Material for PRC
Oct 2014

Dirk Krücker, draft 2.10.14

Data preservation for the HERA Experiments

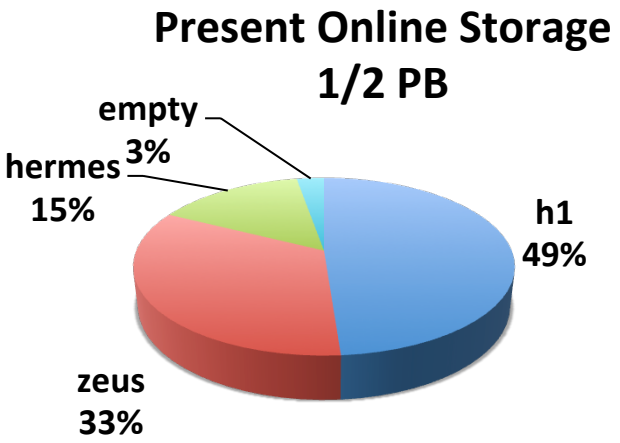
- DPHEP Collaboration is now **officially** installed
  - MoU signed by :
    CERN, DESY, IN2P3, MPI Munich, HEP, IHEP, KEK
- Legacy dataset
  - Defined by now – see next slide
    - but can still be extended
  - Transfer into long term format ~80%
- Documentation
  - Web servers
    - Archive server available for all experiments
    - Transition ongoing
- Software preservation
  - Migration to SL6 successful
  - New software preservation system
    - Integration into IT infrastructure for long term “survival”

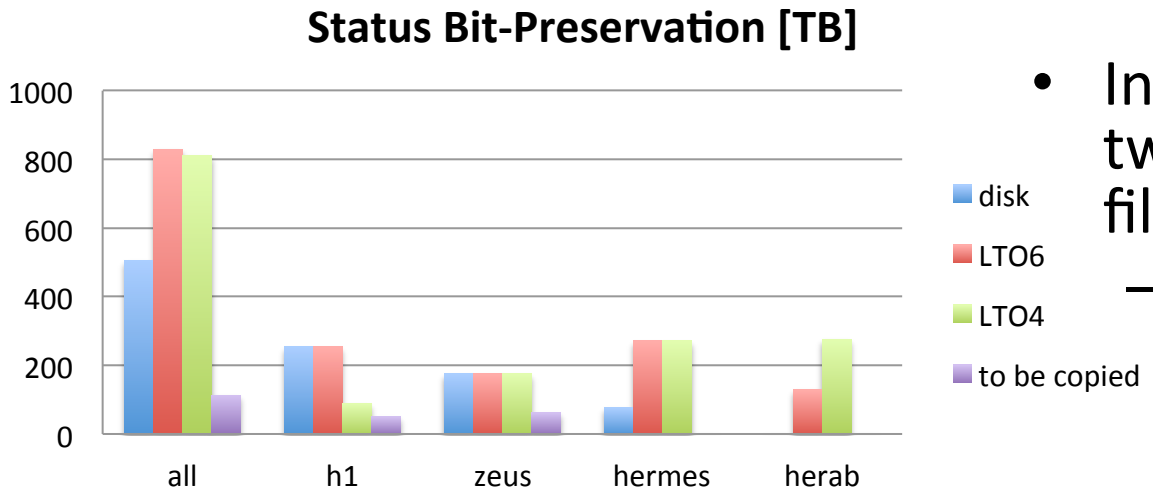
Software Preservation System

- With the leave of Dima Ozerov we lost expertise on the prototype
  - The software was left in an unusable state after an failed approach to migrate to SL6
  - Strong dependence on the expertise of one person
- Lesson learnt
  - The SP system itself is part of the preservation
  - Moving **from an individual to institutional solution**
    - Virtual machines defined by DESY IT Systems Group
    - Managed with DESY IT standard tools (puppet)
    - Ongoing work for an automatic test system
- Strong commitment from H1 and Zeus to continue with this approach despite the tight timeframe and man power situation
- Software preservation strongly depends on the expertise of individuals

Status Bit-Preservation

- The legacy data set is defined
- Additional 115 TB to be copied into DPHEP storage
  - H1 **~50 TB**
  - ZEUS **~60 TB**
  - There will be in total about 620 TB
  - The online storage will be extended accordingly soon





- In addition there will be two tape copies for each file on different tape types
  - About 70% of the present content already on 2 tapes

Backup

Some Statistics on the Present DPHEP Storage

H1	Hermes	Zeus	HeraB	type
509006	2007929	483631	846059	raw files
5784	7269	4856	4104	tar files
106	404	267	393	LTO4 (800G) tapes
79	114	74	53	LTO6 (2.4T) tapes
254	77	175	0	TBytes online
89	271	175	276	TBytes on LTO4 tape
254	271	175	130	TBytes on LTO6 tape