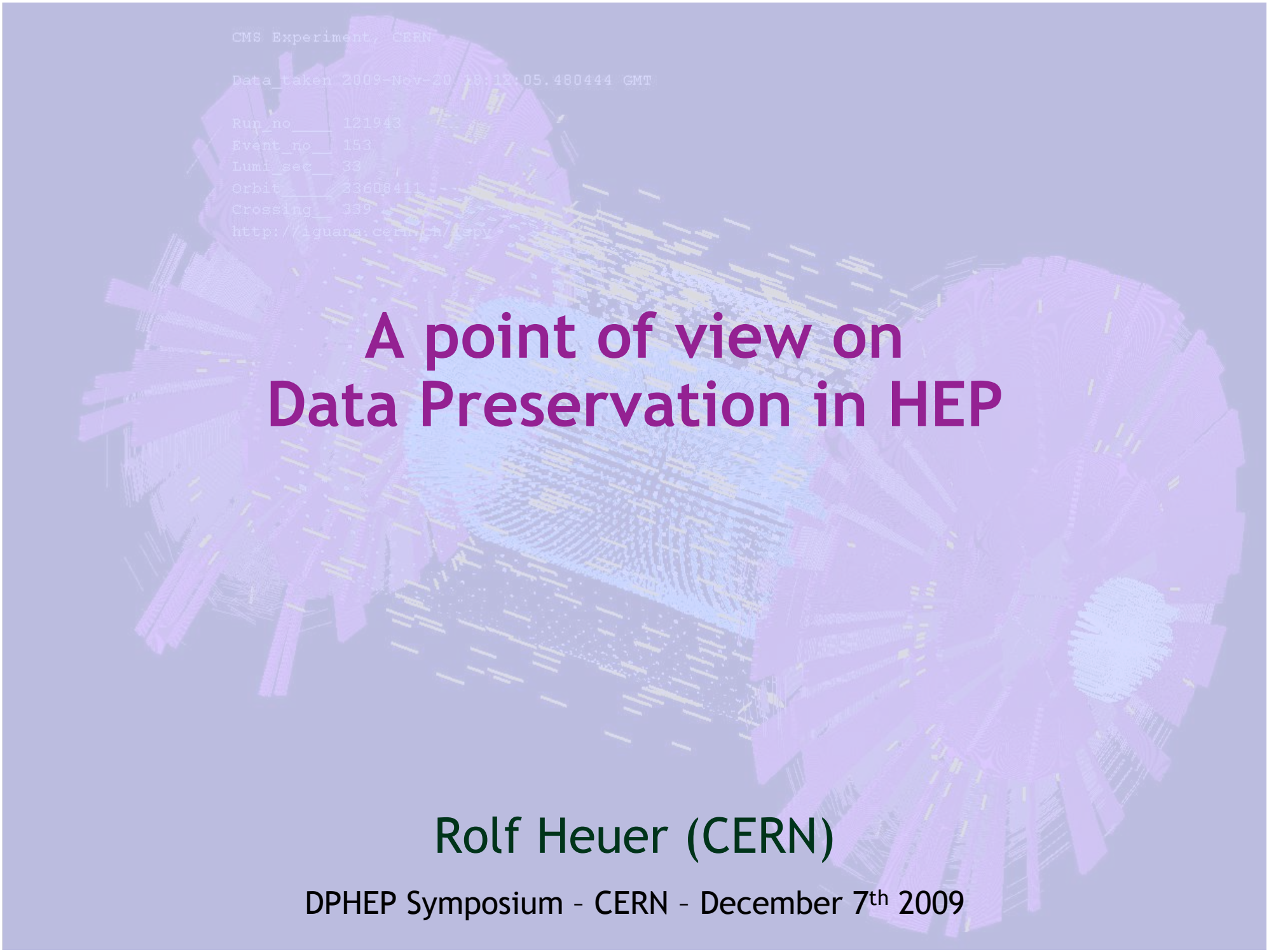
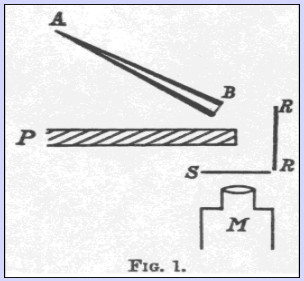


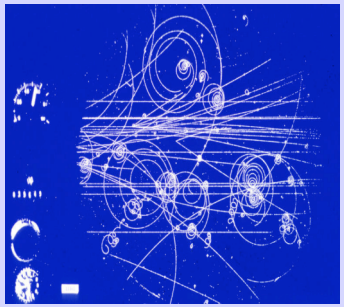
A point of view on Data Preservation in HEP

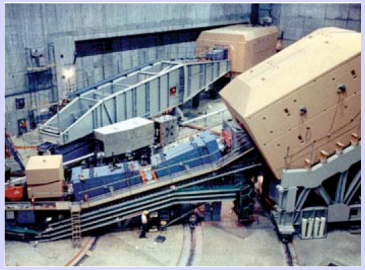
Rolf Heuer (CERN)

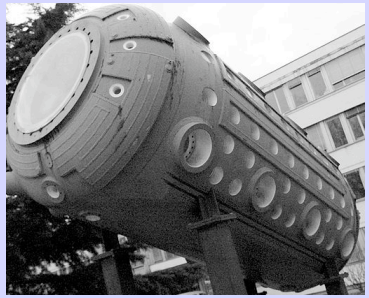
DPHEP Symposium – CERN – December 7th 2009

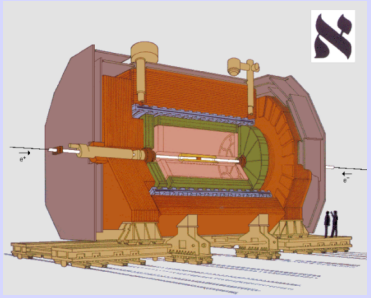
Knowledge is based on increasingly complex data

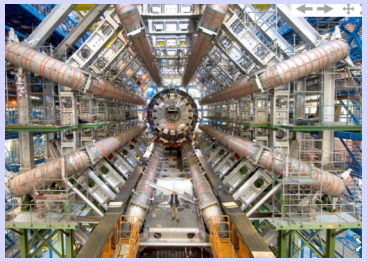


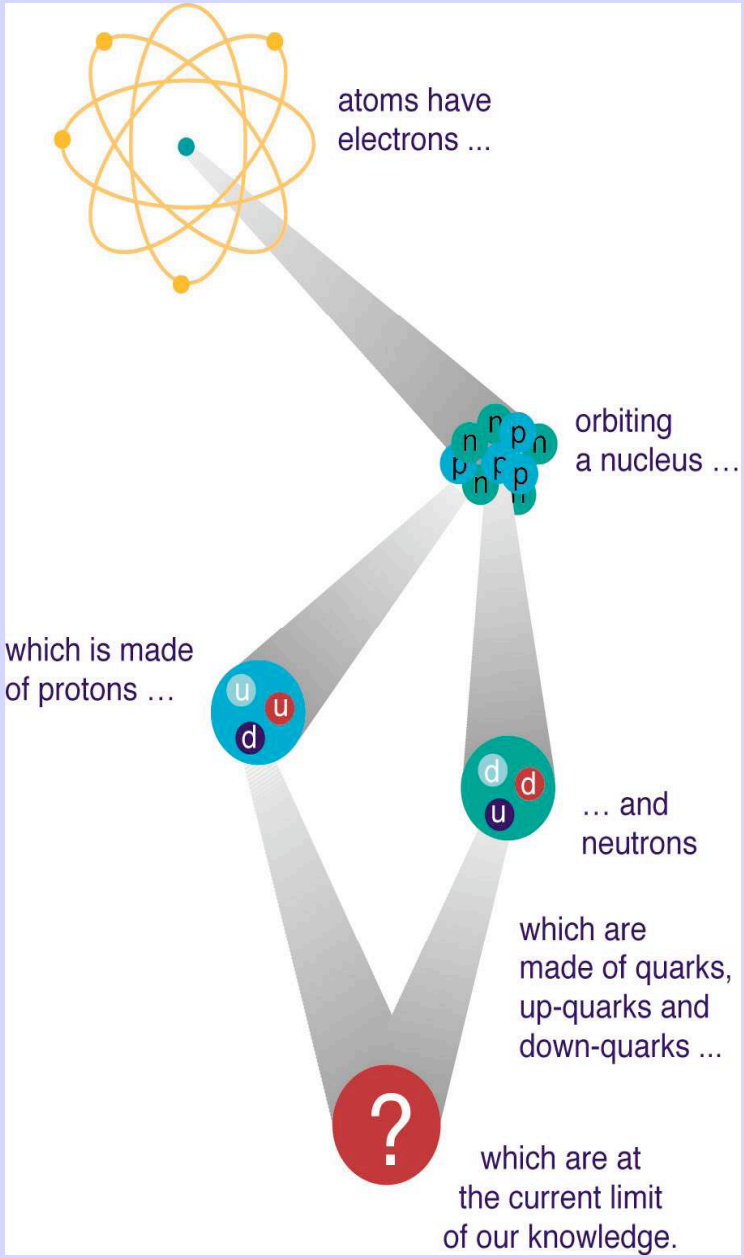




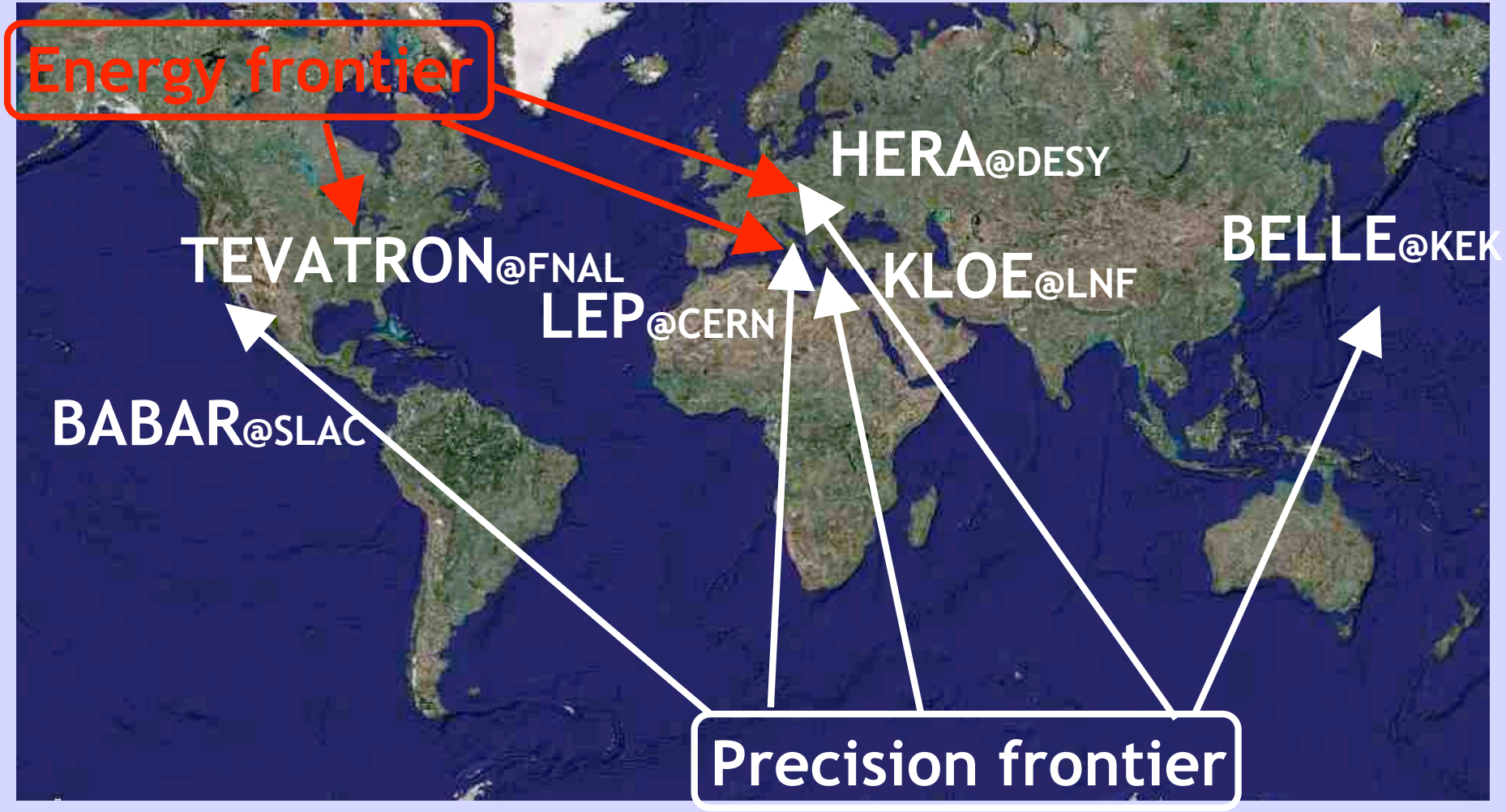








Several frontier HEP facilities

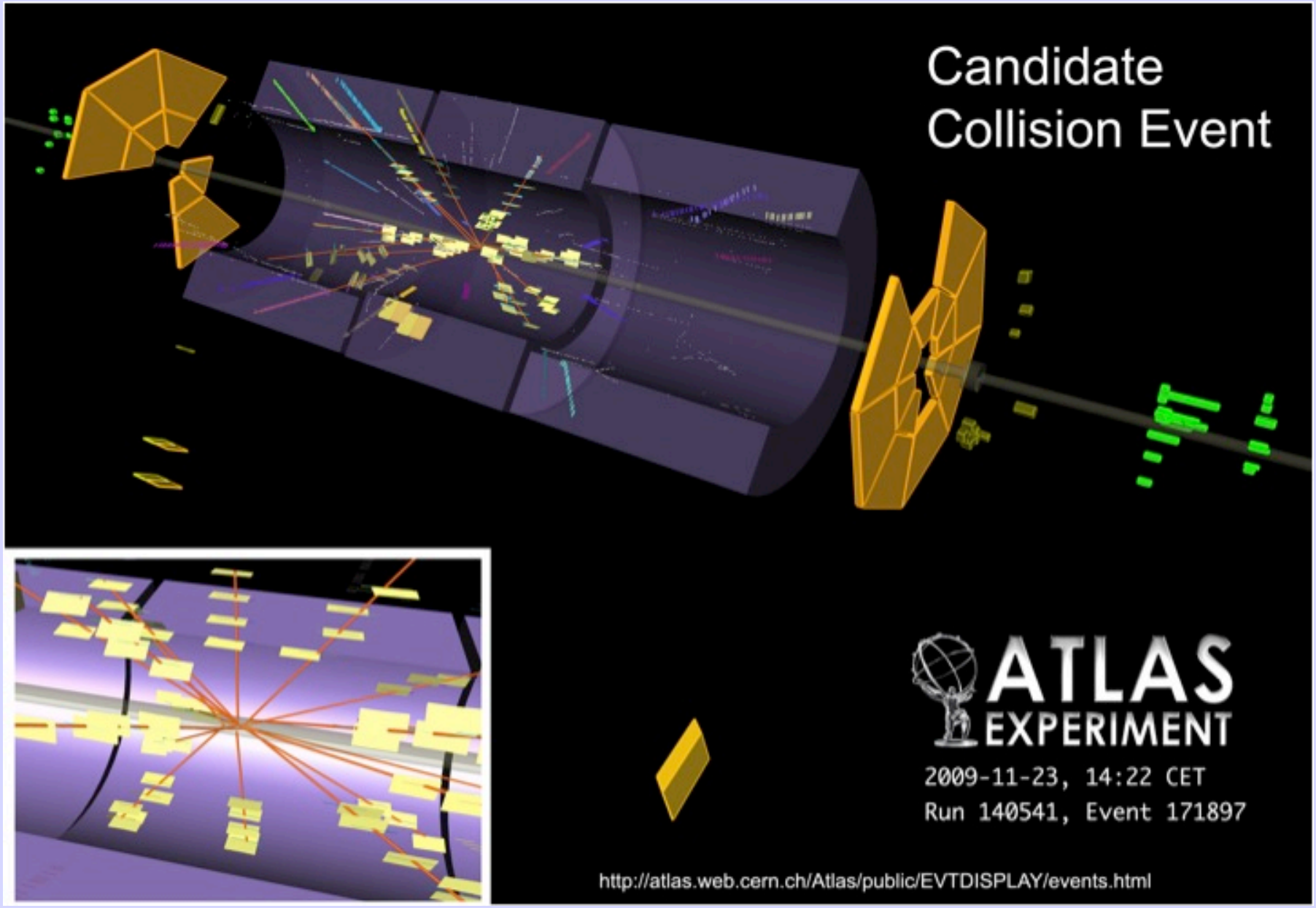


No data preservation strategy, large investments!

Some data sets are unique

Some other will only be (partially) superseded after years/decades

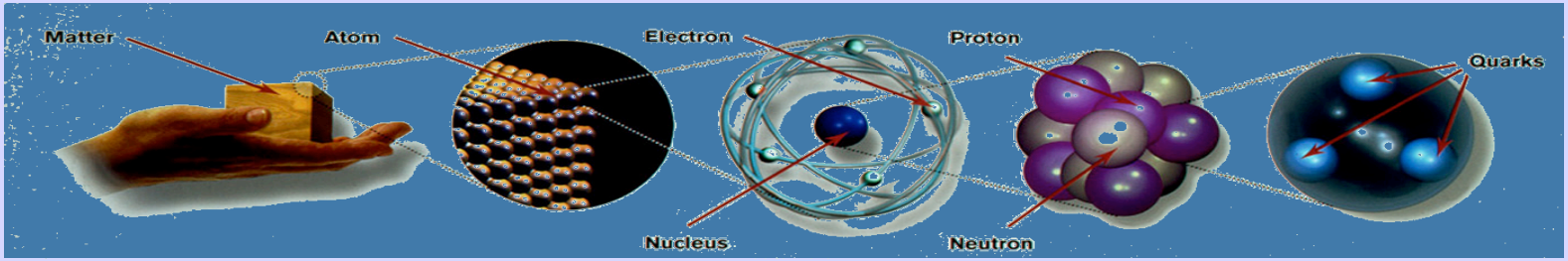
…and the future just started

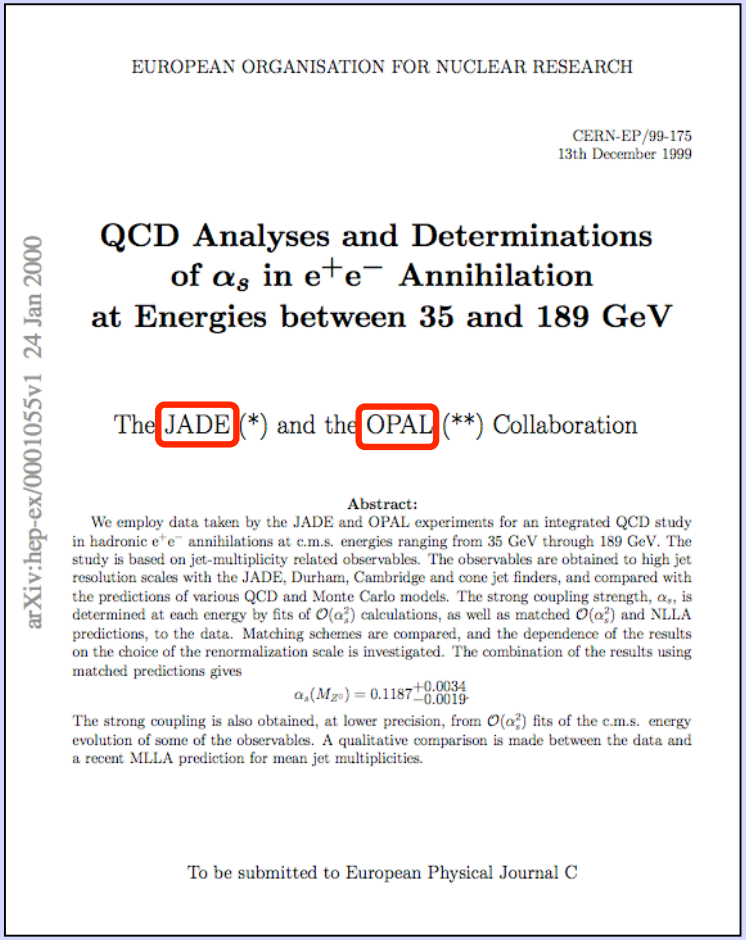


Why not  the data at the end of the experiment?

- **The physics program is completed, but…**
  - Knowledge evolves
  - Complexity of the experiments increases
  - The ability to reproduce/improve experiments diminishes => new paradigm!
- **Preserved data can improve the scientific return of the investement**
  - Improve the science/funding ratio [more science]

An example: the strong coupling saga





EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH

CERN-EP/99-175
13th December 1999

arXiv:hep-ex/0001055v1 24 Jan 2000

QCD Analyses and Determinations of α_s in e^+e^- Annihilation at Energies between 35 and 189 GeV

The JADE (*) and the OPAL (**) Collaboration

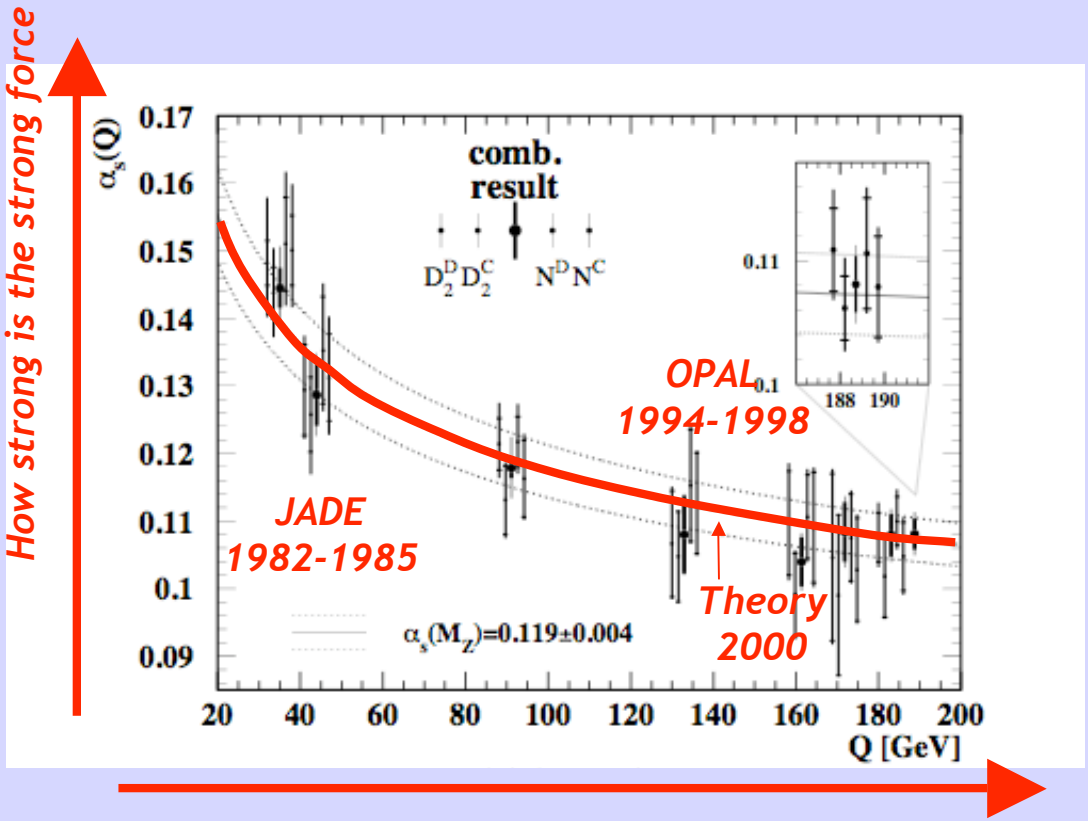
Abstract:

We employ data taken by the JADE and OPAL experiments for an integrated QCD study in hadronic e^+e^- annihilations at c.m.s. energies ranging from 35 GeV through 189 GeV. The study is based on jet-multiplicity related observables. The observables are obtained to high jet resolution scales with the JADE, Durham, Cambridge and cone jet finders, and compared with the predictions of various QCD and Monte Carlo models. The strong coupling strength, α_s, is determined at each energy by fits of $\mathcal{O}(\alpha_s^2)$ calculations, as well as matched $\mathcal{O}(\alpha_s^2)$ and NLLA predictions, to the data. Matching schemes are compared, and the dependence of the results on the choice of the renormalization scale is investigated. The combination of the results using matched predictions gives

$$\alpha_s(M_{Z^0}) = 0.1187^{+0.0034}_{-0.0019}.$$

The strong coupling is also obtained, at lower precision, from $\mathcal{O}(\alpha_s^2)$ fits of the c.m.s. energy evolution of some of the observables. A qualitative comparison is made between the data and a recent MLLA prediction for mean jet multiplicities.

To be submitted to European Physical Journal C



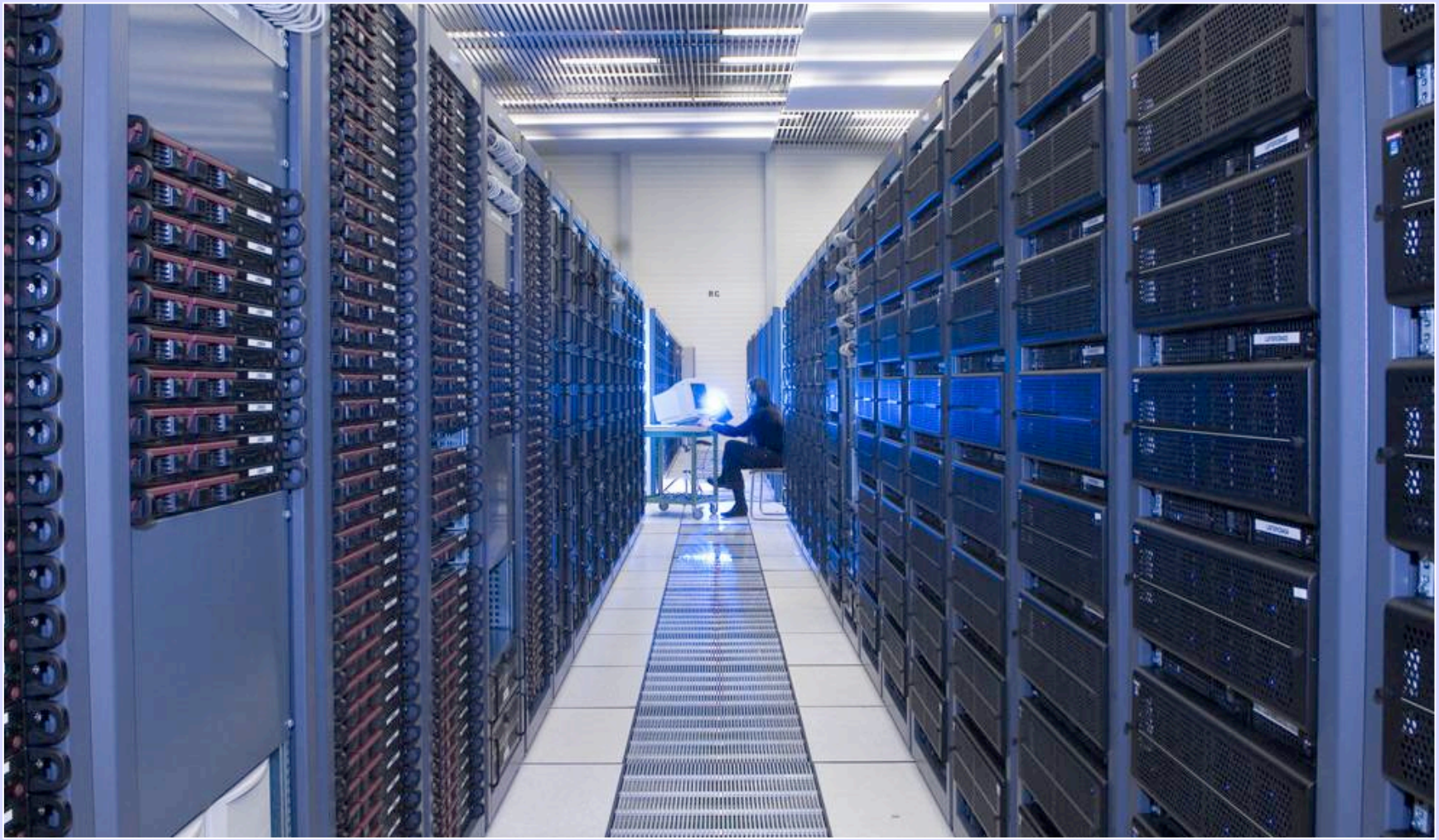
Accelerator energy = how close we study the quarks

… and it continues

Components of Data Preservation

Storing and processing data is feasible

Resources and careful planning are needed



Access and know-how are the real issues

Data Preservation in HEP

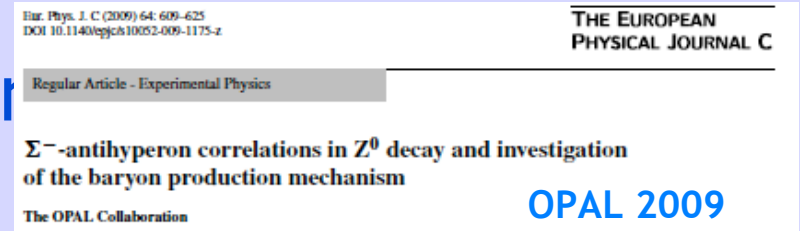
- Highly complex technology and data models
- No standards; *Ad-hoc* formats ; some regularities:
  - raw-> reconstructed-> reduced->analysis -> figures
- No tradition of reuse
  - as in Astronomy or Climate Science
  - Isolated examples but no coherent aproach.
- Data encapsulation generally not pursued
  - Final results depend on calibration constants, human knowledge, internal documentation and...oral tradition!
- Years of training to analyse data

Potential users

- Same researchers who took data, after the closure of the facility
- Researchers at similar facilities at same time
- Researchers of future facilities
- Re-interpretation by theoretical physicists
- Theoretical physicists testing future ideas

Goals of Data Preservation

- Prolong the physics progr
  - Precise data used for further measurements
- Combine experiments
  - Increase the precision
- Check new ideas, test discoveries
  - Compatibility, low/high energy etc.
- Outreach and Education
  - Increase the impact of HEP programs


Eur. Phys. J. C (2009) 64: 609–625
DOI 10.1140/epjc/s10052-009-1175-z

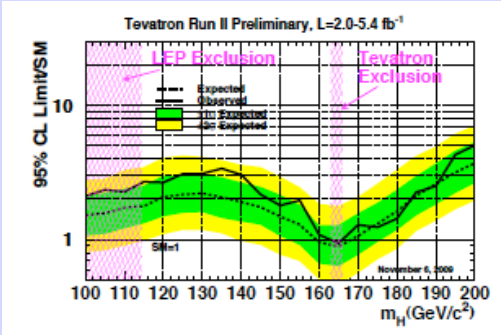
THE EUROPEAN PHYSICAL JOURNAL C

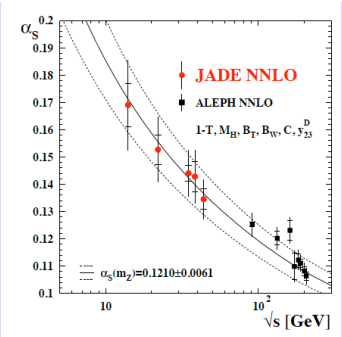
Regular Article - Experimental Physics

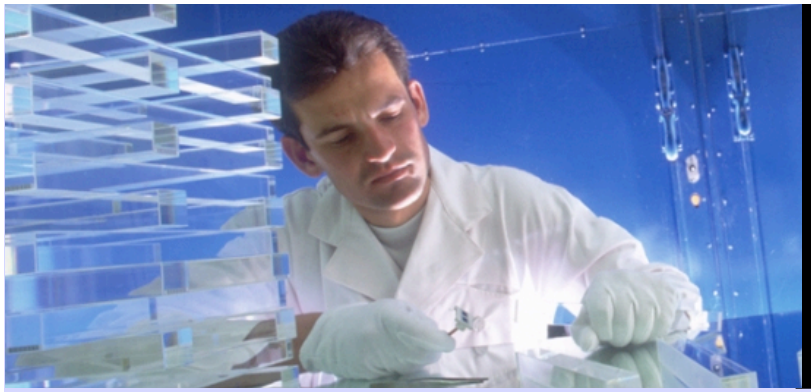
Σ^--antihyperon correlations in Z^0 decay and investigation of the baryon production mechanism

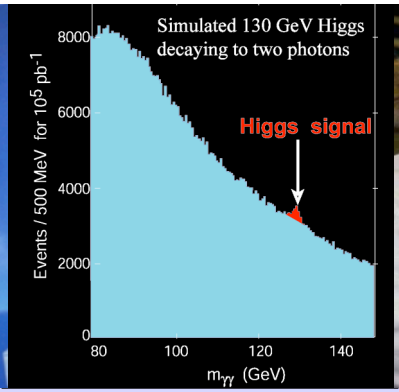
The OPAL Collaboration

OPAL 2009









After long preparation times and exciting physics:
Data preservation should be prepared as a part of the experimental programs

- **Need a strategy: coherent action, global initiative**
- **Need academic incentives and financial stimulus**