



YEARS/ANS CERN



WLCG

DPHEP Introduction

Data Sharing – In Time and Space

Jamie.Shiers@cern.ch

School on Grid & Cloud Computing



International Collaboration for Data Preservation and
Long Term Analysis in High Energy Physics

Training Overview

1. Background to Long-Term Data Preservation in HEP – the DPHEP Study Group
2. “2020 vision” for DP in HEP
- 3. DP in other disciplines – how we can benefit significantly from work (models, standards, procedures, wisdom, tools, services etc.) of others – the bulk of the material comes from other projects / disciplines**
4. A strategy for DP in HEP

DPHEP Introduction

- Historically, HEP has had a bad track record with regard to data preservation
- **In fact, “data destruction” would have been a better description**
 - Old tapes were often de-gaussed or even destroyed; if not typically unreadable due to poor storage conditions / lack of appropriate h/w and s/w
- This began to change in the 1990s, when a number of DP efforts were initiated
- A much more formal approach was initiated as from 2009 through a series of workshops – see DPHEP.org
- This work is well described in the talk of David South at CHEP 2012 (ppt, video), as well as in the DPHEP blueprint.

DPHEP – “post CHEP 2012”

- **Take the time to review the D. South talk and DPHEP blueprint [gap of at least one hour]**
- Shortly after CHEP 2012, the DPHEP group provided input to the update of the European Strategy for Particle Physics that took place around a workshop in Krakow in September of that year
- The outcome of this workshop was captured in a series of recommendations, formally adopted by the CERN Council at a special session in May 2013:
<https://indico.cern.ch/event/244974/page/4>
- Data preservation is included (for the first time) as part of this strategy

From Study Group to Collaboration

- At around the same time, some of the recommendations of the Blueprint began to be implemented
- This included the offer by CERN to provide a Project Manager for an initial period of 3 years (2013 – 2015 inclusive)
- As Project Manager, much of my work has focussed in following the other recommendations of the Blueprint, building partnerships with other DP activities, looking for funding and defining in more detail what DP in HEP really means, how we would go about implementing “it”, as well as measuring our progress / success

2020 Vision for LT DP in HEP

- Long-term – e.g. FCC timescales: **disruptive change**
 - By **2020**, all **archived data** – e.g. that described in DPHEP Blueprint, including LHC data – easily **findable**, fully **usable** by **designated communities** with clear (Open) access policies and possibilities to **annotate** further
 - Best practices, tools and services well run-in, fully documented and sustainable; built in common with **other disciplines**, based on standards
 - **DPHEP portal**, through which data / tools accessed
- **Agree with Funding Agencies clear targets & metrics**

Intro to intro to “EGI training”

- Since more than 10 years, a model for data preservation has existed:
 - Open Archival Information System (OAIS)
- This is the de facto standard (and also an ISO one) across many different disciplines
- There are also “certification procedures” based on OAIS
- DANS (NL), DIN (DE), ISO
- These are covered in the “EGI training” material
- Browse through these and come back....