Preservation of LEP Data

There is still hope

Is there?
LEP?

• Large Electron Positron Collider
• Four general purpose experiments
  – ALEPH, DELPHI, L3, OPAL
  – Varying detector layouts
  – Varying data structures
  – Varying software approaches
    • Optimization
    • Personal preferences
• Data taking: 1989 – 2000
  – ‘89 to ‘95: at and around Z-peak (91 GeV)
  – ‘95 to ‘00 increase in various steps up to 209 GeV
Publications

• Individual publications
  – Several hundred per collaboration
  – Rate decreased ~2004
    • Now in the (long?) tail

• Combined publications
  – Dedicated topics
  – WG ‘s with experts from all collaborations
  – Based of analysis results, not common analysis
    • Topic specific, in general (hbook) ntuples

• Traces on HepData and Rivet
Motivation for DP

• Unique data sets
  – Some ‘overlap’ with SLD data
• Profit from improved theoretical models
• Profit from new results at other facilities
• Allow combining data sets
• Open Access
• Education
• Use case determines preservation needs
New Theoretical Models

• Require full Monte Carlo production chain
  – New four-vectors
    • Standard format 😊
  – Full detector simulation
    • Geometry (Geant3)
    • Digitization
  – Full detector reconstruction
  – Old MC samples for consistency / validation check
  – Analysis
New Results at other Facilities

• Sufficient to access reconstructed objects?
  – Probably, but why limit yourself
Combining Data Sets

- Requires common object definition
- Requires common data format
- Translate assumptions coded in framework
- Only useful at analysis level
  - Or rewrite all simulation & reconstruction code
Open Access

• Policies:

• For education:
  – Often done by former collaborators
  – Using (hbook) ntuples or DSTs
Data

• RAW data
• Processed data
  – Various levels of ‘abstraction’ per experiment
• Selections
  – Tagging
  – Streaming
• DSTs
• File catalogs
• ~100 TB per experiment
• Stored at CERN (castor, dual copies)
• Outside copies desirable
Software

• Mature (dated?)
• Fortran
  – Memory management + IO
    • BOS
    • ZEBRA
  – Little (but some 😐) commercial libraries
  – HEPDB
    • Calibration
    • Status updates
  – HBOOK
• Portable
  – Proven through lifetime of experiments
    • IBM mainframe, Cray, DEC (VMS & Ultrix), Apollo, HPUX, SGI, Linux
    • Maintainable?
  – So far non-disruptive(?)
  – For how much longer?
Tools

• Wrappers
• Scripts
  – Data management
  – MC production chain
    • Aleph & Delphi still exercising it 😊
      – One with one w/o book-keeping
    • OPAL’s died with fatmen 😞
  – Job control
• Environment variables
  – Paths, version numbers, ...

Marcello Maggi, Ulrich Schwickerath, Matthias Schröder, 22.3.13, DPHEP7
Monte Carlo

- Generators
- Detector Geometry & Simulation
  - Geant3
  - Detector specific tuning
- 4-vector files
- Digitized samples (?)
- Processed events
- Production chains
Documentation

- Mainly targeting newcomers & collaborators
  - (Very?) steep learning curve
  - Rarely sufficient w/o experiment expert
    - And their memories also have a half-life

- Format
  - Latex
  - Html
  - PS
  - ?

- Accessible?
  - review documentation considered ‘internal’

- Searchable?

- Probably un-versioned

- Required: Documentation for external dependencies (CERNLIB)
Current Status

• (Very) Few people still doing analysis
  – Using RAW, DSTs and (hbook) ntuples?
• Most experts busy with new experiments
  – Or retired, or…
• No resources
  – Rely on individual contributions
  – No room for experiment specific effort
• Trying to raise awareness for DP…
  – New data is so much more tempting
DP Approaches

• Freeze current environment in VM
  – Until when will hypervisors run them?
  – Have to make all data ‘local’?

• Keep alive by continued porting
  – External dependencies
    • CERNLIB
    • BOS
    • ?
  – When will compilers refuse to accept the code
  – Date formats?
  – Delphi: all sources put on standalone CD

• Migrate to more modern format
  – Information loss
  – Rewrite all software

• Play safe – don’t put all your data in one tier
  – In every aspect!