A quantitative study of TWiki at CERN after ten years of use

Peter L. Jones
CERN
IT/CDA
Geneva 23, Switzerland
+41 22 7674901
Peter.L.Jones@cern.ch

ABSTRACT

The European Organization for Nuclear Research known as CERN [1], is the home of the Large Hadron Collider (LHC) [2] where physicists are recreating conditions close to those at the origin of the Universe. Although the LHC is based at CERN over 10,000 scientists and engineers from over 100 countries and hundreds of universities and laboratories collaborate on this project.

TWiki [3] is an online collaboration platform and was introduced at CERN at the request of the LHC experiments, ALICE, ATLAS, CMS and LHCb have all used TWiki since its introduction.

By analyzing the TWiki website statistics from 2006 until April 2016 this paper presents a quantitative study of how TWiki has been put into practice. The results will give a general picture of who is using the system and how this has evolved over the years.

Categories and Subject Descriptors
H.4 [**Informations Systems Applications**]:

General Terms
Documentation, Human Factors.

Keywords
CERN, LHC, TWiki.

1. INTRODUCTION

TWiki was first released in 1998 and is a simple to use web application platform. Since its introduction at CERN in 2003 it has successfully grown into an important collaboration tool for the LHC experiments and also for groups and individuals for numerous other projects. Over 160 million transactions have occurred since the CERN TWiki installation.

Over 170000 topics have been created to the present day and there have been over 27000 unique editors. Many users have come and gone, new projects start using the tool and other projects migrate from TWiki to some other platform. Through the available data we can analyse the transactions for reading and writing and identify the user groups and trends in use.

2. THE TWIKI SYSTEM
2.1 TWiki webs, topics and attachments.

TWiki webs are the namespaces that group together common documents (topics). Each experiment has its own namespace called a web, for example all CMS related topics will go in the CMS web. They may also have several others sub-webs for more specific work. The TWiki topics are kept inside these web folders and it is possible to attach images and other media documents to the topics. In this study we will only concentrate on webs and their topics.

2.2 The format of the TWiki system logs

There are several mechanisms for recording TWiki use. It is an online tool running on top of an apache webserver that registers every event. TWiki also uses the RCS[4] version control system, whereby each TWiki (topic) has a sister file that contains details of changes to the document including the editors account name, the date and time of any changes made and the version number. The TWiki system also records transactions in its system logs. A new log file is created each month and a new line is appended to the log each time a user visits the website. Table 1 shows the format of the TWiki system log file.

Field	Meaning
datetime	In format YYYY-MM-DD HH:MM:SS
user	The account or email address of the user
command	Transactions like view, save, rename
web.topic	the topic name and web folder
comment	Command specific text
ipaddress	Ipaddress of the URL requested

Table 1. TWiki log file

3. DATA MINING AND RESULTS

The system logs are text files of one line per TWiki transaction but in order to better query the millions of lines of data, all the logs are transformed into MySQL database tables. SQL queries can then be executed to search for statistics of interest. Below are several important examples with data analysed over a 10 year period.
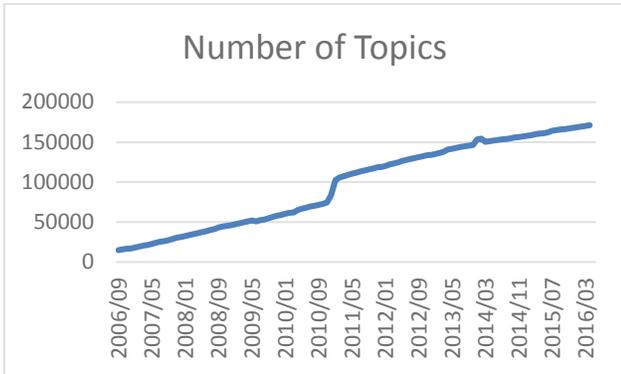

Figure 1. The number of TWiki topics recorded each month
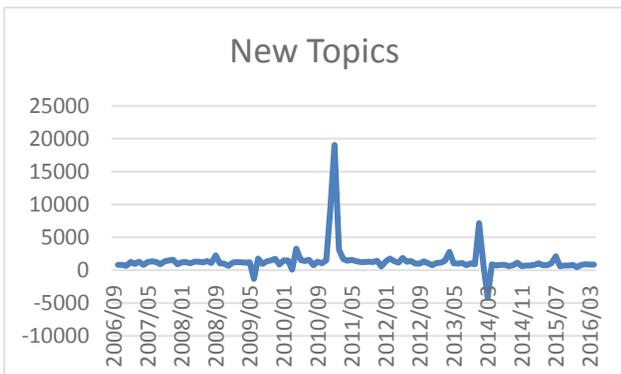
Figure 2. The number of new topics added to TWiki each month
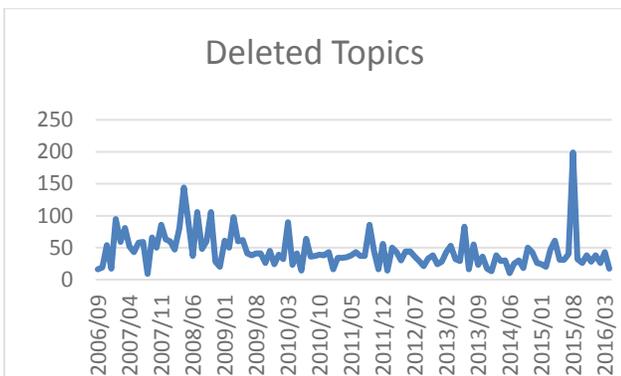

Figure 3. The number of topics moved to the Trash each month
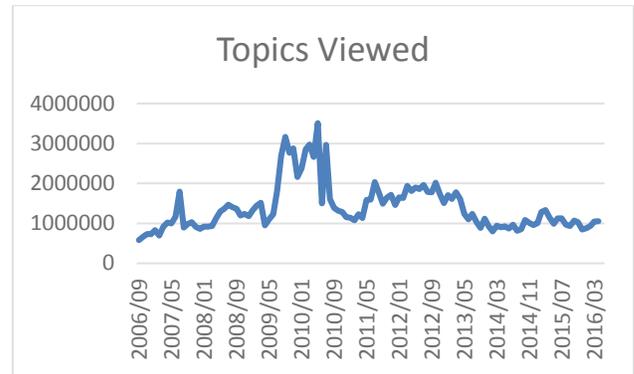

Figure 4. The number of topics viewed each month
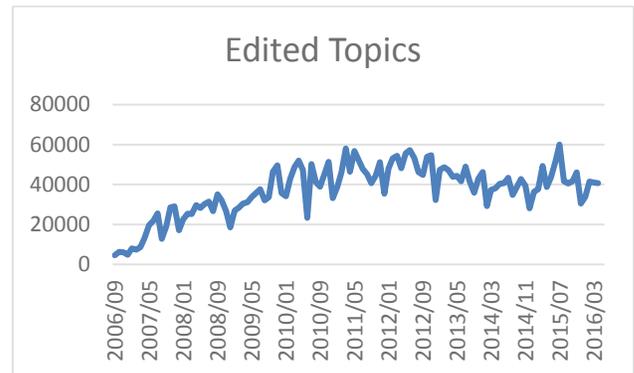
Figure 5. The number of monthly topic modifications
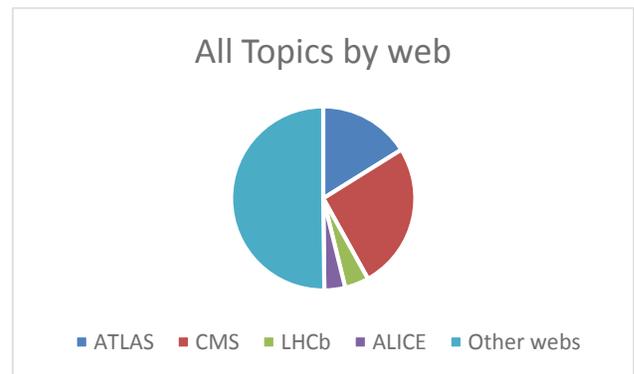
Figure 6. LHC experiments are the main users.
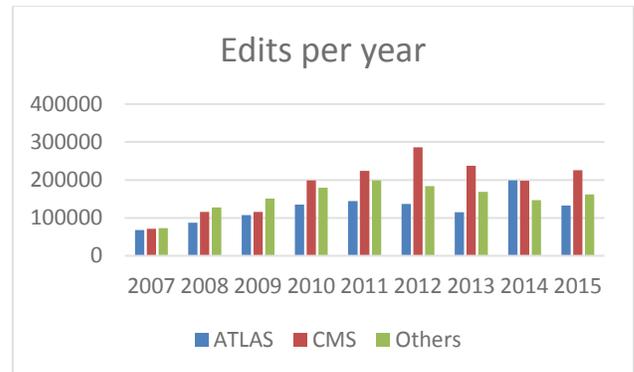
Figure 7. Number changes made by ATLAS and CMS over time.

4. DISCUSSION

Figure 1 shows that the number of TWiki topics has steadily grown over the years but at the same time very few of these are removed from the system (Figure 3). Comparing these metrics show that in 2016 for every topic deletion 50 new topics are created. Further investigation is needed to check for the existence of redundant topics with a possible action necessary for automatic deletion.

Figure 4 indicates that the amount of topics viewed has stabilised since 2013 and figure 5 shows that users are continuing to modify topics at a steady rate however there are some notable spikes especially in 2010 when there was high increase in use for both reading and writing. Further investigation shows that data was imported into the system from other wikis during this time.

Figure 6 reveals that although the LHC experiments make up for half of the TWiki data, there are many smaller groups using it.

Figure 7 confirms that the main users, ATLAS and CMS are still very active in collaborating.

5. CONCLUSIONS

Our main goal was to present the use of TWiki by analysing the available recorded data and we have shown that TWiki has proven to be and continues to be a popular collaboration tool for the LHC experiments and also for many other projects. These logs have enabled us to view the history of TWiki use over the last 10 years and identify the types of transactions made by users. Finally, we identified further areas of research and possible areas for improvement to the service.

6. REFERENCES

[1] CERN. The European Organization for Nuclear Research. URL=http://public.web.cern.ch/public/

[2] The Large Hardron Collider. URL= http://public.web.cern.ch/public/en/LHC/LHC-en.html.

[3] TWiki. An open source enterprise Wiki. URL= http://www.twiki.org