

Cold data support for the CERN Open Data Portal

Jose Benito Gonzalez Lopez^{1,*}, *Pablo Saiz*^{1,**}, and *Zacharias Zacharodimos*^{3,***}

¹CERN

Abstract. 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. The paper describes 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.

1 Introduction

This paper presents the approach implemented on the CERN Open Data Portal [1] to reduce storage costs, thereby increasing the volume of data accessible to the public. It begins with a brief introduction to the CERN Open Data Portal, outlining its goals, evolution, and challenges. The focus then shifts to the storage infrastructure, detailing the requirements for data access and the current service implementation. The third section describes the new approach being adopted to optimize storage efficiency. After that, it will present the future work. Finally, the summary and conclusions will be presented.

2 CERN Open Data Portal

The CERN Open Data Portal is a repository for high energy physics experiments. The content is highly curated and openly available to anyone around the world. Eight different experiments, including the four biggest experiments at the Large Hadron Collider (LHC) [2], have published data in this repository. On top of the datasets, the curators from the experiments

*e-mail: jose.benito.gonzalez.lopez@cern.ch

**e-mail: pablo.saiz@cern.ch

***e-mail: zacharias.zacharodimos@cern.ch

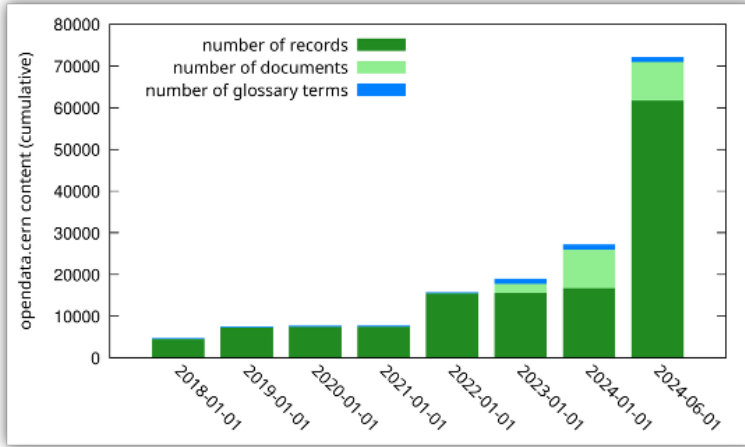


Figure 1. Number of records available on the CERN Open Data Portal

publish the documentation and the frameworks needed to do analysis with the data. All of this makes the content of the CERN Open Data Portal very valuable for multiple objectives: from dissemination and outreach, to education and training, and, of course, for independent research.

At the moment of writing this paper, the repository includes more than 5 PetaBytes of data and more than 80,000 records. The number has increased significantly in the recent years, and it is expected to keep growing (see figure 1). A big portion of those records come from the LHC experiments: ALICE [3], ATLAS [4], CMS [5] and LHCb [6]. DELPHI [7], one of the experiments from the Large Electron Positron Collider (LEP)[8] accelerator, has also recently released part of their data in the repository.

The portal adheres to the FAIR principles [9], ensuring that the data can be found, accessed, it is interoperable and reproducible. Records can be assigned a Digital Object Identifier (DOI) [10], enabling easy citation in publications and ensuring the reproducibility of analysis. Reproducibility requires as well data immutability, long-term preservation policies, and the existence of data replicas. This has to be balanced with the cost of maintaining such an infrastructure: storing the current 5 PetaBytes of data, combined with the new releases planned by the experiments leaves place for several cost-cutting measures.

2.1 Storage

Before this work started, all the space used by the CERN Open Data repository was provided by disks on EOS [11]. The data could be either downloaded from the web portal, or accessed remotely with the xrootd protocol [12].

Out of the 80,000 records, a small subset (around 2,500) are datasets were defined as on demand. In those cases, the information on how to produce the data was stored, instead of data. Users could request access to these data, and the data could be produced on demand. That introduced a delay that could be weeks or months, even for small amounts of data. Following a similar approach, the idea of reducing storage cost by computing costs is one of the cornerstones for the LHCb ntupling service [13].

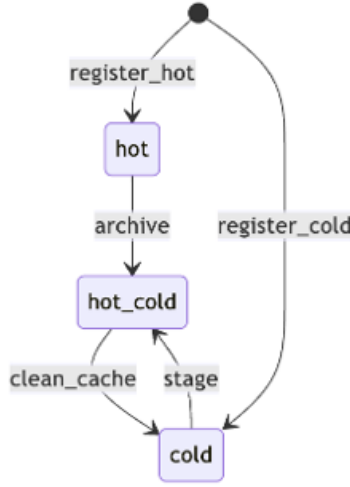


Figure 2. Workflows for cold storage

The approach that this paper describes goes in the direction of having the data stored on different media, favoring cheaper media for datasets that are less demanded, while still giving the possibility to the end user to request the access to the data.

3 Cold Storage

Given the growing volume of data, balancing storage requirements with cost efficiency has become essential. A key advancement enabling this work is the introduction of multiple copies of the same file, each associated with a specific Quality of Service (QoS). As a first step, only two different quality of services have been identified:

1. **Files available online:** These files are stored on disk in the EOS storage system, allowing direct and immediate access for the users. The cost of these copies is higher.
2. **Files available on demand:** These files are stored on a media where the end user does not have direct access. The most common use case, and the one that has been used as a proof of concept, is tape. The files are stored on the CTA [14] system, and they can be fetched using FTS [15].

A record can contain multiple files, and each of those files could be with different QoS. The transitions that a file could go through are depicted in figure 2. Basically, a file could enter the system either in cold or hot storage. A file that is only on cold storage and needs to be accessed will be staged into the hot media. At this point, there will be two copies of the file. Since the space on hot media is limited, at some point the hot copy could be deleted, freeing up the space for other files. In such a case, the only replica left of the file is on cold storage. Note that the cold copy is never deleted.

Since a record contains multiple files, records can have three different status:

1. **Record online:** all the files are online, i.e. in hot storage.

2. **Record partially available:** some of the files of the record are available. This is very useful for cases when records include large datasets, since it allows to give a sample of the data. Users have to request the access to the rest of the files.
3. **Record offline:** all the files are in cold storage. Users have to request access. Depending on the size of all the files of the record, only a subset of the files might be initially brought to hot storage.

It is important to note that the goal of the cold storage is not for backup strategies. The system relies on the different media to address the backup.

3.1 Other cold storage support

As mentioned earlier on, the initial medium used for the cold storage is tape, under the responsibility of the CERN Open Data Portal service managers. A different approach that was also envisioned was to use the data management systems of the experiments to hold the cold copy. This approach brings several big improvements, and at the same time, introduces several challenges. Due to these challenges, it was decided to make a first version with the tape as cold storage could be operational, and only at that point, looking into expanding it with more experiment specific solutions. For the sake of completeness, here are the list of improvements and challenges that the integration with the experiment frameworks would bring:

1. **Benefit: reduce space.** The main benefit of the approach is that there is no need for an extra copy of the data. This solution would take advantage of the space that the experiments already use, and no extra copies of the data will be needed.
2. **Challenge: data sovereignty.** The data would be under full control of the experiment. At the same time, and given the immutability and long-term data preservation mentioned before, the experiment should ensure that the datasets remain unaltered and available on request.
3. **Challenge: integration with experiment framework.** The CERN Open Data Portal would need tighter integration with the framework of the experiments. A Rucio [16] instance would enable interaction with ATLAS and CMS; a similar integration with AliEn [17] would be required for ALICE; and a DIRAC [18] deployment would allow LHCb. All these integrations require that the experiments grant access to the CERN Open Data Portal to fetch data.
4. **Benefit: easier release of data.** At the moment, the first step that the experiments have to do to release their data is to copy it to a place where the CERN Open Data Portal has access to get it. If the experiments could release their data while it is still in their data management system, the data would be transferred only after it is requested for the first time.

3.2 User interface

The CERN Open Data Portal allows any user to browse and download the data. There is no user management, nor the need to login. The introduction of the cold storage brings up the following conundrum: given that the staging of the data can take several days or weeks, what would be the best way of communicating to the users when the requested data is available? On one hand, the access to the data should remain open, and the need of creating accounts (even if they are free and available to everyone) might be seen as a limitation. On the other

hand, asking the users to keep checking the information of the record to see any update left some room for improvement. The approach that was agreed was that users would have the option to leave their emails while requesting the data. If they do, the system would notify them when the data becomes available. If they prefer to not give their emails, they would have to check the same page of the record to see when the data becomes available.

3.3 Admin interface

The requests to stage data will have to be examined, ensuring that they come from real users instead of bots or crawlers. Since it is not clear the number of requests that will be created, the first triage will be done by the service managers of the CERN Open Data Portal. They have direct access to the databases, and they do not require a dedicated interface to see the requests. At some point along the way, the input from the experiment curators might be needed. If this is the case, then a dedicated admin interface will have to be developed where the curators could verify and approve the requests that have been created.

4 Monitoring

Effective monitoring is crucial for ensuring the reliability and performance of the system. On top of the standard key performance metrics that are already collected for the CERN Open Data Portal (like the uptime of the service, number of requests, number of distinct users, etc.), monitoring the available space on the different media becomes critical. The space available on hot media would limit the amount of requests that could be satisfied. On top of that, it should be balanced by the different experiments, ensuring that there is a fair distribution among them.

Since transfers take time, the CERN Open Data Portal has to follow up the status of transfers, rescheduling failed transfers if needed, and updating the repository when the transfers finish successfully. The access of the different files should also be checked, and that would help identifying the datasets that could be deleted from the hot media.

At the same time that the cold storage is being deployed, the monitoring system for the CERN Open Data Portal is also being improved. The first part of this work is to identify the metrics that should be stored long term. The next step would be to identify datasets that are less popular, and therefore good candidates to be moved to cold storage. As a following step, all those moves could be automatized.

5 Future work

The work described in this paper is only the first step in the direction of allowing a seamless integration of cold storage. As stated along this article, several areas have been identified that will require a more in depth evaluation after getting the first feedback on the usage of the system. The following list is a summary of those areas:

1. **Automatic staging and clean up:** Integrating the information from the monitoring with the requests for accessing data in an automatic way would simplify the system maintenance. At the same time, there should be a method to identify the real requests from the ones created by bots and crawlers. The automatic staging will be done as a first step, as long as there is enough available space on the hot media. The automatic deletion will be evaluated later on.

2. **Curator's interface:** A dedicated interface for the curators where they could monitor the usage of hot/cold space and the requests for their data would be very beneficial. This could be expanded at a later stage to facilitate the action of releasing data, giving the required tools and pipelines for verifying the metadata associated with the records.
3. **Interface to follow up requests:** It would bring some added value to have an interface to display the ongoing requests, as well as the completed ones. That might give an estimation of the amount of time that will be needed after users request new transfers.

6 Summary

This paper describes the approach taken by the CERN Open Data Portal to reduce storage costs while ensuring long-term data preservation and accessibility. The CERN Open Data Portal serves as a comprehensive repository for high-energy physics experiments, hosting over 5 petabytes of data and more than 80,000 records from major LHC experiments such as ALICE, ATLAS, CMS, and LHCb. The portal adheres to FAIR principles, ensuring that data is findable, accessible, interoperable, and reproducible.

A key solution to address the growing volume of data is the introduction of multiple copies of the same file with different Quality of Service levels. Files can be stored either online in the EOS storage system for immediate access or on demand in the CERN Tape Archive, retrievable through the File Transfer Service. This approach balances accessibility with cost efficiency by prioritizing storage resources based on file usage patterns.

Additionally, robust monitoring systems are essential for maintaining data availability and performance. Continuous monitoring of file status, retrieval performance, and resource utilization ensures the reliability of both online and on-demand storage. By identifying and addressing potential issues early, the CERN Open Data Portal can provide high-quality service to users while optimizing costs and resources.

This strategy allows CERN to expand the volume of publicly available data, supporting a wide range of scientific, educational, and outreach objectives while maintaining a sustainable storage infrastructure.

References

- [1] <https://opendata.cern.ch/>
- [2] Evans, L., Bryant, P., LHC Machine, JINST 3 (2008) S08001, DOI: 10.1088/1748-0221/3/08/S08001
- [3] ALICE Collaboration, The ALICE experiment at the CERN LHC, JINST 3 (2008) S08002, DOI: 10.1088/1748-0221/3/08/S08002
- [4] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider, JINST 3 (2008) S08003. DOI:10.1088/1748-0221/3/08/S08003.
- [5] CMS Collaboration, The CMS experiment at the CERN LHC, JINST 3 (2008) S08004, DOI: 10.1088/1748-0221/3/08/S08004
- [6] LHCb Collaboration, The LHCb Detector at the LHC, JINST 3 (2008) S08005, DOI: 10.1088/1748-0221/3/08/S08005
- [7] DELPHI Collaboration, The DELPHI detector at LEP, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 1991, volume 303, number 2, DOI: 10.1016/0168-9002(91)90793-P
- [8] Taylor, T., Treille, D., The Large Electron Positron Collider (LEP): Probing the Standard Model, Adv. Ser. Direct. High Energy Phys. 27 (1977) 217-261, DOI: 10.1142/9789814749145_0007

- [9] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- [10] <https://www.doi.org/the-identifier/resources/handbook/>
- [11] Peters, A.J., Janyst, L., Exabyte Scale Storage at CERN, *Journal of Physics: Conference Series* 2011, volume 331, number 5, DOI: 10.1088/1742-6596/331/5/052015
- [12] Dorigo, A., Elmer, P., Furano, F., Hanushevsky, A., XROOTD - A highly scalable architecture for data access, *WSEAS Transactions on Computers*. 4. 348-353.
- [13] Nogga, P., et al. LHCb Open Data Ntupling Service: On-demand production and publishing of custom LHCb Open Data, *Proceedings of the CHEP 2024 conference (to be published)*.
- [14] Davis, M., Afonso, J., Bachmann, R., Bahyl, V., Camarero Vera, J., Leduc, J., Oliver Cortés, P., Rademakers, F., Wardenær, L. and Yurchenko, V. The CERN Tape Archive Beyond CERN An Open Source Data Archival System for HEP 26th International Conference on Computing in High Energy and Nuclear Physics (CHEP 2023). Published in *EPJ Web of Conferences*, volume 295:01048 (2024).
- [15] Murray, S., Patrascioiu, M., Mascetti, L., Lopes, J.P., Misra, S., Silva Junior, E., FTS Service Evolution and LHC Run-3 Operations *EPJ Web of Conf. Volume 295, 2024 26th International Conference on Computing in High Energy and Nuclear Physics (CHEP 2023)*
- [16] Garonne, V., Barisits, M., Beermann, T., Goossens, L., Lassnig, M., Nairz, A., Stewart, G.A., Vigne, V., Serfon, C., Rucio - The next generation of large scale distributed system for ATLAS Data Management, *J. Phys.: Conf. Ser.* 513 (2014) DOI: 10.1088/1748-0221/3/08/S08001
- [17] Saiz, P., Aphecetche, L., Buncic, P., Piskac, R., Revsbech, J. E., Sego, V., AliEn - ALICE environment on the GRID, *Nucl.Instrum.Meth.A* 502 (2003) 437-440, 10.1016/S0168-9002(03)00462-5
- [18] Tsaregorodtsev, A. et al, DIRAC - the distributed MC production and analysis for LHCb, 14th International Conference on Computing in High-Energy and Nuclear Physics, LHCb-PROC-2004-031, CERN-LHCb-PROC-2004-031, DOI: 10.5170/CERN-2005-002.928