

# Exploring JavaScript and ROOT technologies to create Web-based ATLAS analysis and monitoring tools

**A Sánchez Pineda on behalf of the ATLAS Collaboration**

Università di Napoli "Federico II" and INFN, Italy

E-mail: [arturos@cern.ch](mailto:arturos@cern.ch)

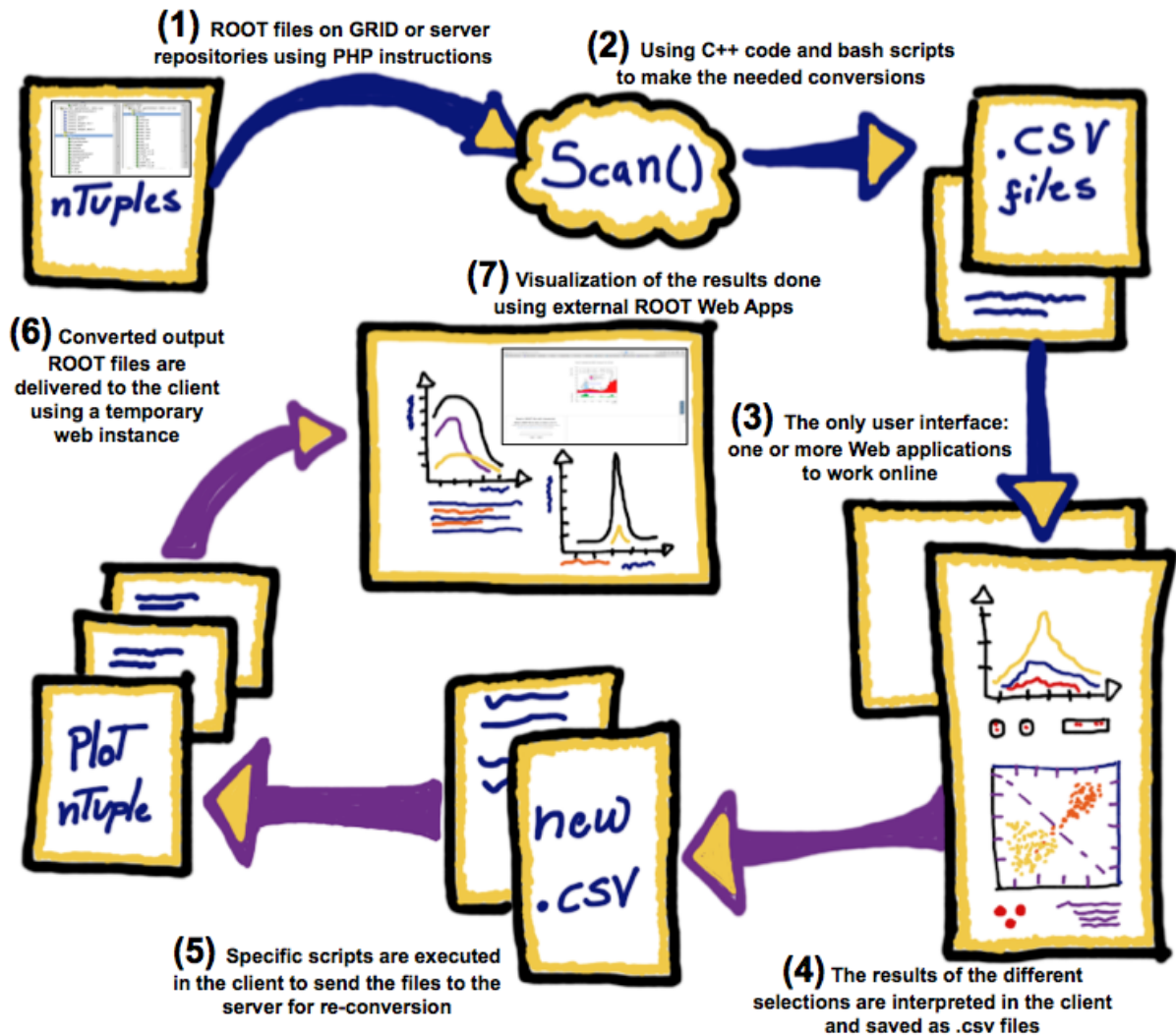
**Abstract.** We explore the potential of current web applications to create online interfaces that allow the visualization, interaction and real cut-based physics analysis and monitoring of processes through a web browser. The project consists in the initial development of web-based and cloud computing services to allow students and researchers to perform fast and very useful cut-based analysis on a browser, reading and using real data and official Monte-Carlo simulations stored in ATLAS computing facilities. Several tools are considered: ROOT, JavaScript and HTML. Our study case is the current cut-based  $H \rightarrow ZZ \rightarrow llqq$  analysis of the ATLAS experiment. Preliminary but satisfactory results have been obtained online.

## 1. Introduction

The cut-optimization procedure is a very time-consuming process during the development of a physics analysis. When immersed in the coding procedure, the researcher often is not able to think about the physics of the data studied until a plot is produced. The idea is to simplify the beginning of the analysis, where the researcher follows the behavior of variables while making cuts in an “on the flight” regime. Nowadays, powerful computers and faster internet connections allow the analysis of large sets of data using a web browser in ways that were impossible just a few years ago. This permits simple but useful analysis with a performance compatible with a ROOT[1] session run on a laptop or an user interface facility.

In order to show the advantages of our online tools, and the possibility to use them for a real analysis, we took Monte-Carlo (MC) samples for Higgs[2] decay  $H \rightarrow ZZ \rightarrow llqq$  and Standard Model background processes and applied a simplified version of cuts from the  $H \rightarrow ZZ \rightarrow llqq$  analysis[3] of the ATLAS[4] Experiment using a website form. In this exercise, the numbers of remaining events after each cut are very close to those from the official analysis (see Table 1). Apart from locating such scripts on single web-pages, it is also possible to include them in TWiki[5] pages, that are widely used for sharing updates, analysis strategies, status of current studies, preliminary results, cutflow comparisons, and so forth within the collaboration. All this, together with commercial analytical services (i.e. provided by Google[6]) can be used not to produce a final paper plot, but to create a faster way to exchange ideas, tutorials, classes and real useful information for the ATLAS analyses for the next data taking at LHC, so-called Run 2, where one of the aspects to take in consideration is the huge amount of data, analyses, tunings and techniques to put under review and test for the coming years.





**Figure 1.** Graphic representation of the different processes and steps in the development of the web-based physics analysis for the ATLAS Experiment.

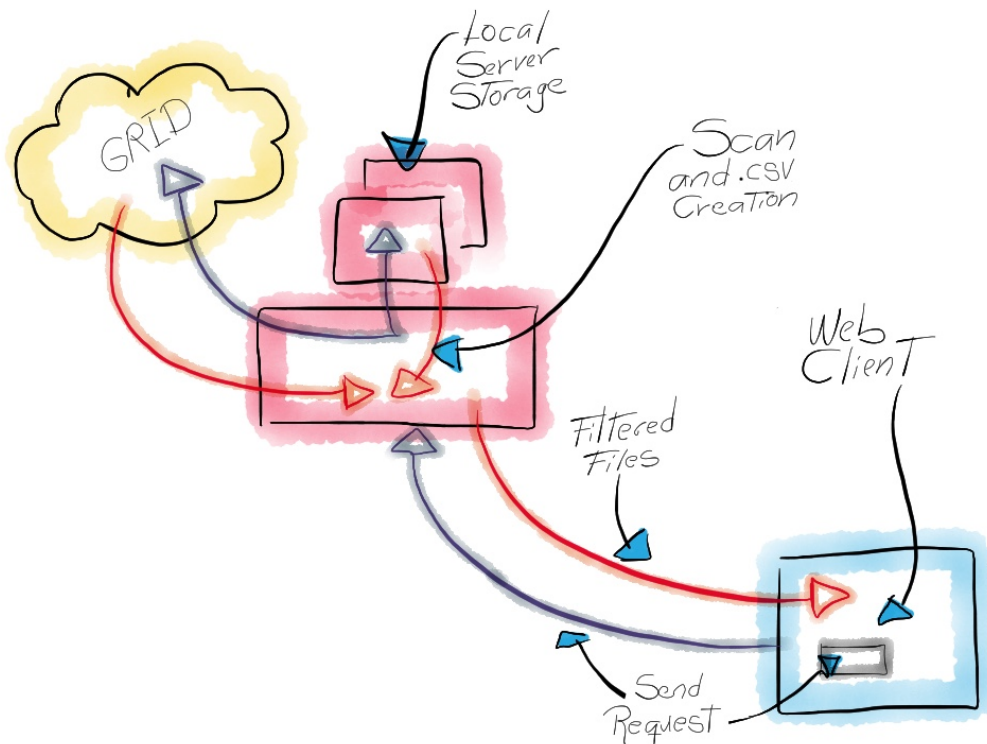
## 2. Objectives and Vision

We can summarise the objectives of this work in the following way:

- Create a web-based system that allows physics or performance analysis selections over real ATLAS samples in ROOT format.
- Generate a collaborative tool where students and researches can learn how to perform new and current physics analysis.
- Generate fast preliminary results for checks of selection performance, sample production integrity or online monitoring.
- Have a pedagogical tool (with simple examples) for a more general public interested in High Energy Physics.
- Take advantage of the new Web applications and languages plus internet performance to present status or results visualizations.

### 3. Workflow of an ideal Web Analysis

The cycle (Figure 1) starts with the reading of the input files from a local disk in the server or an external repository in one of the ATLAS distributed computer facilities (GRID)[7]. The server reads, filters and transforms the files, sending them to the client. The analyzers can work with them in a browser using custom JavaScript[8] applications, performing online selections and send the results to the server, recreating and delivering a ROOT file to the client as a product of the performed work. Due to the fact that this is an ongoing project, some of the procedures mentioned below have been developed successfully, and others are still under construction:



**Figure 2.** Communication between the Web client application and the Web server (See sections 3.1 and 3.2). Here (*bottom-right*) the files are requested, transformed, filtered (*top-left*) and sent to the client, who in a refreshed version of the web, can start to perform cuts on the selected variables.

#### 3.1. Request of input files

The client asks for input files using a menu and/or inserting a string. This request is sent to the server which looks for the files in its internal storage or on the GRID using *https* protocols. In principle, some kind of identification procedure has to be performed, using the GRID certification that every member of the collaboration has, and/or some account for external user (e.g. students). This step is under development.

#### 3.2. Scan of the input files

The ROOT files must be scanned looking just for the variables that are needed in the current web analysis, to perform a transformation to comma-separated values (*.csv*) format files that will be sent to the client as the JavaScript application input. This step has been developed.

### 3.3. Working with the Web-Apps

The creation of the *.csv* files is done and the analyzer can perform different selections or cuts on the variables one by one or in any desired combination to create a physics analysis test and/or optimization without coding and just using bars or mouse selection in the webpage. This step has been developed.

### 3.4. Converting selection in *.csv* files

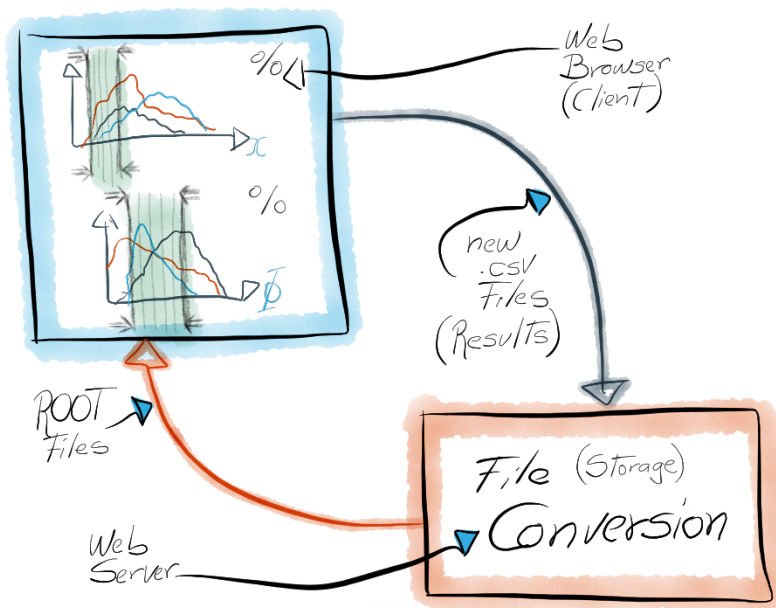
At the end of the online selections (cuts) those results are recorded by the client and saved as *.csv* files using an user-button execution. This will send those files to the server for conversion. This step is under development.

### 3.5. Sending *.csv* output files

Final results for all the variables used in the online analysis are saved in *.csv* files. Using PHP instructions those files are sent to the server where a temporary folder stores the files for conversion. This step is under development.

### 3.6. Transformation to ROOT files

The results in *.csv* files are transformed to ROOT files using the inverse scripts of the step one of this workflow. A temporary HTML page is created to allows the user to download her/his final results. This step has been developed.



**Figure 3.** In a second stage, after the analysis is performed another request to the server is done, sending the analyzed data in form of *.csv* files that are saved, converted and sent back to the client for visualization and/or download (See sections 3.5, 3.6 and 3.7).

### 3.7. Scan of the input files

Visualization of the output ROOT files can be done using it web browser<sup>1</sup> or another web application, as in Figure 4. This step has been developed.

**Table 1.** Number of events remaining after applying the cuts from the official cutflow and in the Multivariable-Analysis webpage for MC signal sample  $H \rightarrow ZZ \rightarrow llqq$  for 400 GeV Higgs mass.

Cut	Official cutflow	Web-based pre-analysis	Difference (%)
OppositeSign	8056	8056	0.00
TwoJets	7282	7282	0.00
DileptonMass	6089	6088	0.02
MET	5763	5758	0.09
NumTagJets0	4263	4261	0.05
NumTagJets1	973	972	0.10
NumTagJets2	513	511	0.39
PtLeadingJet0	4217	4197	0.48
PtLeadingJet1	957	957	0.00
PtLeadingJet2	512	506	1.18
DiJetMass0	2182	2160	1.01
DiJetMass1	438	439	0.29
DiJetMass2	387	383	1.04

## 4. Current Results

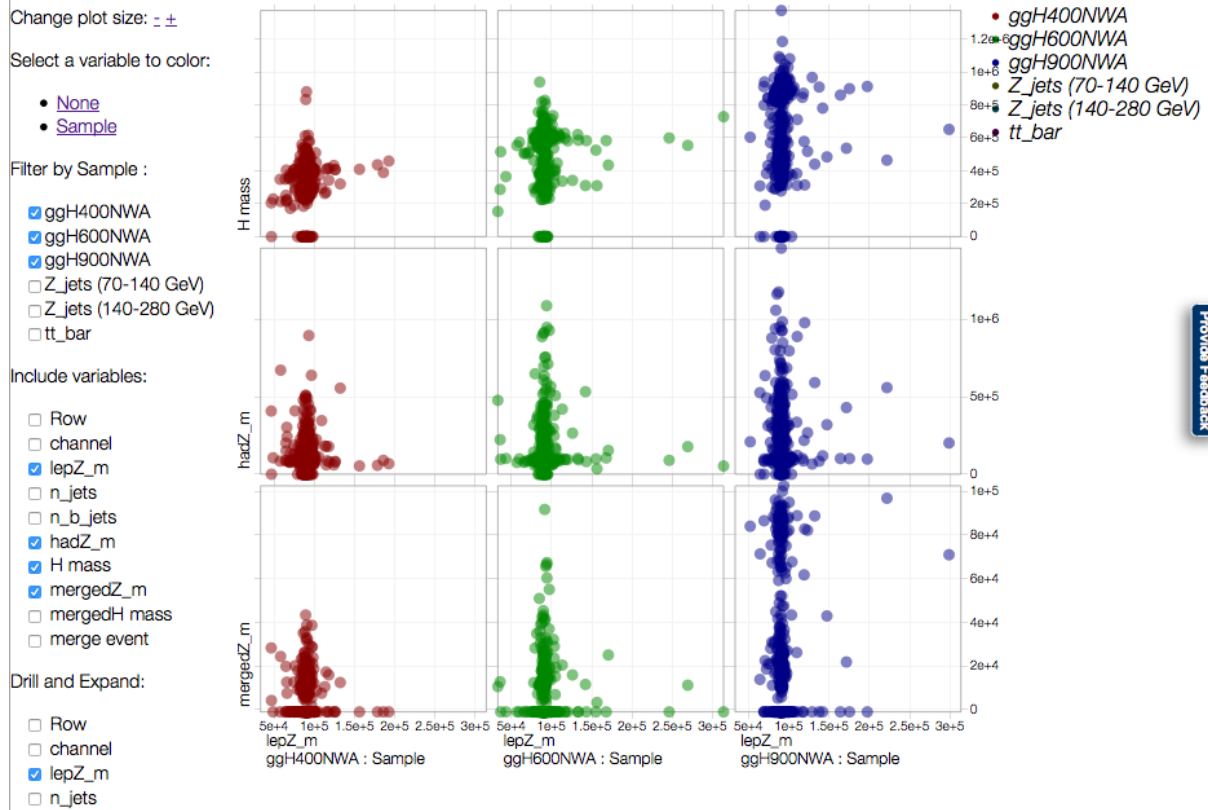
As an example, we considered different signal and background MC samples for the  $H \rightarrow ZZ \rightarrow llqq$  analysis in order to check how large will be the possible differences in the number of events remaining after some basic cuts from official analysis[3] in the Cross-Filter[9] web-page and those from the official cutflows. For example, for the MC signal sample  $H \rightarrow ZZ \rightarrow llqq$  for 400 GeV Higgs (see Figure 5) we applied the following cuts:

- *OppositeSign*: The product of the charges of the two selected muons should be equal to -1.
- *TwoJets*: The number of jets should be at least 2.
- *DileptonMass*: The invariant mass of the two leptons must be between 83 GeV and 99 GeV.
- *MET*: The value of the Missing Transverse Energy should be less than 60 GeV.
- *NumTagJets[0-1-2]*: Number of  $b$ -jets is equal to 0 (or 1 or 2). As a note, the analysis is divided in three categories with respect to the number of  $b$ -jets, so this and the next 2 cuts are applied for each of the cases: 0 or 1 or 2  $b$ -jets into the event.
- *PtLeadingJet[0-1-2]*: The transverse momentum of the leading jet should be greater than 45 GeV.
- *DiJetMass[0-1-2]*: The invariant mass of the two selected jets should be between 70 GeV and 105 GeV.

<sup>1</sup> <https://root.cern.ch/js/>

## ATLAS Higgs llqq Analysis Scatter-Matrix Analyzer

Explore matrix data of the  $H \rightarrow ZZ \rightarrow llqq$  ATLAS Analysis using scatterplot matrix:



**Figure 4.** Web-based scatter matrix analyser example.

Table 1 shows how the number of events changes after the cutflow for the electron channel. The numbers after each of the cuts are close by  $\approx 1\%$ , this is quite good for the pre-analysis. Again, our objective is not to do the complete analysis, but to simplify and speed-up the estimation of the right cuts. These would then need to be clarified in the next stage with the usual coding procedure.

Here you can find the Cross-Filter web-pages for signal and background MC samples for  $H \rightarrow ZZ \rightarrow llqq$  analysis processes<sup>2</sup>:

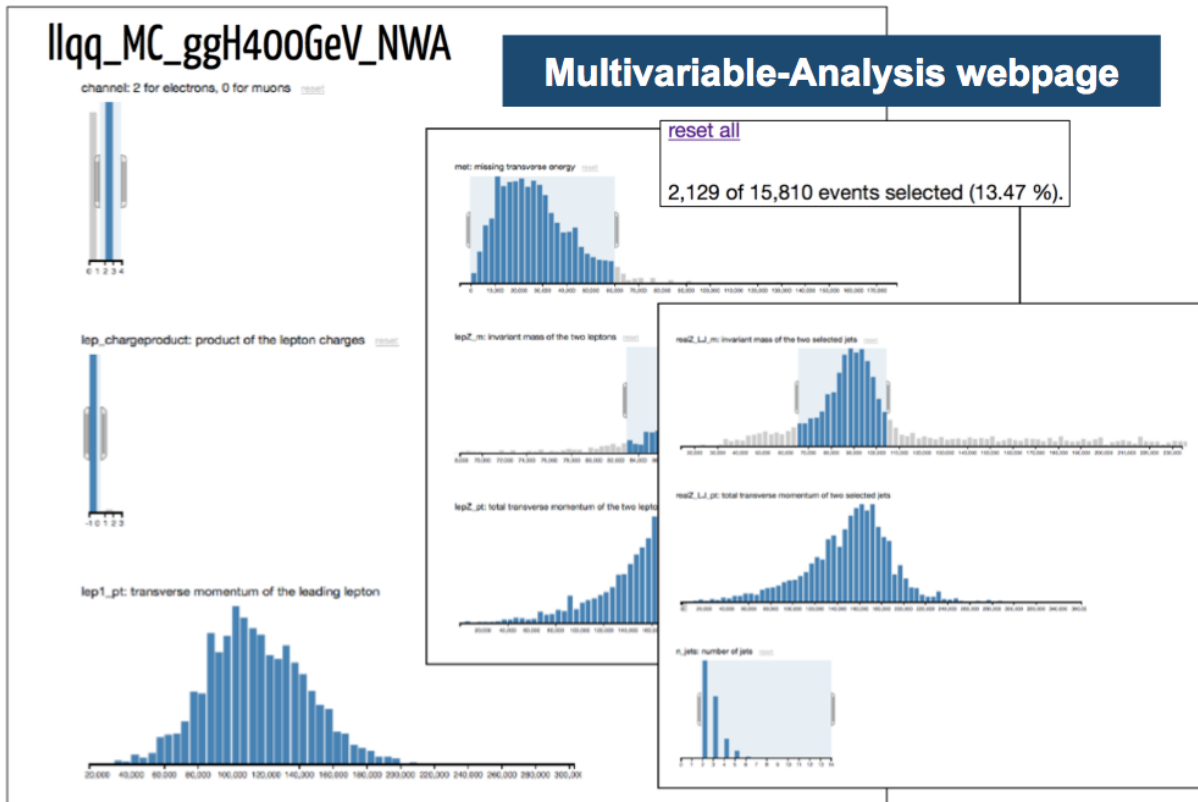
- MC Signal  $H \rightarrow ZZ \rightarrow llqq$ :  $ggH$  400GeV
- MC Signal  $H \rightarrow ZZ \rightarrow llqq$ :  $ggH$  600GeV
- MC Signal  $H \rightarrow ZZ \rightarrow llqq$ :  $ggH$  900GeV
- MC Background  $Top$
- MC Background  $Z + jets$   $70 < pT_Z < 140GeV$
- MC Background  $Z + jets$   $140 < pT_Z < 280GeV$

<sup>2</sup> [https://twiki.cern.ch/twiki/bin/view/Main/SummerStudentProject2014Napoli#August\\_12th\\_20th](https://twiki.cern.ch/twiki/bin/view/Main/SummerStudentProject2014Napoli#August_12th_20th)



## 5. Conclusions

Several tools and examples have been tested and preliminary results show that it is possible to handle a simple cut-based analysis in a browser using web applications. The ATLAS  $H \rightarrow ZZ \rightarrow llqq$  analysis was used to test the performance and the numbers obtained after each of the cuts are rather close[10]. Part of such applications, like the JavaScript Scatter-Matrix Libraries were already developed and we considered them as example tools, mostly, for data visualisation. Another tool, JavaScript Multivariable Cross-Filter, was rewritten and it was found particularly useful for the pre-analysis since it is visual and allows to see the behaviour of the variables in real time. The current project is to complete the workflow (as in 3) and start to test the website using current ATLAS analyses at the so-called Run 2, where already several groups have shown interest to see how this project can be extended for current and new physics analysis, monitoring and outreach of the HEP.



**Figure 5.** Some views of the web-based analysis tool in action. It is showed some cuts in variables relative  $H \rightarrow ZZ \rightarrow llqq$  ATLAS analysis.

## Acknowledgments

This work was developed in the framework of the PRIN Project STOA-LHC 20108T4XTM, CUP: I11J12000080001, and partly supported by it.

## References

- [1] Brun R and Rademakers F 1996 **ROOT** - An Object Oriented Data Analysis Framework *Nucl. Inst. & Meth. in Phys. Res. A* **389** pp 81-86.
- [2] ATLAS Collaboration 2012 *Phys. Lett. B* **716** pp 1-29.
- [3] ATLAS Collaboration 2014 *Search for a heavy Higgs boson in the mass range 200-1000 GeV in the  $H \rightarrow ZZ \rightarrow llqq$  channel*, ATL-COM-PHYS-2014-269.
- [4] ATLAS Collaboration 2008 *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3**.
- [5] TWikiSite: <https://twiki.cern.ch/>.
- [6] Google Apps Script: <https://developers.google.com/apps-script/>.
- [7] Golubkov D, Kersevan B, Klimentov A, Minaenko A, Nevski P, Vaniachine A and Walker R 2012 *ATLAS Grid Data Processing: system evolution and scalability*, ATL-SOFT-PROC-2012-042, ATL-COM-SOFT-2012-115.
- [8] JavaScript programming language: <https://developer.mozilla.org/en-US/docs/Web/JavaScript>.
- [9] Cross-Filter JavaScript library repository: <https://github.com/square/crossfilter>.
- [10] Moskalets T, Sanchez Pineda A, Conventi F 2014 *Web-based pre-Analysis Tools*, **CERN-STUDENTS-Note-2014-122**.