



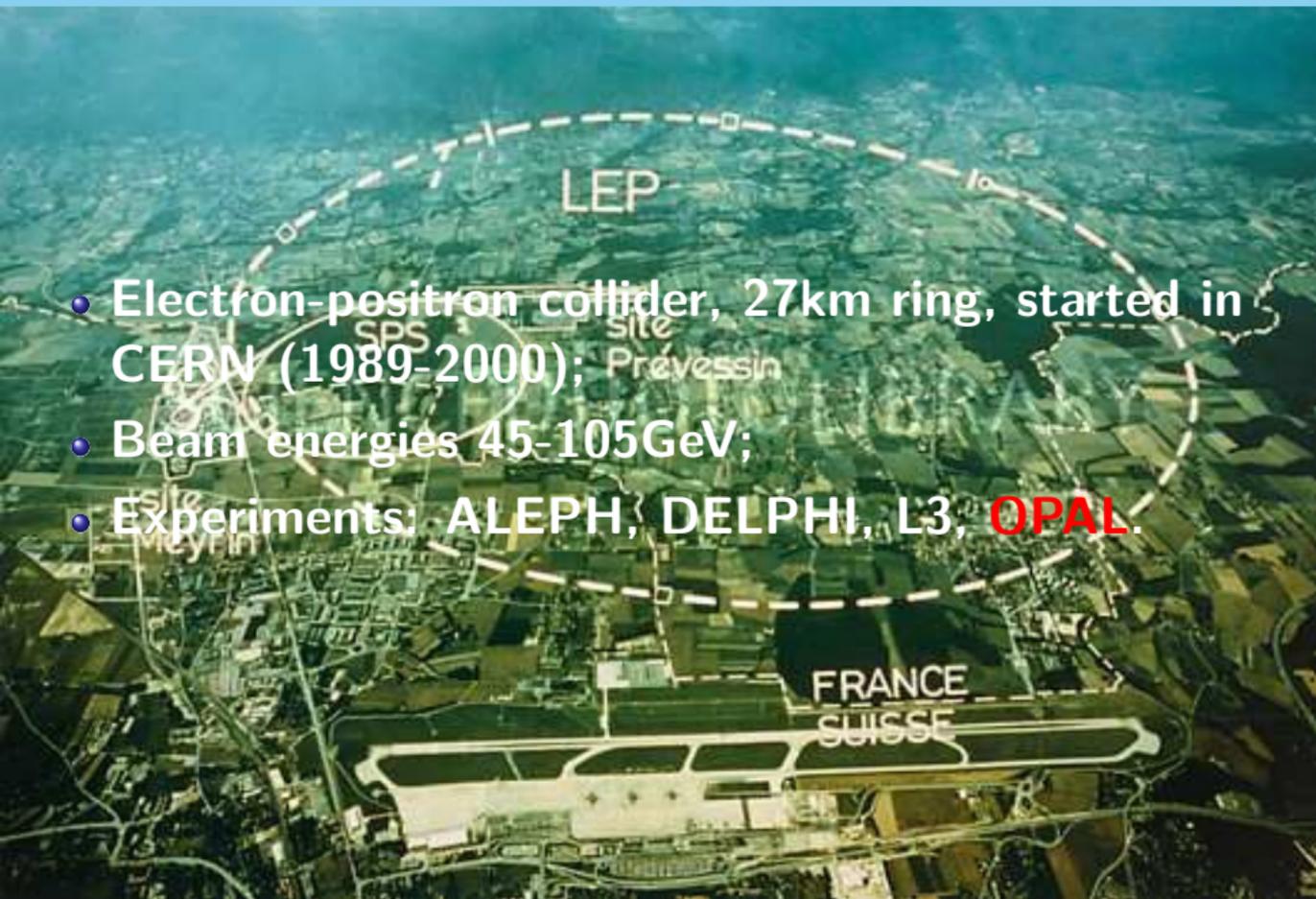
The OPAL long term data preservation project in Max-Planck Institute für Physik and more

Andrii Verbytskyi

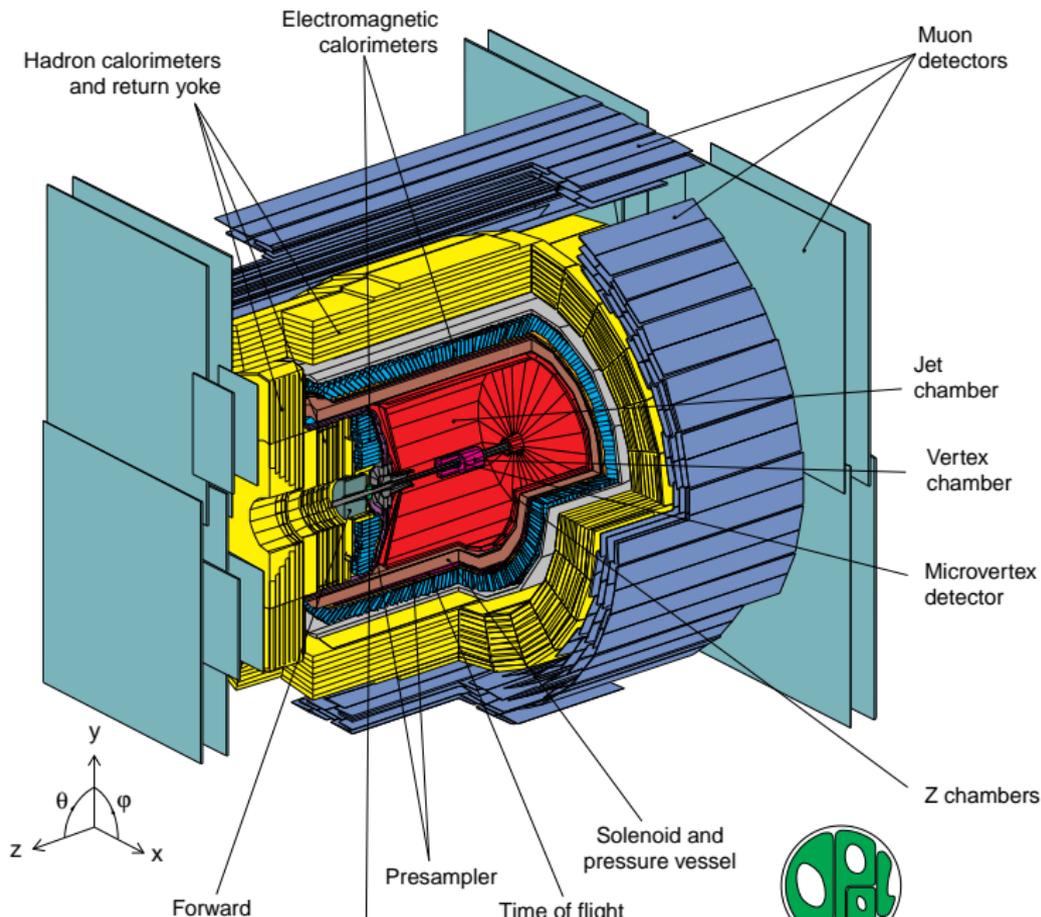
2nd DPHEP Collaboration Meeting
Geneva,
March 14, 2017

LEP

- Electron-positron collider, 27km ring, started in CERN (1989-2000);
- Beam energies 45-105GeV;
- Experiments: ALEPH, DELPHI, L3, **OPAL**.



OPAL



OPAL and data preservation

OPAL@LEP reminder:

- Precision QCD;
- Electroweak physics;
- Flavour physics;
- Searches (also Higgs)...

Motivation for data preservation:

- Future data (re-)analysis with new models and new approaches.
- Modelling for the future experiments.

- 5M events.
- 400k hadronic events at LEP-II (91-210GeV)

MPP DP model for OPAL

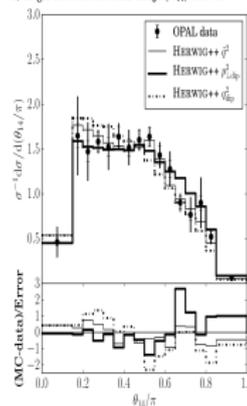
Data preservation is about new and interested results with old data. In addition the Data Preservation experience with OPAL has an extreme importance on itself.

In our model we describe ingredients and tools:

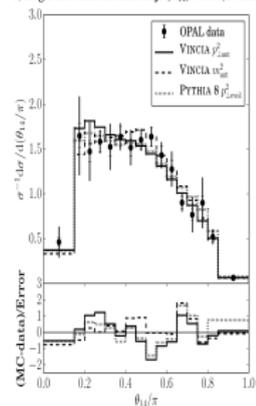
- Data bits
- Software
- But in the end we are interested in **physics** .
- Experiment documentation
- DP policies and documentation

Main idea: enable physics and make it doable with modern methods in modern environments with minimal effort.

a) Angle between 1st and 4th jet, θ_{14} , HERWIG++



b) Angle between 1st and 4th jet, θ_{14} , VINCIA, PYTHIA 8



Eur.Phys.J. C75 (2015)
no.12, 571

MPP DP model for documentation and policies

- OPAL publications are available in InSpire, journals, arXiv or in CERN.
- The non-digital documentation is preserved in CERN.

MPP DP model for OPAL data bits

- OPAL data are stored in multiple copies in CERN and in MPCDF on locally accessible tapes and in disk pool.
- Access via multiple protocols with grid tools worldwide to disk pools.
- Straightforward procedure to add new (MC) samples for MPCDF.

OPAL data in MPP: Bits statistics

Data are ROOT/PAW ntuples, ASCII files, etc. Particle level NLO Monte Carlo is available for multiple generators.

	MPCDF
Files:	112k
Volume:	92TB (91.9 × 10¹² bytes)
Work area:	yes
Access:	Worldwide
Protocols:	Multiple, see list
Auth:	Grid certificate with OPAL VO membership



Available at:

- <gsidcap://grid-srm.rzg.mpg.de:22128/pnfs/rzg.mpg.de/data/opal>
- <grid-gftp2.rzg.mpg.de>
- <davs://grid-dav.rzg.mpg.de:2880//opal>
- ...

MPP model for software preservation

Key ideas:

- Rely on industry, not HEP-only standards.
- Enable integration and compatibility with new physics software, e.g. data bases and Monte Carlo generators.
- **Make software analysis-ready**

Not much to say about MPP specifically. . . But that is good!

- Huge and **successful** effort by Matthias Schröder to keep the code compatible with modern systems.
- High quality code.
- Tests to put software in VM were done.

OPAL software extension

One can go ZEUS/JADE way:

- Add modern MC generators and conversation utility for modern format. Easy: OPAL uses ASCII based HEPEVT-like format.
- Add virtualisation. Things are working, but no image so far. Biggest problem is event display.

Not completed, but one can imagine how things should look like.

MPP Data Preservation summary

- Huge work was done to preserve OPAL data and software.
- The data in ntuples is accessible in MPCDF and is **analysis-ready**.
- Virtualisation is on the way in MPP (and CERN).
- So far MPP is involved mainly in the analysis
 - Jet physics.
 - Collective phenomena.

OPAL Data Preservation applications

- Hadronisation/non-perturbative effects.
- Fundamental questions of QCD.
- Exotic hadronic states.
- Flavour physics.
- ...

- Quality of code is important.
- Compatibility with modern experiments!
- Involvement of experts (M. Schröder)
- Promotion: an anecdotal case of spending two conference evenings advertising preserved LEP data and its availability for re-analysis.

Some opinions

Some thoughts on DP in general

Software:

- Software still requires some efforts. Can we join efforts? Something like software repository with a proper package management system would be a nice idea. The technologies to do that (i.e. rpm/yum/etc) already 20+ years on the market, but still not in HEP.

Some thoughts on DP in general

DP plans:

- Abstraction is good, but there is a huge difference between number of abstract concepts and actual examples of implementation.

Some thoughts on DP in general

Some technical and sociological problems:

- Missing or obscure technical documentation.
- Some very specific details of physics of old experiments can be forgotten.
- The terminology in 1980, 2000 and now is different, old methods are renamed, etc. This slows down communication and understanding of documentation.

Some thoughts on DP in general

More sociological problems:

- Manpower is not actually that severe a problem per se. But the concentration is. With low spatial concentration of knowledge, there are few people that can be asked about specific things, discussions via mail exchange are not that productive. The learning curve becomes too long. Should bigger laboratories like CERN and DESY be more involved? Some events for young physicists? Open data for older experiments?
- Long duration of any potential analysis because of involvement of members in other projects. The duration of any analysis more than 12 months implies that it cannot be advertised as a master or bachelor thesis. This rejects a very important source of manpower or requires more intensive supervision than for the running experiments. Could multiple supervisors and multiple master students per project be a solution? Thought the publication of the result should be mandatory.

Some thoughts on DP in general

More sociological problems:

- Doesn't matter how good are new technologies, it is still hard to convince experienced people to use them. There is a fear the efforts to learn these things will be wasted. Good examples are Grid tools or PROOF.

Some thoughts on DP in general

- Often the investments of manpower in earliest and not very precise measurements are much bigger than investments in the DP or even in the later results. Emphasise of DP importance could fix this imbalance and make later/after-end-of-datataking precise analyses easier. Analysis preservation addresses the issue.
- A bit of sociological problem.