# 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
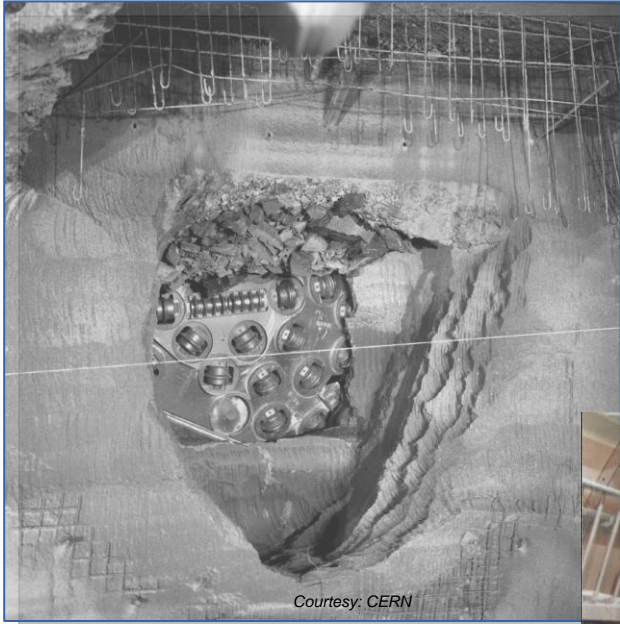- Access policies and open data

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**
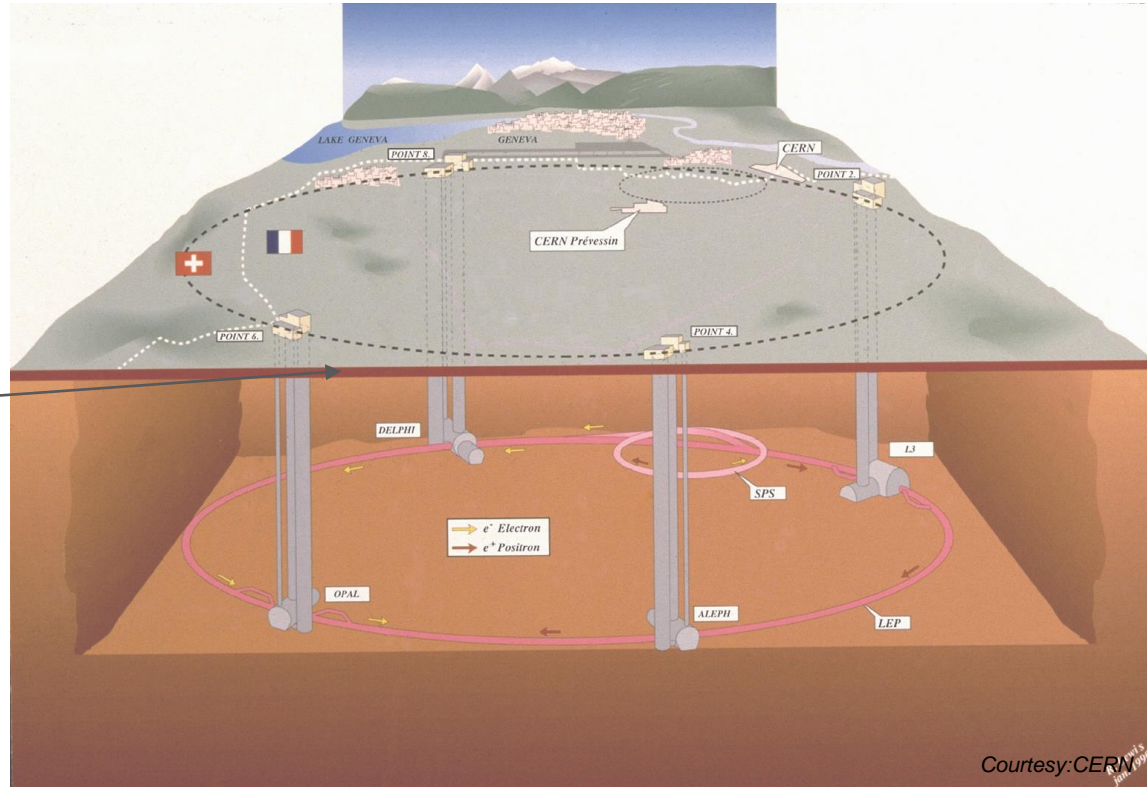
Courtesy:CERN

LEP: some impressions

LEP Tunnel
break through,
1986

Courtesy: CERN

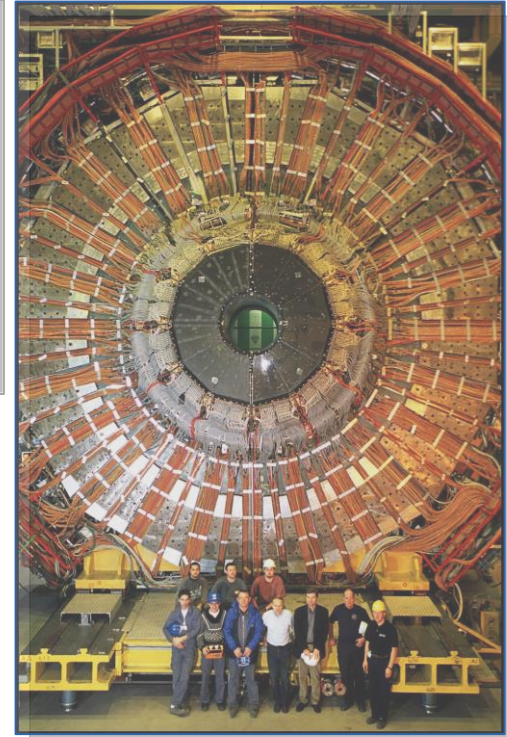LEP RF galleries

Courtesy: W. Liebig

LEP accelerator

Courtesy: W. Liebig

DELPHI end-cap before
dismantling

*Courtesy: CERN*

3

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
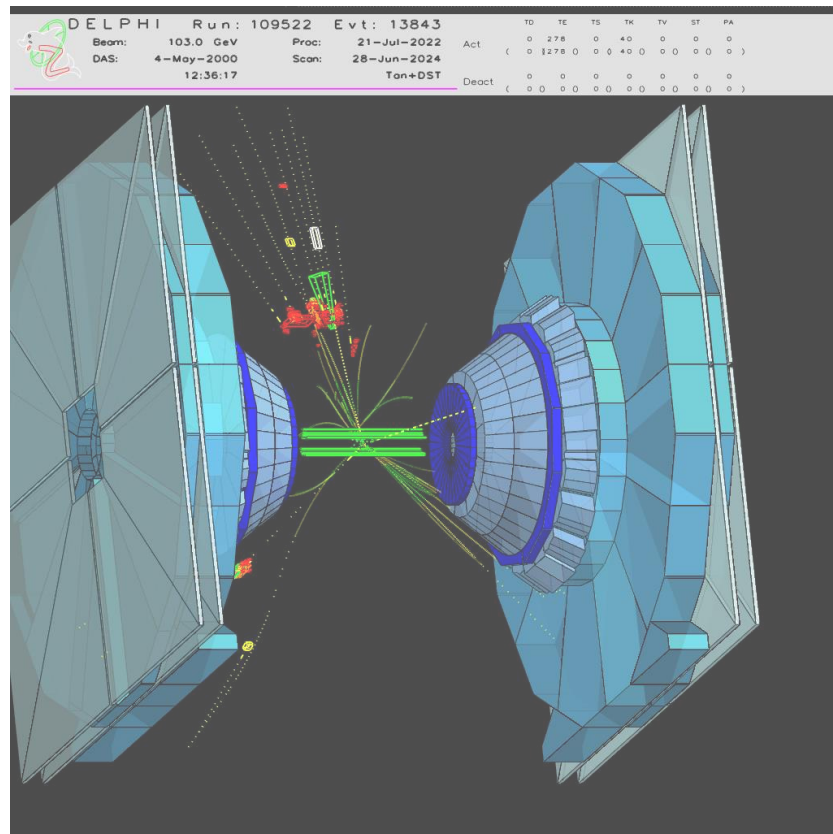
Software preservation: challenges identified during the last workshop

- 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:
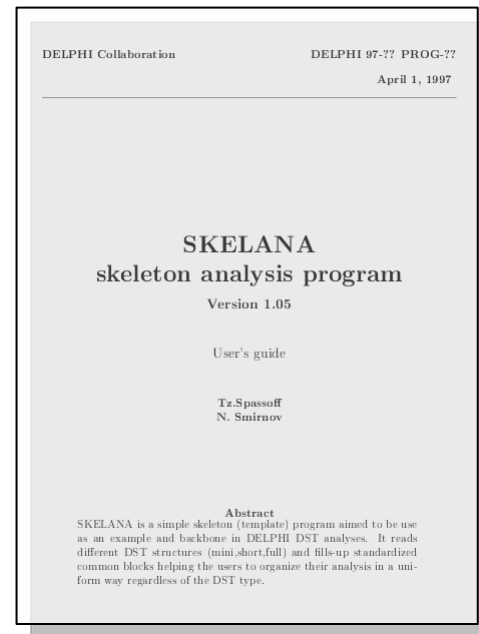# building and testing using gitlab CI

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 Collaboration   DELPHI 97-?? PROG-??

April 1, 1997

SKELANA
skeleton analysis program
Version 1.05

User's guide

Tz.Spassoff
N. Smirnov

Abstract

SKELANA is a simple skeleton (template) program aimed to be use as an example and backbone in DELPHI DST analyses. It reads different DST structures (mini,short,full) and fills-up standardized common blocks helping the users to organize their analysis in a uniform way regardless of the DST type.

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

14

Work in progress

- **AARM64 support**
  - includes RasberryPi (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
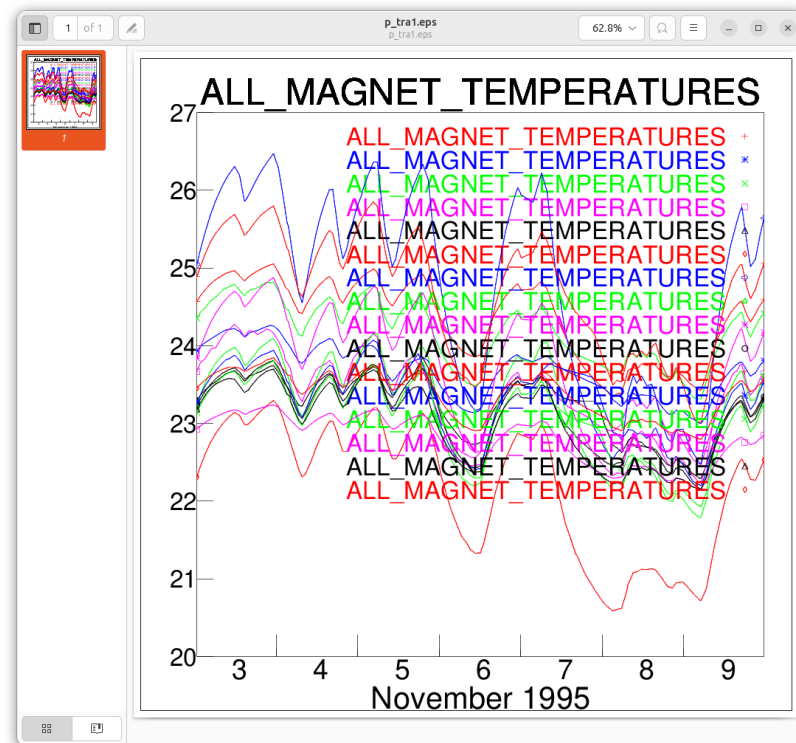
- **More testing and validation needed**
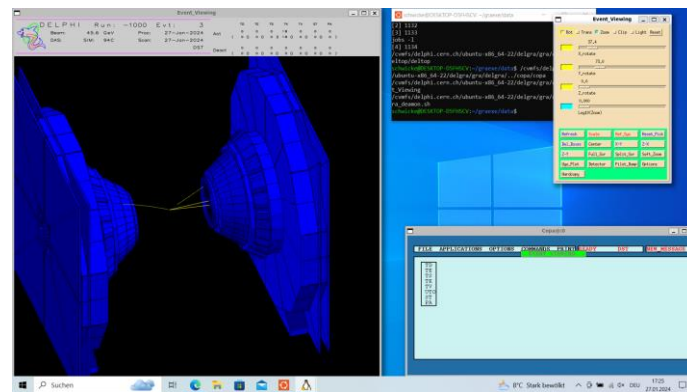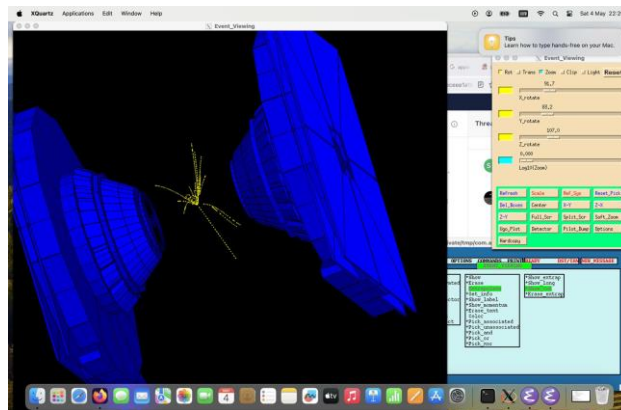
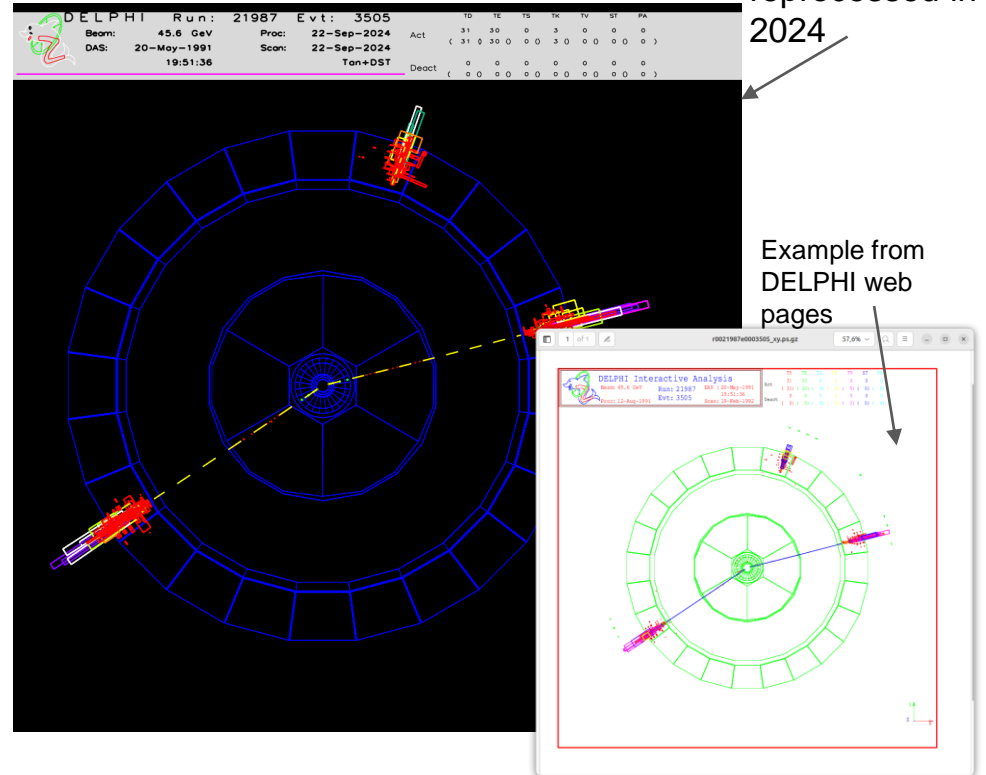# Work in progress

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

reprocessed in 2024

Example from DELPHI web pages

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 …
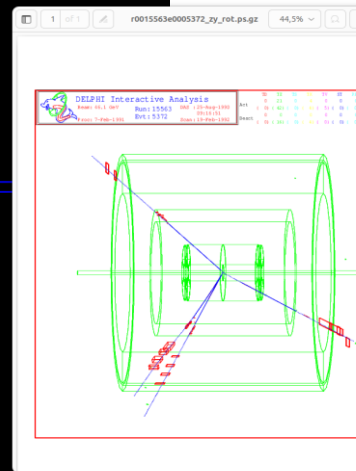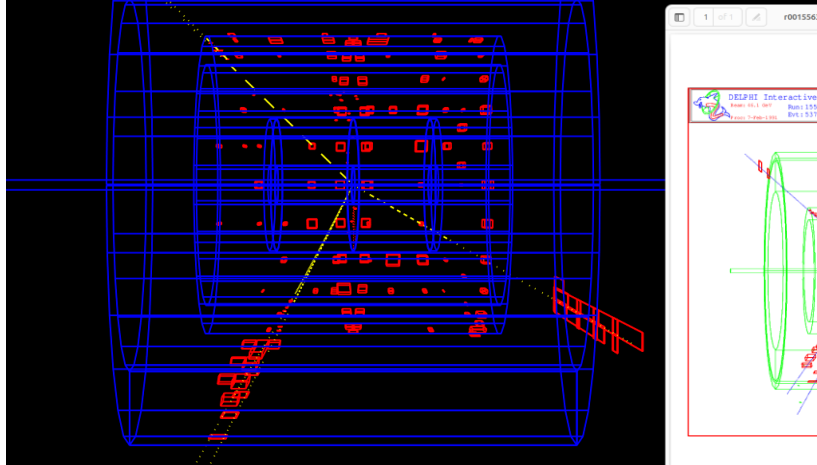
DELPHI Open Data

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

21
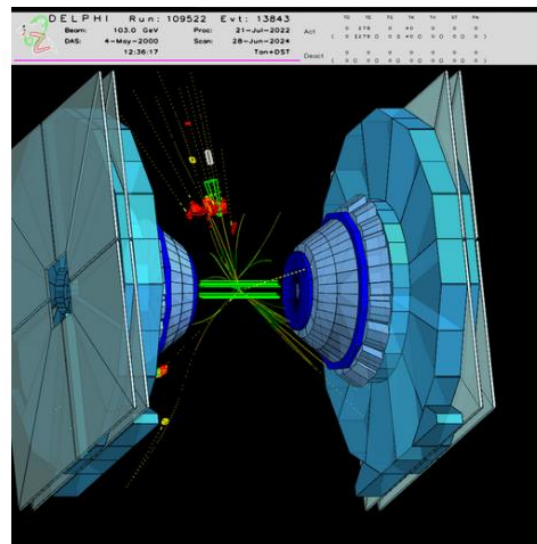
DELPHI data samples

Organised by "nicknames"

- sets of "cassettes" of some equal kind

- e.g. Hadronic Z0 events from 1994

Data samples: example

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

- Eduction
  - 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
  - Out 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**

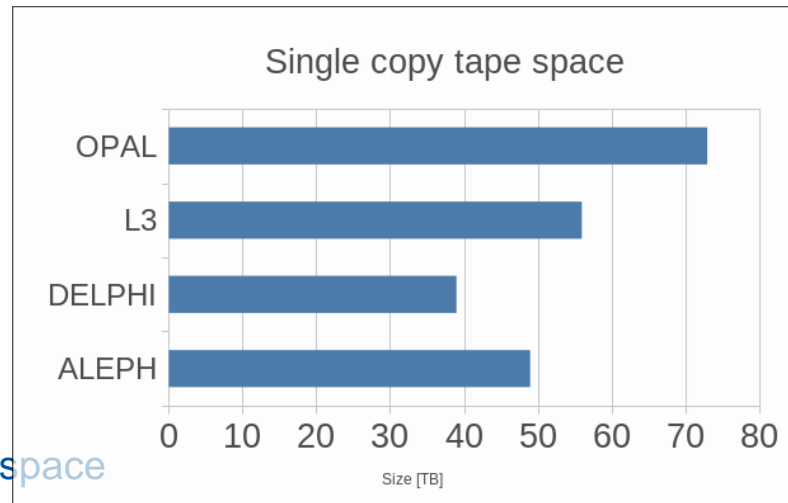 – Both 64bit and 32bit, enabling the experiments to migrate
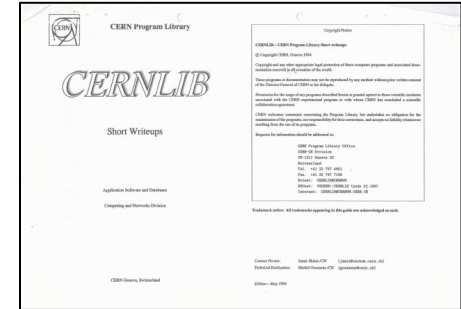
Bit preservation in general

- 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