Data Preservation, Open Access etc

Ian Bird

WLCG Management Board

CERN, 4th December 2018

Data preservation, open access

- ❑ Currently many strands to these activities
  - ▪ Often started independently with different goals and interests behind them
- ❑ Main topics:
  - ▪ Data preservation – building on work of DPHEP
    - • Several aspects from bit preservation, DC certification, metadata, knowledge retention, etc.
    - • Bit preservation is what we already do at Tier 0, 1
  - ▪ Open data:
    - • For example via open data portal
    - • Experiments have different policies of what level, and how much data is made available; and for different intended purposes
    - • This has a cost – today CERN provides ~5 PB disk for this – but growing
      - • Currently not costed as part of pledges
  - ▪ Analysis preservation
    - • E.g. via tools such as REANA, etc.

Open access - Concerns

- ❑ Active groups in IT, EP, experiments on all of these – but not necessarily coordinated
- ❑ No policy for how this should be funded
  - ▪ Today it is essentially CERN
    - • Should it be a shared/distributed problem?
    - • ESCAPE can provide a mechanism for this shared management (data lake)
- ❑ Scale and cost
  - ▪ If the scale increases it will need to be taken from the pledges
- ❑ Today cannot draft an overall coherent policy
- ❑ Propose to organize a workshop to address and coordinate these aspects
  - ▪ And to formulate a strategy for how this should be managed
  - ▪ This should become part of the overall strategy for the future and integrated with the other aspects
- ❑ Need feedback from funding agencies on what is mandated and affordable

Summary

❑ Following discussions at a recent IT Programme of Work meeting in the presence of the DRC and EP-DH, it is foreseen to hold a workshop in early 2019 to discuss and hopefully agree on a strategy for FAIR Data Management (at least in the context of the LHC experiments; perhaps more widely);

❑ FAIR data management, increasingly requested by Funding Agencies, links to both the preservation of data as well as to is re-use and sharing ("as open as possible, as closed as necessary);

❑ As such, the question of resources for (online) open data will also be discussed.