

WORKSHOP GOALS

Data & Analysis Preservation Workshop
26 Nov 2019

Dirk Duellmann

REVIEW STATUS OF

- Data Preservation
- Reproducible Analysis
- Open Data

PRESERVATION MILESTONES... (TOMBSTONES?) - BY JAMIE

- 2012: DPHEP Blueprint published
- 2013: CERN assumes role of DPHEP MGR
 - Many events & w/s [here](#)
- 2014: DPHEP Collab. Agreement by CERN, other labs / FAs
- 2016: CERN services for LTDP@iPRES (Bern)
 - Covers all areas proposed to 2012/13 ESPPU
- 2017: DPHEP w/s on FAIR data, TDRs
 - TDR = Trustworthy Digital Repository
 - Self audit of CERN (ISO 16363) completed
- 2018: Science Europe - Research Data Management
- 2019: Workshop on Sustainable Software - LEP+CERNLIB
- 2019: revision of General Conditions for CERN experiments
 - including Data Management Plans (DMPs)
- 2020: do we complete Certification of CERN or give up?
 - Goes hand in hand with DMP requirement on experiments

CHALLENGES

- Scarce resources
 - eg depending on "preserving memories" as one motivation
- Unclear long-term ownership, Privacy vs Retention
 - for upcoming decisions on preventing data to drop into oblivience
 - next year: account expiration may trigger terminal data/document losses (affects more inactive experiments)
- Growing frustration about slow progress to reach a clear policy

BALANCE BETWEEN BENEFITS AND RISKS

- Benefits
 - reproducible results: quality assurance for science output today
 - obligation to give data back to the public (now, soon, ultimately?)
 - open data: additional science by enabling additional contributors
- Effort/risk
 - extract complex SW and data dependencies and implicit analyser knowledge
 - additional nuisance to overloaded physics analyst?
 - additional effort to "handle" invalid publications based on OD

GOALS

- Can we arrive at a consolidated
 - cost/benefit analysis ?
 - role for hostlab and LCG tiers?
 - forum to exchange experience with preservation services?
 - proposed publishing policies, timeline for OD releases?
- Propose a statement towards managed Data Ownership & Lifetime
 - experiment, hostlab, science community, public
 - principle: (data) ownership comes with responsibility ?
- Give "bottom-up" input to policy makers (CERN management and FAs)
 - present consolidated responses to WLCG OB and CERN SPC

MESSAGE FROM ECKHARD ELSSEN

The topic is highly relevant and important and on everybody's mind. For CERN it is important to find a compromise between Openness, minimisation of confusion and ownership of the data. The latter also has to do with the support of the experimenters. We have seen some good examples recently that combines accessibility and respect for the genuine contributions of the experimenters.

COMMING SOON...

- CERN will host PV2020 conference
 - 12th – 14th May 2020
 - more details [at link](#)

LOGISTICS

- smaller room for discussion with experiment computing management after lunch
- Target
 - short 2-3 page public report by next week