Services for Long-Term DP in HEP

Services, Common Projects, Business Model(s)

Jamie.Shiers@cern.ch

CMS Week, December 2013

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

Introduction

- After several years of study and analysis, the DPHEP Study Group delivered a **Blueprint**

- A summary was input to the ESPP update and **Data Preservation** is now **part** of the European Strategy for Particle Physics

- A **small** set of services / projects have been identified / agreed upon

- These, together with the **associated resources**, should now be considered – and handled – as **FULL PRODUCTION SERVICES**

AND

The Approach (DPHEP@RDA-2)

- Whilst retaining a holistic view, the problem is broken down into a number of key areas. Each is addressed using state-of-the-art techniques, that include:

1. **Digital library** tools (Invenio[2]) & services (CDS[3], INSPIRE[4], ZENODO[5])

2. **Sustainable software**, coupled with advanced **virtualization** techniques[6] and **validation** frameworks[7]

3. Proven bit preservation at the 100PB scale, together with a **sustainable** funding model with an outlook to 2040/50

The Approach (DPHEP@RDA-2)

- Whilst retaining a holistic view, the problem is broken down into a number of key areas. Each is addressed using state-of-the-art techniques, that include:

  1. **Digital library** tools (Invenio[2]) & services (CDS[3], INSPIRE[4], ZENODO[5]) + related tools (HepData, RIVET, ...)

  2. **Sustainable software**, coupled with advanced **virtualization** techniques[6] and **validation** frameworks[7]

  3. Proven bit preservation at the 100PB scale, together with a **sustainable** funding model with an outlook to 2040/50
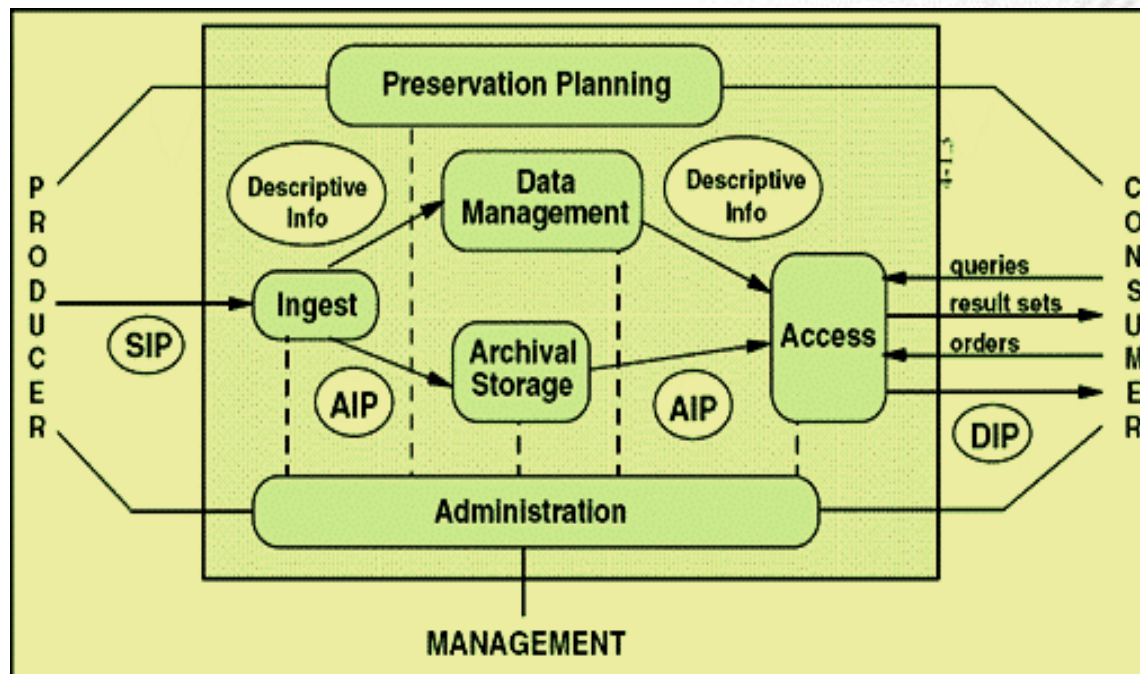
BRIDGES

Digital Preservation Solution

- the ISO OAIS Reference Model for *an* OAIS. This reference model is defined by recommendation CCSDS 650.0-B-1 of the Consultative Committee for Space Data Systems;[1] this text is identical to ISO 14721:2003.

Source: Long-Term Preservation of Digital Documents. 2006. doi:10.1007/978-3-540-33640-2. ISBN 978-3-540-33639-6.
Public Domain.

Digital processes break easily

- Short-period funding
- Software lifecycle: code, interfaces, formats…
- Dependent on expert knowledge
- Thin documentation and metadata

Source: : wikipedia PD Image resources

Bridging Components

- We need to design and build "bridges" between the individual components
- Respecting an overall architecture (OAIS)
- Ensuring the implementation is not tied to "CERN services" – e.g. fully applicable to other HEP experiments / services at other sites
- **As "future-proof" as possible**
- We need to do this **together** – and profiting from extensive existing experience "elsewhere"
- ➢ **A "hurried implementation" now could cost a huge amount in the long term**

Practically

- Get the experts from the service providing groups / teams, plus the users, together **asap**
- **Involve** external expertise early on!
- Having something implemented **well ahead** of (CMS) deadlines
- **Accept** that the first iteration may (will?) require modification / re-design
- This could start in the 1st working week in 2014…
- Continue to use **DPHEP IB** for coordination / information + detailed technical (**multi-party**) meetings to prepare for these

RESOURCES

Resource Requests / Reporting

- **D**ata **P**reservation resources (M+P) will feature in updates of CERN **M**edium **T**erm **P**lan
  - Important / necessary for resourcing and continuity
  - "Zero sum game" is IMHO a "worst case scenario"
- Also expect to use **R**esource **R**eview **B**oard: some significant elements, e.g. "bit preservation" already funded!
- Continue to report to senior CERN management – their buy-in and support essential!
- The January "**F**ull **C**osts of **C**uration" workshop will be fundamental in preparing "Business Plan"

Summary

- **The hardware resources required "immediately" have been agreed as part of a pilot**

- Longer term resource requests need to be formulated as part of overall needs

- ✓ **Key CERN services: strong support and commitment from the departments / groups**

- **Need to bridge these** – and understand (DPHEP) how equivalent services can be offered at other host labs and beyond