LHCb data preservation use cases

- January 8th , 2014 -

S.Amerio (U. Padova)

Use case 1: new analysis on “old” data

New analysis on old (e.g. Run 1) data

Requirements

- All data consistent to the same software (legacy datasets, legacy software release)
- Capability of generating new MC samples (new signals, latest generators) corresponding to old data (problem of interface between “new” software and legacy one)

Target

- LHCb collaboration

Activities:

- Estimate of needed storage
- Definition of legacy release
- Development of a validation system

How much storage is needed?

	Size (TB)		
	RAW	FULL.DST	DST
Data (2010/11/12/13)	2583	4041	787
MC (Sim08+older)	--	--	794
Total	**2583**	**4041**	**1581**

Use case 2: analysis preservation and reproduction

Preservation/reproduction of an analysis

Requirements:
• Framework to collect all analysis components (data, code, documentation, talks and publications)
• Storage to archive data and code
• Not only a container, but a tool which allows to reproduce at least the final part of an analysis, from analysis level ntuples (DST) to published plots. It means we need to preserve the complete computing environment of the analysis, e.g. via virtualization techniques.

Target
• LHCb collaboration
• General public if the collaboration agrees to make the detailed steps of an analysis public

Activities
•Definition of LHCb analysis steps
•Inclusion of LHCb analysis in the analysis preservation framework under development at Cern → http://data-demo.cern.ch/

Comment: all LHC experiments are already collaborating on this subject

Use case 3: open data

Relase LHCb data for general public

Requirements:
• Simplified samples and analysis tools for high school students and teachers (e.g. International Masterclasses exercises)
• Analysis level ntuples, experiment software and computing environment, documentation and analysis examples
• Portal to access data and software
• Storage to archive data and software

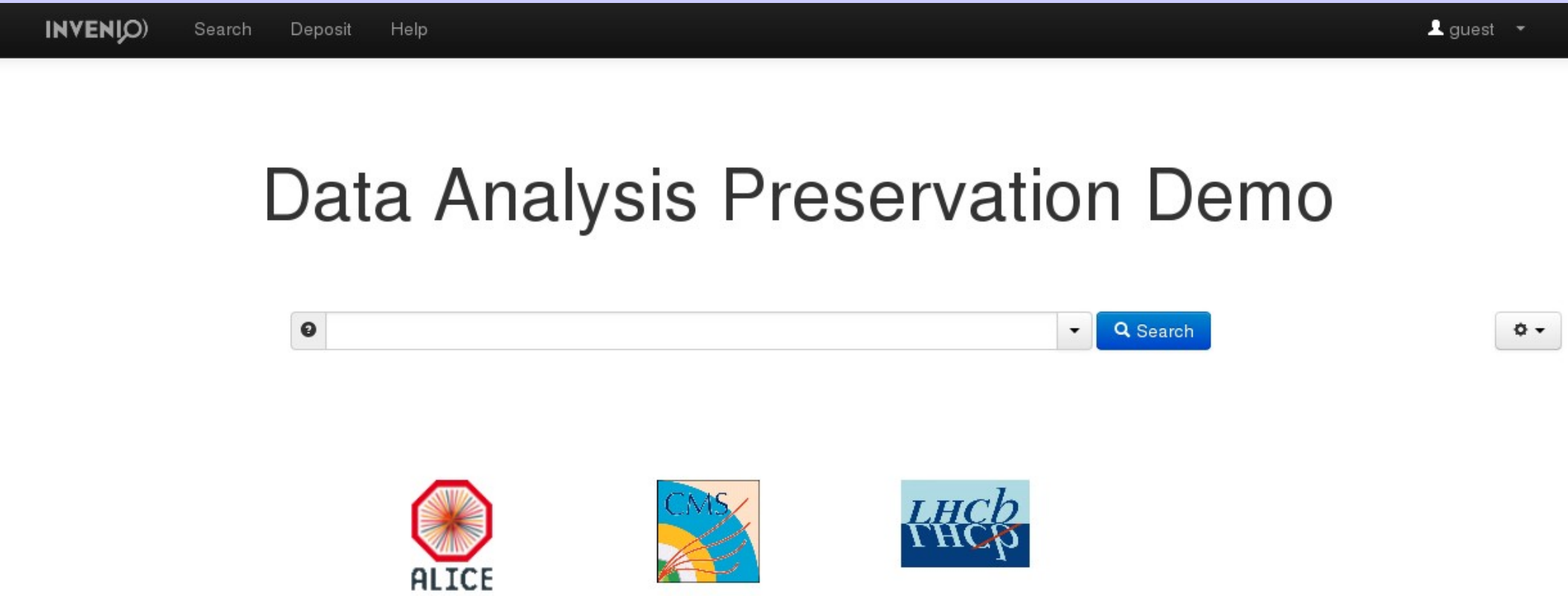
Target
• general public (students, teachers, physicists from other experiments...)

Activities
• Inclusion of LHCb masterclass exercise in the Cern open data portal (http://open-data.cern.ch/)
• New event display in collaboration with Cern MediaLab (TEV, Total Event Visualizer)
• Development of simplified analysis workflows.

Comment: all LHC experiments are already collaborating on this subject and the first LHC open data portal has just been released → http://open-data.cern.ch/

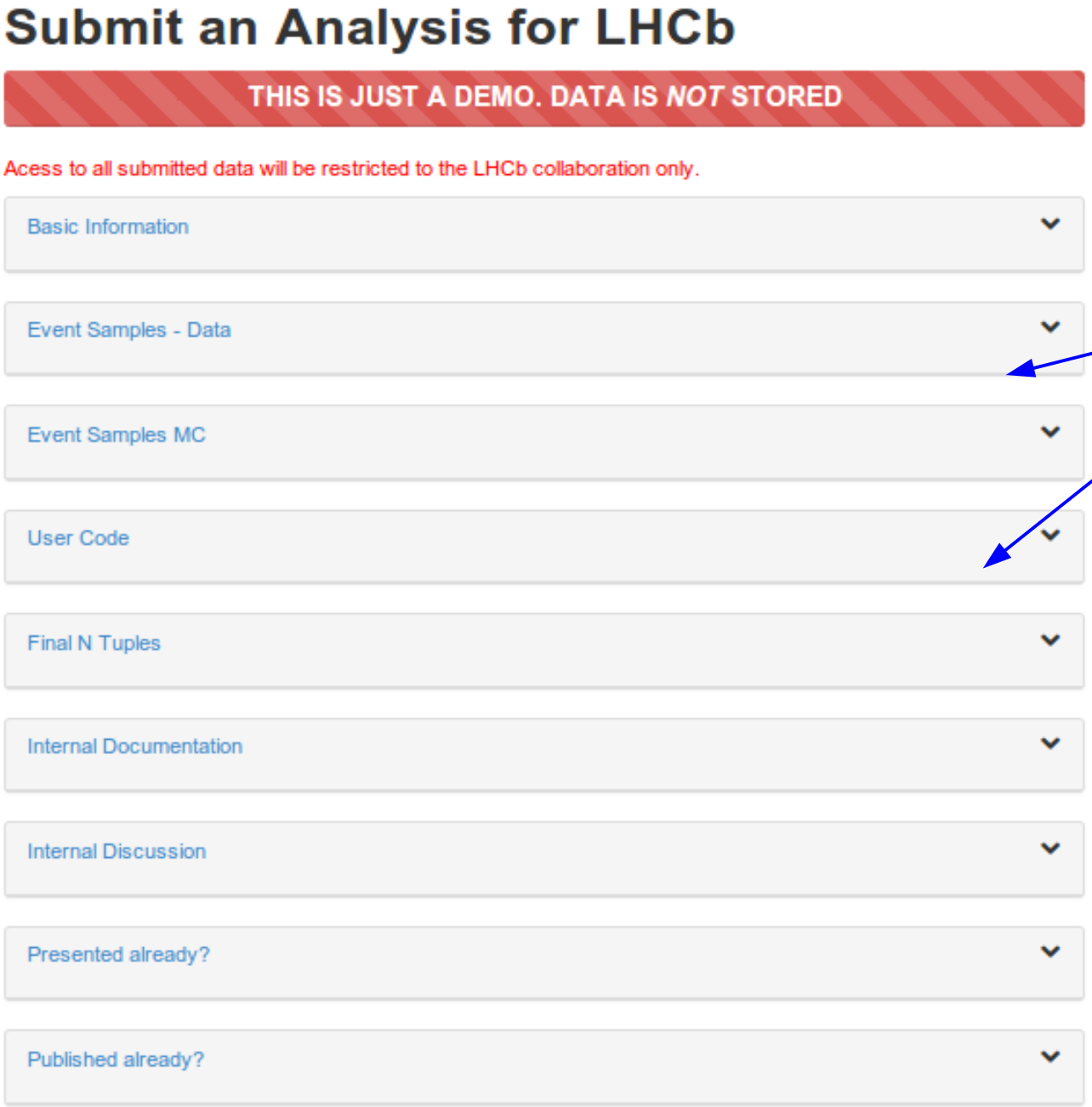
BACKUP

Analysis preservation framework



Goal: create a framework to preserve all the information related to an analysis, from user code to final ntuples to internal and public documentation. This project is developed in collaboration with Cern-IT, CMS and ALICE

Link here: http://data-demo.cern.ch/



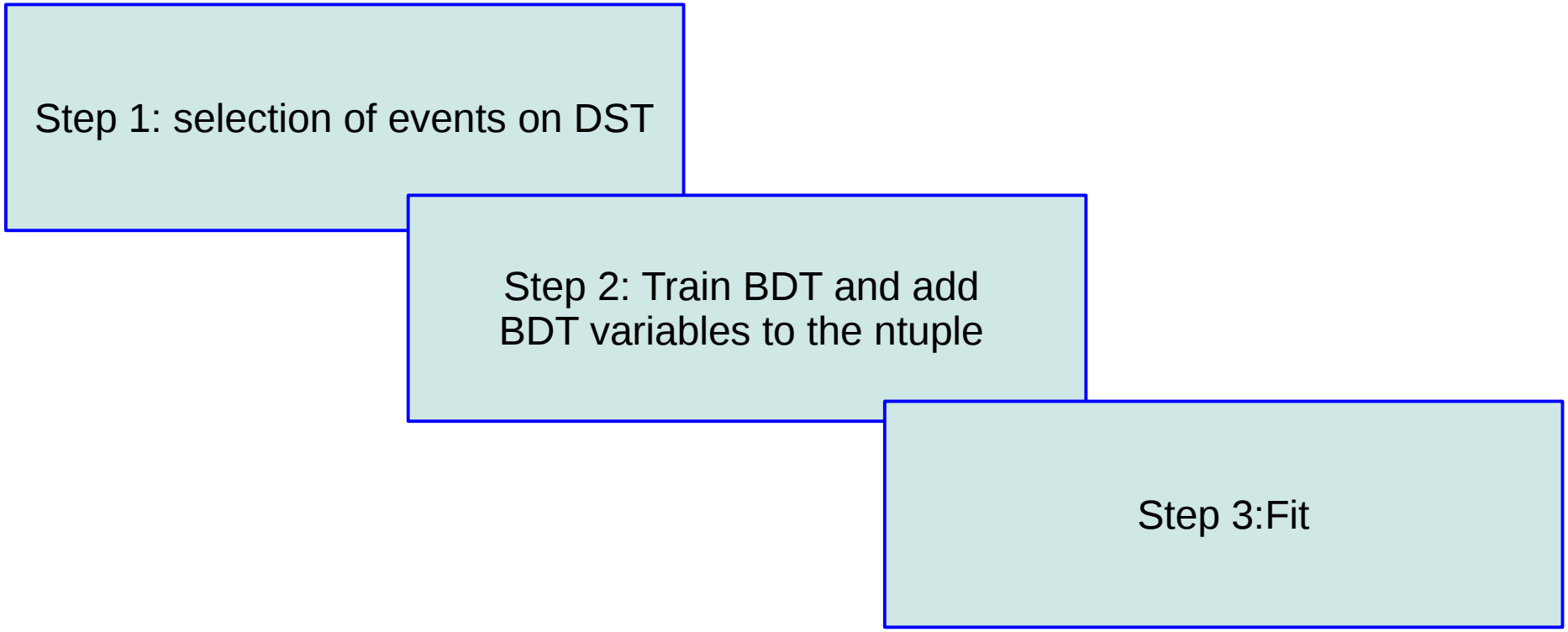
Current version
- Single section for data and MC samples
- Single section for the user code

Very compact, but it does not really follow the flow of the analysis --> difficult for the analysts to compile and for a user to reproduce

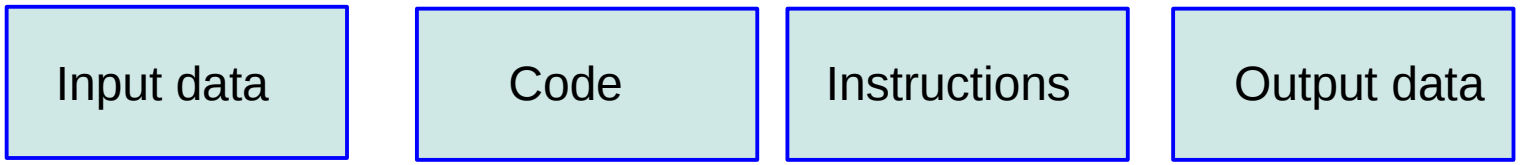
We proposed some changes on the experience with a real analysis (Bs --> DsPi (Ds --> KKPiPi0))

Analysis flow

Each analysis consists of well defined steps. e.g. for Bs --> DsPi (Ds --> KKPiPi0)):



... and each step has a defined “ingredients “



New structure proposed
It will be implemented after open data portal is released.

Open-data portal

http://open-data-demo.cern.ch/

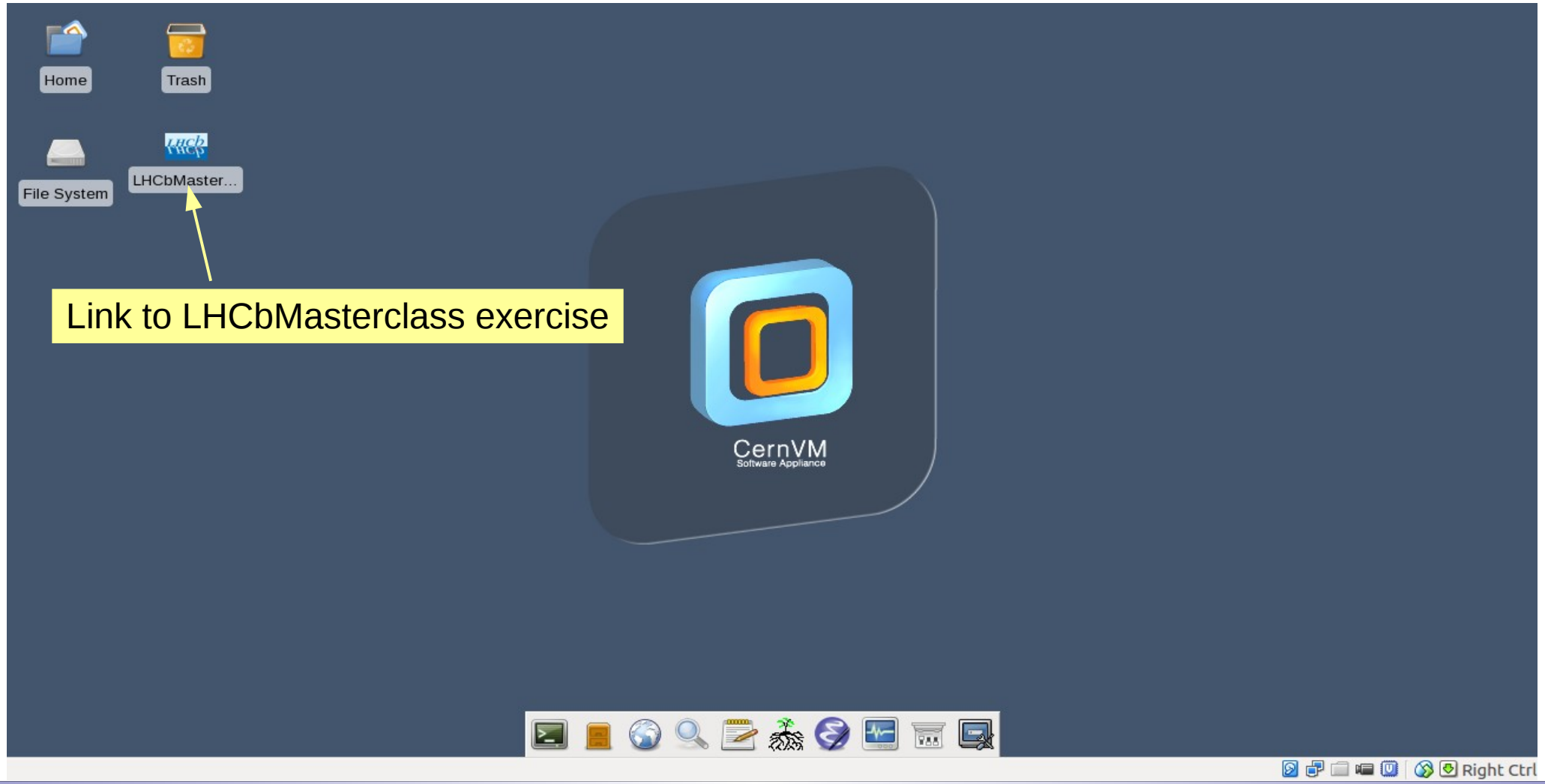
LHCb joined the Open-data portal with the D0 lifetime exercise.
First release expected by early next week.
This portal is intended to collect open data/applications from all the experiments. We started with the masterclass samples/exercise, other samples should follow.

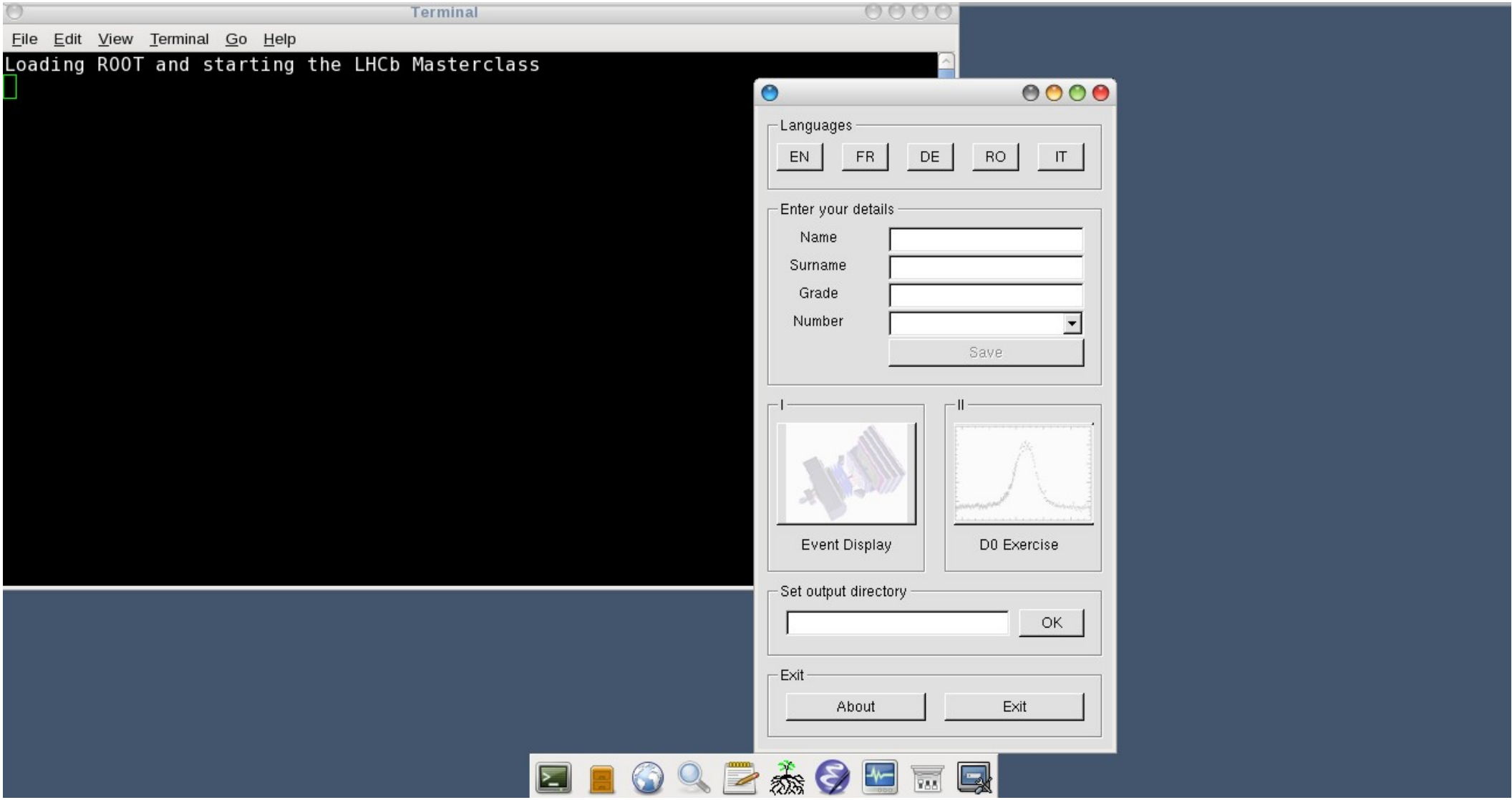
LHCb VM image

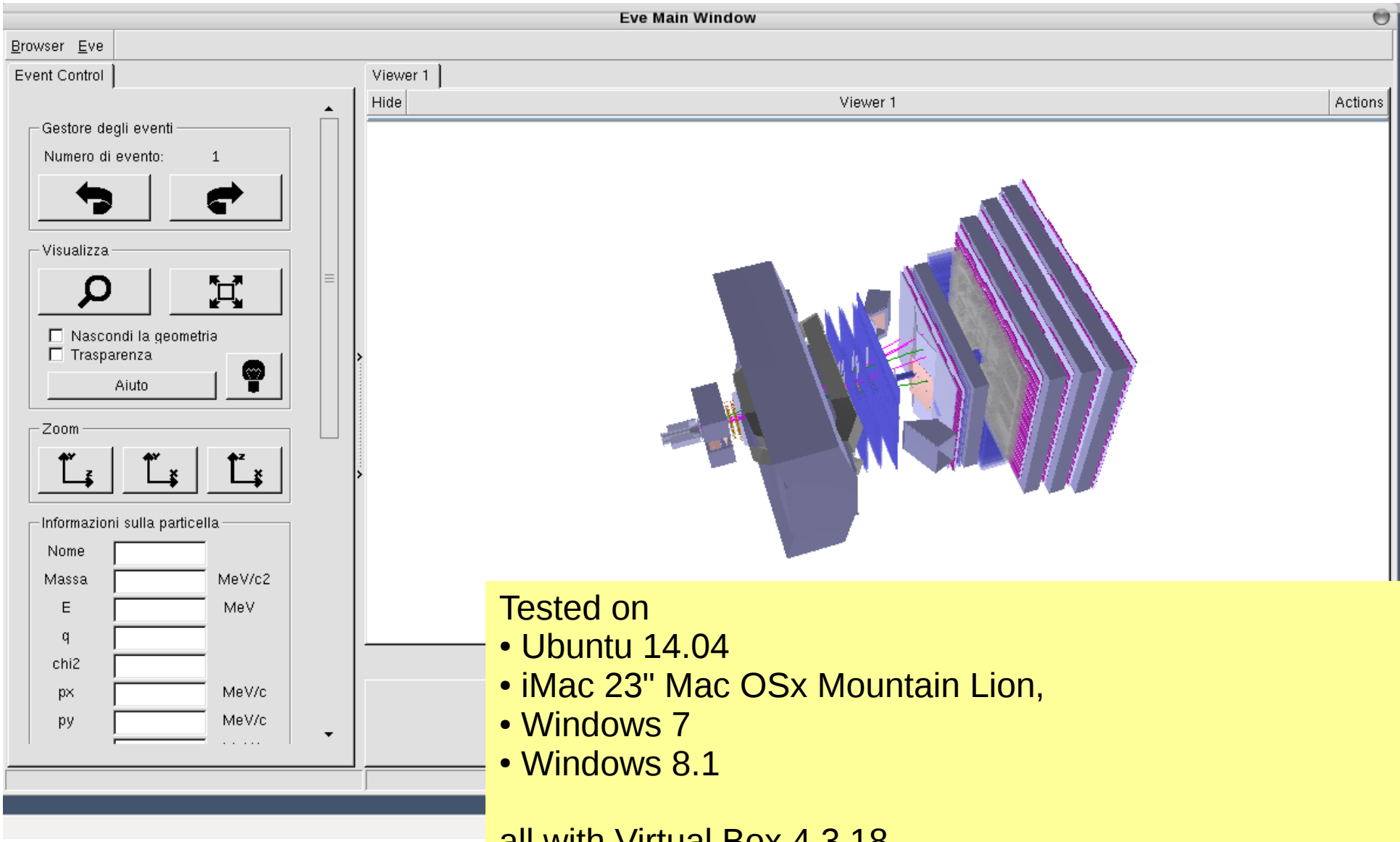
VM image of LHCb D0 lifetime exercise based on CernVM
Very simple to install and use:

- Install VirtualBox
- Download the image from

https://lhcbproject.web.cern.ch/lhcbproject/dist/Masterclass/vm/LHCbMasterclass.ova
and start




Terminal
File Edit View Terminal Go Help
Loading ROOT and starting the LHCb Masterclass
Languages
EN
FR
DE
RO
IT
Enter your details
Name
Surname
Grade
Number
Save
I
Event Display
II
D0 Exercise
Set output directory
OK
Exit
About
Exit



Tested on
- Ubuntu 14.04
- iMac 23" Mac OSx Mountain Lion,
- Windows 7
- Windows 8.1

all with Virtual Box 4.3.18

Good performance.