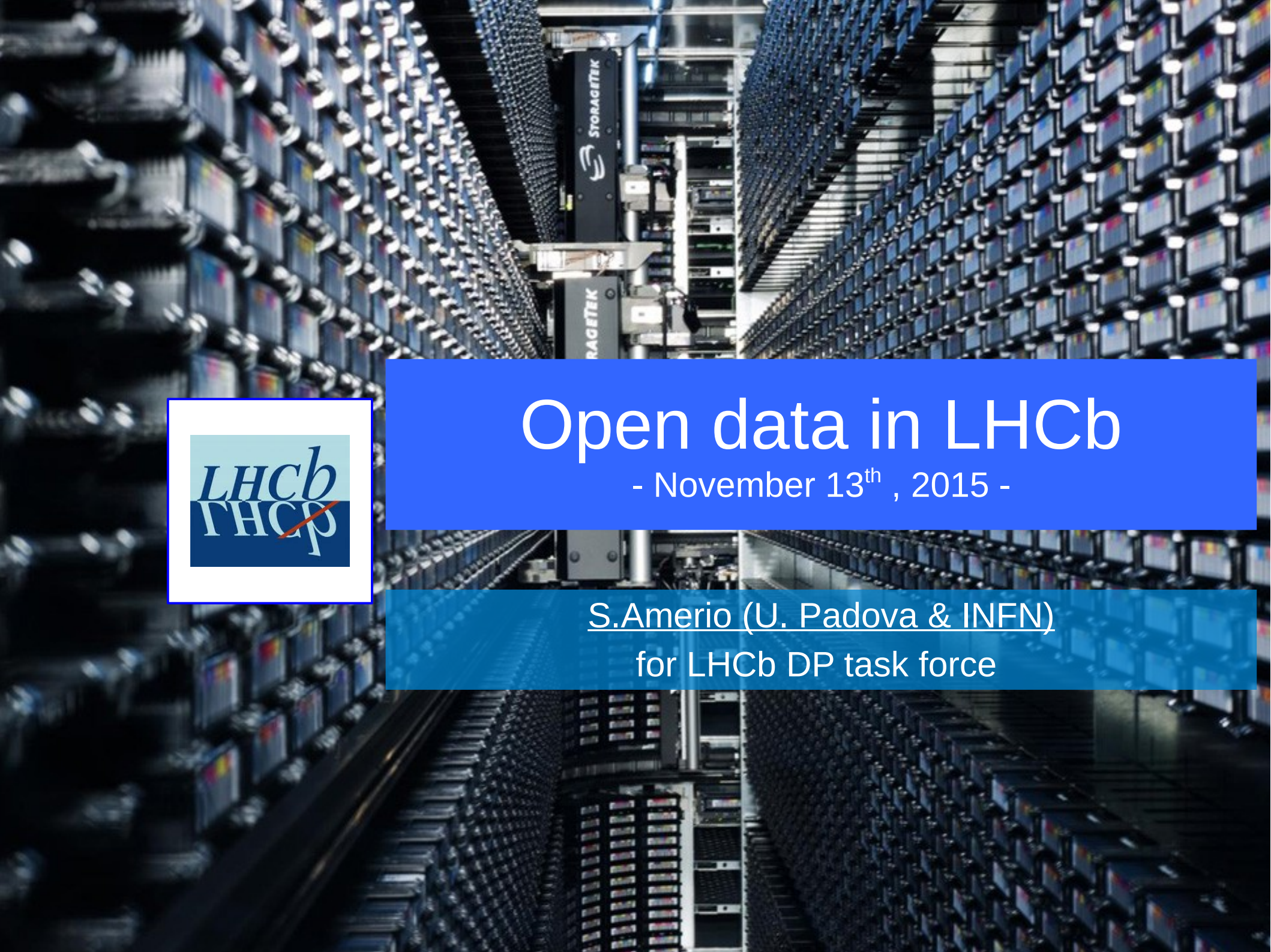


Open data in LHCb

- November 13th , 2015 -

S.Amerio (U. Padova & INFN)

for LHCb DP task force

LHCb Open Access policy

http://opendata.cern.ch/record/410/files/LHCb-Data-Policy.pdf

Official since March 2013.
It follows DPHEP leves of data preservation.

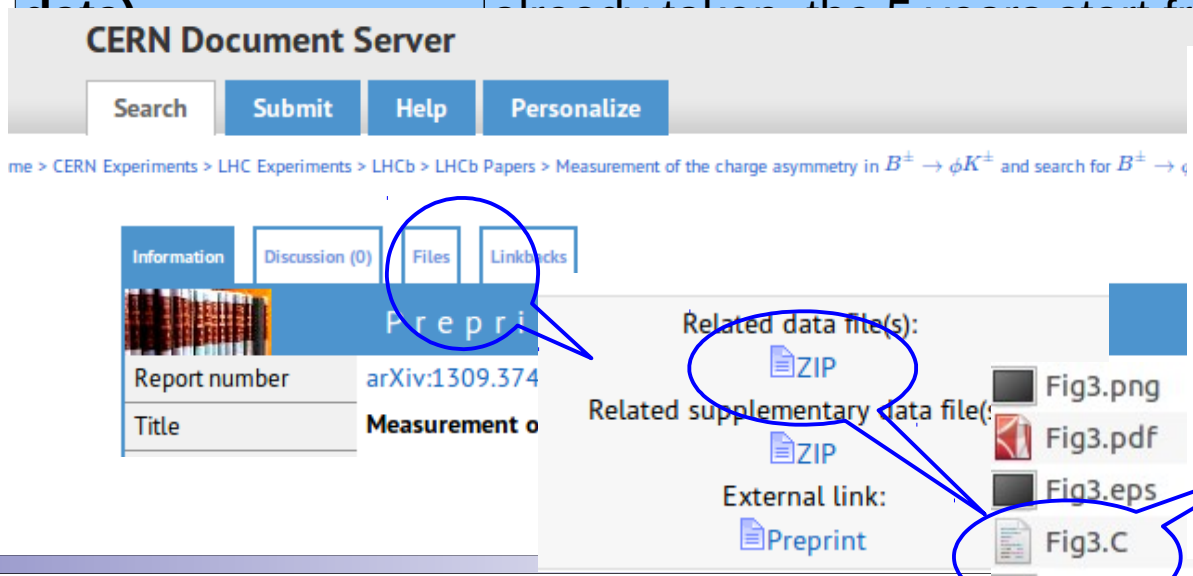
Level 1 (published data)	All scientific results are public. Data associated with the results will also be made available; format and repositories will be decided by the Editorial Board
Level 2 (samples for educational purposes)	LHCb already involved in outreach and education activities. Event displays and simple analysis level ntuples are already available and will continue to be provided to the public. The data are for educational purpose only, not suitable for publication
Level 3 (reconstructed data)	LHCb will make **reconstructed data (DST) available to open public; 50% 5 years after data is taken, 100% after 10 years.**
Level 4 (raw data)	Due to the complexity of the raw data processing stage, the extensive computing resources required and enormous access to tape resources, direct access to raw data is not permitted to individuals within the collaboration. Raw data processing is performed centrally. Due to this, the collaboration is currently **not planning to allow open access to raw data**

Level 1 Open Data

Official since March 2013.
It follows DPHEP leves of data preservation.

Level 1 (published data)	All scientific results are public. Data associated with the results will also be made available; format and repositories will be decided by the Editorial Board
Level 2 (samples for educational p	LHCb already involved in outreach and education activities. Event displays and simple analysis level ntuples are already available and will continue to be provided to the public. The data are for educational
L (reconstructed	**available to open public; ter 10 years.** For the data cation of the policy

- Additional data to HepDATA
- Root macro to create the published plots




```
TGraph *graph = new TGraph(61);
graph->SetName("graph_allcpv_p");
graph->SetTitle("");
graph->SetFillColor(1);
graph->SetLineColor(4);
graph->SetLineWidth(2);
graph->SetMarkerStyle(20);
graph->SetMarkerSize(0.6);
graph->SetPoint(0,-0.06221261421,7.236171847);
graph->SetPoint(1,-0.05650273742,6.969961101);
graph->SetPoint(2,-0.04744748815,6.703843144);
graph->SetPoint(3,-0.03719151822,6.437773779);
graph->SetPoint(4,-0.03218946446,6.304733425);
graph->SetPoint(5,-0.02622174243,6.171732888);
```

Level 2 Open Data

Official since March 2013.
It follows DPHEP leves of data preservation.

Level 1 (published data)	All scientific results are public. Data associated with the results will also be made available; format and repositories will be decided by the Editorial Board
Level 2 (samples for educational purposes)	LHCb already involved in outreach and education activities. Event displays and simple analysis level ntuples are already available and will continue to be provided to the public. The data are for educational purpose only, not suitable for publication
Level 3 (reconstructed data)	LHCb will make **reconstructed data (DST) available to open public; 50% 5 years after data is taken, 100% after 10 years.** For the data already taken, the 5 years start from the ratification of the policy
Level 4 (raw data)	Due to the complexity of the processing stage, the extensive computing resources required and enormous access to tape resources, direct access to raw data is not permitted to individuals within the collaboration. Raw data processing is performed centrally. Due to this, the collaboration is currently **not planning to allow open access to raw data**

Samples for masterclass exercise

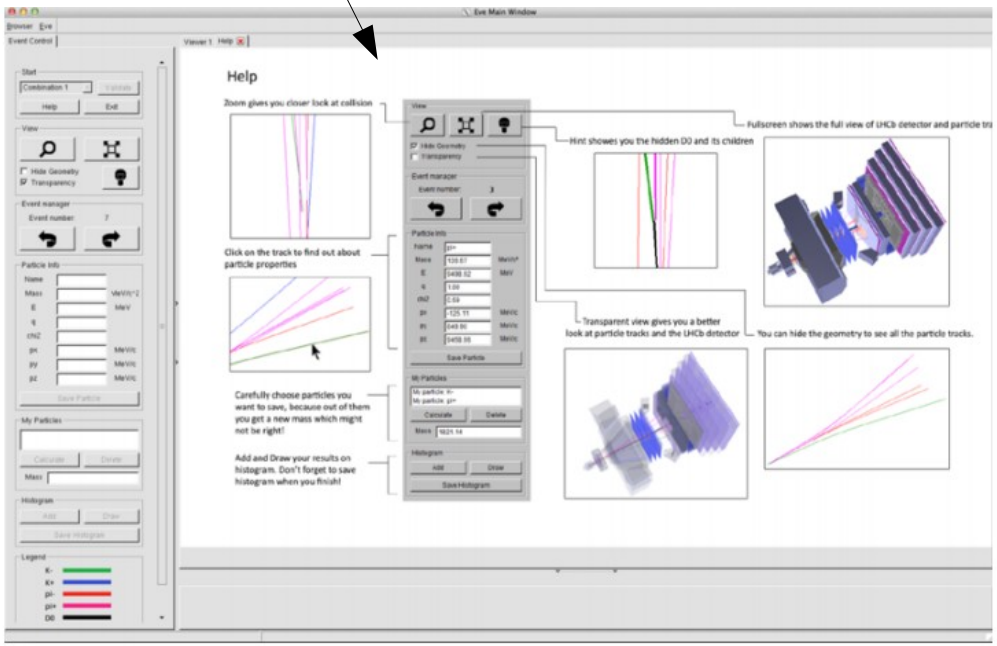
Masterclass exercises

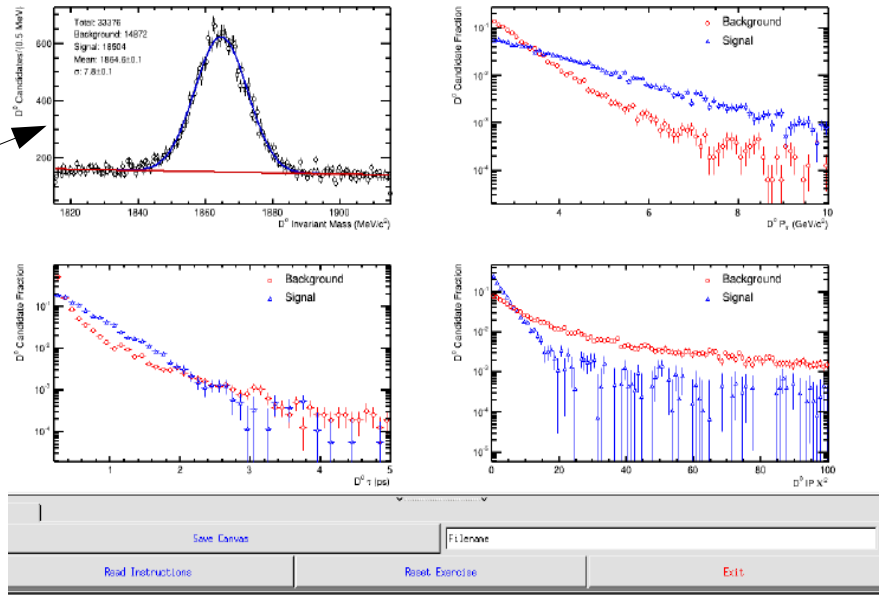
DATA

- Event display data (5k events from 2011 data taking)
- $D^0 \rightarrow$ K pi data (60k events from 2011 data taking)
- Format: ROOT trees

SOFTWARE

- LHCb **virtual machine image**
- **Event display and D^0 lifetime analysis** software (ROOT)

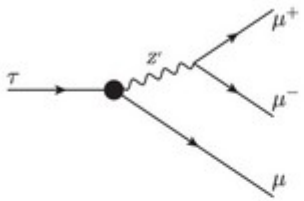




Accessible from a dedicated webpage and from the *For Education* area of the Cern Open Data portal.

Kaggle competition

kaggle | Host | Competitions | Scripts | Jobs | Community ▾ | Sign up | Login



Completed • $15,000 • 673 teams

Flavours of Physics: Finding τ → μμμ

Mon 20 Jul 2015 – Mon 12 Oct 2015 (30 days ago)

Dashboard

Home
- Data
- Make a submission

Information
- Description
- Evaluation
- Rules
- Prizes
- About the Sponsors
- Agreement test
- Correlation test

Competition Details » Get the Data » Make a submission

Identify a rare decay phenomenon

Like last year's Higgs Boson Machine Learning Challenge, this competition deals with the physics at the Large Hadron Collider (LHC). However, the subject of last year's challenge, the Higgs Boson, was already known to exist. The aim of this year's challenge is to find a phenomenon that is not already known to exist – charged lepton flavour violation – thereby helping to establish "new physics".

Working to include the data in the Open Data portal.

https://www.kaggle.com/c/flavours-of-physics

Level 3 Open Data

Official since March 2013.
It follows DPHEP leves of data preservation.

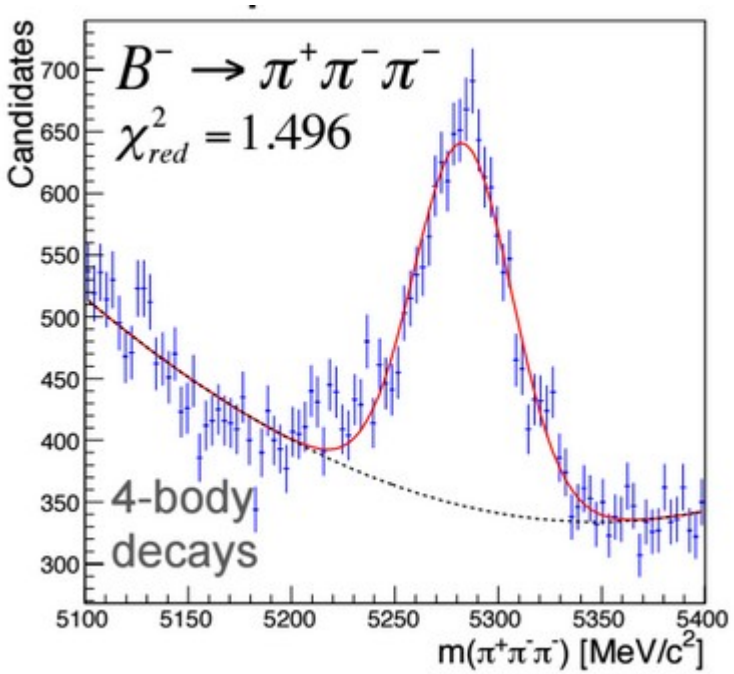
Level 1 (published data)	All scientific results are public. Data associated with the results will also be made available; format and repositories will be decided by the Editorial Board
Level 2 (samples for educational purposes)	LHCb already involved in outreach and education activities. Event displays and simple analysis level ntuples are already available and will continue to be provided to the public. The data are for educational purpose only, not suitable for publication
Level 3 (reconstructed data)	LHCb will make **reconstructed data (DST) available to open public; 50% 5 years after data is taken, 100% after 10 years.**
Level 4 (raw data)	Due to the complexity of the raw data processing stage, the extensive computing resources required and enormous access to tape resources, direct access to raw data is not permitted to individuals within the collaboration. Raw data processing is performed centrally. Due to this, the collaboration is currently **not planning to allow open access to raw data**

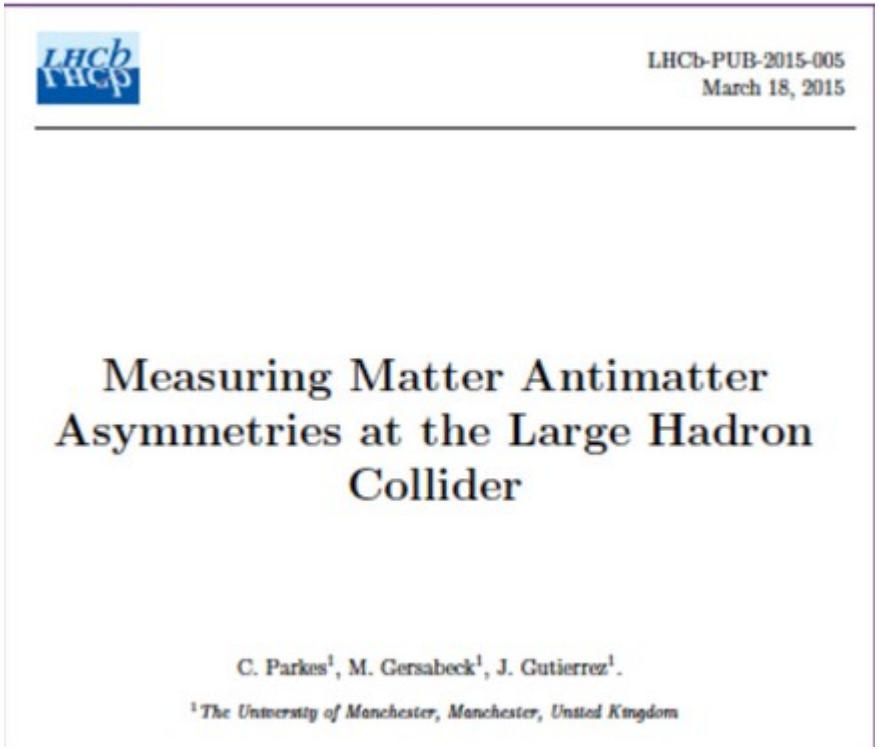
Working towards level 3

Started planning the *For Research area* of the Open Data portal, in view of our public release:

- Example analysis --> Measurement of CP asymmetry
- **Full 2011 data, C++ and ROOT**
- **Ntuples with pre-selected data**
- Mini-particle physics analysis:
  - Observe resonances
  - Make Dalitz plots
  - determine CP asymmetries

http://cds.cern.ch/record/1994172?ln=en





LHCb-PUB-2015-005
March 18, 2015

Measuring Matter Antimatter Asymmetries at the Large Hadron Collider

C. Parkes[1], M. Gersabeck[1], J. Gutierrez[1].

[1] The University of Manchester, Manchester, United Kingdom

Level 4 Open Data

Official since March 2013.
It follows DPHEP leves of data preservation.

Level 1 (published data)	All scientific results are public. Data associated with the results will also be made available; format and repositories will be decided by the Editorial Board
Level 2 (samples for educational purposes)	LHCb already involved in outreach and education activities. Event displays and simple analysis level ntuples are already available and will continue to be provided to the public. The data are for educational purpose only, not suitable for publication
Level 3 (reconstructed data)	LHCb will make **reconstructed data (DST) available to open public; 50% 5 years after data is taken, 100% after 10 years.**
Level 4 (raw data)	Due to the complexity of the raw data processing stage, the extensive computing resources required and enormous access to tape resources direct access to raw data is not permitted to individuals within the collaboration. Raw data processing is performed centrally. Due to this, the collaboration is currently **not planning to allow open access to raw data**

BACKUP