

HEPiX Working Group

Bit Preservation

- Goal of the group (mandate)
- Survey
- Message to DPHEP from the survey
- DPHEP archive storage

wiki: <http://w3.hepix.org/bit-preservation>

mailing list: bit-preservation@hepix.org

Co-chairs: Germán Cancio(CERN)

Dmitry Ozerov(DESY)

Collecting and Sharing the Knowledge on bit preservation among the HEP laboratories

“The goal of the HEPiX Bit Preservation Working Group is to share ideas, practices and experience on bit stream preservation activities across sites providing long-term and large-scale archive services. Different aspects should be covered like: technology used for long-term archiving, definition of reliability, mitigation of data loss risks, monitoring/verification of the archive contents, procedures for recovering unavailable and/or lost data, procedures for archive migration to new-generation technology.”

“The Working Group will produce a survey on existing practices across HEPiX and WLCG sites responsible for large-scale long-term archiving. The collaboration should ideally be extended to other large-scale archive sites from other research fields outside HEP.”

Mandate II

DRAFT 1.0

Technical advisory/hand of DPHEP Collaboration

“The Working Group harmonizes its activity with the strategy of the DPHEP collaboration.”

Recommendation how DPHEP Storage may look like

“Based on best practices and development in storage preservation activities, the Working Group will provide recommendations for sustainable archival storage of HEP data across multiple sites and different technologies.”

Survey

- As first step of the group we sent survey to the sites.
- It's complementary to the survey's done in DPHEP – this time it's view from the resource provider (6 sections, 40 questions)
- Survey may be found on group wiki page:
<http://w3.hepix.org/bit-preservation>
- Detailed report will be presented at HEPiX meeting in two weeks and published on wiki

Message to DPHEP (ownership)

- Majority of data has owner. Most of them are experiments at the end of their lifetime.
- Transfer of the ownership to DPHEP Organisation?
- DPHEP is a bridge between the resource provider and the body/experts of the experiments?

Message to DPHEP (SLA)

- Most of the data preserved/archived without any defined MoU/SLA/... (exceptions are WLCG and few rare cases)
- No defined requirement for the quality of archive, tolerated loss etc.
- Users expectation: 100% safe
- Should it be covered by the DPHEP Collaboration Agreement?

(WLCG MoU: “Tier1 Centres provide a distributed permanent back-up of the raw data, permanent storage and management of data needed during the analysis process”)

Message to DPHEP (loss of data)

- Variation in the data loss during last 24 months: from few/mln's (0.0001%) to 12% of the data
- We know and we learn further what should be done to reduce the data loss on site level.
- Resources (hw/man), Verification, Redundancy (site level), Redundancy (geo)
- Further reduction of data loss -> safe DPHEP storage (architected within the HEPiX group; implementation -> DPHEP)

DPHEP storage

- Just a personal view, input to discussion inside the HEPiX group

