

CERN Open Data Portal **4th DPHEP Collaboration**

Pablo Saiz, on behalf of the CERN Open Data team

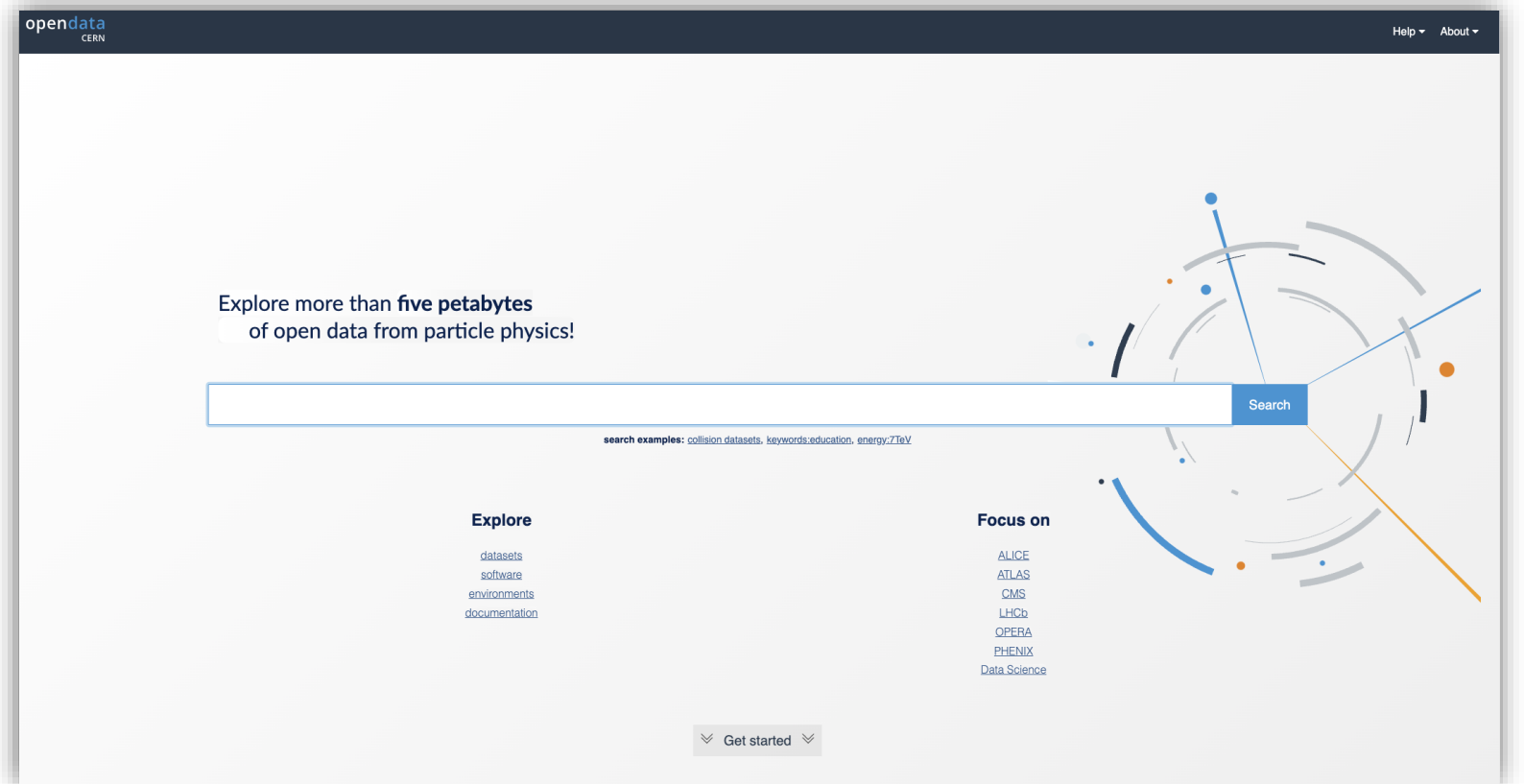
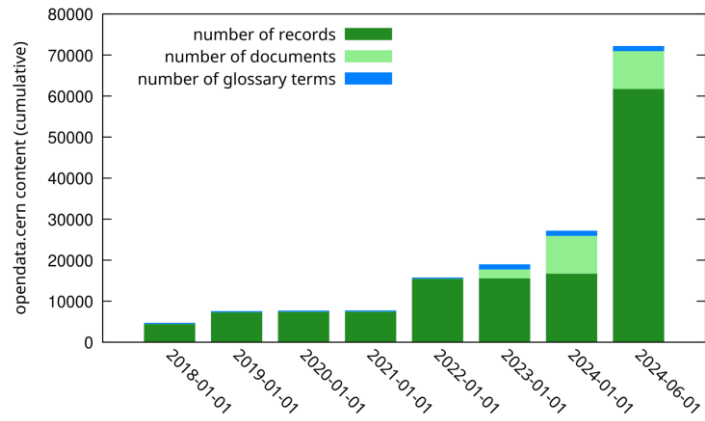
3 Oct 2024

CERN Open Data Portal

<http://opendata.cern>

Repository of HEP data:

- > 5 PT
- > 70k entries
- > 2.5 million files
- 6 HEP experiments



Content news

ATLAS releases 65 TB of open data for research
2024-07-01 by ATLAS Collaboration

[News](#)

Explore over 7 billion LHC collision events – from home

The ATLAS Experiment at CERN has made two years' worth of scientific data available to the public for research purposes. The data include recordings of proton-proton collisions from the Large Hadron Collider (LHC) at a collision energy of 13 TeV. This is the first time that ATLAS has released data on this scale, and it marks a significant milestone in terms of public access and utilisation of LHC data.

"Open access is a core value of CERN and the ATLAS Collaboration," says Andreas Hoecker, ATLAS Spokesperson. "Since its beginning, ATLAS has strived to make its results fully accessible and reusable through open access archives such as Inspire and the ATLAS Open Data Portal. Released under the Creative Commons Attribution-NonCommercial-ShareAlike license, the data represent over 7 billion collision events, representing over 7 billion collision events."

LHCb releases the entire Run I dataset
2023-12-20 by LHCb Collaboration

[News](#)

Today the LHCb collaboration completes the release of the data collected throughout the Run I of the Large Hadron Collider at CERN. The sample made available amounts to approximately 800 terabytes (TB) of data. These data, collected by the LHCb experiment in 2011 and 2012, contains information obtained from proton-proton collisions. The format made available provides pre-filtered data, suitable for a wide range of physics studies. The image below displays an event recorded during 2012.



CMS releases 13 TeV proton collision data from 2016
2024-04-02 by CMS Collaboration

[News](#)

The CMS experiment at CERN is proud to announce the first release of 13 TeV proton-proton collision data collected in 2016. Over 70 TB of 13 TeV collision data and 830 TB of corresponding simulations are now accessible to the global scientific community and enthusiasts alike through the CERN Open Data Portal.



CMS completes Run-1 heavy ion open data collection
2023-09-18 by CMS Collaboration

[News](#)

New release of simulations, proton-lead collision data, and proton reference data

The CMS experiment's extensive collection of heavy ion data from LHC Run-1 (2010-2013) is now accessible through the CERN Open Data portal. Today's release of proton-lead data taken at 5.02 TeV collision energy in 2013 adds to the previously available lead-lead collision data. This accomplishment marks CMS's commitment to open-data sharing, as all proton-proton and heavy ion collision data from the first LHC data-taking period have been made available to the global community within just ten years.



D. Liko: Data preservation and opendata in DELPHI

Important notice: opendata-qa.cern.ch is a quality-assurance server. Please use it for testing purposes only. The content may be erased from time to time. Please use opendata.cern.ch for production.

Running the DELPHI event display

[Documentation](#) [Guide](#)

The DELPHI event display, also known as DELGRA, is used to visualise reconstructed events. It allows to view events as they turn up in the detector, examine tracks, identify jets and much more. This document gives a brief introduction on how to start up this tool.

Important notice: opendata-qa.cern.ch is a quality-assurance server. Please use it for testing purposes only. The content may be erased from time to time. Please use opendata.cern.ch for production.

Raw electron-positron collision event data from the JADE detector (DESY-PETRA)



JADE collaboration
[Dataset](#) [Collision](#) [E+e- physics](#) [JADE](#) [12GeV](#) [e+e-](#) [DESY-PETRA](#)

Description

All data from the JADE detector at the PETRA accelerator taken from 1979 to 1986 with e+e- collision energies ranging from 12 to 46.7 GeV. JADE was one of the detectors responsible for the discovery of the gluon.

R. Hildebrandt: JADE data preservation

Current disk usage

Experiment	Quota (TB)	Used (TB)	%	
ALICE	50	8	15%	
ATLAS	400	165	41%	
CMS	5150	4470	87%	
DELPHI	50	38	76%	
JADE	1	0.65	64%	
LHCb	910	800	89%	
OPERA	0.005	0.004	77%	
PHENIX	0.05	0.00008	0%	

Infrastructure update

7 releases in the current year

- **Streamlined deployment:**
 - Use centralized services DBoD, OpenSearch, Harbor, Monit, ...
 - Decouple content/infrastructure
 - .cern domain
- **Improve data access:**
 - Use eos fuse mount
 - Include new xrootd releases
 - Tune connection timeout for low-bandwidth
- **Frontend:**
 - Move from Angular to Semantic UI
 - Improved facets
 - Improved search

i	OTG0152090	Upgrade of the CERN Open Data Portal	Planned Intervention	Open Science Infrastructure	Open Data Repository	12-09-2024 11:00:00	12-09-2024 11:30:00
i	OTG0150596	Upgrade of the CERN Open Data Portal	Planned Intervention	Open Science Infrastructure	Open Data Repository	27-06-2024 11:00:00	27-06-2024 12:00:00
i	OTG0149747	Upgrade of the CERN Open Data Portal	Planned Intervention	Open Science Infrastructure	Open Data Repository	02-05-2024 10:30:00	02-05-2024 11:30:00
i	OTG0149335	Upgrade of the CERN Open Data Portal	Planned Intervention	Open Science Infrastructure	Open Data Repository	02-04-2024 15:00:00	02-04-2024 16:00:00
i	OTG0149026	Upgrade of CERN Open Data production dat...	Planned Intervention	Open Science Infrastructure	Open Data Repository	26-03-2024 14:00:00	26-03-2024 15:00:00
i	OTG0149023	Upgrade of CERN Open Data development da...	Planned Intervention	Open Science Infrastructure	Open Data Repository	19-03-2024 14:30:00	19-03-2024 15:30:00
i	OTG0147705	Upgrade of the CERN Open Data Portal	Planned Intervention	Open Science Infrastructure	Open Data Repository	29-02-2024 10:00:00	29-02-2024 11:00:00
i	OTG0147250	Upgrade of the CERN Open Data Portal	Planned Intervention	Open Science Infrastructure	Open Data Repository	31-01-2024 11:00:00	31-01-2024 12:00:00

- **Usage statistics:**
 - Dashboard for file access
- **Cold storage prototype**
 - Interface to FTS/CTA
 - Store multiple URI per file

Cold Storage

- **Given:**

- Amount of data: 5 PB, and increasing
- Long term data preservation and data immutability
- Accepted latency
- Multiple copies of data (?) (beware: data ownership)
- Business Continuity and Disaster Recovery

- **Are there any more efficient ways of long term archival?**

- **Possibilities:**

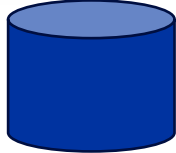
- Computing instead of storage
- Cold Storage. Move part of the data to cheaper storage (Tape, experiment frameworks)

D. Fitzgerald: [LHCb Data preservation and Open Data Activities](#)

Current use case



opendata
CERN



opendata
CERN

HIAllPhysics primary dataset in RECO format from the 2.76 TeV Pb-Pb run of 2010 (/HIAllPhysics/HiRun2010-ZS-v2/RECO)
/HIAllPhysics/HiRun2010-ZS-v2/RECO, CMS collaboration

Cite as: CMS collaboration (2020), HIAllPhysics primary dataset in RECO format from the 2.76 TeV Pb-Pb run of 2010 (/HIAllPhysics/HiRun2010-ZS-v2/RECO). CERN Open Data Portal.
DOI:10.7483/OPENDATA.305G.POEC

There is one publication referring to these data

Dataset Collision Heavy-ion physics CMS 2.76TeV P+P CERN-LHC

Description

HIAllPhysics primary dataset from the 2.76 TeV Pb-Pb run of 2010.

The list of validated runs, which must be applied to all analyses, either with the full validation or with muons only.

Validated runs, full validation

Validated runs, muons only

Dataset characteristics

75482193 events. 28954 files. 151.0 TiB in total.

Dataset characteristics

75482193 events. 28954 files. 151.0 TiB in total.

Options

1. Download files locally

List of files

Please note that the file you are going to download (001DA267-7243-E011-838F-001617C386CE.root) is 4.9 GiB big. On an average ADSL connection, it may take several hours to download it.

Most collaborations provide container images or virtual machine images allowing to perform analyses. If you use one of those, then you do not need to download datasets manually, because all the necessary file chunks will be accessed via the XRootD protocol during the live analysis. Please check the corresponding getting started guides for more details.

Manual download of files via HTTP is only necessary if you would prefer not to use the XRootD protocol for one reason or another.

Cancel Download

00942EEB-1843-E011-BF6C-0025901D5078.root 5.2 GiB

2. Download file index

opendata.cern/record/14011/files/CMS_HiRun2010_HIAllPhysics_ZS-v2_RECO_file_index.txt

```
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/001DA267-7243-E011-838F-001617C386CE.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/006D9FBA-8F42-E011-A129-003048F1CA88.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/00748E08-4642-E011-8701-003048F8F6DC.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/00896E66-7243-E011-8BE9-003048CF99BE.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/00942EEB-1843-E011-BF6C-0025901D5078.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/00C14524-A241-E011-8681-001009F2AF96.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/00CFD683-8F43-E011-9AEB-003048F1C592.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/00EE9498-1443-E011-857F-003048CF65B4.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/02046476-1843-E011-8A5F-003048F8F809.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/020AA340-8541-E011-A913-003048F01118.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/02681178-8042-E011-A691-001009F34488.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/0293F200B-5543-E011-899A-003048D28E06.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/021E6021-8F43-E011-9C4B-003048FCD812.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/027AD9DC-8F42-E011-893A-003048FCD846.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/0298344A-FA44-E011-B08E-0030489454C9.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/023F200B-5543-E011-899A-003048D28E06.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/02AF478F-8042-E011-AACF-0019B9F4A107.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/02C3191F-3443-E011-8893-001009F24FEC.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/02C81108-EE41-E011-9A7A-003048F17C2E.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/021E6021-8F43-E011-9C4B-003048FCD812.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/0419D93F-2D42-E011-941D-003048F174A0.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/041C207D-0341-E011-A719-0025901D5062.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/045A698A-3A42-E011-A797-0025901AF65A.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/04E6F08D-2841-E011-822F-003048F1688C.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/060607D-6C43-E011-9C1A-003048F1DB80.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/06575308-DC43-E011-933A-001009F290CE.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/066F9E00-1F42-E011-8704-003048F11784.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/0686C22B-8E42-E011-8964-003048FD7312.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/069F34AC-6041-E011-A050-001009F23A09.root
root://eospublic.cern.ch/eos/opendata/cms/hidata/HiRun2010/HiAllPhysics/RECO/ZS-v2/0000/069F34AC-6041-E011-A050-001009F23A09.root
```



Cold storage use case

Work in progress



opendata

CERN



Experiment Framework

Dataset characteristics

75482193 events. **28954** files. **151.0 TiB** in total.
50 files currently available. Request access to the rest

Options

1. Download files locally

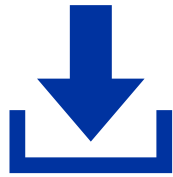
2. Download file index

3. Request staging

```
opendata.com/record/14011/files/CMS_HIRun2010_HIAIPhysics_ZS-v2_RECO_file_index.txt
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/0010A087-7243-E811-836F-8816317C386E.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/0060F8A-8F42-E811-A129-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/00748E08-4642-E811-8718-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/00806156-7243-E811-8818-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/0092E0B-1843-E811-8F1C-8825981D5078.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/00A124-A241-E811-8618-88108925A786.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/00C7083-8F43-E811-9A1B-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/00D4676-1843-E811-8A1F-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/00E83A8-1843-E811-8318-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/00F1178-8042-E811-A631-88108925A786.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/00F81A8-1843-E811-9C3D-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/017A00C-8F42-E811-831A-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/019344-1844-E811-888E-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/02A4F478F-8042-E811-A4C7-88108925A786.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/02B131F-1843-E811-8818-88108925A786.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/02C1188-EE41-E811-9A1D-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/02D130C-8843-E811-950A-88108925A786.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/031903F-2042-E811-9A1D-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/0321C0D-1843-E811-8718-8825981D5078.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/034468A-8A42-E811-4737-8825981D5078.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/0357388-8C43-E811-9C1A-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/036687D-8E42-E811-898A-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/036F980-1F42-E811-878A-883848FC1C88.root
root://eospub.lhc.cern.ch/eos/opendata/cms/hidata/HIRun2010/HIAIPhysics/RECO/ZS-v2/0000/039344C-6841-E811-495B-88108925A786.root
```



Request staging



Ensure that anonymous users can download data



Protect system from abuse



Notify users when data are available

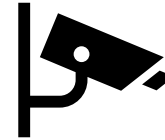
Please, tell us why you want to access these data

Submit

Cold storage challenges

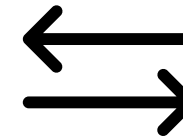
1. Monitoring

1. Identify candidates for cold storage
2. Keep track of cold storage requests



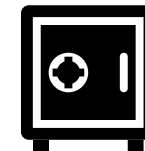
2. Transfer files

- A. CERN Tape Archive: FTS
- B. Experiment frameworks: Dedicated plugins per experiment



3. Data sovereignty

1. Experiment still responsible for data (!)



4. Multiple copies of datasets



Cold storage: current status

 1. Create tools to archive/stage/remove



 2. Store information of multiple copies



 3. Agree on disk/tape quotas



 4. Modify UI: display/search hot copies; request cold copies



 5. Automatic staging



 (Optional) Implement other cold storage alternatives



KPI: Audience

- **KPI for management**
 - Service availability, resources, users,...
- **KPI for service managers/curators**
 - Response time, % errors, accesses per experiment/dataset,
- **KPI for end users**
 - Most cited datasets, most used, ...

New member of Open Data team working on this: D. Horvat

Service Performance Qualification

The screenshot displays the IT SPQ dashboard with the following sections:

- Fabric Services**: Fabric Services at CERN, Fabric Service Dashboards (Print Service IT-FA).
- Computer Security**: Computer Security at CERN, Computing Security.
- Communication Systems**: Communication Systems at CERN, Communication Systems Dashboards (External Connectivity IT-CS, Network Database and Registration Service IT-CS).
- Collaborative Applications**: Collaborative Applications at CERN, Collaborative Applications Service Dashboards (Engineering software service metrics IT-CA, Mail Service IT-CA, **Open Data Service IT-CA** - circled in red).
- CERN Open Data Service**: This dashboard shows the most relevant metrics collected as key performance indicators for the CERN Open Data Service. The Open Data Portal provides users inside and outside of CERN with more than 5 petabytes of publicly available data collected and processed by the experiments through a website. The search portal can be accessed [here](#).
- Base and Analytics**: Analytics at CERN, Analytics Service Dashboards (Reporting and Reporting, On Demand, Storage Service, Storage Service (WIP), Service, J Service, J Service, Arch Service, Oracle Database Service IT-DA).
- Storage Services**: Storage services at CERN, Storage Service Dashboards (Search: AFS Service IT-SD, Ceph Service IT-SD, CERN Tape Archive IT-SD, CERNBox Service IT-SD, EOS Physics Service IT-SD, File Transfer Service (FTS) IT-SD, File Service IT-SD).
- Computing and Device Services**: Computing and Device services at CERN, Computing and Device Service Dashboards (Search: Batch service IT-CD, BOINC service IT-CD, HPC service IT-CD, Linux Operating System IT-CD, Linux Software Building IT-CD, Mac & iOS Service IT-CD, Productivity Software Service IT-CD, Public PC Service IT-CD, Server Provisioning Service IT-CD).

Availability 100%

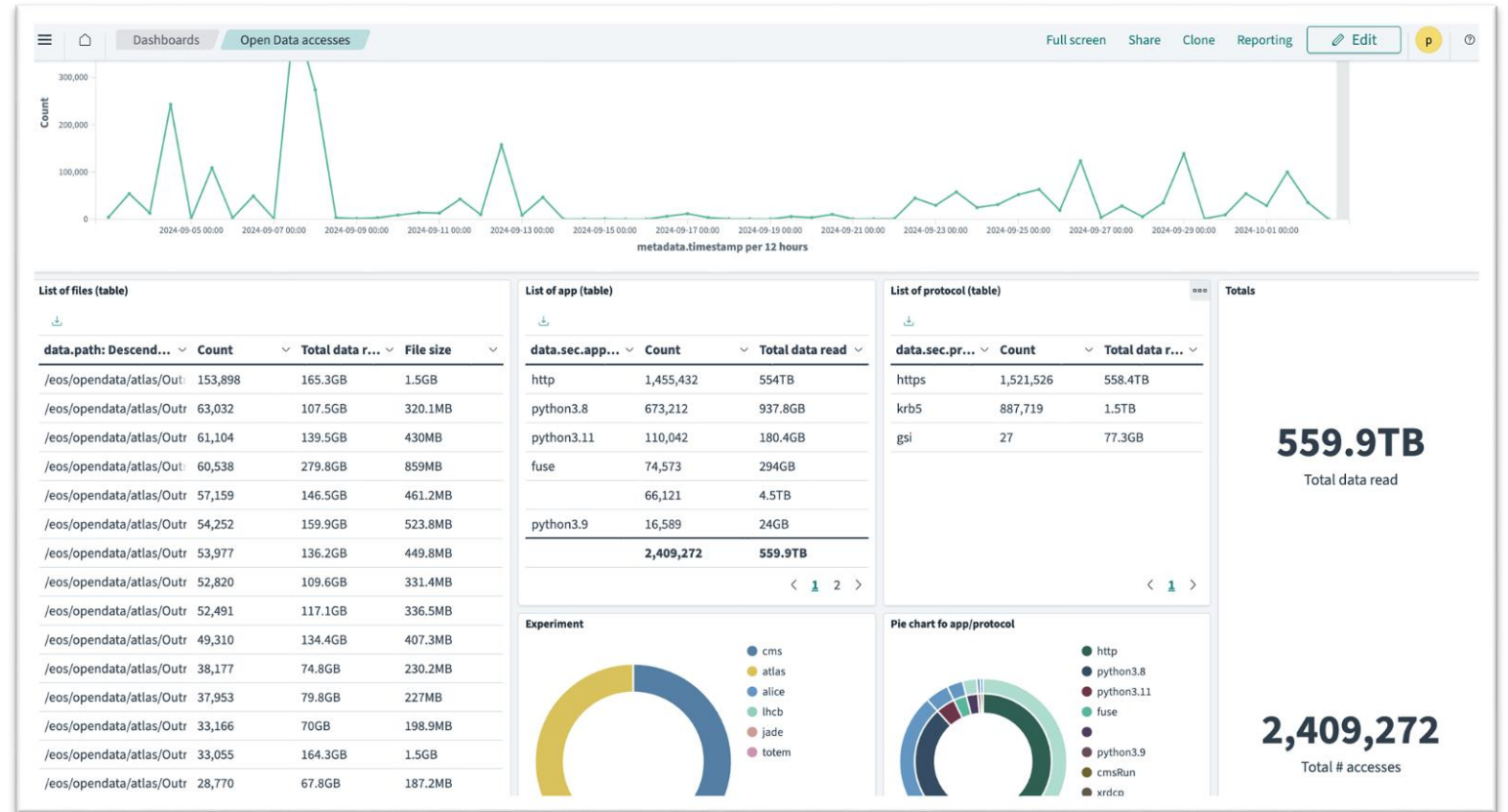
Internal dashboard for Service Admin

Created early this year:

- Using logs from eos
- # accesses
- Volume transferred
- Breakdown per experiment
- Last 30-days

Ongoing:

- Keep long term stats
- Map from file to record
- Open it up
- More breakdown categories
- Prototype with MONIT-GRAFANA



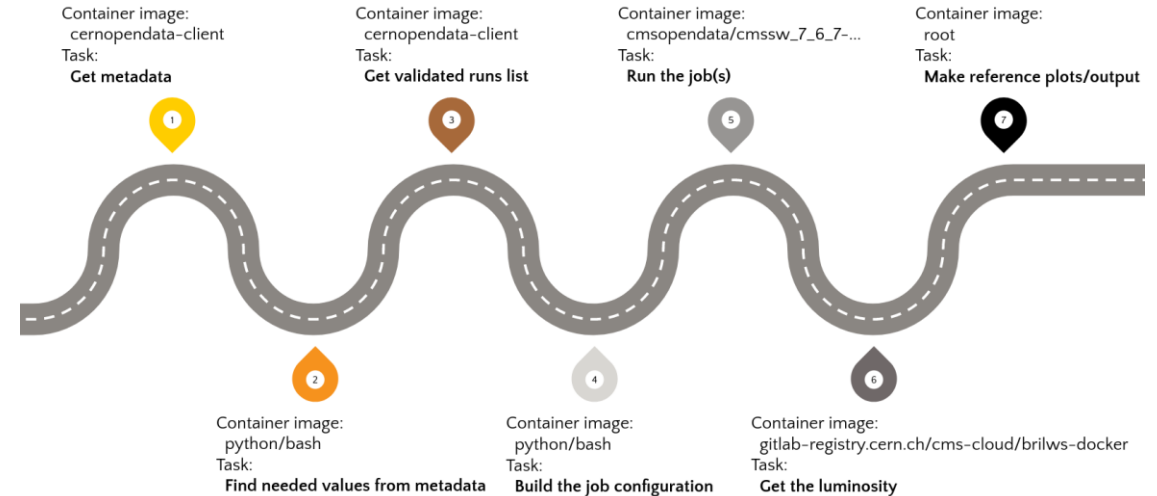
Beyond repositories: usage workflows

cernopendata-client

- Version 0.4.0 released in Aug 2024
- Allows querying metadata relations:
Which gitlab container image for a given dataset?

```
$ cernopendata-client get-metadata \  
--output-value system_details.container_images.name \  
--filter registry=gitlab \  
--doi 10.7483/OPENDATA.CMS.A342.9982
```

Allows to compose data usage workflows





M. Donadoni: [REANA](#) reproducible analysis: status update

CERN Open Data Portal

- Digital repository with 5 PB of data
 - Curated data: datasets, documentation, software...

- Currently used by:      

- 2 more in the process:  

- Current activities:
 - More content
 - Cold data storage
 - KPI

<http://opendata.cern>

opendata-team@cern.ch



home.cern