



CERN (Open) Data Services

Tibor Šimko

@tiborsimko

GL17 · 30 November 2015 · Amsterdam, The Netherlands

Invenio

What is Invenio?

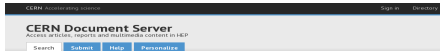
- digital library and document repository software
 - mature platform: first public release in 2002
 - rich data: articles, books, notes, photos, videos, data, code
- originated in high-energy physics
 - institutional repository: CERN Document Server
 - integrated library system: CERN Library
 - disciplinary repository: INSPIRE
- nowadays co-developed by an international collaboration



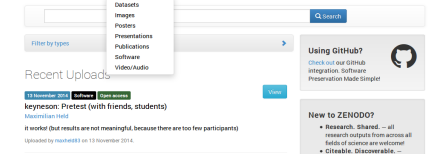
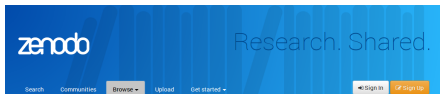
- participating in and collaborating with several EU projects



Invenio applications



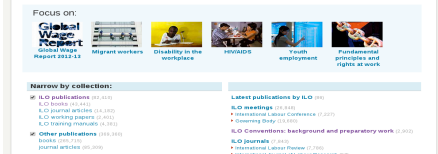
institutional repositories & ILS



long-tail of science



subject-based repositories



world-wide installations

“Small data”

CERN Document Server

CERN Document Server

Access articles, reports and multimedia content in HEP

Search

Submit

Help

Personalize

Search 1,262,525 records for:

Search

[Search Tips](#)

[Advanced Search](#) 

☒ Articles & Preprints (1,086,540)

[Published Articles](#) (534,677) [Preprints](#) (671,836)

[Theses](#) (18,873) [Reports](#) (6,913)

[CERN Notes](#) (36,649)

[Committee Documents](#) (24,269)

☒ Books & Proceedings (100,511)

[Books](#) (71,080) [Proceedings](#) (17,639)

[Standards](#) (11,639) [Design Reports](#) (171)

CERN Articles & Preprints (105,016)

[CERN Published Articles](#) (57,284) [CERN Preprints](#) (17,706) [CERN Theses](#) (4,690)

[CERN Reports](#) (1,180) [Committee Documents](#) (24,269)

CERN Series (18,729)

[CERN Annual Reports](#) (115) [CERN Yellow Reports](#) (1,150) [CERN Theory](#) (13,390)

[Academic Training Lectures](#) (706) [Summer Student Lectures](#) (951) [General Talks](#) (2,577)

CERN Departments (95,148)

“Related data files”

CERN Accelerating science

Sign in Directory

CERN Document Server

Search Submit Help Personalize

Home > Articles & Preprints > Published Articles > First observation of $\bar{B}^0 \rightarrow J/\psi K^+ K^-$ and search for $\bar{B}^0 \rightarrow J/\psi \phi$ decays > Comments

Information Discussion (0) Files

First observation of $\bar{B}^0 \rightarrow J/\psi K^+ K^-$ and search for $\bar{B}^0 \rightarrow J/\psi \phi$ decays - Aaij, R et al - arXiv:1308.5916

Main file(s):

prd.88.e072005

version 1 prd.88.e072005.pdf [1.32 MB] 07 Nov 2013, 15:17 APS Open Acc

Additional file(s):

Related data file(s)

version 1 Related data file(s).zip [10.36 MB] 02 Sep 2013, 16:41

arXiv file(s):

arXiv:1308.5916

version 4 arXiv:1308.5916.pdf [4.36 MB] 26 Oct 2013, 04:15 (cpe:nonunique)

Fig15b.C Fig15b.eps Fig15b.pdf Fig15b.png

Fig16a.C Fig16a.eps Fig16a.pdf Fig16a.png

INSPIRE



Welcome to [INSPIRE](#), the High Energy Physics information system. Please direct questions, comments or concerns to feedback@inspirehep.net.

HEP :: HEPNames :: INSTITUTIONS :: CONFERENCES :: JOBS :: EXPERIMENTS :: JOURNALS :: HELP

HEP Search

High-Energy Physics Literature Database

Use "find" for SPIRES-style search ([other tips](#))

Brief format ▼ Search [Easy Search](#) [Advanced Search](#)

[find j "Phys.Rev.Lett..105*" :: more](#)

HOW TO SEARCH

SPIRES syntax is (mostly) supported (requires "find")

[find a richter, b and t quark and date > 1984](#)
[find j phys.rev.,D50,1140 or j hep,0903,112](#)
[find eprint arxiv:1007.5048](#) (Note the plots available on the detailed record)
[find fulltext "quark-gluon plasma"](#) (Note new "fulltext" operator)
[find a ellis and refersto a witten](#) (Note "refersto")
[find a kane and citedby title SUSY and topcite 200+](#) (Note "citedby")

New techniques:

[1985 richter quark multiplicity](#)
[arXiv:1007.5048](#)
[citedby:author:ellis -refersto:author:witten](#)
[author:randall | author:sundrum cited:450->1350](#)

Additional Help:

[More search tips](#) and [full help](#)

HEP

[Search Tips](#)
[INSPIRE Help](#)
[Corrections](#)
[Additions](#)
[Email Us](#)

INSPIRE

[About INSPIRE](#)
[Recent topics](#)
[HEP Reviews](#)
[SPIRES-HEP \[?\]](#)

RESOURCES

[ADS](#)
[arXiv](#)
[HepData](#)
[PDG](#)

INSPIRE News

2013-03-14 Use INSPIRE
to generate your TeX
bibliography, see
[http://cds.cern.ch/record/1300000](#)

"Data behind plots"



Welcome to INSPIRE, the High Energy Physics information system. Please direct questions, comments or concerns to feedback@inspirehep.net.

HEP :: HEPNAMES :: INSTITUTIONS :: CONFERENCES :: JOBS :: EXPERIMENTS :: JOURNALS :: HELP

Information References Citations Files Plots HepData

Search for new phenomena in final states with large jet multiplicities and missing transverse momentum at $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS experiment - ATLAS Collaboration (Aad, Georges *et al.*) JHEP 1310 (2013) 130 arXiv:1308.1841 [hep-ex] CERN-PH-EP-2013-110

THIS DATA COMES FROM DURHAM HEPDATA PROJECT

DATASETS:

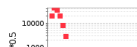
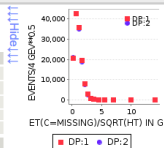
Description: MET/sqrt(HT) distributions for the multi-jet + flavour stream with PTmin=50 GeV, and exactly eight jets, with the signal region selection, other than that on MET/sqrt(HT) itself. A selection of zero b-jets is applied.

[Go to the record](#)

Plain

$|ETARAP(B^-jet)| < 2.5$
 $N - BJETS(p_T > 40 \text{ GeV}) = 0.0$
 $N - JETS(p_T > 50 \text{ GeV}) = 8.0$
 $p p \rightarrow JETS MM$
 $= dATA = MC$
 $|ETARAP(jet)| < 2.0$
 $ET_{MISSING}/\sqrt{HT} \text{ (GeV}^{0.5}\text{)}$ $EVENTS/4 \text{ GeV}^{0.5}$

0.25	20504.0	21017.74 ± 0.25
0.75	42632.0	42806.83 ± 0.00
1.25	35848.0	35159.43 ± 0.00
1.75	19376.0	18926.27 ± 0.00



HEPDATA

“Interactive data behind plots”

[Submit](#)
[Sandbox](#)
[Help](#)
[Your Account](#)

[Browse all](#)
[Cohen, Martin et al.](#)

Ultra- and Hyper-compact HII regions at 20 GHz

Cohen, Martin, Murphy, Tara, Ekers, Ronald D., Green, Anne J., Wark, Robin, Moss, Vanessa

Abstract
We present radio and infrared observations of 4 hyper-compact HII regions and 4 ultra-compact HII regions in the southern Galactic plane. These objects were selected from a blind survey for UCH-II regions using data from two new radio surveys of the southern sky/ the Australia Telescope 20 GHz survey (AT20CG) and the 2nd epoch Molonglo Galactic Plane Survey (MGPS-2) at 843 MHz. To our knowledge, this is the first blind radio survey for hyper- and ultra-compact HII regions. We have followed up these sources with the Australia Telescope Compact Array to obtain H70-alpha recombination line measurements, higher resolution images at 20 GHz and flux density measurements at 30, 40 and 95 GHz. From

Data Abstract
CERN LHC: Measurements of the cross section for ZZ production using the 4l and 22nu decay channels in proton-proton collisions at a centre-of-mass energy of 7 TeV with 4.6 fb⁻¹ of data collected in 2011. The final states used are 4 electrons, 4 muons, 2 electrons and 2 muons and missing transverse momentum, and 2 muons and missing transverse momentum (MET). The cross section values reported in the tables should be multiplied by a factor of 1.0141 to take into account the updated value of the integrated luminosity for the ATLAS 2011 data taking period. The uncertainty on the global normalisation (1 sigma) remains at 5.8%. See Eur.Phys.J. C73 (2013) 2518 for more details. The 4l channel fiducial region is

corrected systematic uncertainty excluding luminosity, the second is the...

[passed review](#)

Table 3

Normalized ZZ fiducial cross section (multiplied by 10⁻⁶ for readability) in bins of the leading reconstructed dilepton pT for the...

[passed review](#)

Table 4

Normalized ZZ fiducial cross section (multiplied by 10⁻⁶ for readability) in bins of the dilepton pT for the 22nu channel...

[passed review](#)

Table 5

Normalized ZZ fiducial cross section (multiplied by 10⁻⁶ for readability) in bins of deltaPhi between the two leptons of the...

[passed review](#)

Table 8

Normalized ZZ fiducial cross section (multiplied by 10⁻⁶ for readability) in bins of the transverse mass of the ZZ system...

[passed review](#)

Normalized ZZ fiducial cross section (multiplied by 10⁻⁶ for readability) in bins of deltaPhi between the two leptons of the leading dileptons for the 4l channel. The first systematic uncertainty is detector systematics, the second is background systematic uncertainties. UPDATE (30 APR 2014): extra significant digit added for first bin

[energies 7000](#) [observables DSIGDPHI](#) [reactions PP--> Z0 Z0 X](#)

Data

RE	P P --> Z0 < LEPTON+ LEPTON- -> Z0 < LEPTON+ LEPTON- -> X
SQRT(5)	7000
Leading dilepton DELTA(PHI)(LEPTON+, LEPTON-) GEV	10**6 * 1/SIG(fiducial) * D(SIG(fiducial))/D(DELTA(PHI)(LEPTON+, LEPTON-))
0.5 - 1	280000 100000, 100000 stat 9000, 9000 sys,detector 400, 400 sys;background
1 - 1.7	260000 80000, 80000 stat 10000, 10000 sys,detector 300, 300 sys;background

[Download Data](#)

Visualize

Deselect variables or hide different error bars by clicking on them.

Variables
10**6 * 1/SIG(fiducial) * D(SIG(fiducial))/D(DELTA(PHI)(LEPTON+, LEPTON-))

stat error
sys,detector error

Copyright - 1975-Present. HEP Data | Powered by Invenio

[About](#) [Team](#) [Contact](#)

Zenodo

The screenshot shows the Zenodo.org website. The header has the Zenodo logo and the tagline "Research. Shared.". Navigation links include Search, Communities, Browse (with a dropdown menu open), Upload, and Get started. The 'Browse' dropdown menu lists: Datasets, Images, Posters, Presentations, Publications, Software, and Video/Audio. On the right, there are "Sign In" and "Sign Up" buttons. Below the navigation bar is a search bar with a "Search" button. A "Filter by types" button is also visible. The main content area features a "Recent Uploads" section with a card for a software upload titled "keyneson: Pretest (with friends, students)" by Maximilian Held, dated 13 November 2014. The card includes a "View" button and a description: "it works! (but results are not meaningful, because there are too few participants)". To the right of the uploads are two informational boxes: "Using GitHub?" and "New to ZENODO?".

zenodo

Research. Shared.

Search Communities Browse Upload Get started Sign In Sign Up

Datasets
Images
Posters
Presentations
Publications
Software
Video/Audio

Search

Filter by types

Recent Uploads

13 November 2014 Software Open access

keyneson: Pretest (with friends, students)
Maximilian Held

it works! (but results are not meaningful, because there are too few participants)

Uploaded by [maxheld83](#) on 13 November 2014.

View

Using GitHub?
Check out our GitHub integration. Software Preservation Made Simple!

New to ZENODO?

- **Research. Shared.** — all research outputs from across all fields of science are welcome!
- **Citeable. Discoverable.** —

"Code you can cite"

The image is a collage of screenshots from the Zenodo website, illustrating the process of creating a citable code repository. The main components are:

- Zenodo Homepage:** Shows the 'Research. Shared.' header and navigation links. The 'Get started' section includes a '1 Flip the switch' step, which involves enabling the repository for automatic generation of releases.
- Releases Page:** Displays the 'Releases' section for the 'instan-cers/decouple' repository. It shows a list of releases, including 'v1.1.3' with a DOI of '10.5281/zenodo.8345'. The 'v1.1.3' release is highlighted, showing its details and download links for 'zip' and 'tar.gz'.
- .zenodo.json File:** A sample JSON file for the repository, showing the 'name', 'affiliation', 'description', 'access_right', 'license', 'related_identifiers', and 'identifier' fields.

Arrows indicate the flow from the 'Get started' guide to the 'Releases' page, and from the 'Releases' page to the '.zenodo.json' file.

<https://guides.github.com/activities/citable-code>

“Data ↔ Code ↔ Paper”

The image displays three overlapping screenshots illustrating the workflow between data, code, and paper:

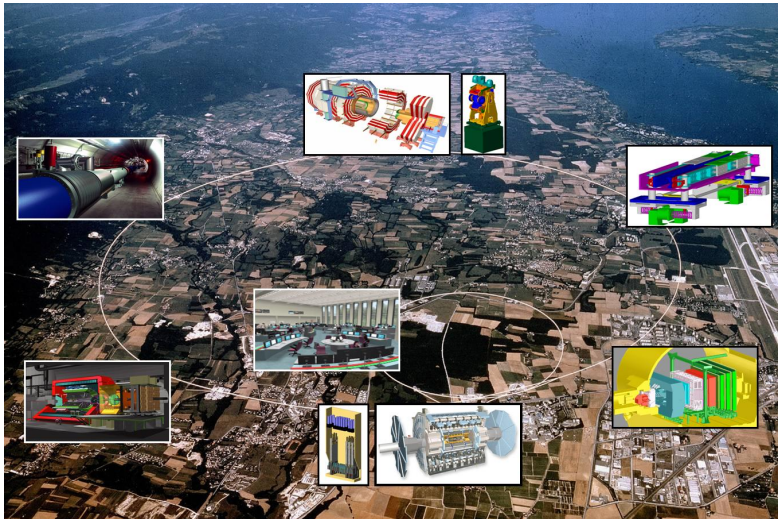
- Top Left (Zenodo):** A screenshot of the Zenodo website showing the 'decouple software associated to arXiv:1401.0080'. It includes a table with columns for Name, Date, and Size, listing various software releases.
- Top Right (INSPIRE):** A screenshot of the INSPIRE website showing the 'decouple software associated to arXiv:1401.0080'. It includes a description of the software and a link to the Zenodo repository.
- Bottom (arXiv):** A screenshot of the arXiv website showing the paper 'A Novel Approach to Higgs Coupling Measurements' by Kyle Cranmer, Sven Kreiss, David Lopez-Fuertes, and Tizian Plehn. The paper is dated Dec 30, 2013, and is 39 pages long. It includes a PDF link and a table of contents.

data (DATAVERSE) ↔ code (ZENODO) ↔ paper (arXiv)

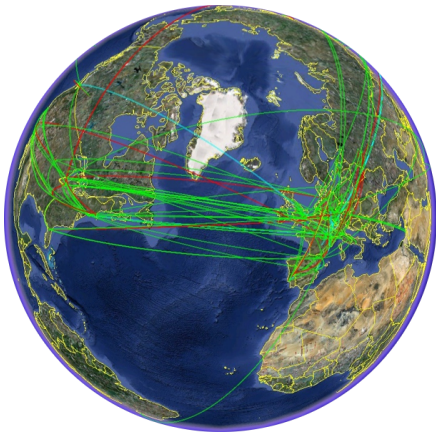
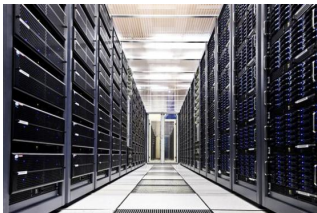
arXiv:1401.0080 · hep-ex/0011057

“Big data”

CERN LHC Experiments

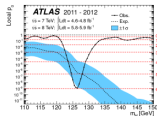
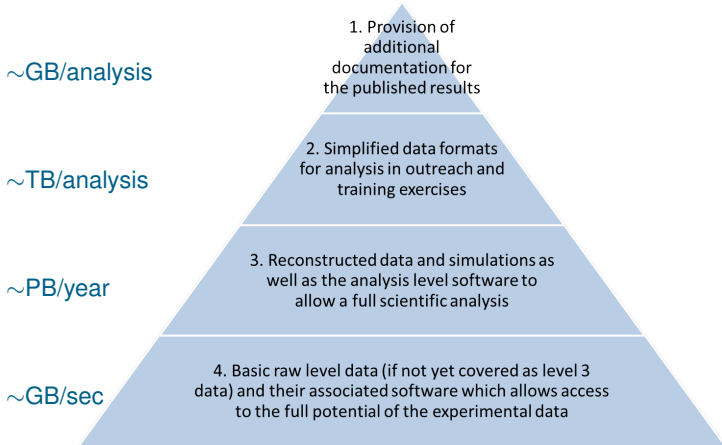


Large Scale Solutions

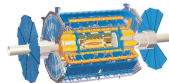


Primary site: 100k cores (10k nodes), 100k disks (50 PB), 21k NIC
Grid: 13 Tier-1 sites, 155 Tier-2 sites, 10 Gbps optical fibre

LHC Data Pyramid

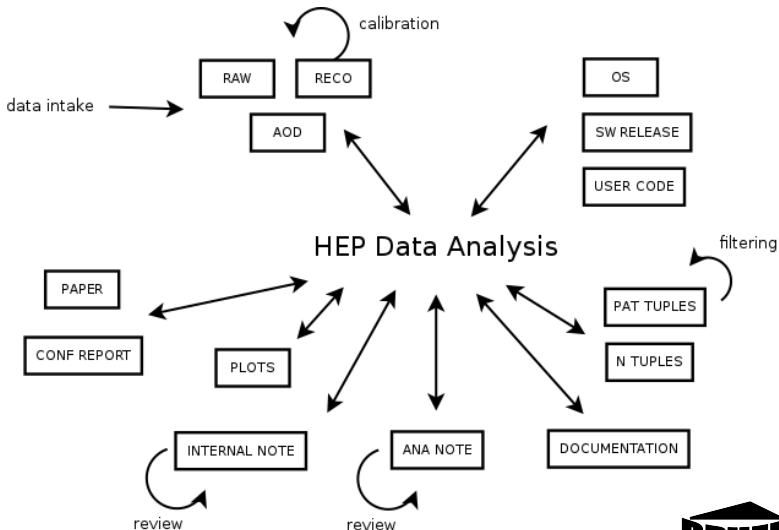


analysis

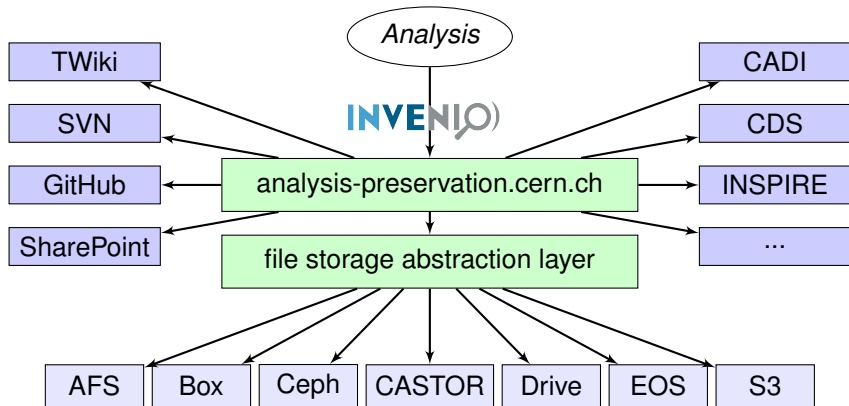


CERN Analysis Preservation

Preserve an analysis?



System architecture



Pilot project

CERN Accelerating science

Sign in

Directory

CERN Analysis Preservation

Demo (your data will NOT be preserved!)

[Search](#) [Deposit](#)

+

⚙

🔍



Example: LHCb analysis

 Delete

 Save

 Submit

Submit an Analysis for LHCb

THIS IS JUST A DEMO. DATA IS *NOT* STORED

Access to all submitted data will be restricted to the LHCb collaboration only.

Basic Information 

 **Analysis Name ***

Please enter Analysis Name

E.g. Bs2psiKS

 **Analysis Number ***

Please enter Analysis Number

E.g. LHCb-ANA-2012-049

Event Samples - Data 

 **DST BK Path**

Please enter path to DST BK

e.g: simz/LHCb/Collision12/Beam4000GeV-VeloClosed-MagDo wn/RealData/Reco14/Stripping20 /90000000 (Full stream)/BHADR

 **Data**

☐ 2010

☐ 2012

My uploads

Unsubmitted

Untitled just now 

-  Submit an Analysis for LHCb
-  Basic Information
-  Event Samples - Data
-  Event Samples MC
-  User Code
-  Final N Tuples
-  Internal Documentation
-  Internal Discussion
-  Presented already?
-  Published already?
-  Other Information
-  Submit

Example: CMS statistics

CERN Analysis Preservation

Demo (your data will NOT be preserved!)

[Search](#) [Deposit](#)

 Delete

 Save

 Submit

CMS Statistics Questionnaire

THIS IS JUST A DEMO. DATA IS NOT STORED

Access to all submitted data will be restricted to the CMS collaboration only.

CMS Statistics Questionnaire

The present questionnaire has the purpose of informing the Statistics Committee on the details of ongoing analyses, in the interest of a quicker and simpler path to approval of physics results. We recommend that you submit your answers at your earliest convenience, regardless of the status of your analysis (in progress, or in advanced stage, or about to be pre-approved). Changes in the statistical analysis after submission can be sent as a simplified "update", using a link you will receive via email after submitting this form.

Please note: each analysis with a separate CADI entry should fill out a separate entry. It is highly advisable that the questionnaire be filled by an author who knows the details of the full analysis, if necessary in consultation with colleagues.

My uploads

Unsubmitted

[Untitled](#)

[just now](#)



 CMS Statistics
Questionnaire

 Other Information

 Submit

Knowledge capture

Final selection step

OS

SLC 5.x

Analysis software

CMSSW

5_3_x

User code

Please enter URL to your code

Tag? e.g.v2.3

Example of supported repositories:
git@github.com:johndoe/myrepo.git
svn@svnweb.cern.ch/cern/wsvn/myrepo

Harvest?

☒ yes, harvest user code

☐ no, create link only

Input Data Files

☒ taken from output of pre-selection step

☐ taken from output of custom analysis step

Output Data Files

Please enter path to data files used

+ Add another file

Example of supported protocols:
xroot://castorpublic.cern.ch/castor/cern.ch/user/j/johndoe/mydir/myfile.root
root://eospublic.cern.ch/eos/lhcb/.../myfile.root
file:///tmp/myfile.root
http://john.doe.example.org/myfile.root

Harvest?

☒ yes, harvest files

☐ no, create link only

Knowledge representation

■ prototype: extended MARC21

- “technical” metadata: beyond bytes

e.g. 256 “computer file characteristics”

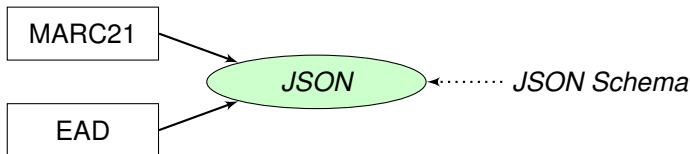
\$a characteristics	\$e events	\$t text
\$b bytes	\$f files	...

- “knowledge” metadata: semantics

e.g. 505 “formatted contents note” CSV column information

\$t title	\$g miscellaneous
-----------	-------------------

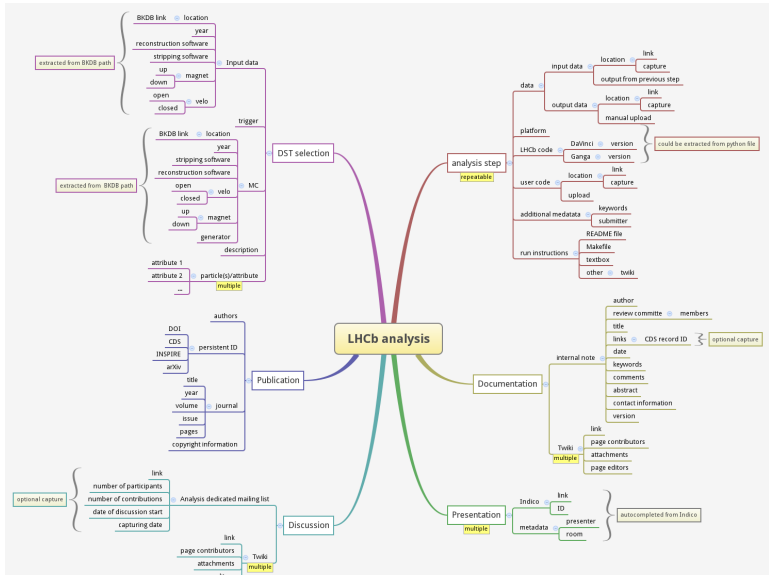
■ internal format: JSON



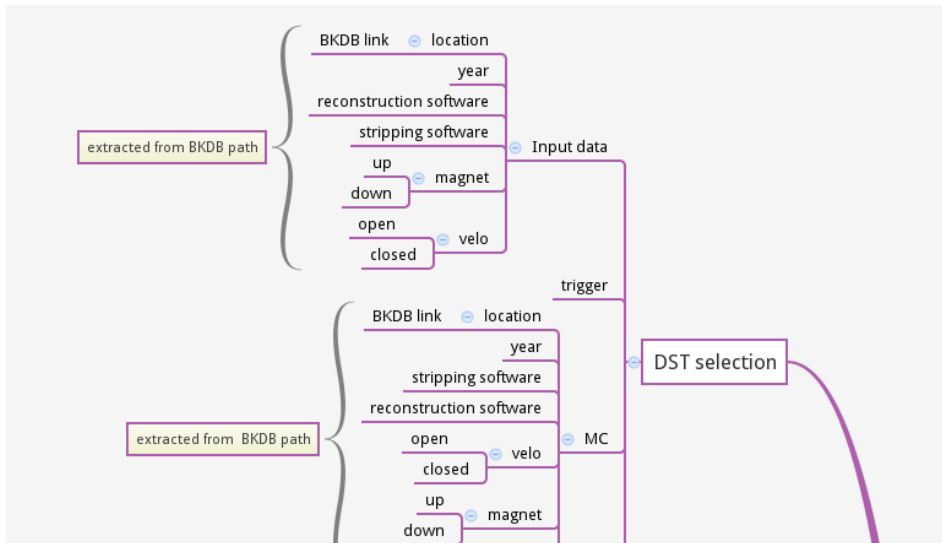
json-schema.org

```
{
  "title": "Example Schema",
  "type": "object",
  "properties": {
    "firstName": {
      "type": "string"
    },
    "lastName": {
      "type": "string"
    },
    "age": {
      "description": "Age in years",
      "type": "integer",
      "minimum": 0
    }
  },
  "required": ["firstName", "lastName"]
}
```

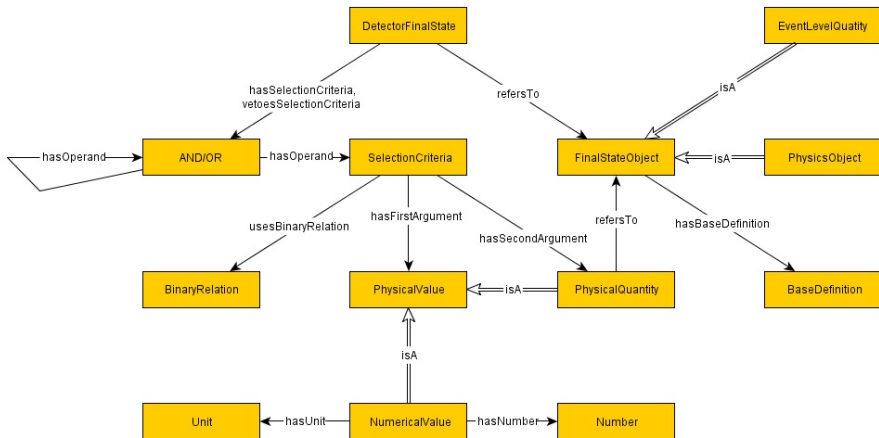
Knowledge modelling



Knowledge modelling



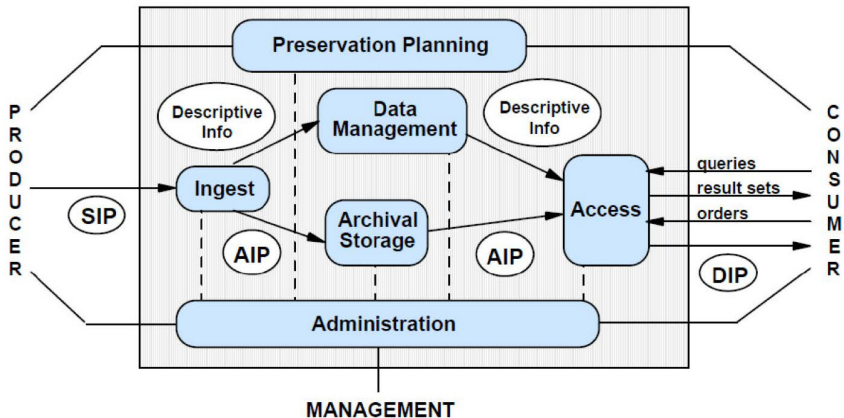
DASPOS



JSON Schema based description

```
"primary_dataset": [  
  {  
    "@type": "dcat:Dataset",  
    "title": "/Mu/Run2010B-Apr21ReReco-v1/AOD",  
    "description": "Mu primary dataset in AOD format from",  
    "licence": "CC0 waiver",  
    "persistent_identifiers": [  
      {  
        "identifier": "10.7483/OPENDATA.CMS.B8MR.C4A2",  
        "scheme": "DOI"  
      }  
    ],  
    "issued": "2011-04-26 11:32:43",  
    "modified": "2011-05-02 21:22:30",
```

Open Archival Info System



SIP = Submission Information Package · AIP = Archival Information Package · DIP = Dissemination Information Package

CERN Open Data

Open data policies

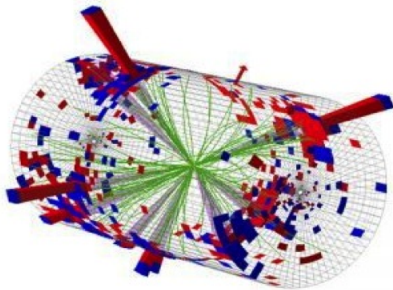
Data policies

- restricted → embargo period → open

“[...] Data with high abstraction, such as AOD, will be conditionally made publicly available after an embargo period of 5 years after publication for 10% of the data and 10 years for 100% of the data [...]” —ALICE Data Policy

Challenges

- audience
 - general public
 - high-school students
 - citizen scientists
 - data miners
- computing
 - exploring in the browser
 - specialised VMs



opendata.cern.ch



[ABOUT](#) [SEARCH](#) [EDUCATION](#) [RESEARCH](#)

Education

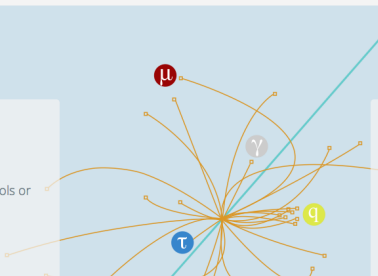
Visualise events, check reconstructed data, run tools or build your own!

Start learning

Research

Get the genuine working environments, virtual machines and datasets to start your research

Start analysing



Education



The CMS (Compact Muon Solenoid) experiment is one of two large general-purpose detectors built on the Large Hadron Collider (LHC). Its goal is to investigate a wide range of physics such as the characteristics of the Higgs boson, extra dimensions or dark matter.

[Explore CMS >](#)



ALICE (A Large Ion Collider Experiment) is a heavy-ion detector designed to study the physics of strongly interacting matter at extreme energy densities, where a phase of matter called quark-gluon plasma forms. More than 1200 scientists are part of the collaboration.

[Explore ALICE >](#)



The ATLAS (A Toroidal LHC Apparatus) experiment is a general purpose detector exploring topics like the properties of the Higgs-like particle, extra dimensions of space, unification of fundamental forces, and evidence for dark matter candidates in the Universe.

[Explore ATLAS >](#)



The LHCb (Large Hadron Collider beauty) experiment aims to record the decay of particles containing b and anti-b quarks, known as B mesons. The detector is designed to gather information about the identity, trajectory, momentum and energy of each particle.

[Explore LHCb >](#)

For education purposes, the complete primary data need to be processed into a format (examples below that is good for simpler applications). Get in touch if you wish to build your own applications similar to those shown here



[Visualise events >](#)



[Visualise histograms >](#)



[Learning Resources >](#)

Research



To analyse CMS data, a Virtual Machine with the CMS analysis environment is provided. The data can be accessed directly through the VM. In the primary datasets, no selection nor identification criteria have been applied. For this release, no simulated Monte Carlo datasets are provided.

[Explore CMS >](#)



According to the ALICE data preservation strategy, reconstructed data and Monte Carlo data as well as the analysis software and documentation needed to process them will be made available on a time scale of 5 years (for 10% of the data). Thus, the first release of ALICE research data will happen in 2016.



According to the ATLAS Data Access Policy, reconstructed data and accompanying tools will be released after reasonable embargo periods.



According to the LHCb External Data Access Policy, reconstructed data and accompanying tools will be released after reasonable embargo periods.

For research purposes, specific software environments and tools need to be deployed to analyse these complex primary data. In addition to the data below, you will find introductions for setting up your working environments here

[Install your Virtual Machine >](#)



[Start analysing the data >](#)



Terms of

@tiborsimko

38 / 50

Visualise detector events

opendata

ABOUT SEARCH EDUCATION RESEARCH

Home > Education > Visualise Events > CMS

Explore CMS open data and visualise detector events

Need HELP?

/Mu.JgEvents/Run_146436/Event_90626440

Detector Model

- Tracker Barrels
- Tracker Endcaps
- ECAL Barrel
- ECAL Endcaps
- ECAL Preshower
- HCAL Barrel
- HCAL Endcaps
- HCAL Outer
- HCAL Forward
- Drift Tubes (muon)
- Cathode Strip Chambers (muon)
- Resistive Plate Chambers (muon)

Tracking

- Tracks (reco.)

ECAL

- Barrel Rec. Hits
- Endcap Rec. Hits
- Preshower Rec. Hits

HCAL

- Barrel Rec. Hits
- Endcap Rec. Hits
- Forward Rec. Hits
- Outer Rec. Hits

Muon

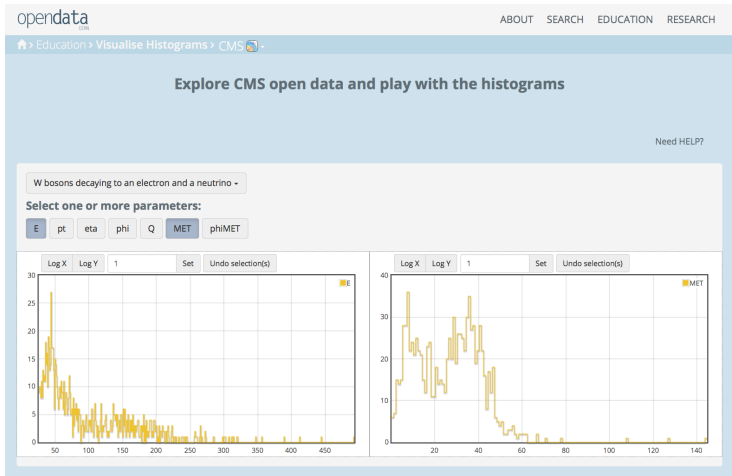
- Matching muon chambers

Physics Objects

- Electron Tracks (GSP)
- Tracker Muons (Reco)
- Stand-alone Muons (Reco)
- Global Muons (Reco)
- Calorimeter Energy Towers
- Jets
- Mixing Bz (Barrel)

The image shows a 3D visualization of the CMS detector, which is a large cylindrical structure. The detector is composed of several layers, including the Tracker Barrels, Tracker Endcaps, ECAL Barrel, ECAL Endcaps, ECAL Preshower, HCAL Barrel, HCAL Endcaps, HCAL Outer, HCAL Forward, Drift Tubes (muon), Cathode Strip Chambers (muon), and Resistive Plate Chambers (muon). The visualization shows a central point where a particle collision occurred, with numerous tracks radiating outwards. The tracks are colored in various colors (blue, green, red, yellow) and are shown as lines connecting points on the detector layers. The detector is shown in a perspective view, with the front face of the barrels visible. A 3D coordinate system (x, y, z) is shown in the bottom right corner of the visualization area.

Basic histogramming



CMS primary datasets

Photon primary dataset in AOD format from RunB of 2010 (/Photon/Run2010B-Apr21ReReco-v1/AOD) 2014

/Photon/Run2010B-Apr21ReReco-v1/AOD

CMS collaboration

Cite as: CMS collaboration (2014). Photon primary dataset in AOD format from RunB of 2010 (/Photon/Run2010B-Apr21ReReco-v1/AOD). CERN Open Data Portal. DOI: [10.7483/OPENDATA.CMS.QKAX.PSWG](https://doi.org/10.7483/OPENDATA.CMS.QKAX.PSWG)

Collection CMS Primary Datasets Collision Energy 7TeV Accelerator CERN-LHC Experiment CMS

Description

Photon primary dataset in AOD format from RunB of 2010

Characteristics

Dataset: 25465895 events 2814 files 2.6 TB in total

System Details

Software release: CMSSW_4_2_1_patch1

Indexes

CMS_Run2010B_Photon_AOD_Apr21ReReco-v1_0002_file_Index.txt Description: Photon AOD dataset file index (3 of 6) for access to data via CMS virtual machine	Size: 41.8 kB	Download index
CMS_Run2010B_Photon_AOD_Apr21ReReco-v1_0001_file_Index.txt Description: Photon AOD dataset file index (2 of 6) for access to data via CMS virtual machine	Size: 55.8 kB	Download index
CMS_Run2010B_Photon_AOD_Apr21ReReco-v1_0004_file_Index.txt	Size: 46.6 kB	Download index

CMS primary datasets

0072FAED-2471-E011-B7D2-0018FE2930C6.root xrootd

Size: 527.0 MB

Download

Files 1 - 5 out of 2814

How were these data selected?

Events stored in this **primary dataset** were selected because of presence of at least one high-energy photon in the event.

How were these data validated?

During data taking all the runs recorded by CMS are certified as good for physics analysis if all subdetectors, trigger, lumi and physics objects (tracking, electron, muon, gamma, jet and MET) show the expected performance. Certification is based first on the offline shifters evaluation and later on the feedback provided by detector and Physics Object Group experts. Based on the above information, which is stored in a specific database called Run Registry, the Data Quality Monitoring group verifies the consistency of the certification and prepares a json file of certified runs to be used for physics analysis. For each reprocessing of the raw data, the above mentioned steps are repeated. For more information see:

- ☐ CMS data quality monitoring: Systems and experiences
- ☐ The CMS Data Quality Monitoring software experience and future improvements
- ☐ The CMS data quality monitoring software: experience and future prospects

How can you use these data?

You can access these data through the CMS Virtual Machine. See the instructions for setting up the Virtual Machine and getting started in

- ☐ How to install the CMS Virtual Machine
- ☐ Getting started with CMS open data

Issues & Limitations

IPython notebooks

[nbviewer](#)[FAQ](#)[IPython](#)[Jupyter](#)

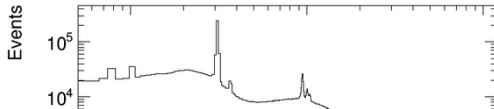
Write the analysis code

Loop on CMS dimuon events in a CSV file and calculate the dimuon invariant mass. These events were extracted from the CMS Mu Primary Dataset on the CERN Open Data Portal: <http://opendata.cern.ch/record/14>. Thanks to <http://openstack.cern.ch> for providing a CernVM running SL5 where I could set up CMSSW 4.2.8 and to <https://github.com/tpmccauley/dimuon-filter> for generating the CSV from CMS AODs.

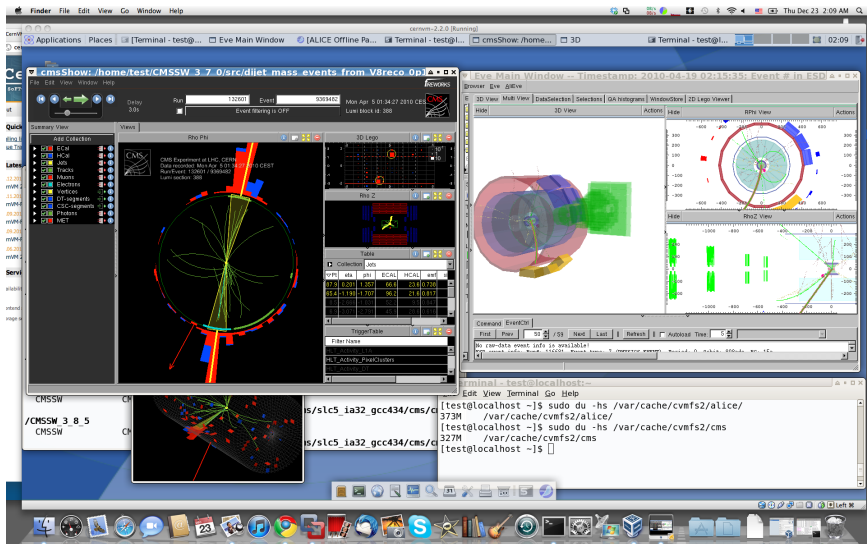
```
In [3]: def analysis(filename):  
        import csv  
        from math import sqrt  
        from rootpy.plotting import Hist  
        from root_numpy import hist2array  
  
        h = Hist(1500, 0.5, 120)  
  
        with open(filename, 'rb') as csvfile:  
            reader = csv.reader(csvfile, delimiter=',')  
            header = reader.next()  
            column = dict([(header[i], i) for i in range(len(header))])
```

```
In [8]: plot(arrays, Hist(1500, 0.5, 120))
```

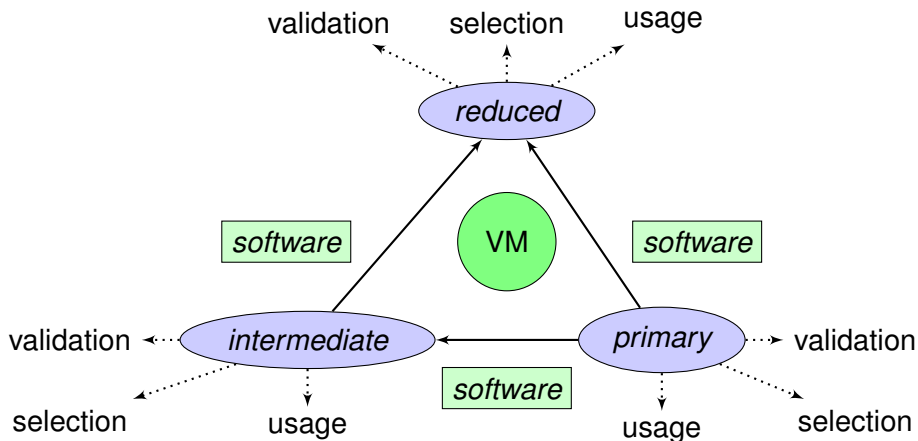
Out[8]:



CernVM virtual machines



OS $\leftrightarrow$ Data $\leftrightarrow$ Software

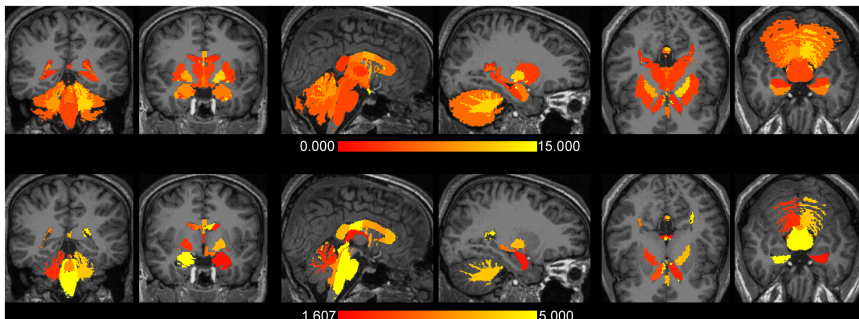


OS/Software influence in medecine

The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements

Ed H. B. M. Gronenschild , Petra Habets, Heidi I. L. Jacobs, Ron Mengelers, Nico Rozendaal, Jim van Os, Machteld Marcelis

Published: June 1, 2012 • DOI: 10.1371/journal.pone.0038234



$8.8 \pm 6.6\%$ (volume) and $2.8 \pm 1.3\%$ (cortical thickness)

Open data? Who cares?



Run your own experiment using CERN's public LHC data

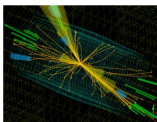
18:30 24 November 2014 by Jacob Aron
For similar stories, visit the [The Large Hadron Collider](#) Topic Guide

Why build your own particle accelerator when you can borrow CERN's? The home of the Large Hadron Collider near Geneva, Switzerland, has started putting data from its experiments online for anyone to use. They hope it could fuel education, art and perhaps even physics discoveries.

"It's very important that we keep this data open and usable," says Kati Lassila-Perini of the CMS experiment at the LHC, which has uploaded 27 terabytes of data to the new [CERN Open Data Portal](#). A web interface lets you visualise the paths of particles created by collisions at the LHC, or you can work directly with the data for more serious analysis.

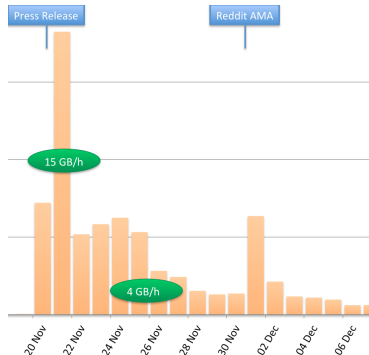
Other LHC experiments have uploaded smaller data sets for educational purposes, allowing budding particle physicists to try their hand at massive-scale physics. The data could also be turned into art or music, as was done previously at a [CERN arts festival](#).

It's possible that physicists might search through the data and make discoveries, but most people with the necessary skills already work with similar experiments, says Lassila-Perini. "There are not so many spare physicists around."



Maybe you'll be able to spot a particle in data from the LHC (image: CERN)

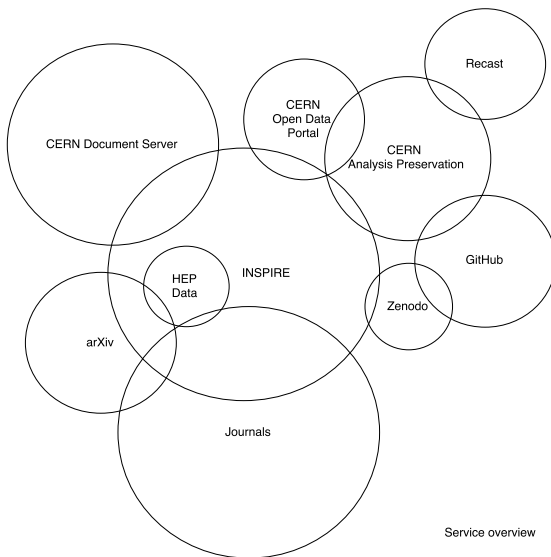
ADVERTISEMENT



82,000 distinct users visited the site
21,000 distinct users viewed data records
16,000 distinct users used event display
3,000 distinct users used histogramming

Conclusions

CERN (open) data services



Service overview

CERN (open) data services



Invenio

- <http://invenio-software.org>
- <http://github.com/inveniosoftware>
- [@inveniosoftware](#)
- info@invenio-software.org



CERN Analysis Preservation

- <http://analysis-preservation.cern.ch>
- <http://github.com/cernanalysispreservation>
- analysis-preservation-development@cern.ch



CERN Open Data

- <http://opendata.cern.ch>
- <http://github.com/cernopendata>
- opendata-development@cern.ch

CERN IT J. Delgado Fernandez, J. Kunčar, T. Smith, T. Šimko · **CERN Library** S. Dallmeier-Tiessen, P. Fokianos, P. Herterich · **ALICE** M. Gheata, C. Grigoras · **ATLAS** K. Cranmer, L. Heinrich, D. Rousseau, F. Socher · **CMS** A. Calderon, A. Huffman, K. Lassila-Perini, T. McCauley, A. Rao, A. Rodriguez Marrero · **LHCb** S. Amerio, B. Couturier, A. Trisovic · **CERN CernVM** J. Blomer · **CERN EOS** L. Mascetti · **DASPOS** M. Hildreth · **DPHEP** F. Berghaus, J. Shiers