

A Perspective of User Support for the CMS Experiment

S Malik,^{1,6} K Lassila-Perini², B Hegner³, A Vedaee⁴ M Stankevicius⁵

¹ University of Nebraska-Lincoln, Lincoln, NE 68588-0299, U.S.A.

² Helsinki Institute of Physics, PO Box 64, FI-00014 University of Helsinki, Finland

³ CERN-PH, 1211 Geneva 23, Switzerland

⁴ INFN – Section of Perugia, Via A. Pascoli, 06123, Perugia, Italy

⁵ CERN-PH, 1211 Geneva 23, Switzerland

E-mail: malik@fnal.gov

Abstract. The CMS (Compact Muon Solenoid) experiment is one of two large general-purpose particle physics detectors at the LHC (Large Hadron Collider). An international collaboration of nearly 3500 people operates this complex detector whose main goal is to answer the most fundamental questions about our universe. The size and globally diversified nature of the collaboration and the Petabytes/year of data being collected, presents a big challenging task in bringing users up to speed to contribute to the physics analysis. The CMS User Support performs this task by helping users quickly learn about the CMS computing and the needed physics analysis tools. In this presentation we give an overview of its goals, organization and usage of collaborative tools to maintain the software and computing documentation and conduct year around tutorials on several physics tools needed as a pre-requisite for physics. We also talk about the user feedback evaluating its work.

1. Introduction

The CMS experiment [1] is one of two large general-purpose particle physics detectors built on the LHC at CERN in Geneva, Switzerland. In order to meet the challenges of designing and building a detector of the technical complexity of CMS, a globally distributed collaboration has been assembled with different backgrounds, expertise, and experience. An international collaboration of nearly 3500 people from nearly 200 institutes in 40 different countries built and now operates this complex detector. The energy and luminosity of the LHC collisions brings us to a new frontier of potential physics discoveries. We must answer the key questions concerning the form and the limits of the Standard Model - How can anything have mass? Are there deeper symmetries and particles beyond the Standard Model? Why is the Universe predominately made of matter, and not antimatter? All of these topics require the physics analysis of the very rare particles and decay processes. However, the distributed nature of the collaboration makes things complicated. The users need to understand the

⁶ Sudhir Malik, MS 121, Fermilab, PO Box 500, Batavia, IL 60510, U.S.A.

complex CMS computing environment and the physics tools as a pre-requisite task. The CMS User Support engages the entire collaboration in this important task by using the collaborative tools [2]. This paper describes the goals, challenges, working strategy and activities of the User Support.

2. Challenges

The distributed environment of the CMS collaboration brings many challenges in terms of the data analysis for physics. The growing collaboration size and the life span of high-energy physics (HEP) experiments has brought in a need for an organized effort to support users in meeting their physics needs. The model where the collaboration size is relatively smaller than CMS and people could do physics analysis efficiently only at the hub of detector activity is no longer practical. The financial and logistic constraints may not permit everyone to travel frequently or for extended periods to CERN. The users come from different work cultures and experiences. In addition, the complex computing environment and non-trivial tools needed to do physics add to the time it takes for a user to be fully equipped to do physics analysis. These scenarios pose a new, unique and a big challenge to support the physics needs of this community.

To realize the full physics potential of the CMS detector and contribute effectively towards the physics analyses, it is very crucial that users quickly come up to speed by learning about the CMS computing environment and the available physics and software analysis tools. The CMS Experiment has put in a lot of effort to organize a User Support structure that plays a vital role to help the collaborators achieve the goal above. To engage a big collaboration like CMS in physics analyses activity, the User Support uses the available collaborative tools supported by CERN. It provides a general physics support by maintaining, supervising and improving the CMS software and computing user documentation in the form of hundreds of web wikis, conducts dozens of tutorials year around on the tools needed as a pre-requisite to perform physics analyses and answers user questions via emails or in person. The big workforce required for the User Support activity comes from the users themselves. The users learn, gain experience, provide feedback, help improve the techniques adopted and in turn help new users learn.

3. Organization

The CMS Experiment is organized into several tasks for efficient management in several areas. Among them is the CMS Computing that is further divided into the Data, Analysis and Facility Operations, and the User Support. They are geared towards helping users analyze data and produce physics results. The Data Operations team gets the data out of the CMS detector (online system), processes and re-processes and distributes them for analysis all over the world to different computing centers. Analysis Operations focus on the operational aspects of enabling physics data analysis at the globally distributed computing centers worldwide by managing data movement, access and validation. The Facility Operations provide and maintain a working distributed computing fabric with a consistent working environment for the Data Operations and the users. The CMS User Support provides the general computing software and physics support structure to enable users to do physics analysis.

The CMS User Support is lead by two co-conveners. There is a very small number of dedicated personal doing User Support and therefore the manpower is drawn from the CMS user community itself. The users who are expert in different physics analysis tools help new users learn and master them. They in turn become experts and provide feedback to further improve the instruction material and style. These new experts get recruited to help the User Support effort further. In this way the cycle of teaching and learning goes on in a self-sustained manner. This mode of operation is a clear manifestation of the big collaborative spirit without which it is not possible to run such a diverse and globally distributed community as CMS.

4. Collaborative Tools

The goal of the User Support is to engage the collaboration in learning the tools needed to perform the physics analysis and bring them up to the speed to contribute to the same. It is highly desirable to distribute the expertise besides CERN to other collaborative institutions. To meet these challenges the User Support uses the collaborative tools available and supported by CERN. It does not take part in developing any collaborative tools, however, the CMS community serves as a very active test bed for them. It maintains and supervises the CMS documentation and periodically organizes tutorials on learning computing and physics tools and workshops on physics analyses. It answers general computing and physics questions via emails and online chat sessions. The specific questions go to the respective discussion forums and get answered in a collaborative manner by the experts. Whereas tools like Indico [3], EVO (Enabling Virtual Organizations) [4], Twikis [5] and Hypernews [6] enable it to communicate with the users, for a dedicated purpose, like setting up a tutorial forum, it uses espace Sharepoint technology [7]. These tools are described below.

4.1. Indico

Indico is web-based tool used to schedule and organize events, conferences, meetings, workshops with sessions and contributions. The tool also includes an advanced user delegation mechanism; allow paper reviewing, archival of conference information and electronic proceedings. The User Support uses it to display the agenda for a tutorial, upload presentations and video recordings of the tutorials.

4.2. EVO

EVO is heavily used by CMS and other LHC experiments for video conferencing. The User Support utilizes EVO to enable successful participation of the remote collaborators during the tutorials. The video/audio recording feature of EVO is used to record the tutorials.

4.3. Twikis

CMS uses web twikis for the purpose of its knowledge base and documentation management system. The flexibility to edit and modify a twiki by any CMS user and instantly publish it using only a web browser (no programming required) makes them very suitable for a collaborative environment.

4.4. HyperNews

Hypernews (HN) is a discussion management system used by all LHC experiments including CMS. It bridges the use of e-mail and forums. The user may subscribe to forums to be informed by e-mail or look at the forums of his choice via the web interface. He may reply by e-mail or through the web interface. Discussions are grouped into Forums and threads, and Forums are in turn grouped into Categories. Currently, CMS has about 20 such Hypernews categories with about 700 forums. HyperNews is used to send announcement about tutorials and documentation review and to respond to questions/comments from the users about documentation.

4.5. SharePoint Technology (espace)

The User Support uses espace an online collaboration tool based on Microsoft SharePoint technology, to prepare and conduct tutorials. This is used to plan and build a course workflow that goes from announcing a tutorial to getting feedback. It serves as central platform that has link to the tutorial announcement, indico agenda for the tutorial lectures and exercises, the submission exercises from, discussion forum and the feedback form.

5. Activities

In this section we will describe how do we use the collaborative tools CMS community-wide to spread the knowledge of analysis tools and related communication in order to successful do physics analysis. In particular these tools are used for documentation, tutorials, software reference and presentations.

5.1. Documentation

CMS has about 17000 twikis serving the entire CMS documentation. The usability of this documentation has been tested [8]. The challenge of such an enormous amount of information is to track the information that may be stale or outdated. To deal with this and improve the performance of the twiki usage, a reminder service of twiki pages has been tested and deployed. As a result, each twiki responsible gets a reminder about the pages that he or maintains and is encouraged to remove the outdated information. To reduce effort on part of the responsible, the reminder service in addition offers a possibility of a delete request. The twiki user only needs to select the topics to be deleted and they will be deleted centrally.

The User Support manages and periodically reviews structured documentation described below. Many topics are also reviewed by the end-users to make the information as best as possible.

5.1.1. WorkBook

WorkBook [9] serves as the first entry point for a user on how to go about approaching physics analysis. It is in the form of about a dozen chapters spread over about 100 twiki pages. The twiki page on each topic has a responsible person associated with it. The WorkBook covers all aspects to help and equip a user jump-start physics analysis. One learns about how to get computer accounts and learn CMS computing framework, common physics tools and examples on advance analyses.

5.1.2. Software Guide

The Software Guide (SWGuide) [10] has details of each domain belonging to the CMS software. It contains details about the data formats, software framework and physics analysis software. There are about 1350 twiki pages and each topic has a responsible. For example, if a user looks at the WorkBook topic on how to use Physics Analysis Toolkit (PAT) [11], he can find more details on its concept and coding structure in the SWGuide.

5.1.3. The Data Formats and Software Reference Manual

CMS uses Doxygen [12] as a tool to generate and maintain reference manual for data formats and CMS software. With this tool one can generate the documentation from the software code for each release of the CMSSW. The User Support has provided software scripts to better integrate the reference manual with the rest of the documentation. The user, while reading the document, can easily refer to the actual code. The software manual contains cross-links to the main documentation in the SWGuide and lists all the packages connected to each software domain.

5.2. Tutorials

The tutorials are organized [13] throughout the year on common physics analysis tools. They are given by the experts and rely on the collaborative spirit. The number of attendees in a given tutorial is limited by the practical factors such as space in the conference room and number of tutors available. A typical number is around 25, though some additional participants attend the tutorials remotely. Some topics on which the tutorials are held are mentioned below along with their importance towards accomplishing physics analysis. The tutorials are mostly held at CERN but also frequently at other collaborative labs like FNAL (Fermi National Accelerator Lab), DESY (Deutsches Elektronen-Synchrotron) etc. The presentations and video recordings of the tutorials are uploaded to Indico as a persistent storage for the users to view later. Frequent tutorials ensure an up to date documentation.

5.2.1. Software Framework and Python

The CMS software framework is very complex. It is essential to work on any computing aspect of the CMS Detector Studies or Physics Analysis. It uses configuration files scripted in Python language to configure the initial settings for computer programs. Thus having CMSSW tutorial along with the usage of Python is required. This tutorial is now a pre-requisite for part Physics Analysis Toolkit (PAT) tutorial described below.

5.2.2. Physics Analysis Toolkit

Analyzing physics data at LHC experiments is a complicated task involving multiple steps, sharing of expertise, cross checks and comparing different analyses. To maximize physics productivity, the CMS experiment at LHC has developed a collection of analysis tools called the PAT (Physics Analysis Toolkit). A comprehensive training program was designed and setup on using PAT software as an integral key part of the analysis of data from CMS experiment. This Software framework and Python are now a part PAT tutorial as pre-requisite exercises.

5.2.3. Statistical Tools – RooStat and RooFit

To keep CMS community abreast with statistical tools and modeling of the physics event distribution in data analysis, a combined RooStats [14] and RooFit [14] tutorial is organized. One learns to implement techniques like - frequentists, Bayesian and likelihood based methods for statistical calculations.

5.2.4. Grid Computing

Analysis of CMS data is performed in a distributed way using grid infrastructure. CMS Analysis Remote Builder (CRAB) [15] is a utility to create and submit CMSSW jobs to the distributed computing resources. It interacts with the local user environment, the CMS Data Management services and with the Grid middleware. It is essential to learn this tool in order to successfully perform one's data analysis.

5.2.5. Event Display

The 3D-accurate display to visualize proton-proton collisions is very essential to understand the physics and the outcome of the physics event. The CMS event display, called Fireworks [16], is specialized for this purpose. The collision data is presented via user-friendly graphical and textual views. This tool is easy to install and use. To benefit maximum to one's need does require some detailed coaching in the form of tutorial.

5.2.6. CMS Orientation for CERN summer students

To give orientation about CMS, series of lectures are organized every summer for the CERN summer students and the new comers to CMS. These lectures give them an overview of CMS detector, physics and software.

5.2.7. Analysis Workshops

Holding analysis workshops has also been a great way of introducing new members of the collaboration - students, post docs and faculty - to CMS software and data analysis. User Support organizes workshop on analysis problems on collision data as well as data simulating the conditions at LHC. The experienced volunteers from the CMS facilitate team of tutors and lecturers. Teams of participants and attendees on the order of about 100 present their simulated physics measurements in "conference" at the end of the workshop. The material prepared for this workshop serves as a persistent resource for CMS after the close of the Workshop.

5.2.8. Office Hours

It also provides in-person help where a user can walk in and ask questions during specific office hours.

6. User Feedback

In September of 2010, the CMS computing project conducted a survey to see how CMS worked for the users with respect to their computing and analysis needs. This survey included specific questions about the User Support. About 120 participants took part. The survey indicated that the users find SWGuide and WorkBook as a very valuable as shown in figure 1. The users also felt that more topics for tutorials should be added. The users prefer twiki pages for self-study and HyperNews to direct

their questions. Every tutorial session is followed by feedback survey. It helps improve the quality of the instruction material and its organization of the next tutorial.

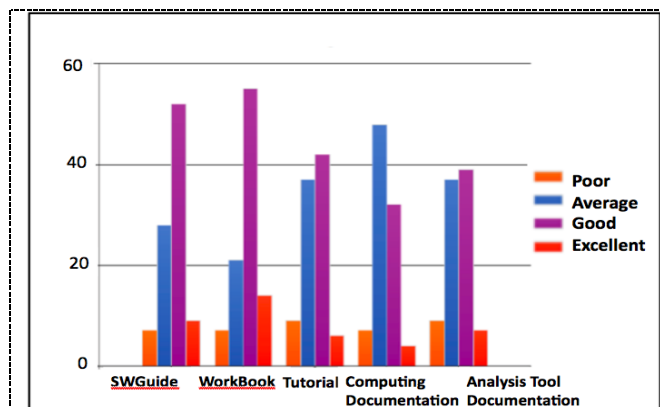


Figure 1. User rating for the documentation

7. Conclusions

The CMS User Support model has proved to be successful compared to its predecessors in other HEP experiments. It has put systematic and organized effort to equip users do physics analysis. It has made impact in awareness and usability of the structured documentation suite, facilitating the usage of common physics tools. It relies on the collaborative effort in maintaining, contributing and improving the documentation and holding tutorials. The user feedback is essential to maintain and improve the CMS specific knowledge base. The collaborative spirit is the key to its success.

8. References

- [1] The CMS Collaboration, “The Compact Muon Solenoid Technical Proposal”, CERN/LHCC 94-38, 1994
- [2] Goldford S *et al* 2010 J. Phys.: Conf. Ser. **219** 082001
- [3] López J *et al* 2010 J. Phys.: Conf. Ser. **219** 082002
- [4] **EVO** stands for Enabling Virtual Organizations. More info on <http://evo.caltech.edu>
- [5] Amran N *et al* 2010 J. Phys.: Conf. Ser. **219** 082008
- [6] **Hypernews** is web-based threaded discussion board software written in Perl language. More info on <http://hypernews.org>
- [7] **Espace** is a Self-service site creation for collaboration workspaces based on SharePoint - MOSS 2007. More info on <http://espace.cern.ch>
- [8] Lassila-Perini K and Salmi L 2010 J. Phys.: Conf. Ser. **219** 082011
- [9] <http://twiki.cern.ch/twiki/bin/view/CMS/WorkBook>
- [10] <http://twiki.cern.ch/twiki/bin/view/CMS/SWGuide>
- [11] Adam W *et al* 2010 J. Phys.: Conf. Ser. **219** 032017
- [12] **Doxygen** is a documentation generator for the programming languages. More info on <http://www.doxygen.org>
- [13] Lassila-Perini K See “Planning and Organisation of an e-learning ...” in this journal
- [14] **RooStats** is a project to create statistical tools built on top of **RooFit** and distributed in ROOT. More info on http://root.cern.ch/root/html/ROOFIT_ROOSTATS_Index.html
- [15] CMS - The Computing Project Technical Design Report, CERN/LHCC 2005-023, CMS TDR 7 (2005)
- [16] D Kovalskyi *et al* 2010 J. Phys.: Conf. Ser. **219** 032014