



Contribution ID: 181

Type: Talk

Cold data support for the CERN Open Data Portal

Monday 21 October 2024 14:42 (18 minutes)

The CERN Open Data Portal holds over 5 petabytes of high-energy physics experiment data, serving as a hub for global scientific collaboration. Committed to Open Science principles, the portal aims to democratize access to these datasets for outreach, training, education, and independent research.

Recognizing the limitations of current disk-based storage, we are starting a project to expand our data storage methodologies. Our approach involves integrating hot storage (such as spinning disks) for immediate data access and cold storage (such as tape, or even interfaces to the experiment frameworks) for cost-effective long-term preservation. This innovative strategy will significantly expand the portal's capacity to accommodate more experiment data. However, we anticipate challenges in navigating technical complexities and logistical hurdles. These challenges include the latency to access cold data, monitoring and automatizing the transitions between hot and cold and ensuring the long-term preservation of data in the experiment frameworks. The strategy is to integrate existing solutions like EOS, FTS, CTA and Rucio.

In our presentation, we will discuss these challenges, present our prototype solution, and outline future developments aimed at enhancing the accessibility, efficiency, and resilience of the CERN Open Data Portal's data ecosystem.

Primary authors: GONZALEZ LOPEZ, Jose Benito (CERN); SAIZ, Pablo (CERN); ZACHARODIMOS, Zacharias (CERN)

Presenter: SAIZ, Pablo (CERN)

Session Classification: Parallel (Track 8)

Track Classification: Track 8 - Collaboration, Reinterpretation, Outreach and Education