



PPA INFORMATION SUMMIT

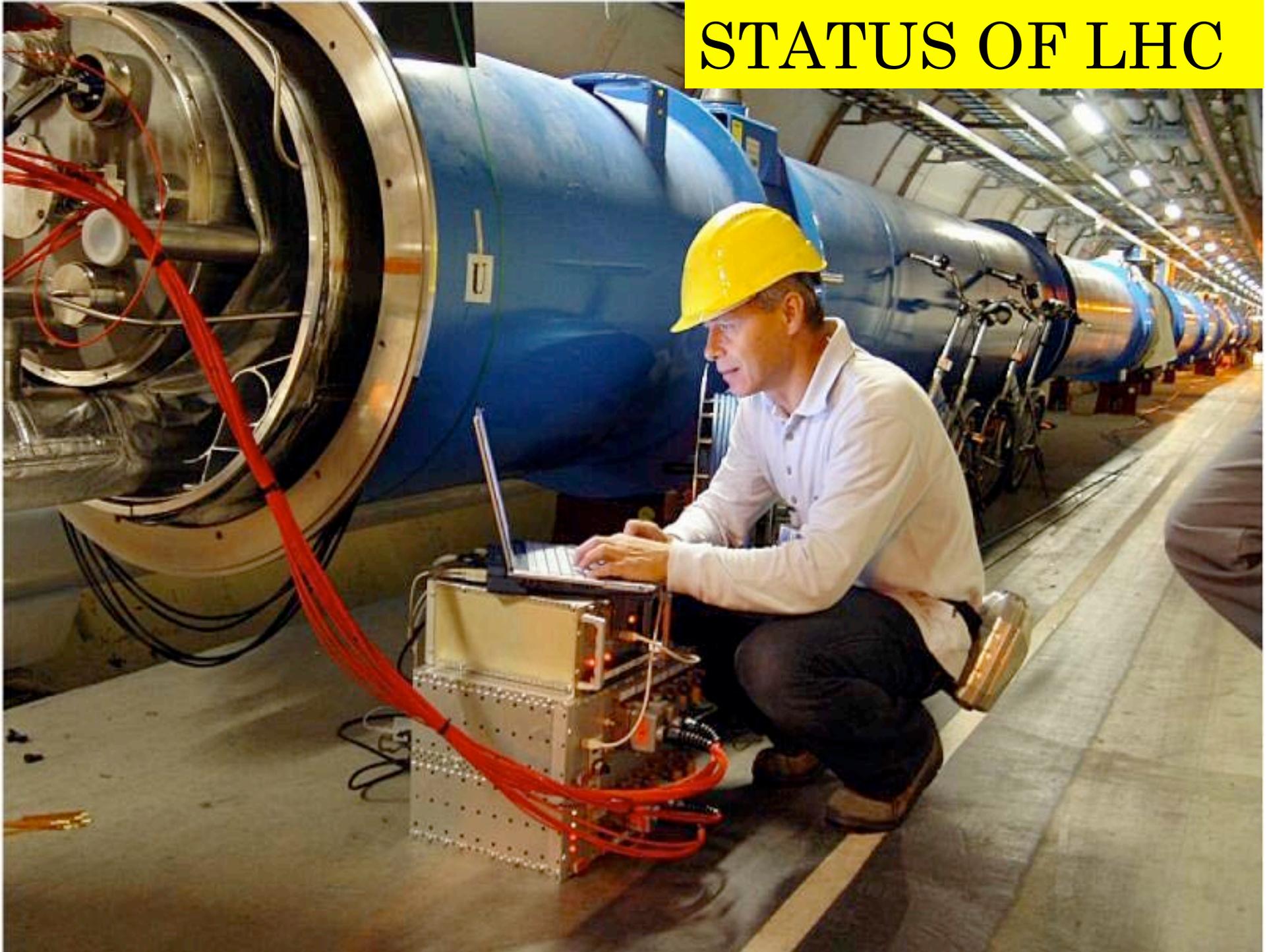
- ◆ Where does Experimental High Energy Physics goes in the next 5-10 years?
- ◆ How it will change the information needs researchers in that field will have ?

Future of Experimental HEP



- ◆ Long term future of HEP is the most debated theme in the community (together with the date of the 1st 10pb^{-1} @LHC).
- ◆ However the short term road map is reasonably defined: My best guess is that TEVATRON will run until in competition with LHC who will be then the ONLY frontier collider for the following 10 years.

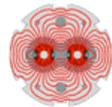
STATUS OF LHC





Status of LHC : Cryodipoles

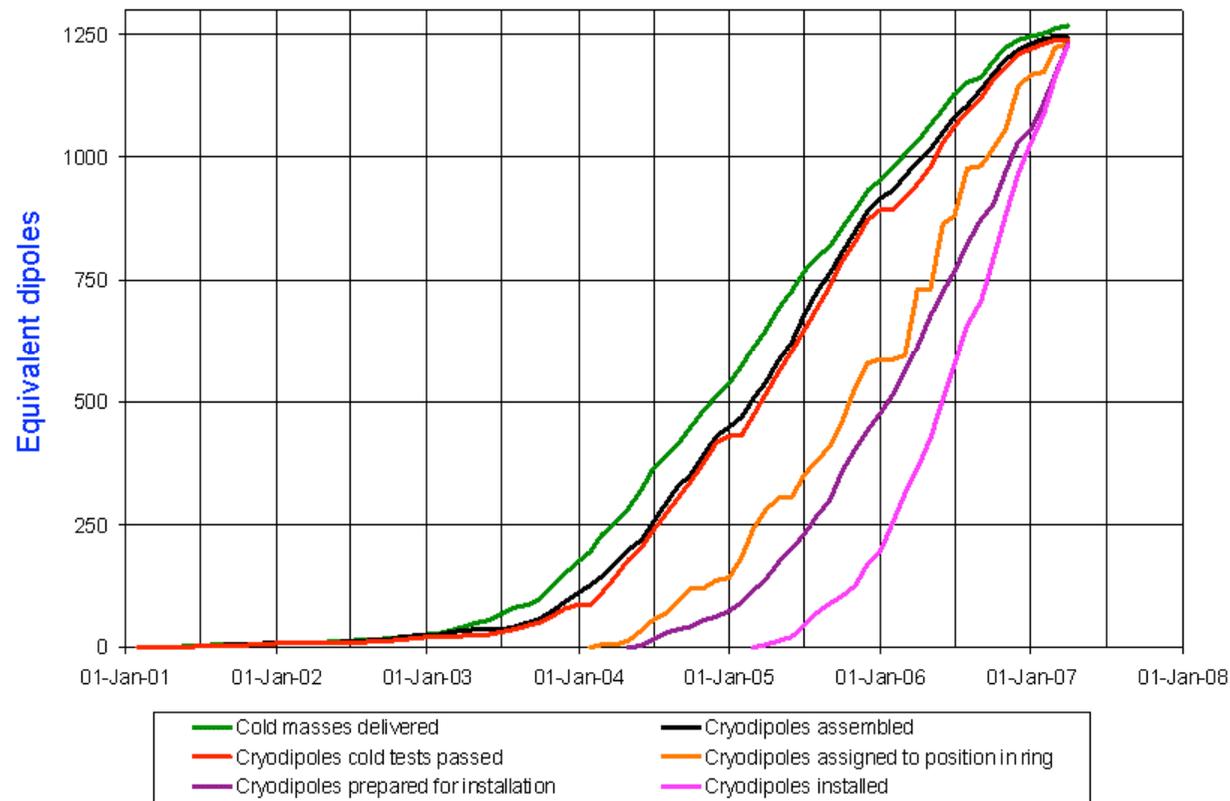
Install Last Dipole - end March



LHC Progress
Dashboard

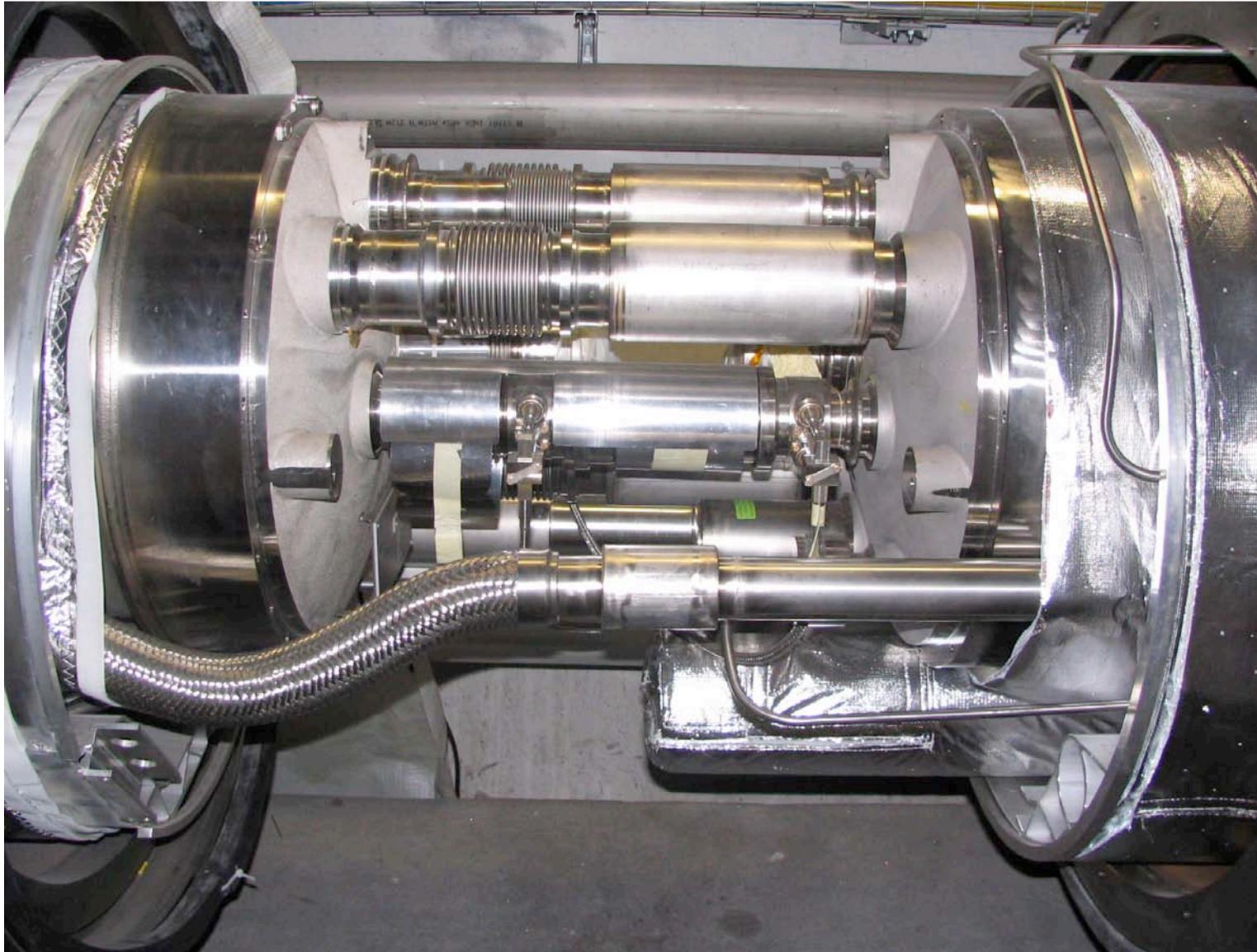


Cryodipole overview





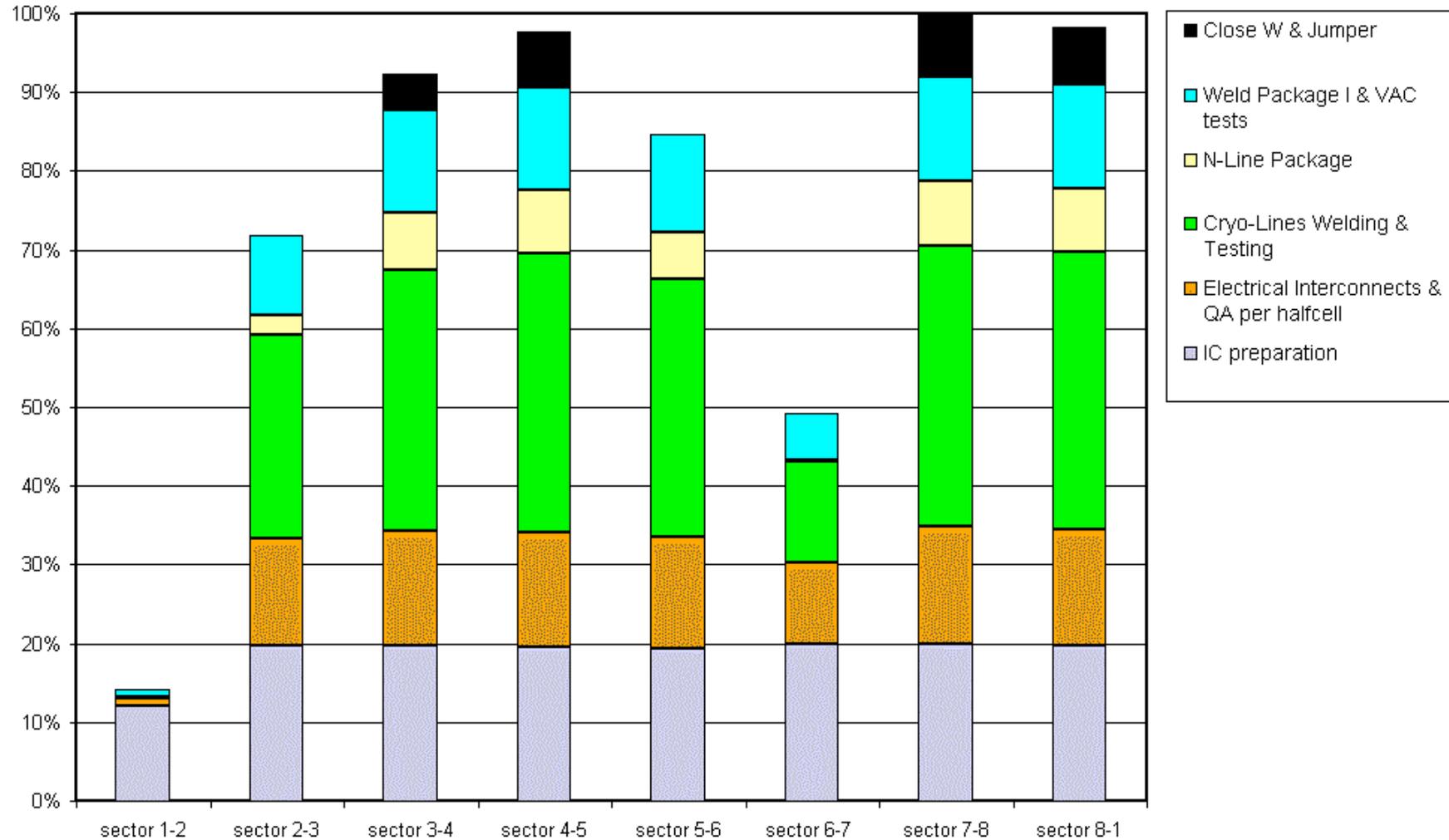
LHC Dipole-Dipole Interconnect





Interconnect Status

General Advancement of Interconnects per Sector 16-Apr-2007

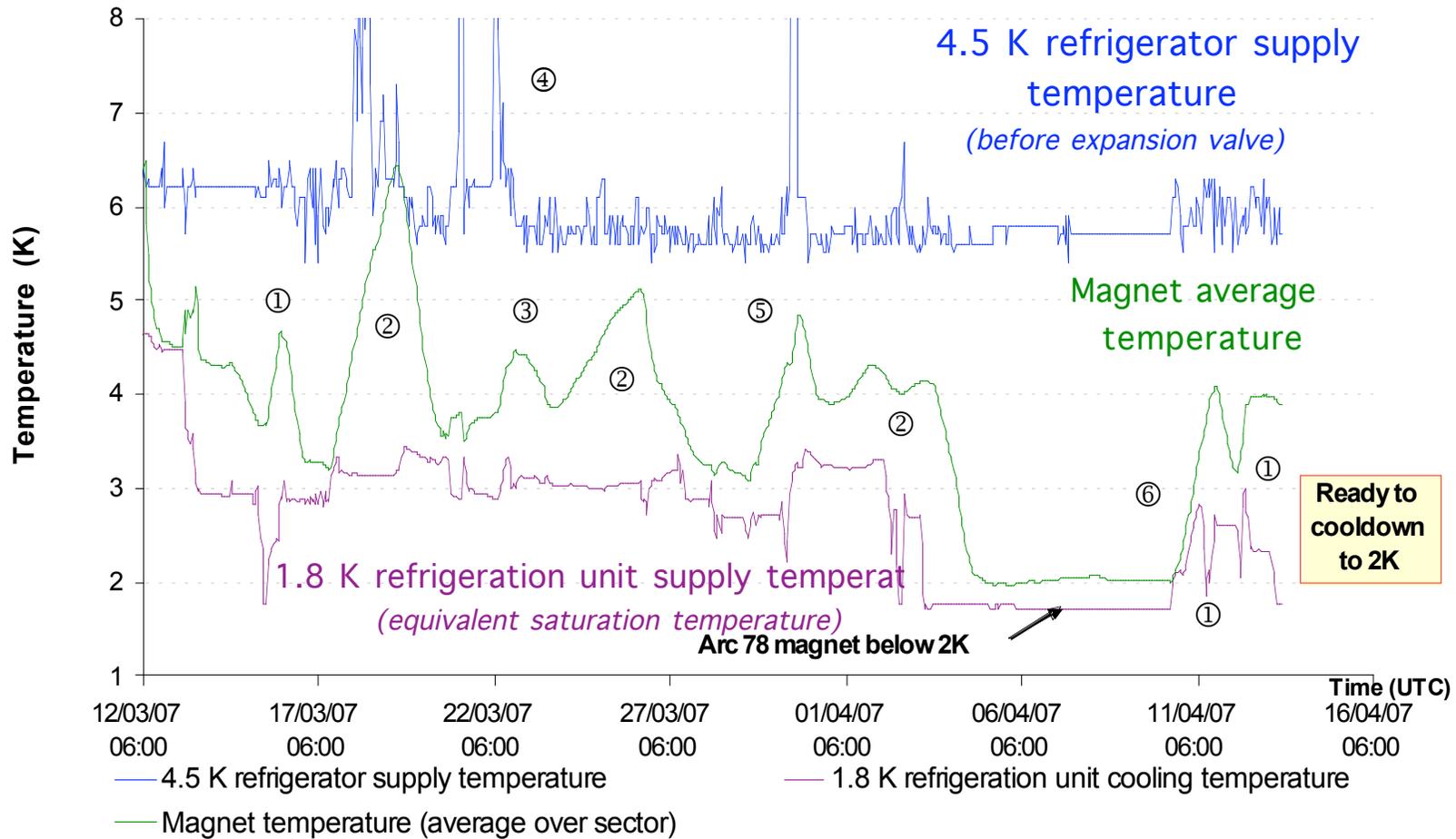
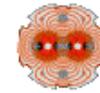




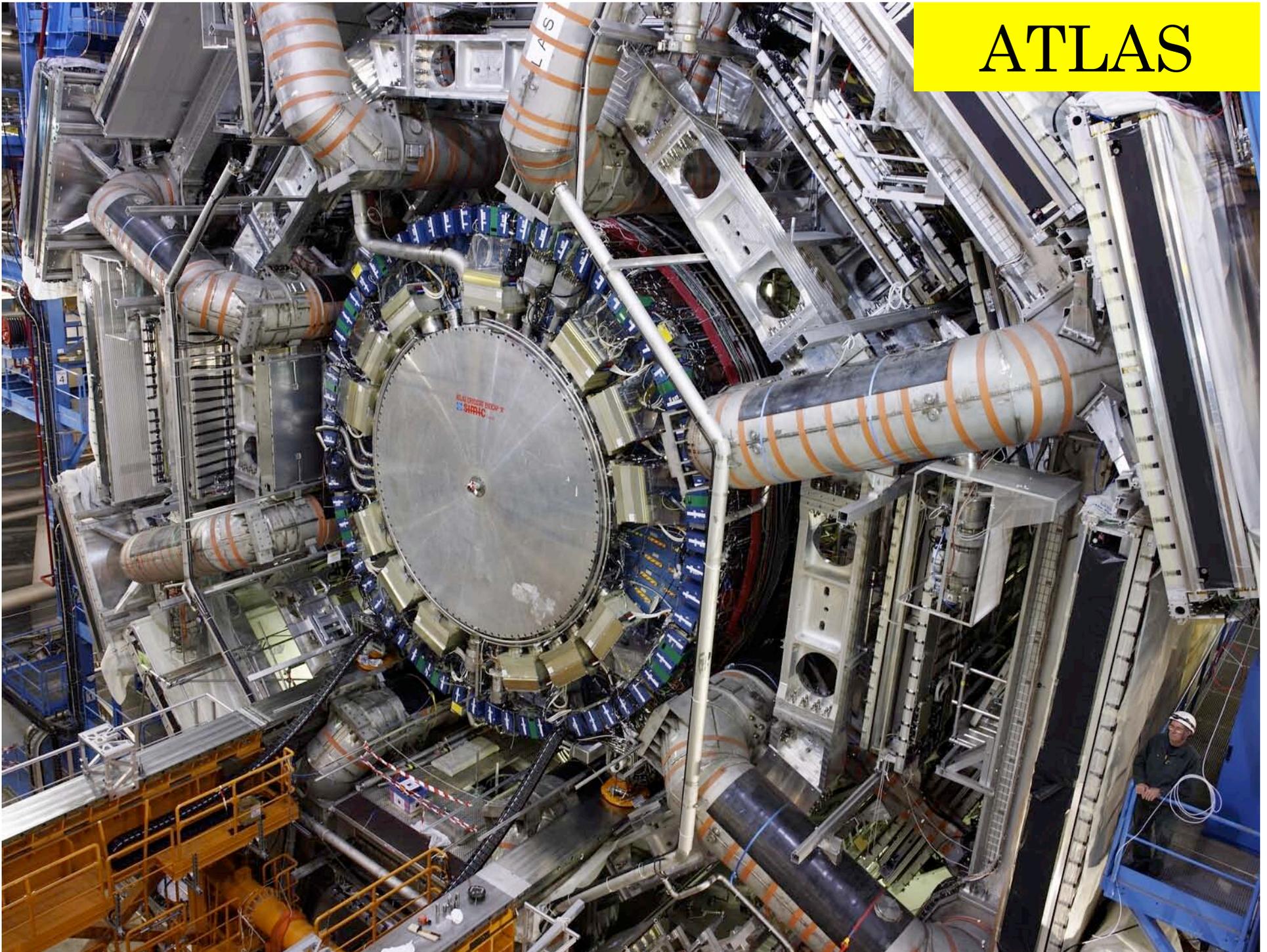
Sector 7-8 Cooldown



LHC sector 78 - First cooldown - Phase 4.5 K to 1.9 K

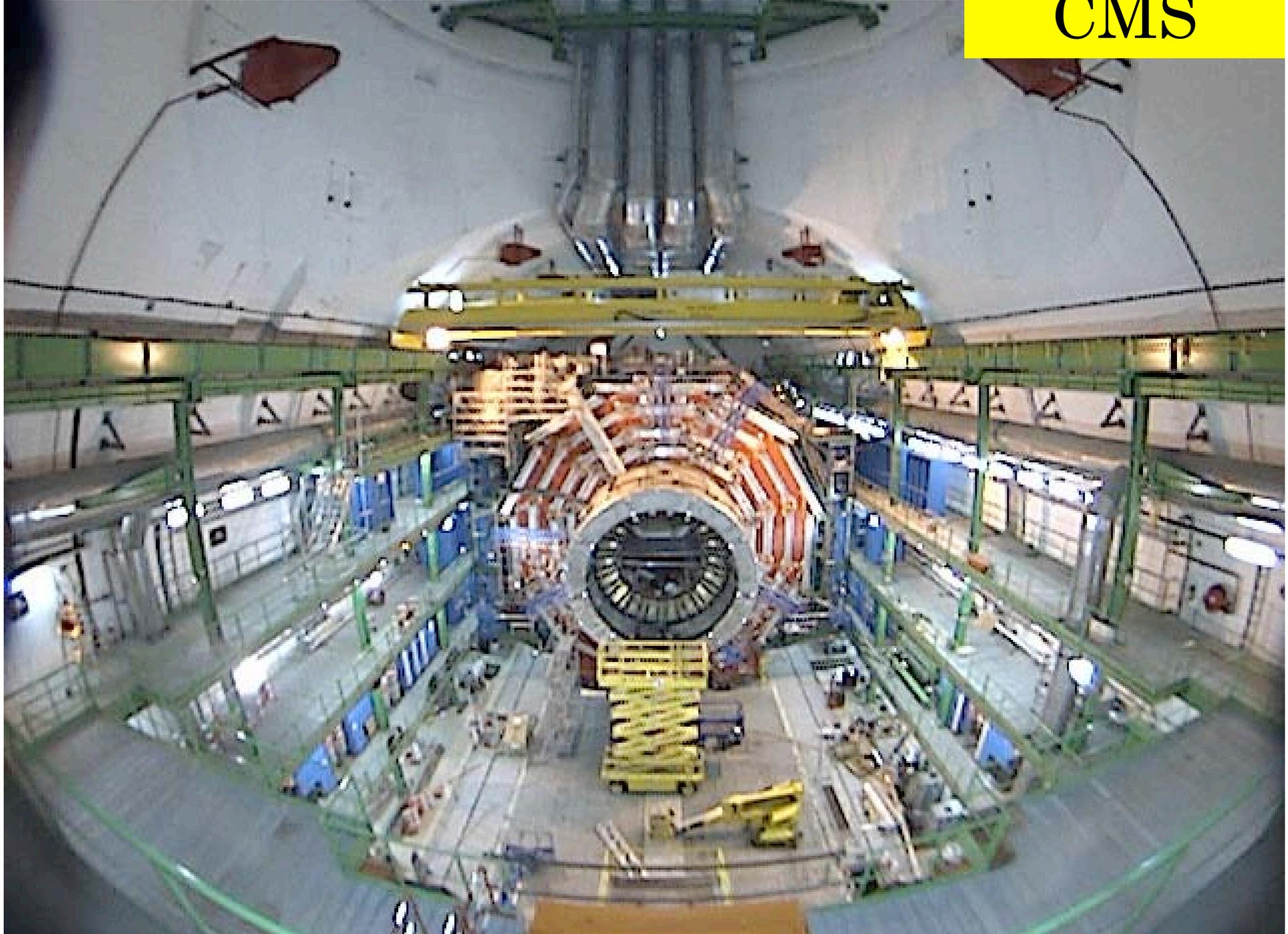


ATLAS



cmseye07 2007-04-09 19:15:58

CMS



Information needs



- ◆ Changes in the information needs are influenced more by new opportunities driven by technology than by changes in the needs of researchers. **Today we can do more.....**
- ◆ In the last few years we became more globally informed, more used to get what we need in one or two clicks and much faster than we ever dreamed only 10 years ago.....



- ◆ however the available on-line information has now increased by orders of magnitude and we face today a different problem: how to reduce the level of signal to noise when looking for what we need.... I know that what I am looking for is there, but I do not know how to find it.
- ◆ How I am going to find **THIS** ?

Timely access to new preliminary results



- ◆ Papers published on peer-reviewed journal will stay as official Collaboration blessed final results.
- ◆ However preliminary results typically shown at Conferences are “published” much earlier than the journal paper. Often we need to **search for slides of talks**, that appear on the net only few days after the presentation. Especially major winter and summer conferences when HEP Collaboration traditionally update their results.
- ◆ This should be an “almost online” service

Searching for talks



- ◆ Becomes very common not only for timely reference for a “new preliminary” result but also for retrieving plots, figures to be reused when preparing a new talk.
- ◆ In this respect it is important to retrieve not only the .pdf but also the .ppt

All instances of a scientific result



- ◆ A scientific result will appear in many preliminary forms: internal notes of collaborations; slides shown in internal meetings; conference contributions (both slides and proceedings). It will then become public in its final form in a seminar, possibly recorded in multi-media format, and as a short letter to a journal (appearing first as a pre-print and then as a published article).
- ◆ Will it be possible to retrieve all instances of the same result in one click ?

Evolution in time of a “number”



- ◆ What was the top mass and its error in 2002 ?
- ◆ This has an interest that goes beyond the “history of physics” . Today one has to browse several papers in order to make the relevant plot.

Smart ranking



- ◆ Papers are ranked by number of citations. In some cases one wants to classify them on the number of **selected citations**.
- ◆ I have to give an experimental course on LHC and I want to give few “easy” references on BSM physics. What are the BSM “generalist” papers most cited as reference **in former undergraduate lectures** ?

Generic text search



- ◆ My student asks me how to compute the angular distribution of tau to mu gamma...I saw it few years ago in a paper but this is all what I know. How do I find it ?



DATA PRESERVATION
AND
PUBLICATION OF RAW
DATA in EXPERIMENTAL
HEP



HOW AM I GOING TO ANALYZE ALEPH DATA IN 2020 ?

The problem is NOT data and software
preservation .

The problem is the lack of a simple data model.

THE PROBLEM of HEP data preservation



- ◆ The HEP data model is a **highly complex data model** (from the start difficult to export to OA a` la astronomy)
- ◆ Raw data -> calibrated data -> skimmed data -> high-level objects
- ◆ **Final results depend on all the grey-literature on constants, human knowledge, algorithms which are needed for each pass**
- ◆ Experiment lifetimes > computing environment lifetimes. Many migrations within the lifetime of an experiment (in this sense preservation is not an issue !)

Lesson learned from LEP



- ◆ **Apart from publication of numbers or tables, no real OA Either little useful or little usable (with small exceptions): continuous need for additional knowledge, difficult to encode and store.**
- ◆ **Regardless of community openness in pre-printing, wide-spreading of preliminary results at conferences and insider information, little priority on OA bringing to partial failures of LEP data archiving for the "general" public.**
- ◆ **Need force-majeure (Discovery at LHC of something we should have seen at LEP?) to access data again.**
- ◆ **Final results (containing additional unpublished information) but also high-level objects have been already combined (LEP Electroweak vs LEP Higgs)**

The "Parallel way" to archiving and publishing data



- ◆ In addition to internal data models, elaborate a **parallel format** for useful and usable high-level objects
- ◆ Publish high-level objects behind each scientific paper (after a time lapse?)
- ◆ Publish all high-level objects after end of collaboration
- ◆ Address issues of accountability, reproducibility of results, "careless discovers", "careless measurements"

A possible R&D program



- ◆ **Use LEP or Tevatron or BABAR as a case study for information retrieval to better assess the different methods**
- ◆ **Define some high-level object to make a OA-based analysis possible for an "external" but "motivated" researcher of the field**
- ◆ **Propose strategies to define "parallel" high-level objects to be included in the LHC data model, that is not post-mortem but aim to make it part of the data-model designing process. This is very timely.**
- ◆ **Imagine solutions to expand digital-library records of experimental results to include the OA data behind the results**
- ◆ **Initiate a discussion on priority issues and time-delays in making these "parallel" high-level objects available. This is very timely.**