Cleanup of existing data

Paul Dauncey
Imperial College London
Cosmics and DESY beam data

- DAQ was **not** final
  - **Prototype**: several bugs/features in firmware and software
  - Data are not ideal in several ways
- Try to fix bugs before next round of data-taking
  - DAQ work over the next four months
- Question is how to deal with existing data
  - Correct DAQ files, or records “on-the-fly” when reading DAQ files?
  - Correct in LCIO converter?
- **Proposal**: do the first of these
  - Expertise about what is missing and what is needed is on DAQ side
  - Mistakes = responsibility is on DAQ side
  - LCIO conversion will look similar for old and new data
- Already implemented the most important of these
  - DaqFixer corrects run/event numbers where necessary
Run numbers

• Original implementation used epoch time for run numbers
  • I.e. Run number = start time in seconds since 1 January 1970
  • All cosmic runs and early DESY runs had these run numbers
  • Allows easy interleaving of slow readout data taken outside of runs

• Requested to change this to contiguous numbering
  • Started at 100000, went to 100224 at DESY
  • HCAL tests now using 110000 upwards
  • Imperial test stand using 120000 upwards

• Proposal: convert epoch numbered files to contiguous numbers
  • Early DESY data 099975 to 099999, so continuous range 099975-100224
  • Cosmics data 090000-0900036
  • Original raw files not removed, but new (bin) versions made

• New scheme for slow readout data being implemented
  • Will link run files to epoch numbered files also; best of both worlds
ECAL stage readout

- **ECAL movable stage** controls position of ECAL
  - Changes relative alignment to the beam and tracker
  - Changed almost on a run-by-run basis during DESY tests
  - Stage controller readout not implemented at DESY (but now fully tested)

- Raw data files have **no information** on stage position
  - Only record is in (human-written) logbook file
  - Nominal alignment needs to be done per run…
  - …unless someone has extracted the logbook values already?

- **Proposal**: add stage data to records when reading
  - Stage readout subrecord now defined
  - Need to code logbook numbers into EmcFixer and call it
  - Data will look the same as in future runs for LCIO converter

- **Should mean only one alignment needed for all DESY runs**
  - In principle; must check for other effects (and mistakes in copying logbook)
Subrecord class changes

- All data stored in the record is in **subrecords**
  - Each has a class which maps exactly onto the data
  - Each subrecord class has a **version number** to handle changes to the class which affect the data layout
  - Can access both versions in one job, but they are separate classes, so need to know this is necessary
- **Classes will** need to change with time
  - Only one has changed since DESY so far; control information for CRC readout (so not urgent for LCIO conversion)
  - Should consider how to handle this in general
- **Proposal**: convert old to new versions when reading
  - Have to put some defaults/null values into missing data
  - Maybe cases where need to convert new to old? Unclear
  - Always have single compatible set for LCIO conversion…
  - …but converter will need to be modified to handle newer class data
Warning

• Need to reorganise disk space at Imperial
  • Have new PCs for DAQ, including 2.5TBytes of local disk
  • Will move all old data to these and do repairs there
  • Should recopy complete structure to DESY when finished
  • Existing data area at Imperial will disappear
  • Do not do updates from data area from now on

• Also moving code area
  • Need to have central repository on DAQ machines
  • Currently accessible area will disappear
  • Do not do updates from code area from now on
  • When code changes settled down, will dump everything into cvs

• You should assume frozen versions for a while
  • If there are any outstanding updates, send to me to include in central repository asap