ATLAS exercise with Escape datalake

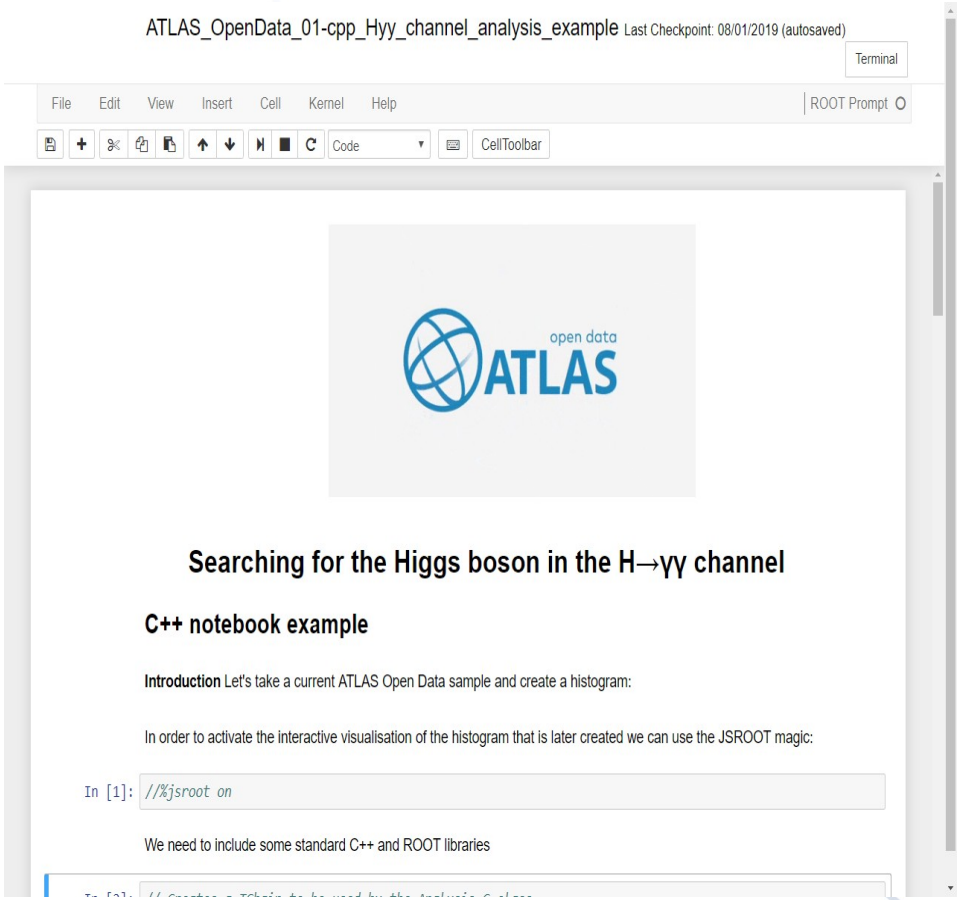
Stéphane JEZEQUEL

Lapp/CNRS

6 May 2020

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

ATLAS analysis demo



- ATLAS Exercise
  - Mockup of analysis activity using public ATLAS files
  - C++ code
  - Stream input ROOT files located at CERN on Web server
  - Produce H → yy plot
- ESCAPE exercise :
  - Upload files to ESCAPE datalake with rucio client
  - Adapt file access to ESCAPE datalake
  - Produce plots

Funded by the European Union's Horizon 2020 - Grant N° 824064

Upload files

Files hosted on LAPP local machine

```
(rucio) lappsl7h.in2p3.fr> cd data_ATLAS_Hyy/
(rucio) lappsl7h.in2p3.fr> ls
data_A.GamGam.root  data_C.GamGam.root  mc_341081.ttH125_gamgam.GamGam.root.1  mc_345041.VBFH125_gamgam.GamGam.root.1       mc_345319.ZH125J_Zincl_gamgam.GamGam.root
data_B.GamGam.root  data_D.GamGam.root  mc_343981.ggH125_gamgam.GamGam.root.1  mc_345318.WpH125J_Wincl_gamgam.GamGam.root.1  rucio_upload.json
(rucio) lappsl7h.in2p3.fr>
```

Apply procedure described in Rucio tutorial

```
(rucio) lappsl7h.in2p3.fr> rucio list-file-replicas  --protocol root atlas:exercise_Hyy_v0
```

SCOPE	NAME	FILESIZE	ADLER32	RSE: REPLICA
atlas	data_A.GamGam.root	81.310 MB	5f4a738e	LAPP-DPM: root://lapp-testse01.in2p3.fr:1094/dpm/in2p3.fr/home/escape/rucio/lapp_dpm/atlas/a1/a9/data_A.GamGam.root
atlas	data_B.GamGam.root	289.966 MB	ed3610ce	LAPP-DPM: root://lapp-testse01.in2p3.fr:1094/dpm/in2p3.fr/home/escape/rucio/lapp_dpm/atlas/56/b2/data_B.GamGam.root
atlas	data_C.GamGam.root	424.844 MB	de1ab403	LAPP-DPM: root://lapp-testse01.in2p3.fr:1094/dpm/in2p3.fr/home/escape/rucio/lapp_dpm/atlas/fb/ae/data_C.GamGam.root
atlas	data_D.GamGam.root	686.890 MB	86bc40ea	LAPP-DPM: root://lapp-testse01.in2p3.fr:1094/dpm/in2p3.fr/home/escape/rucio/lapp_dpm/atlas/c5/38/data_D.GamGam.root
atlas	mc_341081.ttH125_gamgam.GamGam.root.1	217.916 MB	17e85173	LAPP-DPM: root://lapp-testse01.in2p3.fr:1094/dpm/in2p3.fr/home/escape/rucio/lapp_dpm/atlas/f3/17/mc_341081.ttH125_gamgam.GamGam.root.1
atlas	mc_345041.VBFH125_gamgam.GamGam.root.1	108.678 MB	a1c2ed8a	LAPP-DPM: root://lapp-testse01.in2p3.fr:1094/dpm/in2p3.fr/home/escape/rucio/lapp_dpm/atlas/88/66/mc_345041.VBFH125_gamgam.GamGam.root.1
atlas	mc_345318.WpH125J_Wincl_gamgam.GamGam.root.1	29.802 MB	2d06f3ca	LAPP-DPM: root://lapp-testse01.in2p3.fr:1094/dpm/in2p3.fr/home/escape/rucio/lapp_dpm/atlas/29/1b/mc_345318.WpH125J_Wincl_gamgam.GamGam.root.1
atlas	mc_345319.ZH125J_Zincl_gamgam.GamGam.root	58.784 MB	c5c51eba	LAPP-DPM: root://lapp-testse01.in2p3.fr:1094/dpm/in2p3.fr/home/escape/rucio/lapp_dpm/atlas/a9/3a/mc_345319.ZH125J_Zincl_gamgam.GamGam.root

Files hosted french DPM federated storage

(LAPP/LPSC/LPC/CPPM : FR-ALPAMED)

Funded by the European Union's Horizon 2020 - Grant N° 824064

Installing code

```
lappsl7h.in2p3.fr> ls^C
lappsl7h.in2p3.fr> git clone https://github.com/atlas-outreach-data-tools/atlas-outreach-cpp-framework-13tev.git
Cloning into 'atlas-outreach-cpp-framework-13tev'...
remote: Enumerating objects: 1080, done.
remote: Counting objects: 100% (1080/1080), done.
remote: Compressing objects: 100% (389/389), done.
remote: Total 1080 (delta 756), reused 1002 (delta 686), pack-reused 0
Receiving objects: 100% (1080/1080), 519.17 KiB | 0 bytes/s, done.
Resolving deltas: 100% (756/756), done.
lappsl7h.in2p3.fr> cd atlas-outreach-cpp-framework-13tev
lappsl7h.in2p3.fr> source welcome.sh
Welcome to ATLAS 13 TeV Open Data C++ framework!!!
Input your option now
1 = create all output directories (do it at the very beginning just once)
0 = remove all output directories (in case you want to clean and begin from zero)
1
Creating the directory: Analysis/WBosonAnalysis/Output_WBosonAnalysis
Creating the directory: Analysis/ZBosonAnalysis/Output_ZBosonAnalysis
Creating the directory: Analysis/TTbarAnalysis/Output_TTbarAnalysis
Creating the directory: Analysis/SingleTopAnalysis/Output_SingleTopAnalysis
Creating the directory: Analysis/WZDiBosonAnalysis/Output_WZDiBosonAnalysis
Creating the directory: Analysis/ZZDiBosonAnalysis/Output_ZZDiBosonAnalysis
Creating the directory: Analysis/HWWAnalysis/Output_HWWAnalysis
Creating the directory: Analysis/HZZAnalysis/Output_HZZAnalysis
Creating the directory: Analysis/ZTauTauAnalysis/Output_ZTauTauAnalysis
Creating the directory: Analysis/HyyAnalysis/Output_HyyAnalysis
Creating the directory: Analysis/SUSYAnalysis/Output_SUSYAnalysis
Creating the directory: Analysis/ZPrimeBoostedAnalysis/Output_ZPrimeBoostedAnalysis
Creating the directory for the Plotting code: Plotting/histograms
lappsl7h.in2p3.fr> cd Analysis/HyyAnalysis/
lappsl7h.in2p3.fr> ls
HyyAnalysis.C  HyyAnalysis.h  HyyAnalysisHistograms.h  main_HyyAnalysis.C  main_HyyAnalysis_web.C  Output_HyyAnalysis  run.sh  run
lappsl7h.in2p3.fr>
```

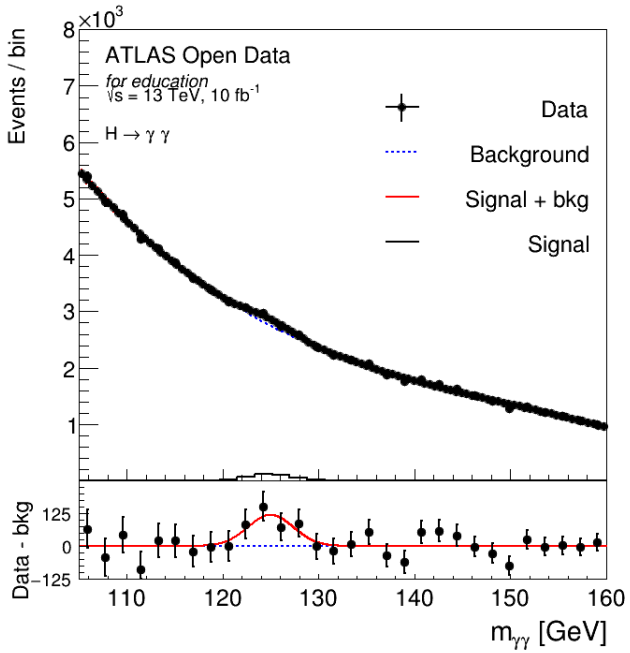
```
////////////////////////////////////////////////////////////////////////////////////////////////////
// Creates a TChain to be used by the Analysis.C class
#include "TROOT.h"
#include "TChain.h"
#include "TFile.h"
// #include "TProof.h"

void main_HyyAnalysis(int proof = 0, int option= 0)
{
  // path to your local directory *or* URL, please change the default one!

  /* Local path example */
  //TString path = "/eos/project/a/atlas-outreach/projects/open-data/OpenDataTuples/renamedLargeRJets/GamGam/";

  /* Path link to the ATLAS Open Data website repository */
  // BEFORE ESCAPE : TString path = "https://atlas-opendata.web.cern.ch/atlas-opendata/samples/2020/GamGam/";
  TString path = " root://lapp-testse01.in2p3.fr:1094/dpm/in2p3.fr/home/escape/rucio/lapp_dpm/atlas/";

  //**************************************************************************************************//
  // We run only over data and Higgs MC
  // option==0 will run all one by one
  // Currently 1 option for MC (2) and 1 for data (1) which can be run in parallel
  //**************************************************************************************************//

  //  if (proof == 1)  TProof::Open("");

  if (option==1 || option==0){

  //data
  TChain* chain_data = new TChain("mini");
  // BEFORE ESCAPE : chain_data->AddFile(path+"Data/data_A.GamGam.root");
  chain_data->AddFile(path+"a1/a9/data_A.GamGam.root");
  chain_data->AddFile(path+"56/b2/data_B.GamGam.root");
  chain_data->AddFile(path+"c5/38/data_D.GamGam.root");
  if (proof == 1)  chain_data->SetProof();
  chain_data->Process("HyyAnalysis.C+","data");
  } // option 1

  if (option==2 || option==0){
  // Higgs MC, ggH production
  TChain* chain_ggH125 = new TChain("mini");
  chain_ggH125->AddFile(path+"5a/17/mc_343981.ggH125_gamgam.GamGam.root");
  if (proof == 1)  chain_ggH125->SetProof();
  chain_ggH125->Process("HyyAnalysis.C+","ggH125_gamgam");

  // VBF Higgs production
  TChain* chain_VBFH125 = new TChain("mini");
  chain_VBFH125->AddFile(path+"88/66/mc_345041.VBFH125_gamgam.GamGam.root");
  if (proof == 1)  chain_VBFH125->SetProof();
  chain_VBFH125->Process("HyyAnalysis.C+","VBFH125_gamgam");

  // WH production
  TChain* chain_WH125 = new TChain("mini");
  chain_WH125->AddFile(path+"29/1b/mc_345318.WpH125J_Wincl_gamgam.GamGam.root");
  if (proof == 1)  chain_WH125->SetProof();
-UU-:----F1  main_HyyAnalysis.C    Top L1     Git:master  (C++/l Abbrev)
```

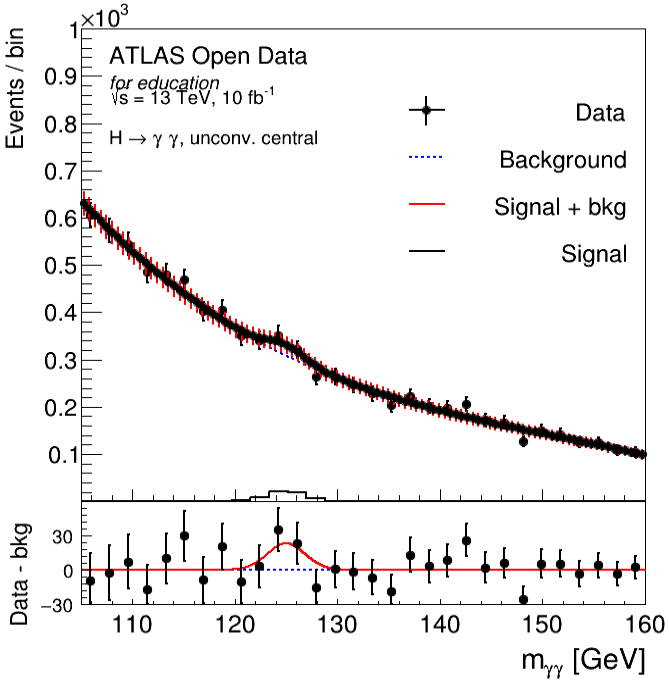
- Path for input files :
  - Can be automatised decoded by just providing dataset and location
  - Pattern can nbe built to include xcache in the path

Funded by the European Union's Horizon 2020 - Grant N° 824064

Running the code

- Just submit a bash script which runs job interactively during few minutes
  - Processing time dominated by access speed to data
  - Output root file stored locally
- Post processing root code to make plots





Funded by the European Union's Horizon 2020 - Grant N° 824064

Prospects

- Use IAM authentification
- Stream data through xcache (caching + read-ahead) :
  - Require patched version of ROOT distribution provided by WLCG through cvmfs
- Build the file path live depending on location and protocol
  - Resilience against storage in downtime → QoS
- Exercise could serve as a model for experiences without examples

Funded by the European Union's Horizon 2020 - Grant N° 824064

Acknowledgement to

- Rucio developers
- Support from ATLAS experts over last 10 years
- Support from Aris to document rucio@escape installation and debug my approximations
- Support from Riccardo for first xcache attempts

Funded by the European Union's Horizon 2020 - Grant N° 824064