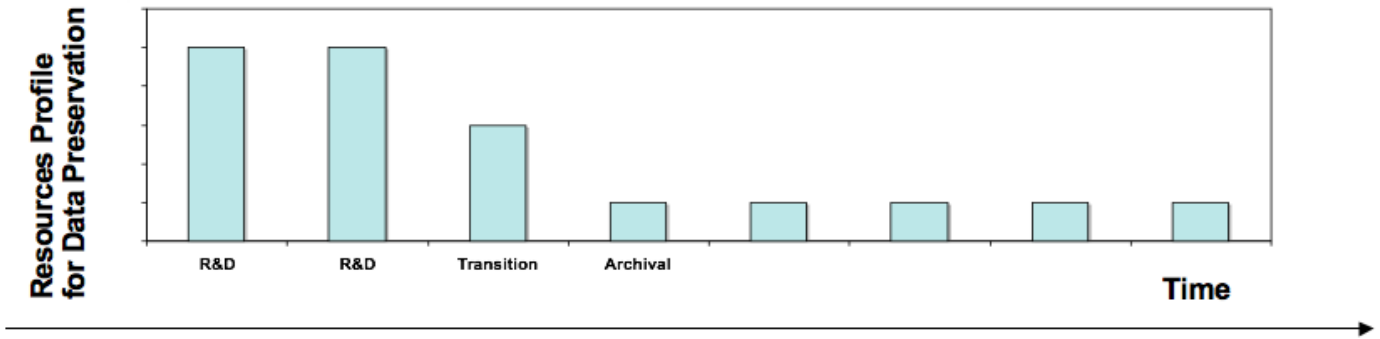


- 170M events (3.6 TB) produced in 6 days
- Amazon Spot Instances → 0.20 USD / 10k events

- Success of the GRID during first year of LHC; Belle buy time on the Cloud
- Where does analysis of preserved data take place and under which protocols?

Governance issues and resources at experiment level





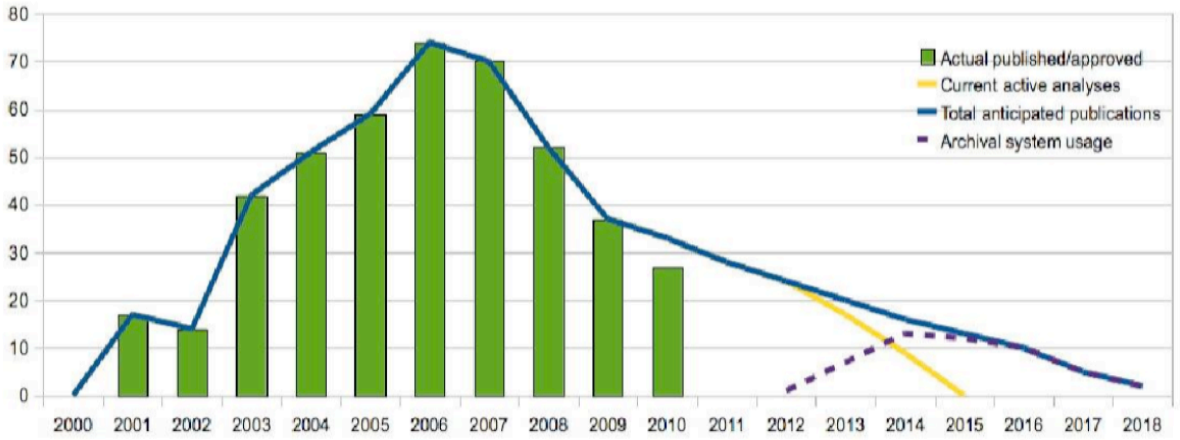
Typically a surge of 2-3 FTEs for 2-3 years, followed by steady 0.5-1.0 FTE per exp./lab

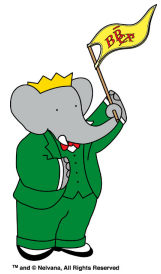
- This should be compared to 300-500 FTEs for many years / experiment!

Cost estimates : << 1% of the original investment

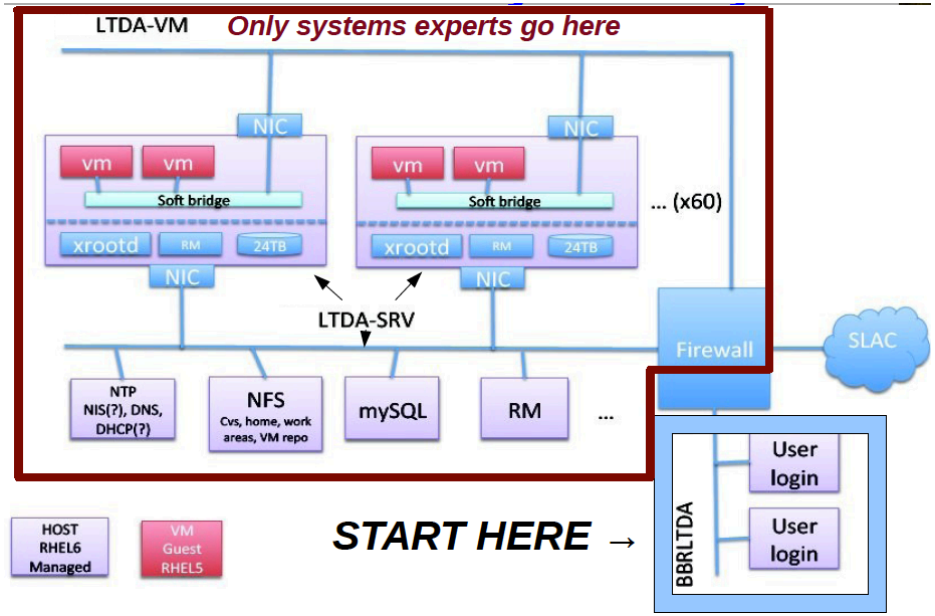
Scientific return : O(10%) in number of publications

Data Preservation at BaBar





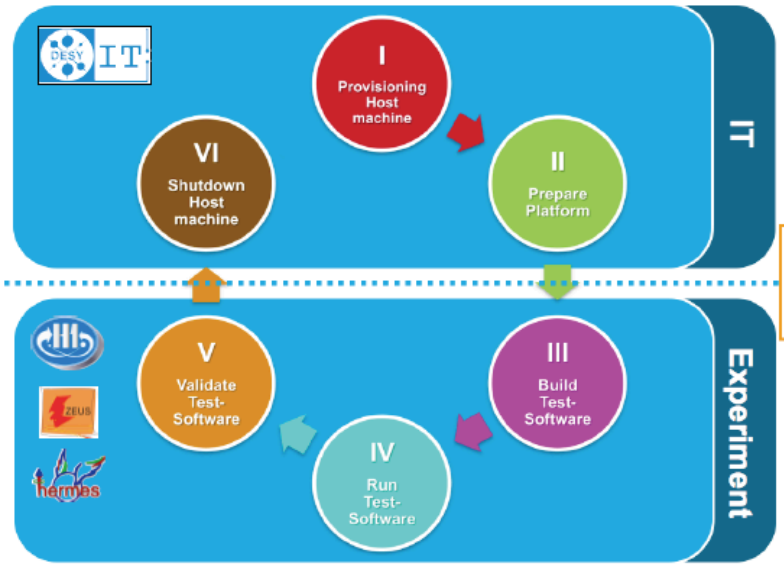
- BaBar moving to an "Archival Mode", preserving analysis ability beyond 2012
  - Virtualisation and cloud computing techniques
  - 2011-2012: Hardware purchase, commissioning

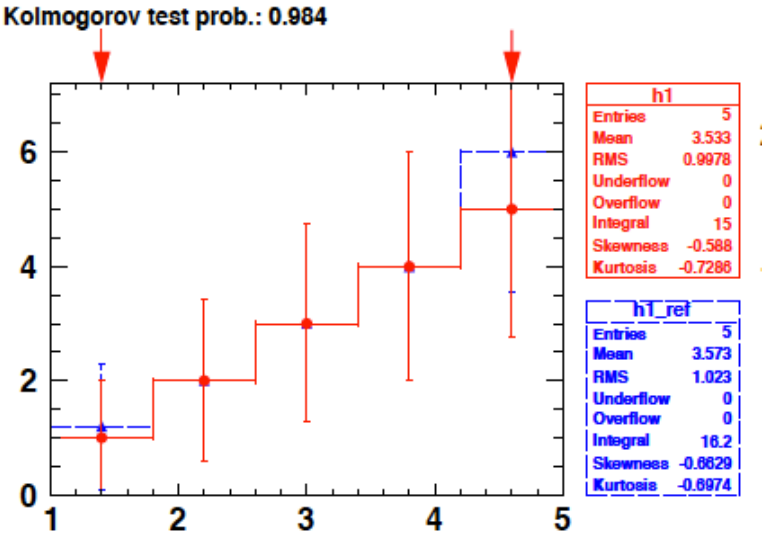


Resources taken into account in funding model during analysis phase

Data Preservation project at DESY

- Validation of experimental software using a virtual environment
  - Roll over the technology steps (OS, h/w, etc.)





- Generic solution, for all HERA experiments: **validate the whole analysis chain**
  - Pilot project being implemented
  - Multi-center cooperation is envisaged : include other experiments

LHC Data Preservation

- Reflection just started in ATLAS, ALICE, CMS, LHCb
  - Common understanding that starting earlier will consolidate the long term future
  - Strong wish to develop a common policy at CERN and within DPHEP

    Make LHC data future-proof: start now!
  - Already concrete thinking on specific cases

    low energy LHC data, trigger configurations, shutdown, versions etc.

ATLAS

Any physics result published by ATLAS must be reproducible by the collaboration at any point in time *without the need to obtain information from the original authors* of the analysis.

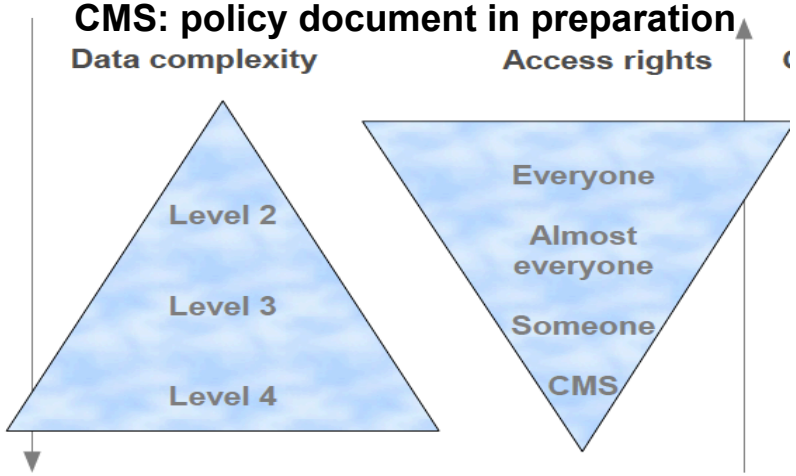
LHCb

LHCb phase I is supposed to collect 5fb^{-1}, recorded in 2010: 0.04fb^{-1}. One might think, no need to make big efforts to preserve 2010 data.

HOWEVER:

- Large statistics of quasi unbiased data with no pile-up, due to low luminosity at startup
- Unique samples of data taken at 0.9TeV, 2.76TeV and 7TeV center of mass energy

CMS: policy document in preparation



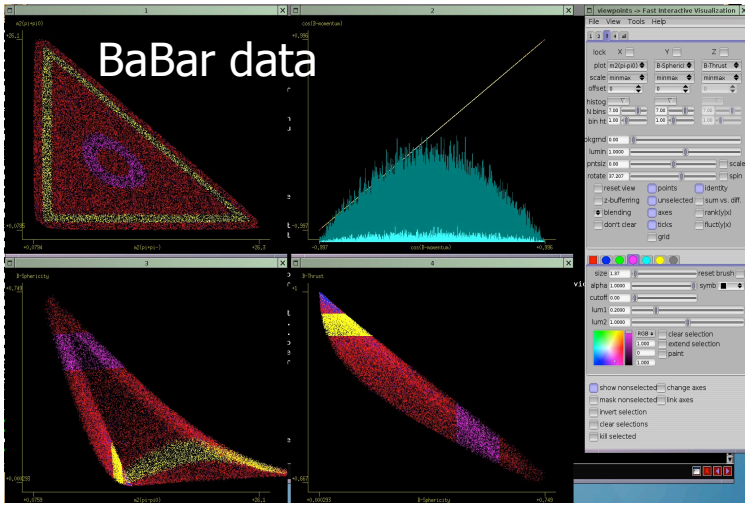
Outreach

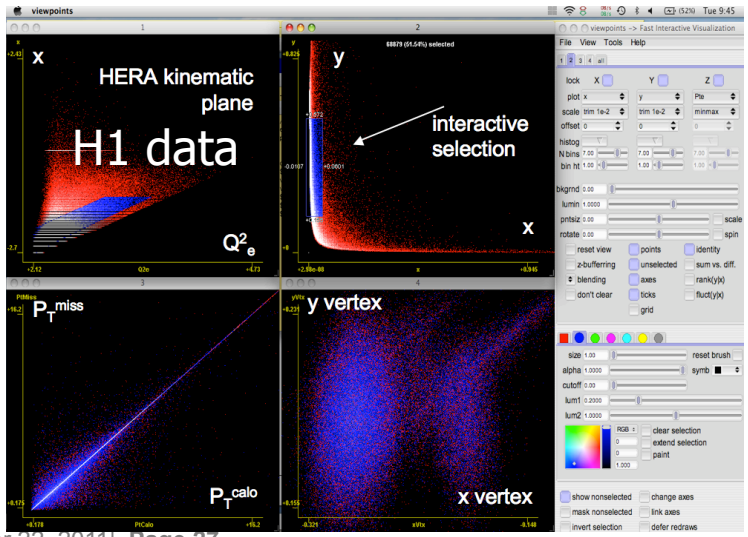
- Use **real** data to enhance HEP education worldwide
- Simple data format: input using text file of kinematics of HEP events

# Q2e	y	x	PtCalo	PtMiss	Ex	Ptx	Phix	Thx	Ee	Pte	Phie	The	Empz	xVtx
377.673	0.174	0.021	2.769	2.769	109.685	15.153	-11.780	8.231	26.226	17.665	164.138	137.656	50.542	-0.237
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
187.320	0.211	0.009	2.584	2.584	51.742	9.773	78.869	13.682	23.482	12.168	-106.349		140.813	55.164
264.266	0.580	0.005	0.238	0.360	35.343	11.738	-138.270		64.975	15.904	11.407	41.034	134.465	57.043
229.056	0.043	0.052	4.204	5.067	65.601	19.196	72.870	17.842	28.485	14.085	-98.351	148.685	58.941	-0.237
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
240.102	0.183	0.013	3.513	3.434	67.134	17.402	85.049	17.281	24.719	14.004	-92.840	145.491	56.060	-0.266
451.996	0.289	0.021	1.723	1.723	49.126	17.196	66.018	24.927	25.936	18.913	-114.452		133.180	55.818
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
391.944	0.000	0.000	2.107	2.107	183.513	21.270	75.875	6.693	31.602	19.959	-108.713		140.834	58.375
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
335.801	0.052	0.064	16.769	16.769	29.256	1.142	-90.021	2.250	29.219	17.048	83.461	142.349	52.723	-0.242
286.039	0.009	0.315	2.514	2.514	194.568	18.922	-83.365	5.616	29.944	16.837	92.126	145.787	56.826	-0.254
207.703	0.137	0.015	8.095	8.095	84.993	21.487	82.549	15.129	25.701	13.389	-95.886	148.685	53.237	-0.258
387.371	0.358	0.011	1.272	1.272	70.266	15.468	-91.456	17.173	21.232	15.772	93.071	132.827	55.684	-0.236
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
154.527	0.667	0.002	3.589	3.589	72.273	8.818	174.458	101.604	18.598	7.176	-28.067	137.379	92.478	-0.248
304.756	0.025	0.121	1.622	1.622	120.028	17.765	-145.756		8.522	29.678	17.240	39.272	144.486	55.298
270.950	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
456.769	0.204	0.022	1.621	1.568	48.542	18.231	-134.045		26.029	26.095	19.063	50.043	133.070	55.126
275.593	0.050	0.055	0.296	0.296	39.621	16.380	-71.596	25.608	28.728	16.184	109.363	145.712	58.077	-0.249
890.585	0.853	0.010	7.314	4.893	193.928	16.386	173.467	6.104	12.119	11.435	-5.872	70.653	48.847	-0.248
328.267	0.018	0.105	2.294	2.294	200.323	16.451	-16.314	4.712	30.890	17.959	169.425	143.356	54.999	-0.258
156.729	0.466	0.003	2.340	1.529	64.710	10.664	-88.809	17.271	16.150	9.146	93.350	145.507	59.919	-0.257
270.064	0.025	0.105	8.909	8.909	304.993	24.646	-177.466		4.653	29.349	16.225	10.869	146.439	56.009

- Discussions about common formats ongoing
  - B-lab (KEK) example considered
  - Experience at LHC
  - Connect to existing projects (master classes etc.)

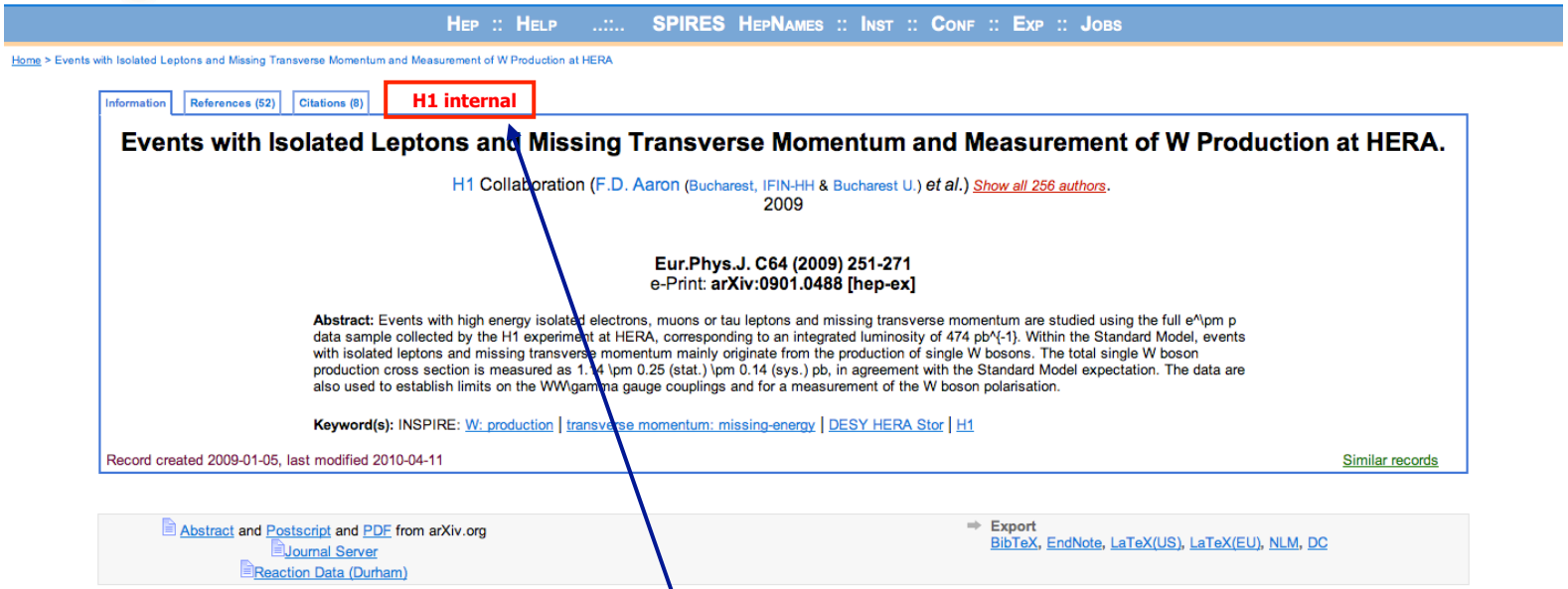
Viewpoints (NASA)





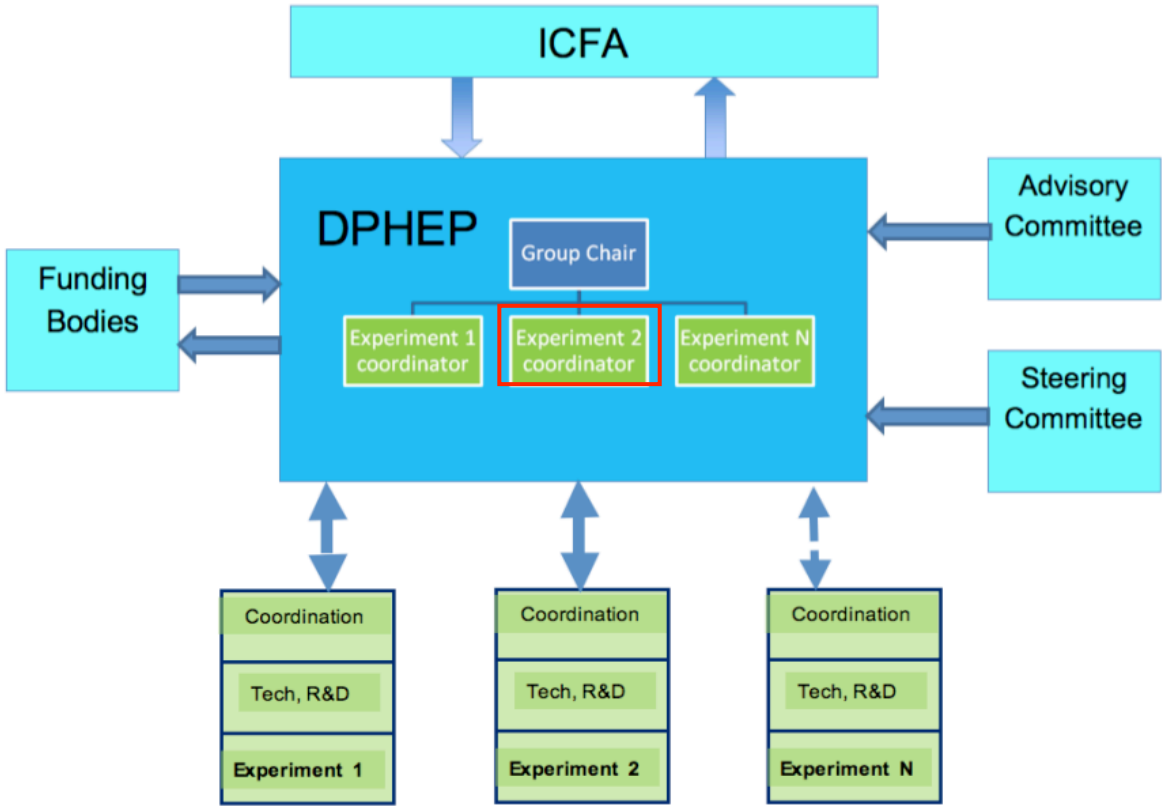
Common project on documentation: INSPIRE

http://inspirebeta.net/



- Envisage an additional link for the collaboration members only
- Provides additional information (notes, slides etc.)
- Reduced data and macros also possible

DPHEP Organisation



- Support expressed by major laboratories and committies: ICFA, HEPAP, FALC
- Funding plan in preparation: a **Project Manager** is needed
  - Ensure collaborative continuity, fund rising and connections, project overview

From concepts to implementation: recent progress

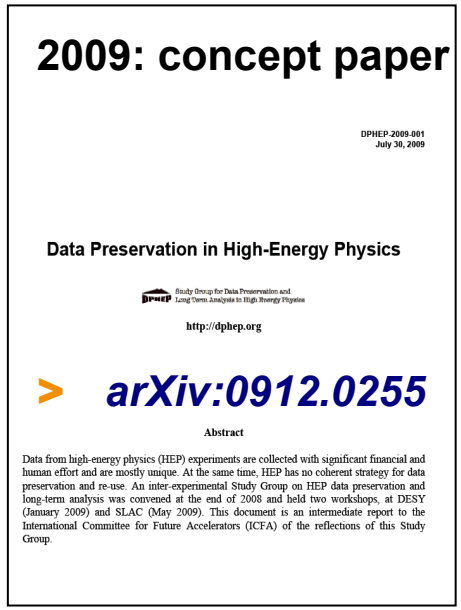
- Data preservation plans/projects started by experiments
  - Technology and Organisation within the experiments
  - 2 dedicated projects funded (DESY, SLAC)

- March 2011: LHC experiments joined DPHEP

- Progress in multi-experiment projects
  - Preservation technologies
  - Documentation (Inspire+experiments)
  - Outreach (common outreach formats etc.)

- Blueprint in preparation: status, proposals,costs

2009: concept paper



DPHEP-2009-001
July 30, 2009

Data Preservation in High-Energy Physics

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

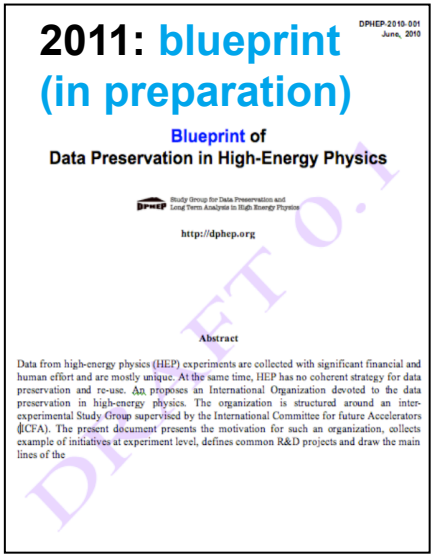
http://dphep.org

arXiv:0912.0255

Abstract

Data from high-energy physics (HEP) experiments are collected with significant financial and human effort and are mostly unique. At the same time, HEP has no coherent strategy for data preservation and re-use. An inter-experimental Study Group on HEP data preservation and long-term analysis was convened at the end of 2008 and held two workshops, at DESY (January 2009) and SLAC (May 2009). This document is an intermediate report to the International Committee for Future Accelerators (ICFA) of the reflections of this Study Group.

2011: blueprint (in preparation)



DPHEP-2010-001
June, 2010

Blueprint of
Data Preservation in High-Energy Physics

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

http://dphep.org

DRAFT 0.1

Abstract

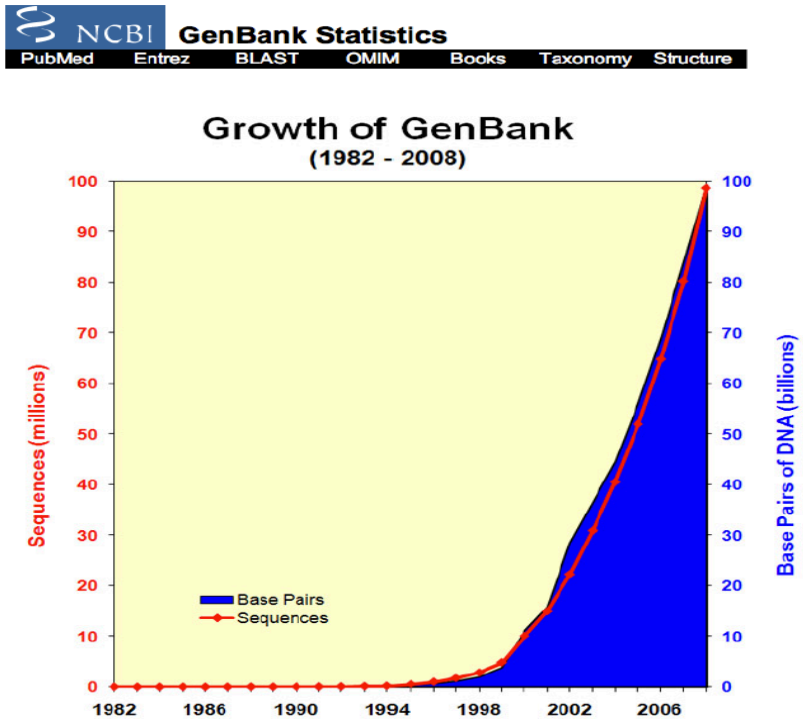
Data from high-energy physics (HEP) experiments are collected with significant financial and human effort and are mostly unique. At the same time, HEP has no coherent strategy for data preservation and re-use. An proposes an International Organization devoted to the data preservation in high-energy physics. The organization is structured around an inter-experimental Study Group supervised by the International Committee for future Accelerators (ICFA). The present document presents the motivation for such an organization, collects example of initiatives at experiment level, defines common R&D projects and draw the main lines of the

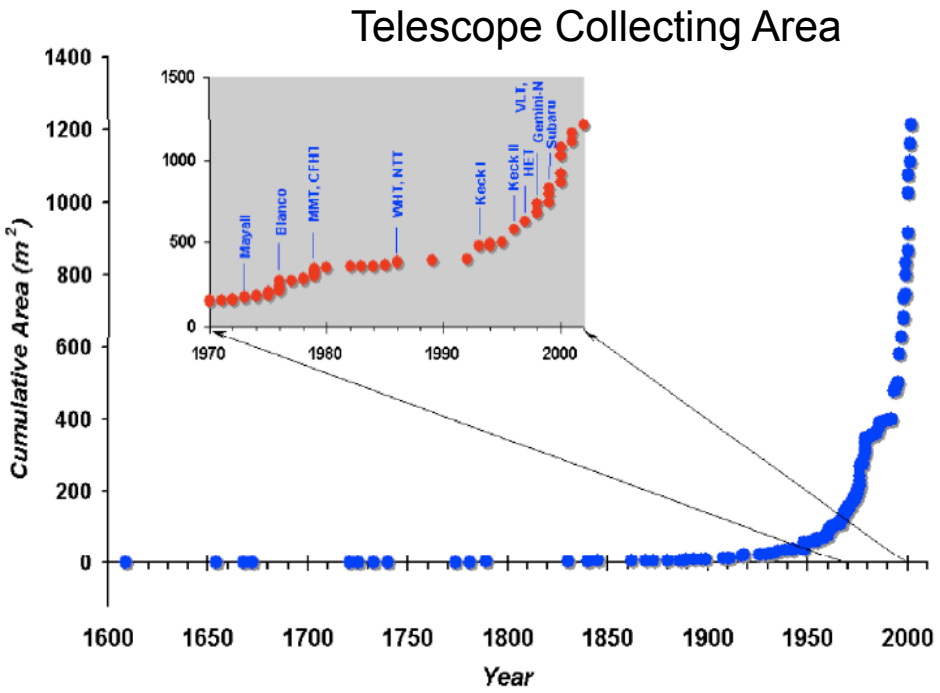
Priorities in HEP Data Preservation

- **Priority 1: Experiment Level Projects in Data Preservation.**
  - **2-3 FTEs /exp**
- **Priority 2: International Organisation DPHEP.**
  - **Project Manager (1 FTE)** is needed
- **Priority 3: Common R&D projects**
  - **each involving 1-2 dedicated FTE, across several laboratories**
- These priorities could be enacted with **a funding model** implying contributions from the three regions (Europe, America, Asia) and strong connections with laboratories hosting the data samples.

We are not alone….

- Other fields observe a dramatic increase in data and are questioning the long term future of this data





Other fields

- Task forces already in place to address this issue in a generic way (standards)
  - e.g. Blue Ribbon, APA, DPC, eSciDir, …

http://www.alliancepermanentaccess.eu
http://brtf.sdsc.edu
(intermediate report and references)

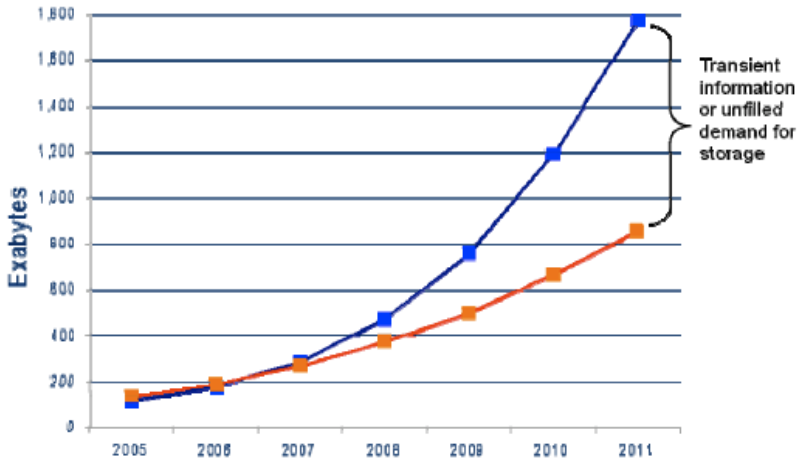


FIGURE 1.3: **Information and Storage**
Source: J. Gantz January 2008 (revised). Used with permission.

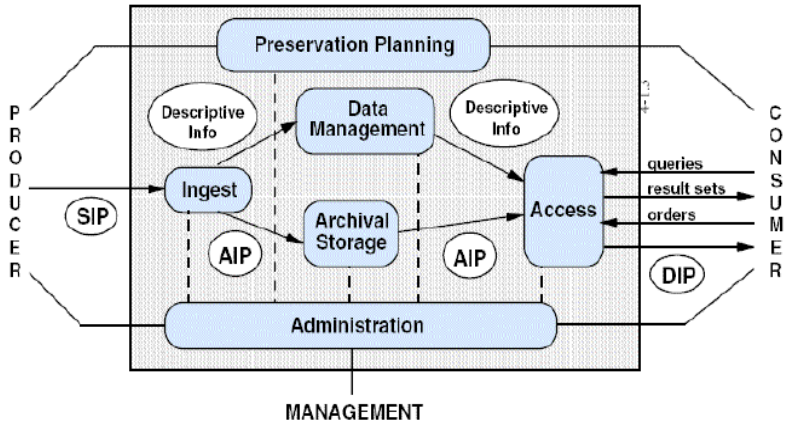
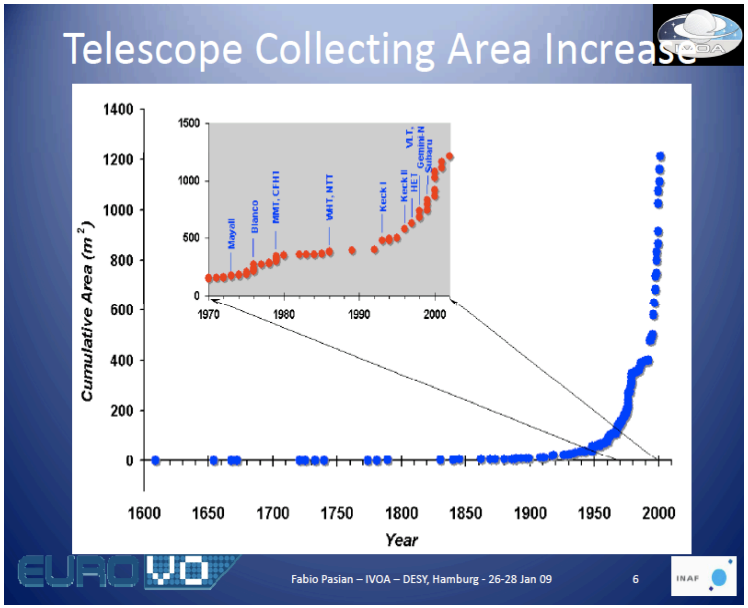
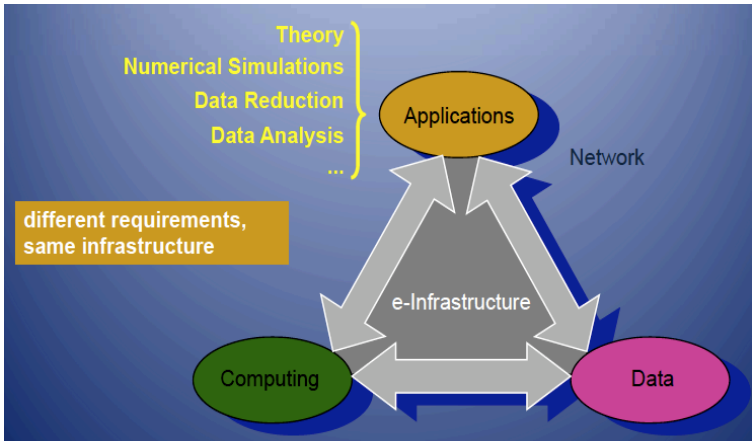


FIGURE 2.1: **The OAIS Reference Model**
http://public.ccsds.org/publications/archive/650x0b1.pdf, Page 4-1.
Source: Consultative Committee for Space Data Systems January 2002.

- Scientific Data is a major component of the ongoing efforts (complexity)
- Some scientific fields are well advanced : astrophysics

Virtual Observatories in Astrophysics





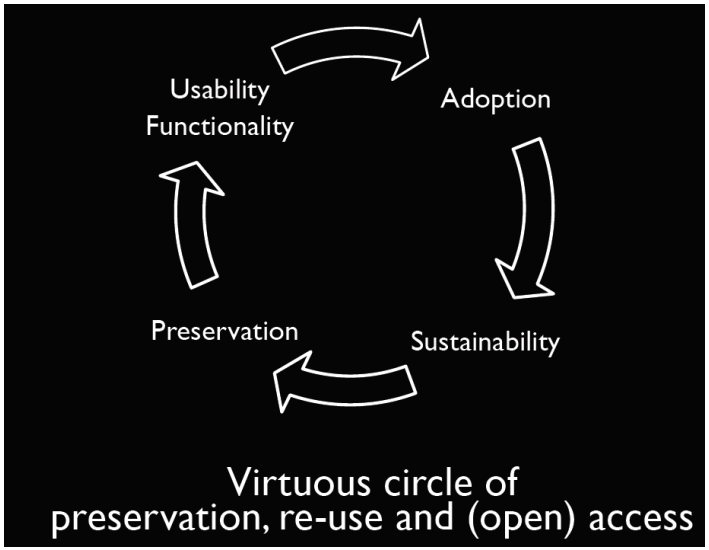
F.Pasian

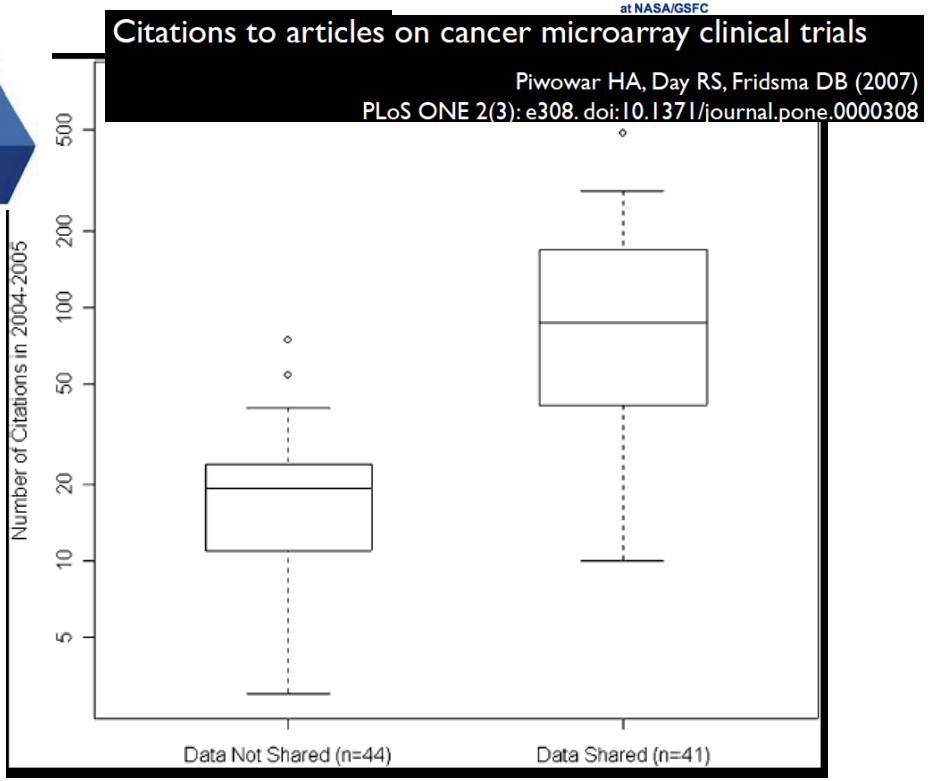
- Data Archives Inter-operable
- Work on standards and access to
  - Data, simulation, mining techniques
- International, multi-experiment

C.Diaconu, DESY, February

Scientific data preservation as a common goal?

- Common issues on:
  - Data integration
  - Technologies for data curation
  - Work on standards
  - (Open) access
  - Policy issues
  - Funding models





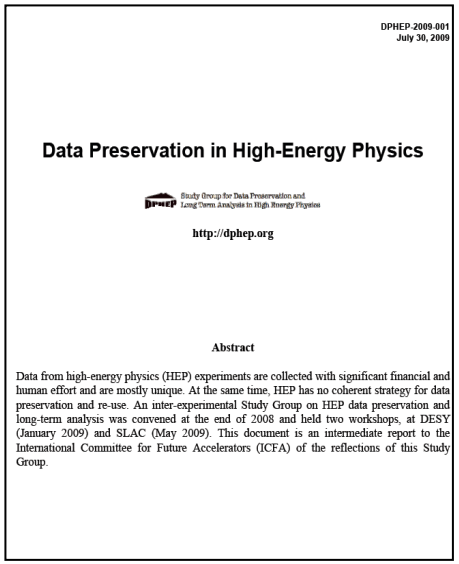
Conclusion and Outlook

- **Data preservation in HEP have a true scientific potential**
  - Relevant **physics cases** for future use can be made
  - It is **timely**, given the current experimental situation and plans
  - It may **enhance the return** on the initial investment in the experimental facilities
  - It provides additional research at particularly **low cost**
- It requires a strategy and well-identified resources
  - Synergic action of experiments, laboratories and funding agencies
- International cooperation is the best way to proceed
  - Support a coherent approach and a common structure: the DPHEP organization
- Potential for collaborations beyond HEP: many issues are already addressed in other fields
  - HEP has a true specificity: complex data, large scale analysis, long term projects

Backup

DPHEP Intermediate Recommendations (end 2009)

- ***arXiv:0912.0255***



DPHEP-2009-001
July 30, 2009

Data Preservation in High-Energy Physics

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

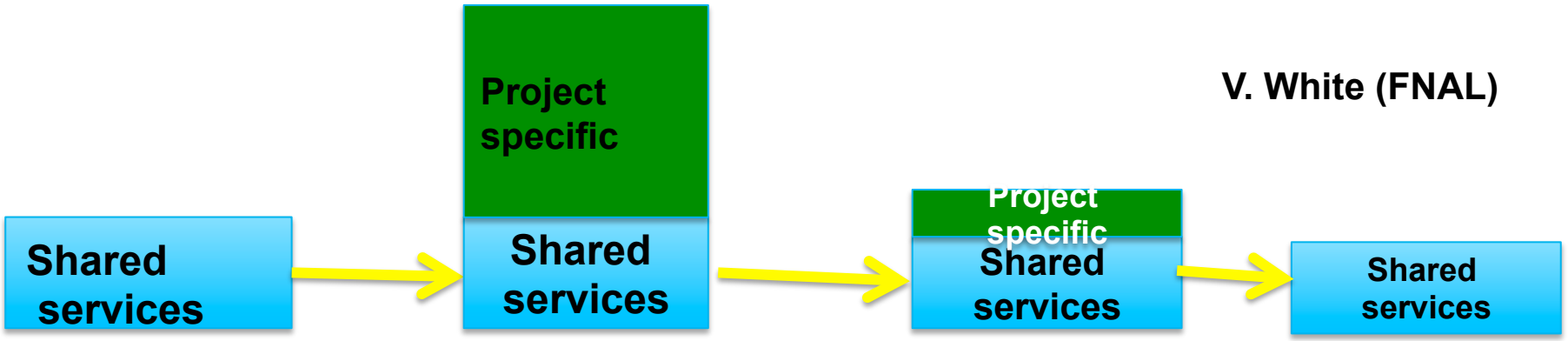
http://dphep.org

Abstract

Data from high-energy physics (HEP) experiments are collected with significant financial and human effort and are mostly unique. At the same time, HEP has no coherent strategy for data preservation and re-use. An inter-experimental Study Group on HEP data preservation and long-term analysis was convened at the end of 2008 and held two workshops, at DESY (January 2009) and SLAC (May 2009). This document is an intermediate report to the International Committee for Future Accelerators (ICFA) of the reflections of this Study Group.

- An urgent and vigorous action is needed to ensure data preservation in HEP
- The preservation of the full analysis capability of experiments is recommended, including the preservation of reconstruction and simulation software
- An interface to the experiment know-how should be introduced: **data archivist** position in the computing centres
- The preservation of HEP data requires a synergic action: collaborations, laboratories and funding agencies
- An International Data Preservation Forum is proposed as a reference organisation. The Forum should represent experimental collaborations, laboratories and computing centres

Data cycle



V. White (FNAL)

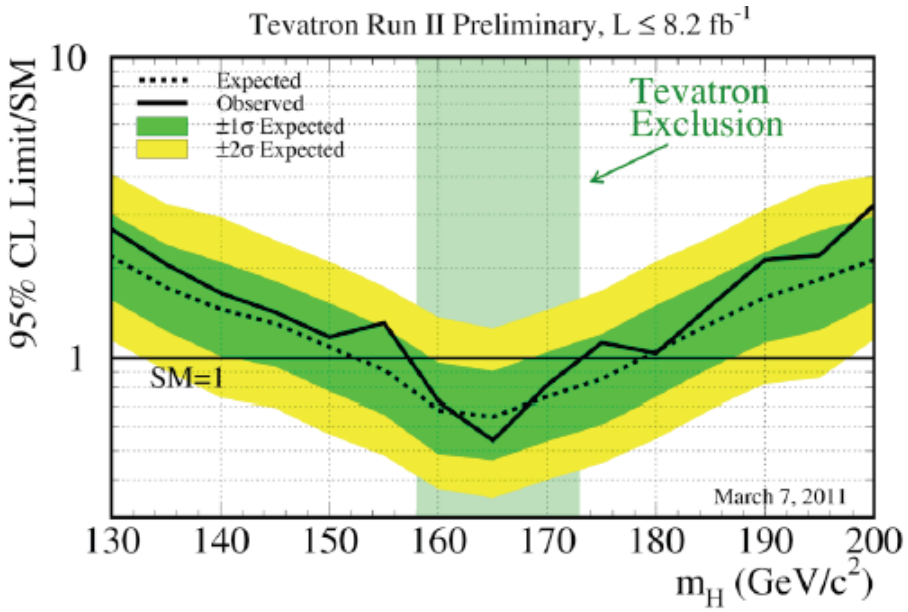
Early Period
R&D, Ideas, Simulations
LOI, TDR, Proposals

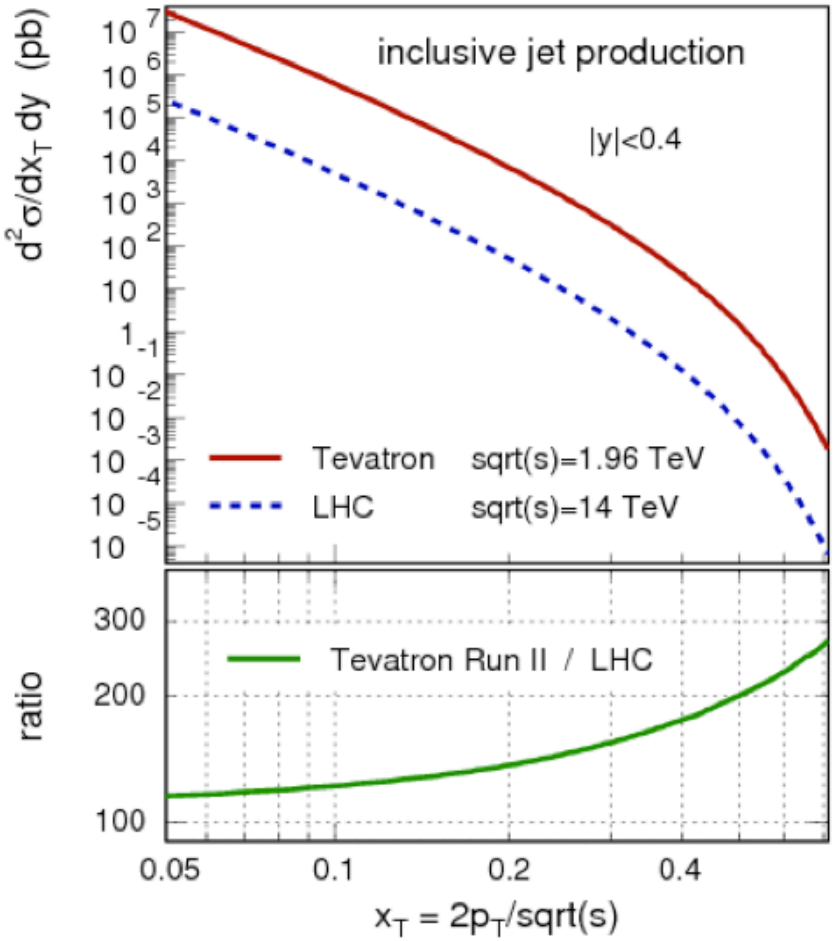
Mature phase
Construction, operations, analysis

Final data-taking and beyond
Final analysis, data preservation and access

Example: LHC will not completely take over Tevatron physics

- Tevatron proton-antiproton collisions are in fact unique
- There is a physics case for Tevatron-LHC/LEP combinations





Many more examples are available….

There is a solid physics case to prolong the HEP data's lifetime

Better theory, better methods

OPAL [2011] arXiv:1101.1470 [hep-ex]

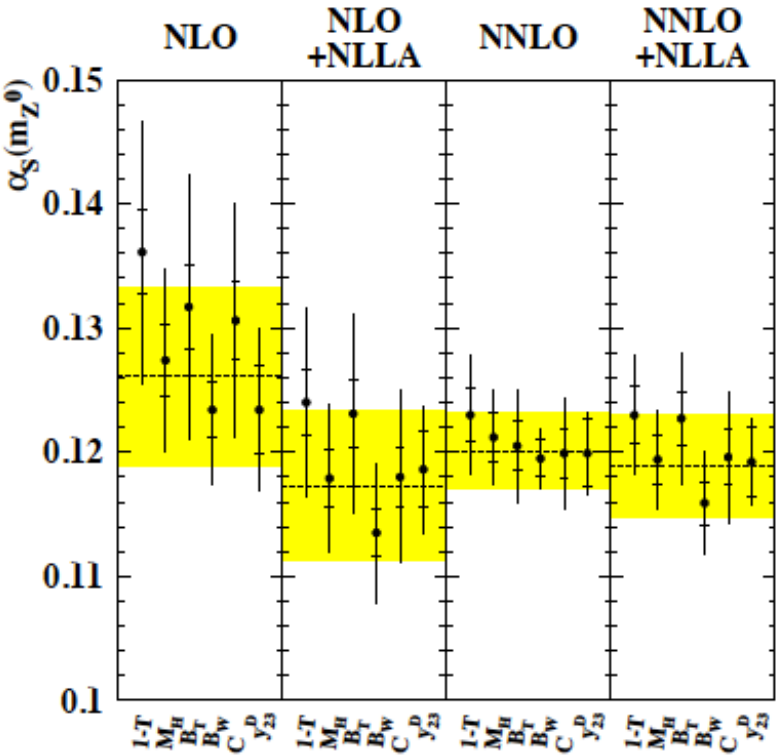


Figure 6: α_S results combined over all OPAL c.m. energies for different event shape variables and different QCD calculations as indicated on the figure. The shaded bands and dashed lines show the values of $\alpha_S(m_{Z^0})$ combined from these values with total uncertainties. The inner and outer uncertainty bars show the combined statistical and experimental and total uncertainties, respectively.

DELPHI, Eur.Phys.J. C71 (2011) 1557

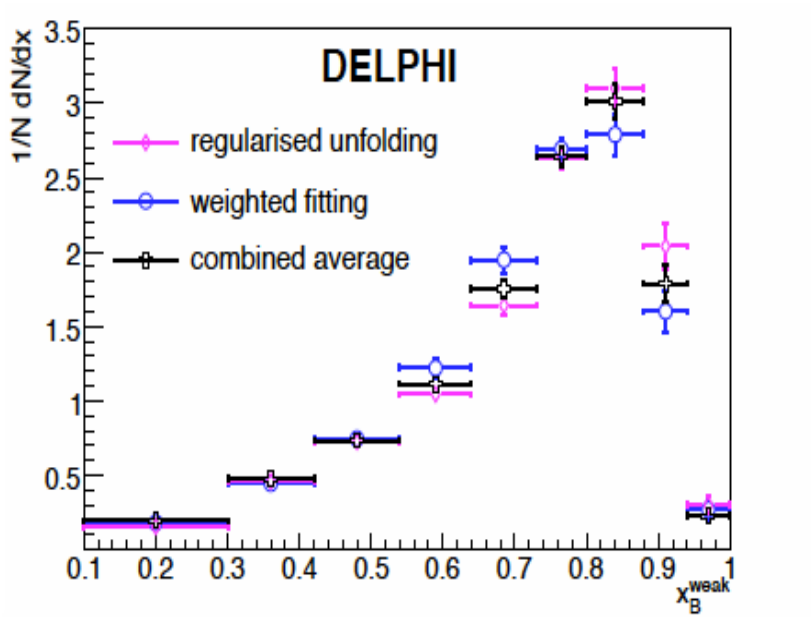
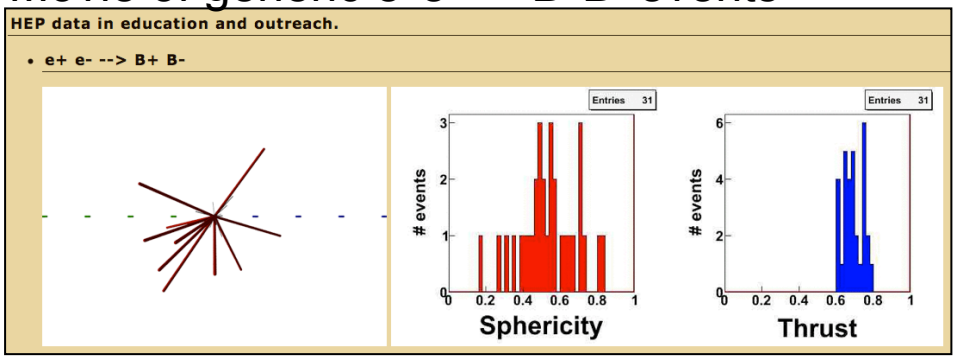


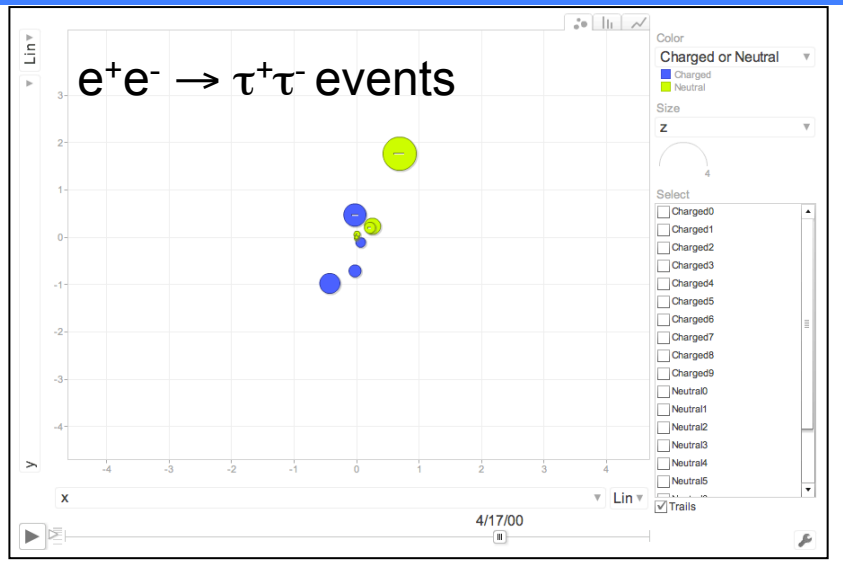
Figure 6: Measured fragmentation distributions in the two analyses and their combined average. Uncer-

Outreach Data and Tools

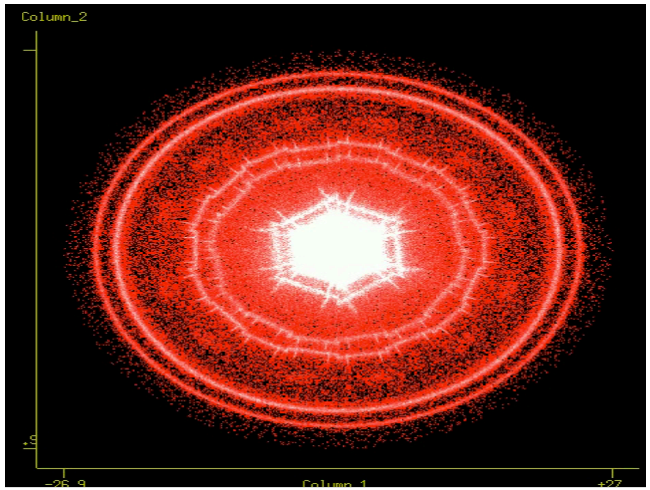
http://www.slac.stanford.edu/~bellis/HEP_data.html

Movie of generic $e^+e^- \rightarrow B^+B^-$ events

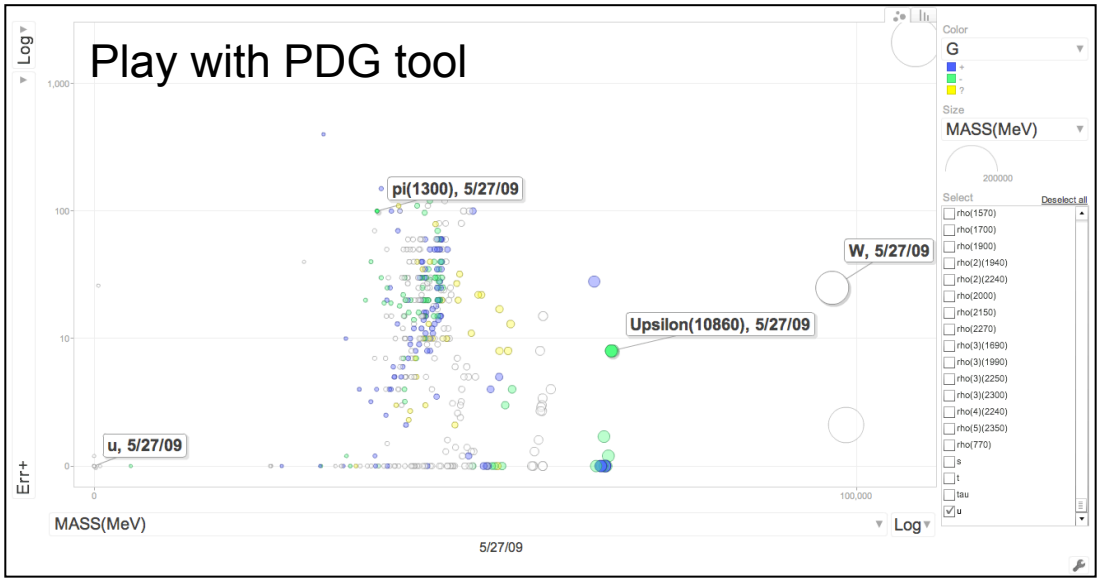




Several outreach tools already being used in classrooms

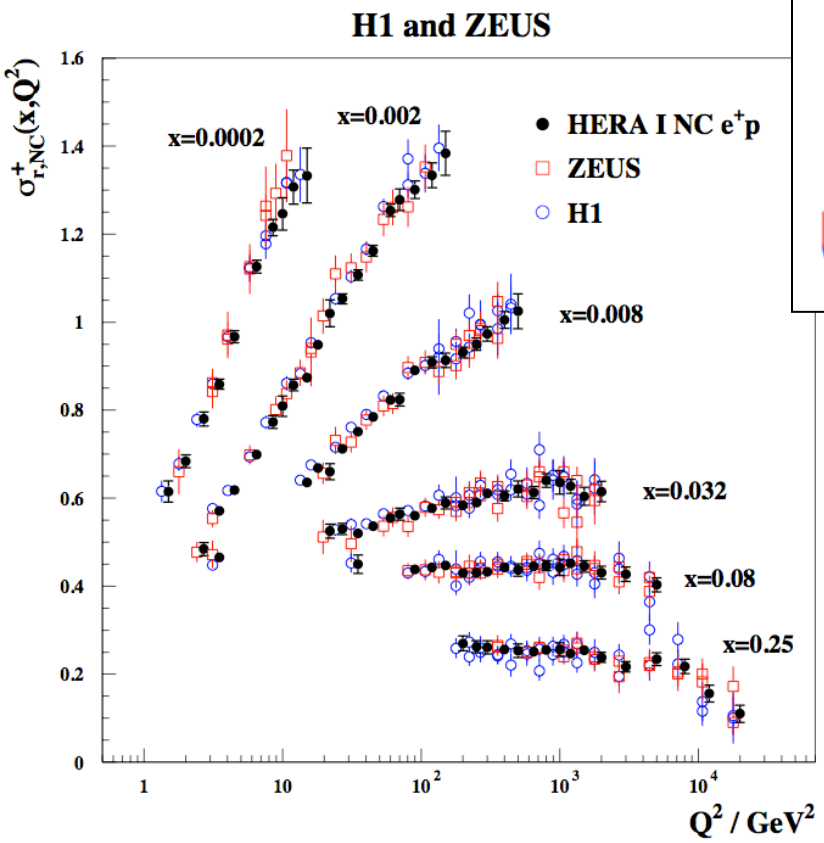


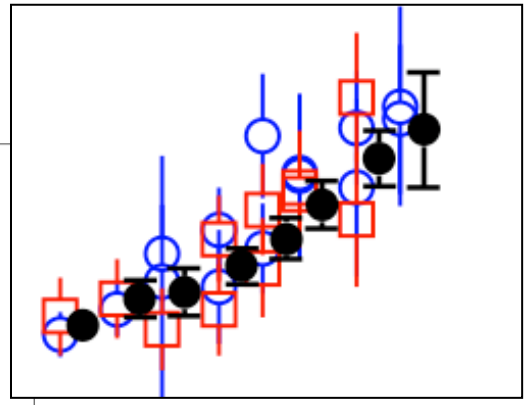
Tomography studies using converted photons in BaBar Silicon Vertex-Tracker

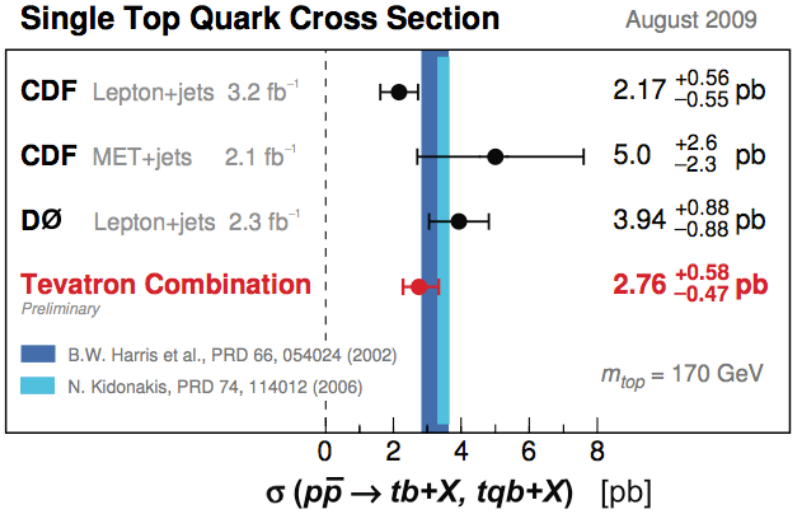


Cross Collaboration and Combinations

- Combined results already exist from LEP, Tevatron, HERA as well as BaBar+Belle (in progress)







- Preserved data would make possible more combined analyses across experiments

New theories, new interpretations

CERN-PH-EP-2011-080
May 20, 2011

Test of the τ-Model of Bose-Einstein Correlations and Reconstruction of the Source Function in Hadronic Z-boson Decay at LEP

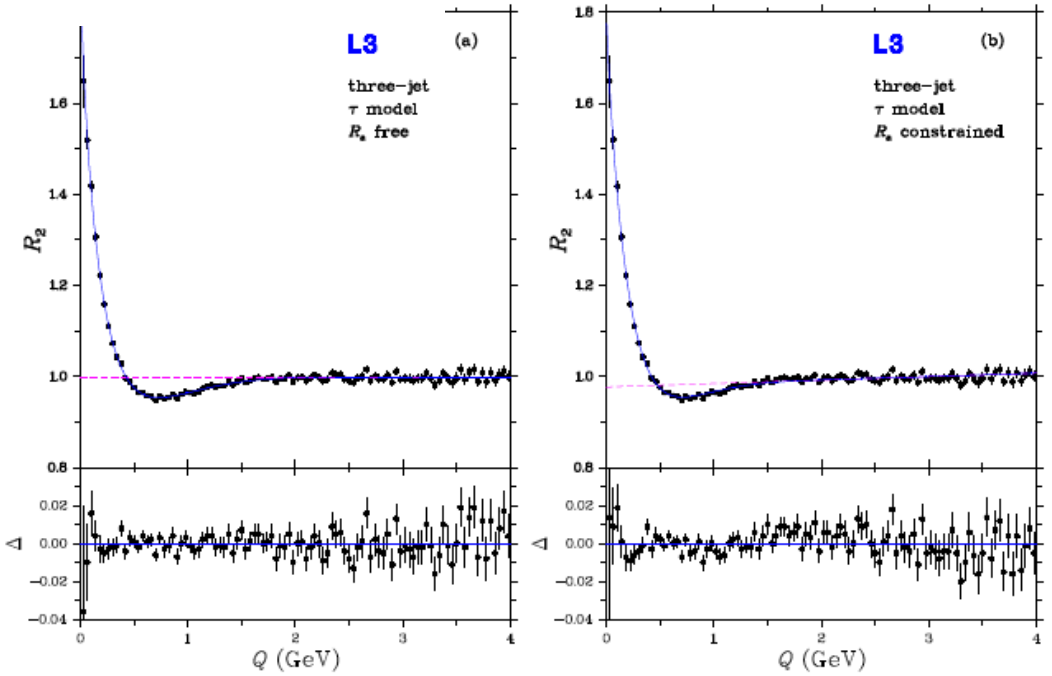


Figure 5: The Bose-Einstein correlation function R_2 for three-jet events. The curve corresponds to the fit of the one-sided Lévy parametrization, Eq. (13), with the parameter R_s (a) free and (b) constrained by Eq. (14). The results of the fits are given in Tables 1 and 2, respectively. Also plotted is Δ, the difference between the fit and the data. The dashed line represents the long-range part of the fit, *i.e.*, $\gamma(1+\epsilon Q)$.

Dark photons: subject is new, data is old

Figure from Bluemlein, Brunner (2011)

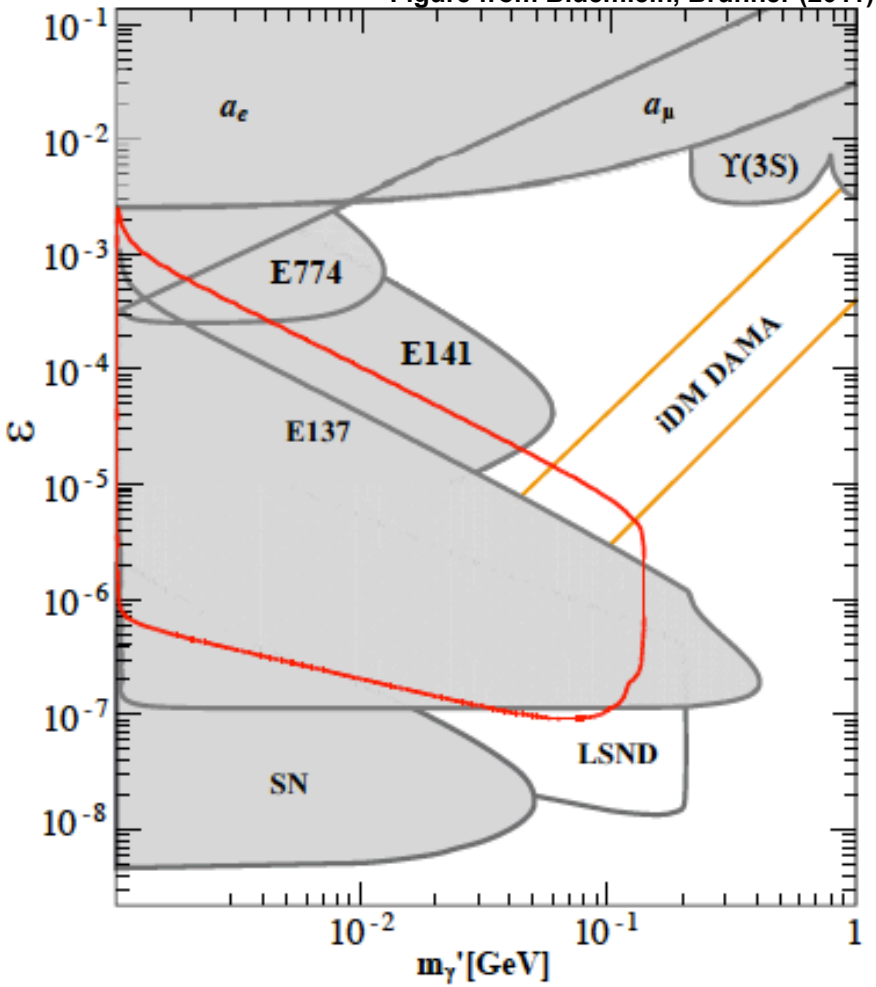
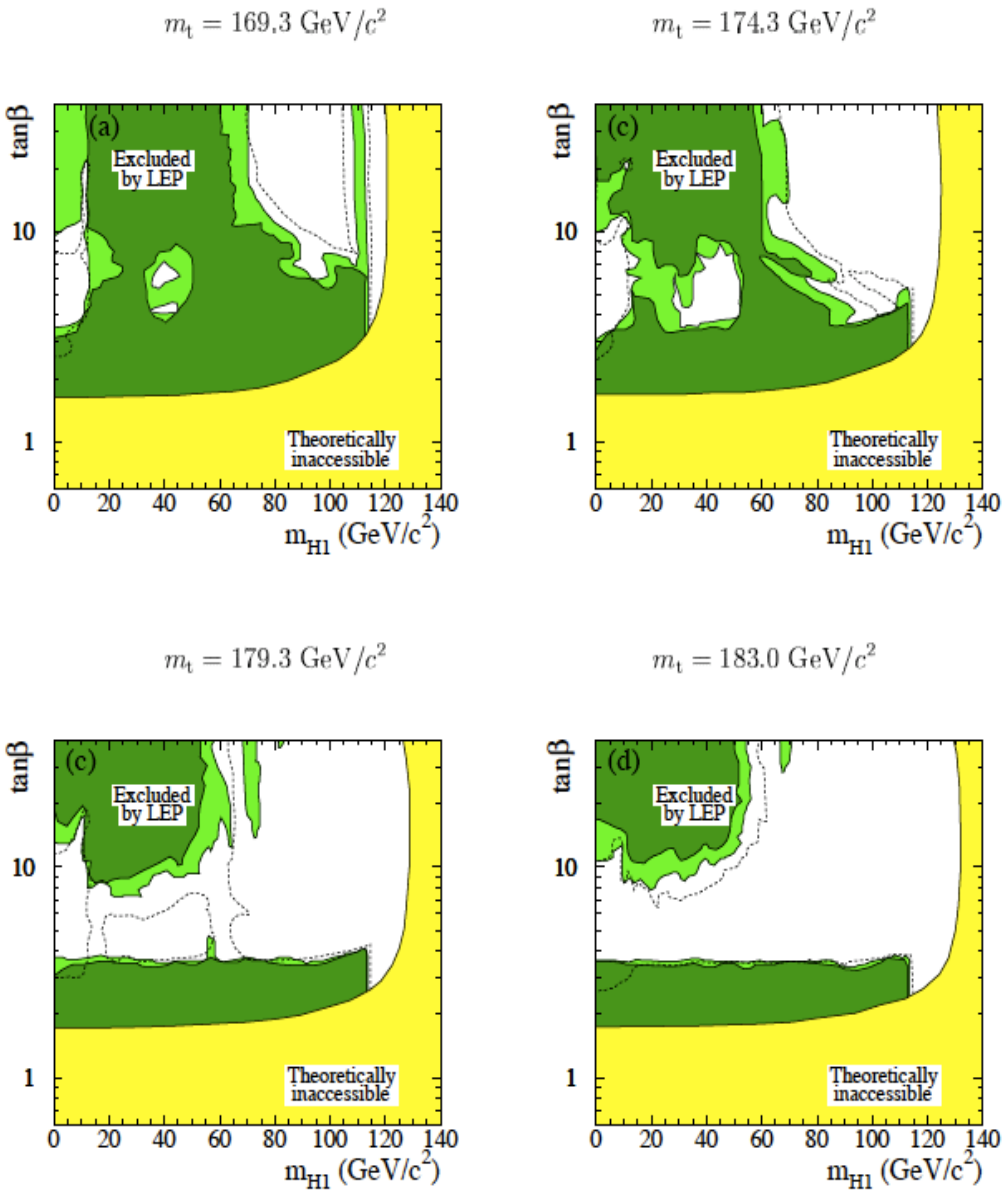


Figure 5: Comparison of the present exclusion bounds (red line) with other limits from the measurement of the anomalous magnetic moments a_e and a_μ [19], $\Upsilon(3S)$ decay [20], the beam dump experiments E137, E141, E774 [21–23], and supernovae cooling [4,24]. We indicate the prospects for LSND [7,25] (open grey-bounded area), and the DAMA/LIBRA region (open orange bounded area) [26]. The limits for $\epsilon > 10^{-7}$ have been taken from Ref. [6].

Excluded?

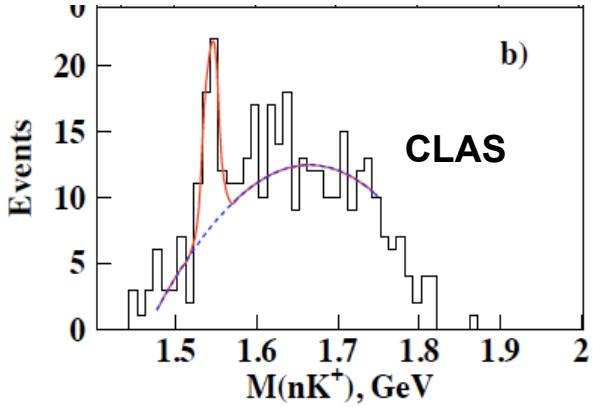
- Some external parameters may be not well known
- Re-optimisation may be a case for re-analysis

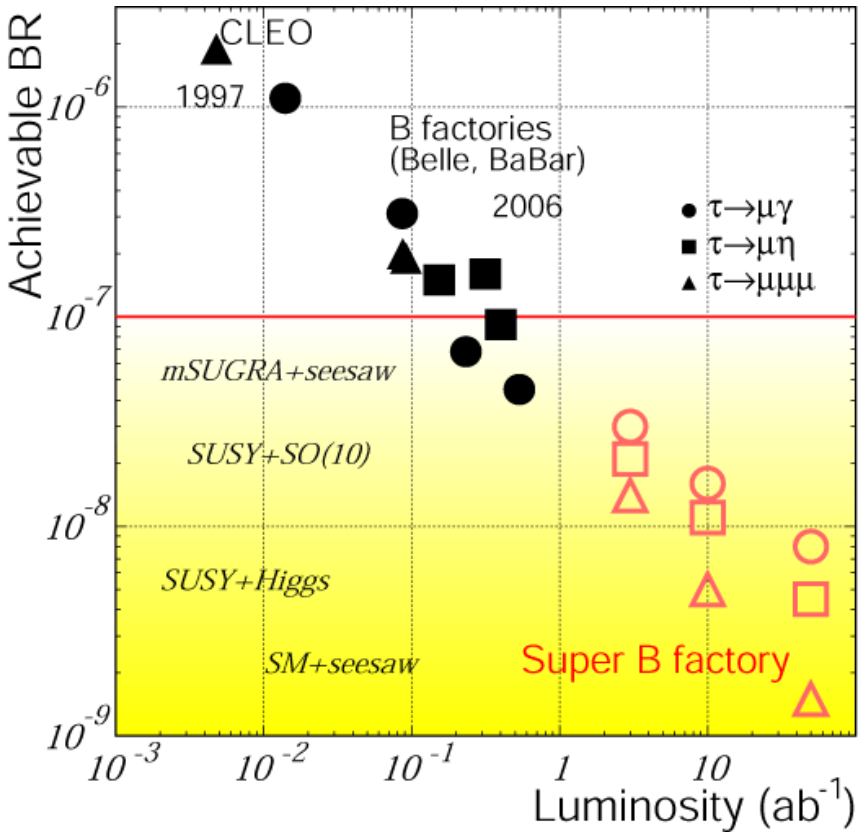


More examples…

- B- and SuperB-factories
- Low energy
- …and many others
  - your favourite?

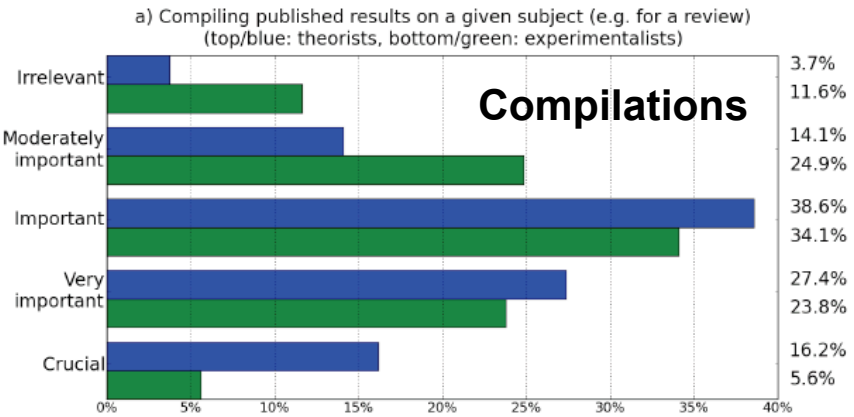
...surprises can occur at lower energies too

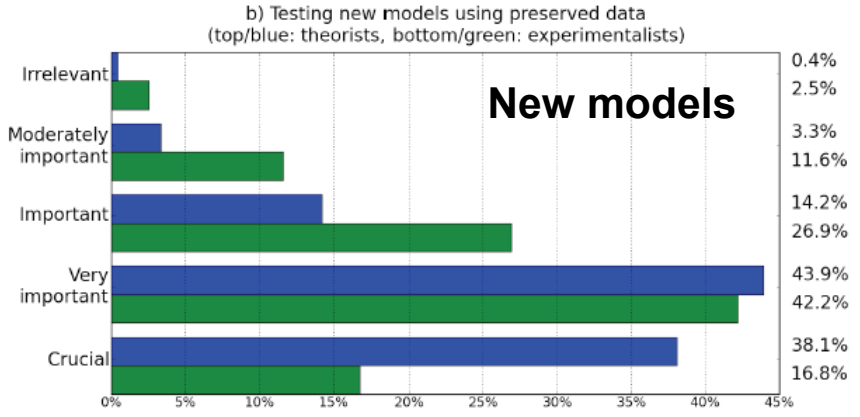


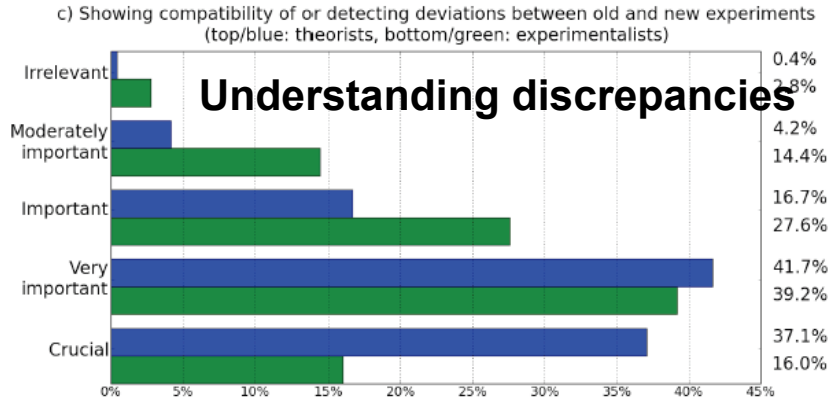


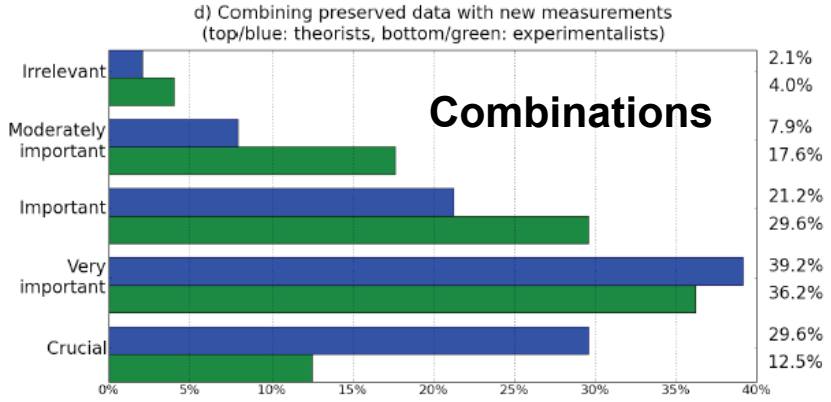
Physics case: opinions in the HEP community

Preserving HEP data is important for:



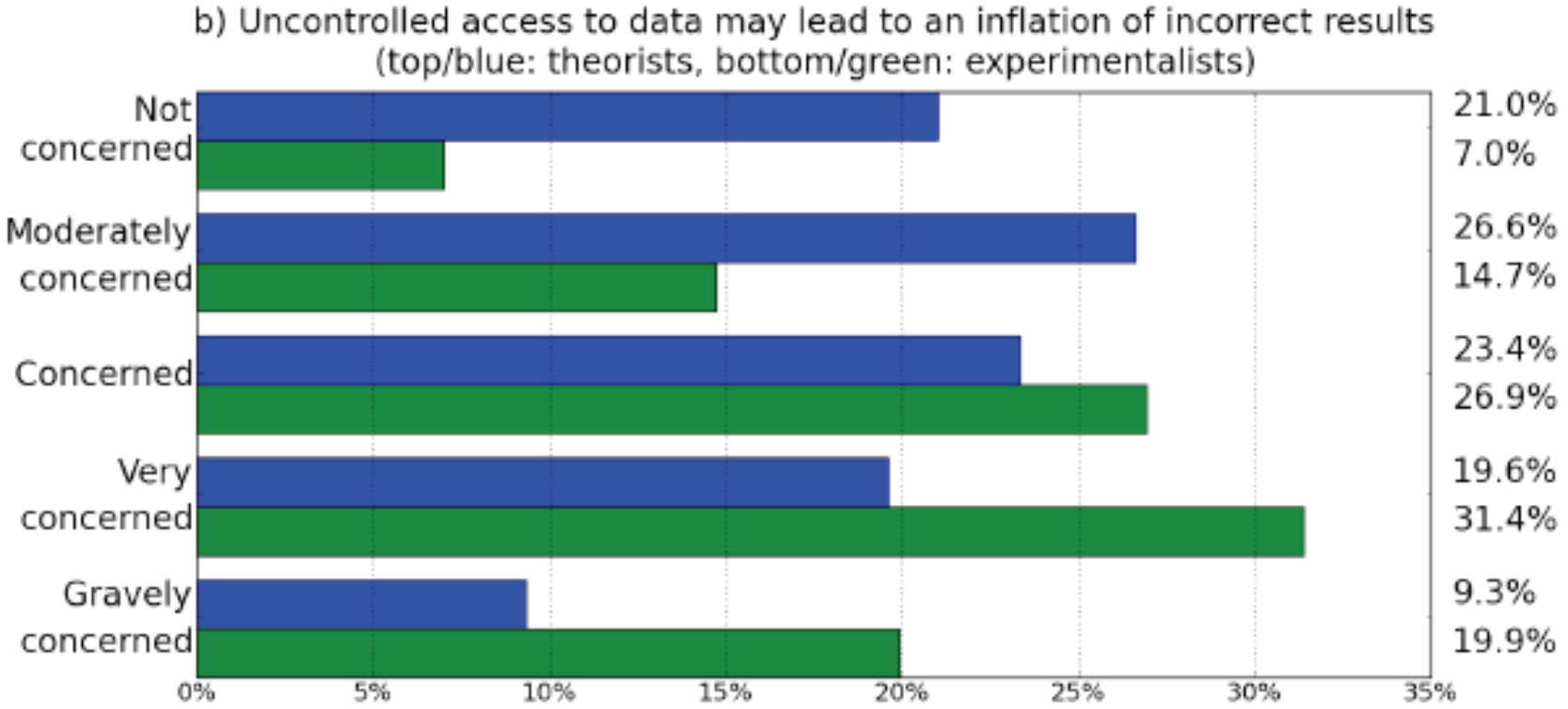






Risks of re-use?

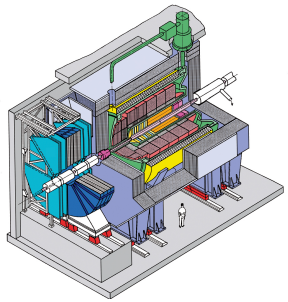
Parse.insight



Governance issues are very important

"Errors using inadequate data are much less than those using no data at all."
Charles Babbage

An example: The H1 Data Analysis Model



HERA delivered $e^{\pm}p$ collisions 1992-2007 and the H1 Collaboration collected 0.5 fb^{-1} of data, ~ 10^9 events

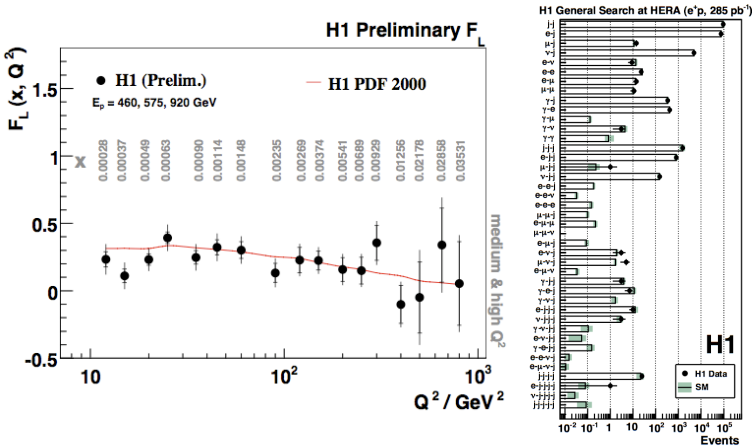
119.977679	129.534731	124.739135	176.316414
130.46875	135.839924	130.84732	168.289658
135.895502	149.510531	140.795689	120.686833
134.127052	140.495868	132.823819	206.138393
129.851598	137.880438	124.888856	189.675642
123.797241	131.84633	126.146789	202.496855
118.435374	130.691651	112.877008	140.366234
112.401212	121.561443	114.237637	125.298579
112.388488	128.496503	113.302591	192.223669
129.011813	138.880759	128.517198	108.701884
127.077465	139.289941	129.528986	127.406576
124.9785	135.363241	127.454638	129.669126
124.294035	133.242253	124.704841	244.567067
125.663717	135.159011	125.476984	169.271991
123.704853	127.612613	124.25382	170.401964
118.926697	122.818967	115.379664	134.970308
116.588208	121.798711	116.018173	323.148148
119.458869	124.788744	119.103839	204.736734
120.081967	124.847434	120.425321	289.50681
123.462329	127.367029	123.298233	287.632974
124.442179	128.115374	125.592252	362.764329
125.490169	128.448761	124.411031	382.978361
124.446597	128.898705	126.602473	358.369956

The raw data output from the detector is written to tape

Raw data transformed into DST format using Fortran based software, regular re-processing

Analysis level data format and software written in C++ and based on ROOT

Regular common data and MC production, calibrations and analysis performed using central computing resources



H1 publishes physics results