EOSCpilot
The European Open Science Cloud for Research Pilot Project

Science Demonstrator

High Energy Physics

Data Preservation An Open Data Portal pop up

Brief Overview

Funding agencies today require (FAIR) Data Management Plans, explaining how data acquired or produced will be preserved for re-use, sharing and verification of results.

The preservation of data from CERN's Large Hadron Collider poses significant challenges: not least in terms of scale. The purpose of this demonstrator is to show hBrief overviewow existing, fully generic services can be combined to meet these needs in a manner that is discipline agnostic, i.e. can be used by others without modification.

Objectives

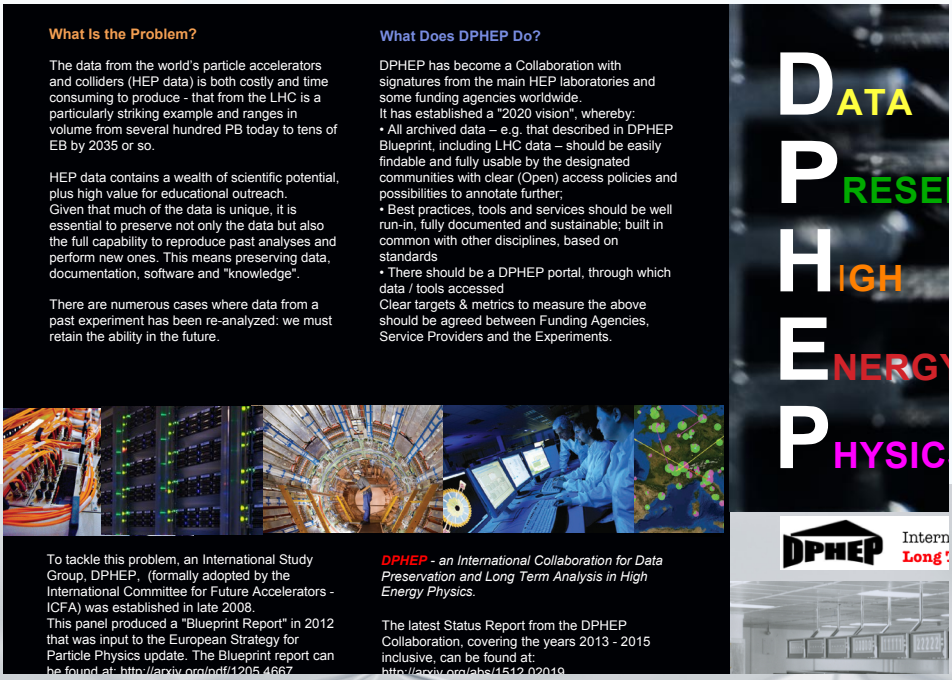
The DPHEP science demonstrator wants to deploy services that tackle the following functions:

» Trusted and certified digital repositories where data is referenced by a Persistent Identifier (PID);

» Scalable "digital library" services where documentation is referenced by a Digital Object Identifer (DOI);

» A versioning file system to capture and preserve the associated software and needed environment; A virtualised environment that allows the above to run in Cloud, Grid and many other environments.

The final goal is to use non-discipline specific services combined in a simple and transparent manner (e.g. through PIDs) to build a system capable of storing and preserving Open Data (at a scale of 100TB or more).



EOSCPilot.eu has received funding from the European Commission's Horizon 2020 research and innovation programme under the Grant Agreement no 739563.

EOSCpilot
The European Open Science Cloud for Research Pilot Project

Main achievements

Some limited success was achieved with the individual services identified (Zenodo, CVMFS, a Trustworthy Digital Repository), but it was not possible to integrate them into an usable service.

Recommendations for the implementation

The EOSC Pilot integrates services from three well-established e-infrastructures, mentioned above. Equivalent services are used in production by the CERN Open Data Portal, which is available via anonymous access over the Internet worldwide.

While it was possible to upload a documentation file into the EUDAT B2SHARE test instance and while software from the LHC experiments is stored in the RAL CVMFS instance, there have been significant delays in finding a site that could act as a TDR for this pilot.

There were numerous misunderstandings regarding the scope, duration and scale of the demonstrator; no bulk upload of existing "Open Data" was achieved, anonymous access was not addressed, nor were the 3 services successfully integrated.

Partners of the SD

CERN

"Showing that these three building blocks could be used together in the context of the EOSC Pilot to deliver a discipline-agnostic "open data portal" would have been a powerful vindication of the Pilot. More work needs to be done to address the stumbling blocks, particularly around the area of data ("bit") repositories. The CERN production version runs at the PB scale, so it is known that this can be achieved!"

Contacts
Jamie.Shiers@cern.ch

EOSCPilot.eu has received funding from the European Commission's Horizon 2020 research and innovation programme under the Grant Agreement no 739563.