



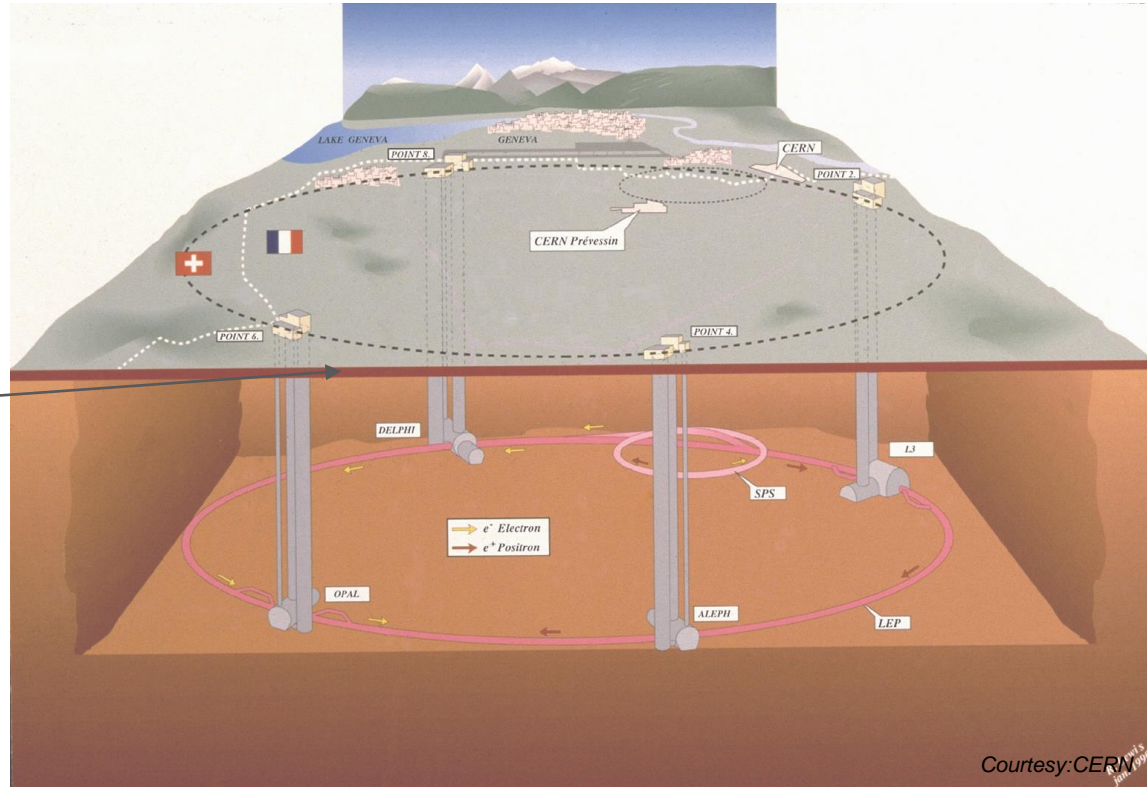
Data preservation and open data status in DELPHI

- What was LEP again?
- Status of DELPHI data and software
 - Bit preservation
 - Software preservation
 - Documentation
 - Analysis preservation
- Lessons learned
- Work in progress
- Access policies and open data

What was LEP again ?

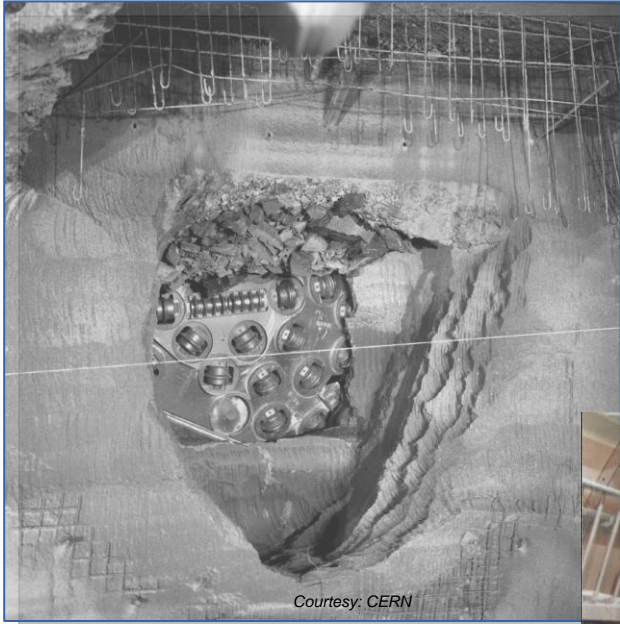


- e^+e^- collider at CERN
 - Data taking 1989-2000
 - Energy 90-209 GeV
- 4 experiments:
 - ALEPH
 - DELPHI
 - L3
 - OPAL
- **Largest circular lepton collider so far**



Courtesy: CERN

LEP: some impressions



Courtesy: CERN

LEP Tunnel
break through,
1986

LEP RF galleries

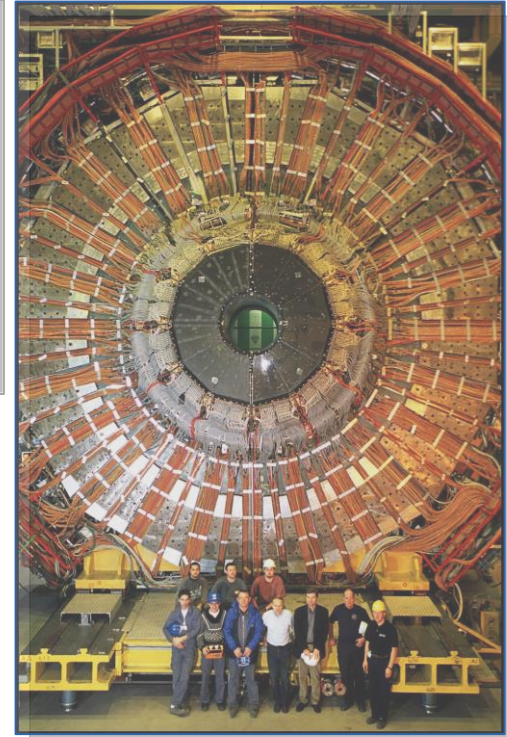


Courtesy: W. Liebig



Courtesy: W. Liebig

LEP accelerator



Courtesy: CERN

DELPHI end-cap before
dismantling



DELPHI: Bit preservation



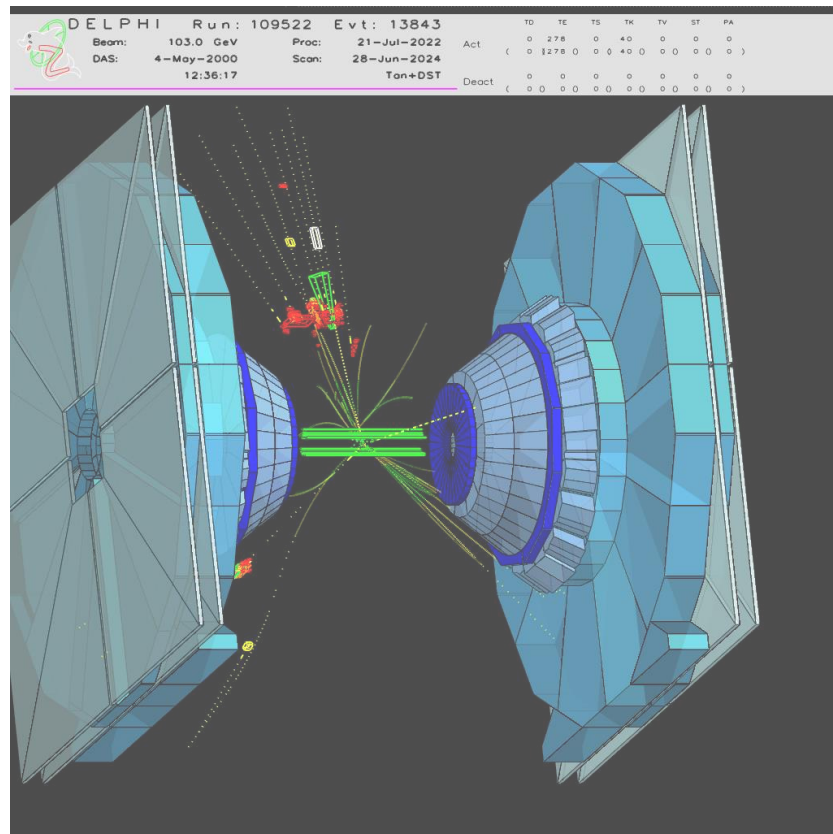
- DELPHI data on CTA at CERN
 - 2 copies on tape
 - Raw data, detector data bases and simulations
- Additional copies on EOS
 - Direct file based access
 - Accessible from outside CERN
- External copies
 - E.g. DELPHI in Santander (Spain)



DELPHI software preservation

What is preserved ?

- **Analysis** frame work
 - User-facing
 - Loops over events (data or simulated)
 - Runs specific modules on them and
 - Executes user defined analysis code
- **Simulation**
 - Build-in event generation or external files
 - Simulation of the detector response
- **Reconstruction**
 - Reconstruct event based on detector response
 - Raw data or simulated raw data
- **Visualisation**
 - Event display, for interactive inspection and analysis of events





Software preservation: challenges identified during the last workshop

- CERNLIB dependency
 - Extensive use of ZEBRA
 - Last official version 2006b
 - **32bit** only
- Fortran compiler support
 - g77 replaced by gfortran
- X11 Motif widgets
 - affects CERNLIB
- X11 itself
 - Wayland
- Commercial 3d libraries
 - Event display



Automation: building and testing using gitlab CI

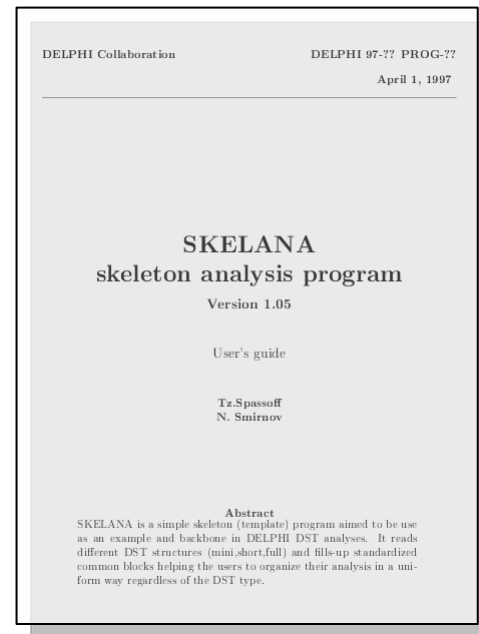
The screenshot displays the GitLab CI Pipelines interface for a project named 'delphi'. The breadcrumb navigation shows 'delphi / deployment / Pipelines / #7579237'. The interface includes a top navigation bar with 'Pipeline', 'Needs', 'Jobs 135', and 'Tests 0'. A search bar is present below the navigation. On the left, a sidebar lists project navigation options: deployment, Pinned, Issues (0), Merge requests (1), Manage, Plan, Code, Build, Pipelines (selected), Jobs, Pipeline editor, Pipeline schedules, Test cases, Artifacts, Secure, Deploy, Operate, Monitor, and Analyze. The main content area shows a grid of pipeline jobs organized into columns representing stages: common, cernlib, dstana, ddb, idea, and delgra. Each job is represented by a green checkmark icon, indicating successful completion. The jobs are listed as follows:

Stage	Job Name	Status
common	common_al9_64	Success
cernlib	cernlib_al8_32	Success
	cernlib_al8_64	Success
	cernlib_al8_aarch64	Success
	cernlib_al9_32	Success
	cernlib_al9_64	Success
	cernlib_al9_aarch64	Success
	cernlib_cc7_32	Success
	cernlib_cc7_64	Success
	cernlib_debian12_64	Success
	cernlib_ubuntu18_64	Success
	cernlib_ubuntu20_64	Success
	cernlib_ubuntu22_64	Success
cernlib_ubuntu24_64	Success	
dstana	dstana_al8_32	Success
	dstana_al8_64	Success
	dstana_al8_aarch64	Success
	dstana_al9_32	Success
	dstana_al9_64	Success
	dstana_al9_aarch64	Success
	dstana_cc7_32	Success
	dstana_cc7_64	Success
	dstana_debian12_64	Success
	dstana_ubuntu18_64	Success
	dstana_ubuntu20_64	Success
	dstana_ubuntu22_64	Success
dstana_ubuntu24_64	Success	
ddb	ddb_al8_32	Success
	ddb_al8_64	Success
	ddb_al8_aarch64	Success
	ddb_al9_32	Success
	ddb_al9_64	Success
	ddb_al9_aarch64	Success
	ddb_cc7_32	Success
	ddb_cc7_64	Success
	ddb_debian12_64	Success
	ddb_ubuntu18_64	Success
	ddb_ubuntu20_64	Success
	ddb_ubuntu22_64	Success
ddb_ubuntu24_64	Success	
idea	idea_al8_32	Success
	idea_al8_64	Success
	idea_al8_aarch64	Success
	idea_al9_32	Success
	idea_al9_64	Success
	idea_al9_aarch64	Success
	idea_cc7_32	Success
	idea_cc7_64	Success
	idea_debian12_64	Success
	idea_ubuntu18_64	Success
	idea_ubuntu20_64	Success
	idea_ubuntu22_64	Success
idea_ubuntu24_64	Success	
delgra	delgra_al8_32	Success
	delgra_al8_64	Success
	delgra_al8_aarch64	Success
	delgra_al9_32	Success
	delgra_al9_64	Success
	delgra_al9_aarch64	Success
	delgra_cc7_32	Success
	delgra_cc7_64	Success
	delgra_debian12_64	Success
	delgra_ubuntu18_64	Success
	delgra_ubuntu20_64	Success
	delgra_ubuntu22_64	Success
delgra_ubuntu24_64	Success	



DELPHI Knowledge preservation

- Documentation - Manuals
 - CERN Document server (<https://cds.cern.ch>)
 - DELPHI web pages
- Papers, internal notes, theses, photos...
 - CDS (<https://cds.cern.ch>)
 - Inspire (<https://inspirehep.cern.ch>)
- Sample code, examples
 - <https://gitlab.cern.ch/delphi/examples>





DELPHI Analysis preservation

No general analysis preservation was done ...

- Some code and executables preserved, along with their output
 - Vanishing expertise on how to run these
 - Versioning issue: Not always obvious if the conserved code really matches to conserved output
 - May be useful for validation of re-compiled stacks
- **Missing documentation** and metadata on how to interpret these sets



Lessons learned

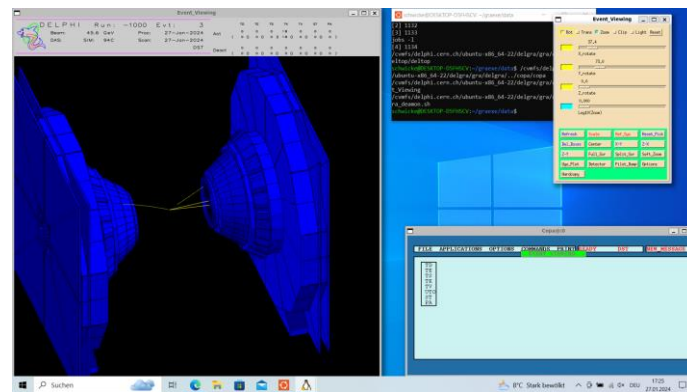
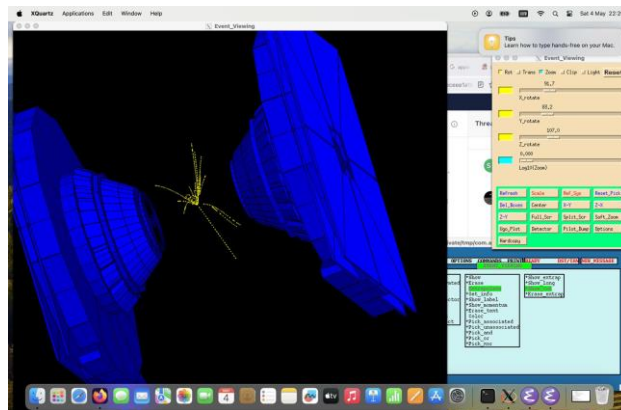


- Data preservation should be taking into account from the start
 - Software versioning and detector changes over the years
 - Analysis preservation
- Avoid commercial and closed source software
 - Companies disappear or get bought by others
 - No funds to pay license cost at long term
 - No guarantee that new builds can be made available
- Embrace new technologies
 - Containers allow to run old executables at least for a while
 - **Continuous integration and continuous testing** help a lot in long term maintenance



Work in progress

- **AARM64 support**
 - includes RaspberryPi (64bit) and Apple M* CPUs
- **DARWIN (MacOS) support**
 - CERNLIB OK, no issues known right now
 - No binaries, requires rebuilding from source
 - issues with OpenGL
 - Issues with 1992 short DST creation (crash), not understood
 - Debugging Fortran code on Apple is difficult
- **Windows**
 - No need for a generic build
 - Tested with WSL2, using Ubuntu 20.22 and CVMFS
- **More testing and validation needed**

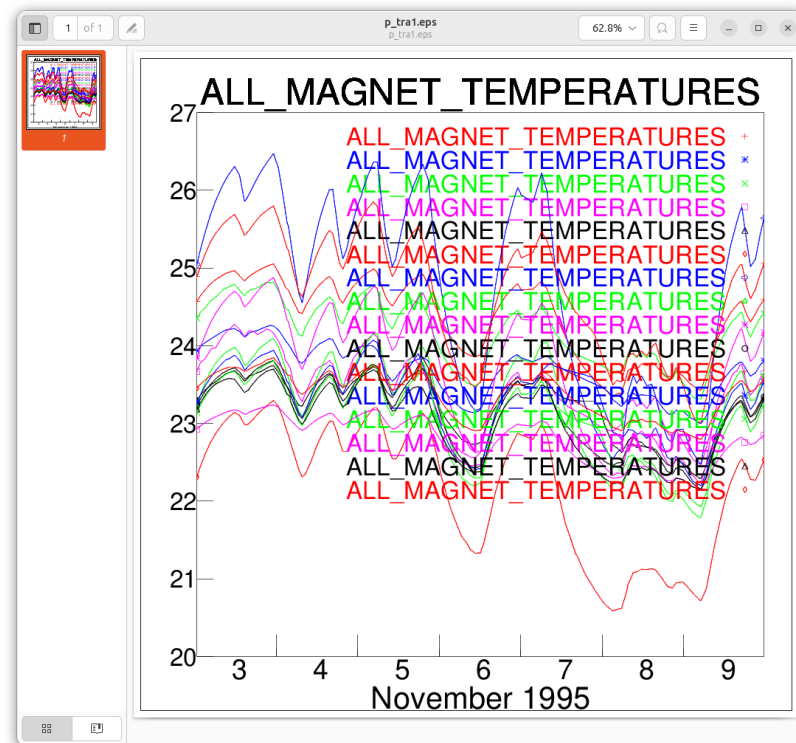




Work in progress

Recovery of database tools

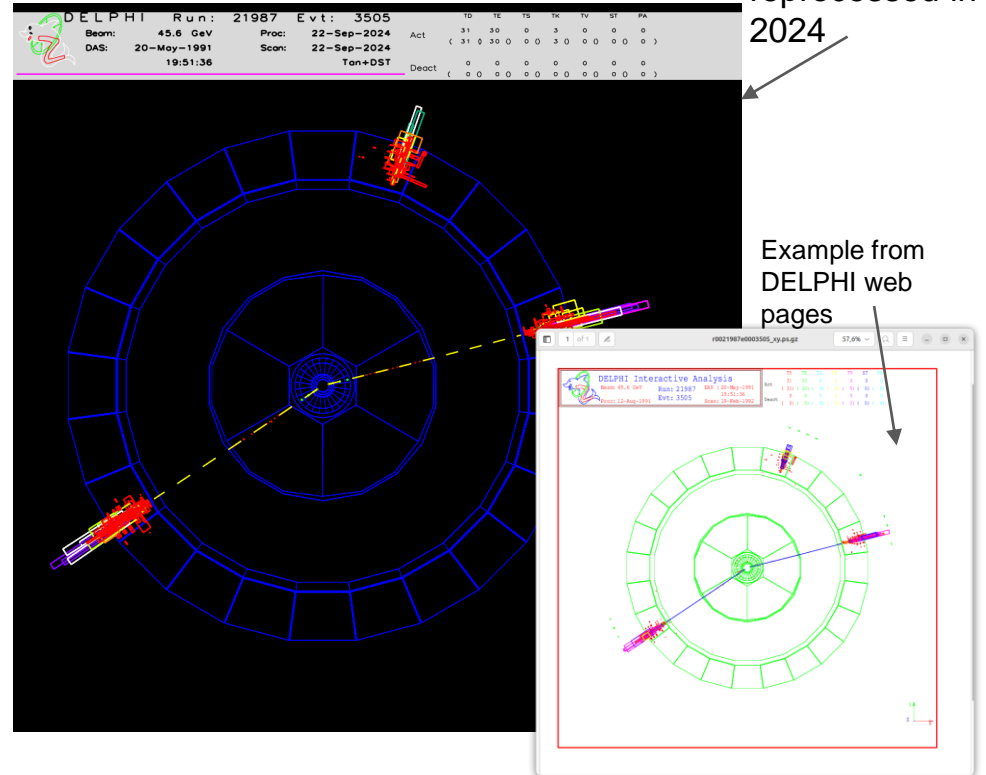
- Extract and plot information from the database, e.g. LEP beam energy but also stuff like air pressure, magnet temperature etc.
- Tools to read parts of the database and create snapshots. These were used for MC production and are useful for data recovery
- **Needs more testing**



Work in progress: 1991 data recovery

Z0 scan data, abandoned by the collaboration

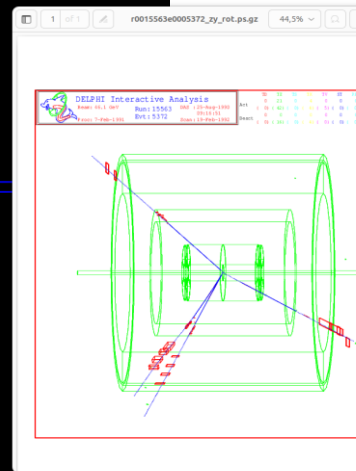
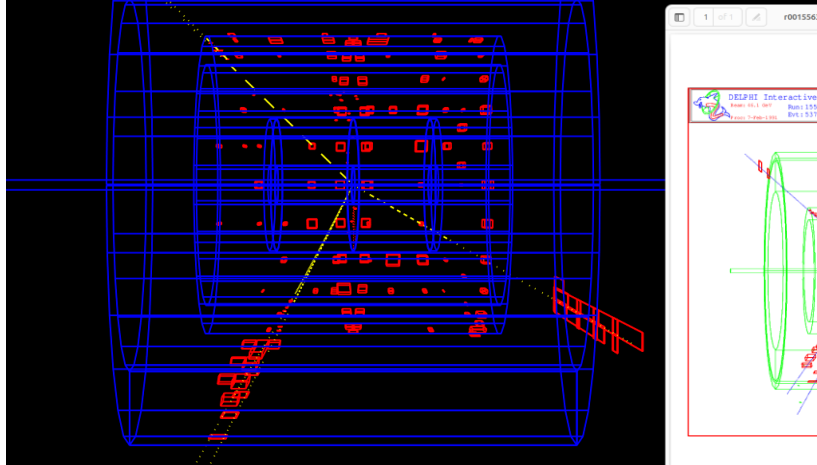
- Original raw data, short DST and some simulation samples exist
- Calibration files could be recovered
- Resurrected, with some uncertainties and guess work
- **Needs validation** work and is NOT to be used but for educational purposes



Work in progress: 1990 data recovery

Abandoned due to low data quality and small statistics

- Only raw data is conserved
- Calibration files seem to be lost, still searching in archives ...
- Some events are published on the DELPHI web pages
- Reconstruction of individual events via the event server works
- Still a long way to go ...





DELPHI Open Data



DELPHI data access policy

- Changed by decision of the collaboration board in **March 2024**
 - replaces the original, more restrictive rules
- Available from the DELPHI web pages, [direct link](#)
- Key points:
 - Covers data, software and documentation
 - Implementation of FAIR principles on DELPHI data

DELPHI Open Data

- Available in the QA instance for now
 - work in progress on the data samples
- Software binaries
 - Container image
 - CVMFS
- Works best from a Linux desktop or VM

DELPHI Collaboration releases its entire data collection

2024-06-15 by DELPHI Collaboration

[View](#)

The DELPHI Collaboration decided to open their data and released it to the physics community for public access, following FAIR principles. DELPHI was one of four detectors at the LEP collider, which collided electrons and positrons, at until now unreached energies. Since the two particles annihilate entirely into pure energy, the total energy of the events is precisely known, and the end states are much cleaner than this is the case for hadron colliders like LHC.

Along with the data, the collaboration decided to open the original software stack, including simulation and reconstruction. This enables scientists to pass the output of decent generators through the detector simulation of the DELPHI detector, and compare to what was actually measured at the time. It is also possible to use this data for educational purpose. The available data and software allow to either pick and reprocess individual events, from raw data, or create simulated events of a specific end state, pass them through the detector response, and then scan them using the event display.

The publication includes:

DELPHI Data sets:

- The complete raw data set, for the years 1990-2000, about 6 TB
- The complete reconstructed data set, years 1991-2000, about 14 TB
- The complete set of simulated data, about 16 TB of data
- The full DELPHI conditions database

Software stack components:

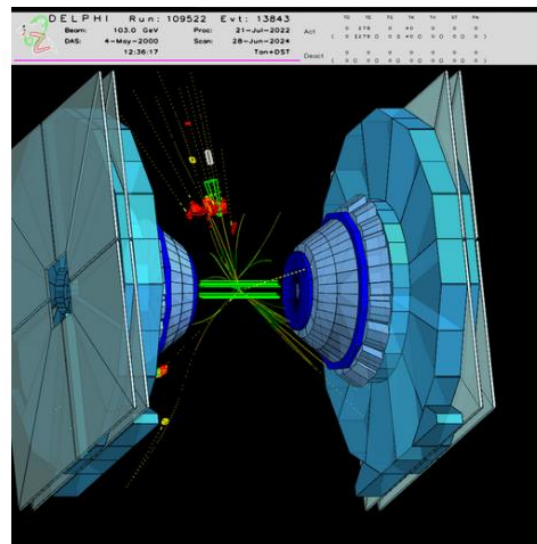
- The original DELPHI analysis framework
- The simulation and reconstruction code for the years of data taking 1992-2000
- The DELPHI event server, a tool to read and reconstruct individual events from raw data.
- The event display

Documentation:

- Basic manuals needed to get started
- Internal notes and publications are available from CDS

See also:

- About DELPHI
- DELPHI data access policy
- DELPHI web site
- DELPHI quickstart guide





DELPHI data samples

Organised by “nicknames”

- sets of “cassettes” of some equal kind
- e.g. Hadronic Z0 events from 1994



Data samples: example

```
uschwick@dtschwicke: ~  
uschwick@dtschwicke:~$ fatfind sh_apacic105_e91.25_w94_2l_c2  
This is what I found:  
  NICK : SH_APACIC105_E91.25_W94_2L_C2  
  GNAME: //CERN/DELPHI/P01_SIMD/SHORT/APACIC105/E91.25/WUPPERTAL/SUMT/C001-943  
  DESC : Short DST simulation 94c2 done at ecms=91.25 , Wuppertal  
  COMM : in total 2760474 events in 943 files time stamp: Thu Aug  1 15:11:25 2002  
  
1 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2001.sdst  
2 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2002.sdst  
3 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2003.sdst  
4 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2004.sdst  
5 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2005.sdst  
6 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2006.sdst  
7 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2008.sdst  
8 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2009.sdst  
9 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2010.sdst  
10 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2011.sdst  
11 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2012.sdst  
12 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2013.sdst  
13 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2014.sdst  
14 /eos/experiment/delphi/castor2015/MCprod/wupp/apacic105/v94c/91.25/apacic105_91.25_2015.sdst
```



Data samples: work in progress

- An Initial set of data set is available
- The plan is to add more meta-data, e.g. :
 - (Average) collision energy per set
 - Cross sections for simulation samples, where still available
- Once done,
 - Ask for one DOI per nickname
 - Release to production at <https://opendata.cern.ch>



Probable future pain points ...

- Future of X11
 - Mainly on CERNLIB side
 - Used by HIGZ, paw, ...
- Motif
 - Used in CERNLIB by paw++, kxterm
 - Used as well in OpenPHIGS
 - for OpenPHIGS moving to something else may be feasible
 - GTK may be an alternative
- OpenGL
 - Used by OpenPHIGS
 - Vulkan is more modern.
- DARWIN
 - Apple is changing the linker.
 - CERNLIB links but ZEBRA does not work when using the new linker and crashes



Possible future work ...

- Improve documentation for OpenData
 - add more manuals
 - add more examples
- Automated testing
 - add more quantitative tests
- Recover any still existing analysis code
 - query former collaboration members for anything that may be left
 - hope they are willing to contribute



What is the interest of the community

- Education
 - High school, University
- Technical studies
 - Machine Learning
 - Real Data for FCCee studies ?
- Citizen Science



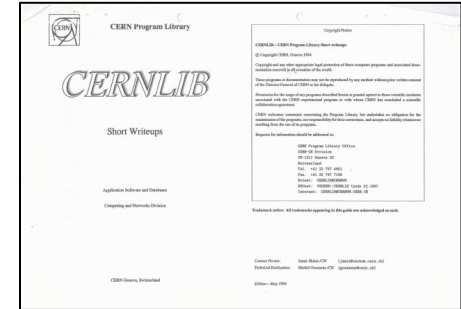
What is the interest of the community

- Not everybody will want to learn Fortran and Zebra
- Some suggestions
 - EDM4HEP (like ALEPH)
 - Ntuples (if you prefer NanoAOD)
 - Python Interface
- We can help in such transition
 - If there is interest and concrete plans
 - Our manpower is limited

Questions ?

LEP: the 64bit challenge

- **32bit support** is vanishing
 - At LEP time, **Linux based computers were running in 32bit mode only** (and usually single-core)
 - Nowadays, everything is 64bit (apart from some older gaming applications)
 - Example: **Motif 32bit libraries gone** from Ubuntu 20.04 and newer
 - Most LEP experiments rely on CERNLIB which is no longer supported for a while now
- **Community CERNLIB effort**
 - Recent effort to revive CERNLIB, in collaboration with MPG (Germany)
 - Good progress, aiming at a **first release in the next months**
 - Both 64bit and 32bit, enabling the experiments to migrate



Commits to master

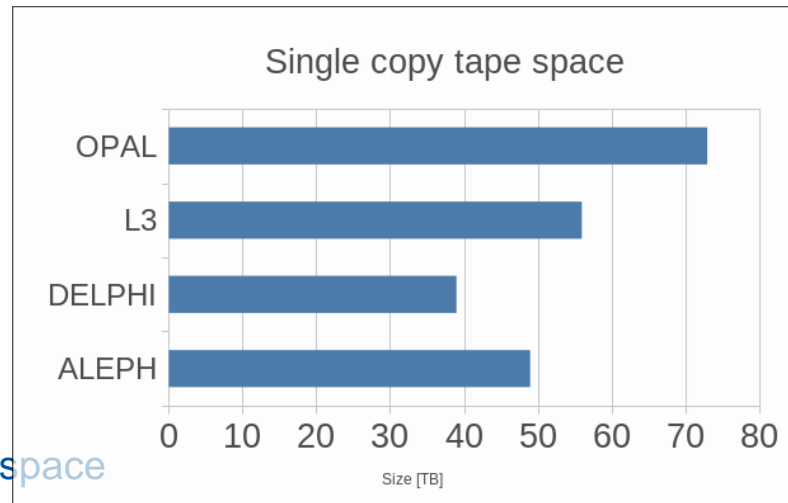
Excluding merge commits. Limited to 6,000 commits.



Bit preservation in general



- Long term storage of data, including
 - Raw data
 - Reconstructed data
 - Simulations
 - Databases, ...
- Data format and representation
 - Typically **compressed binary data**
 - E.g. BOS, ZEBRA, ROOT
 - Human readable form would take too much space
- Technically considered to be a **solved problem**



Tape storage in use for LEP data

Knowledge preservation

- Manuals
 - Detector specifications
 - Software, interfaces, ...
- Internal and technical notes
- Theses
- Conference contributions
- Publications
- Web pages and documentation

