Logbooking Software for Science

F.P. van der Meulen, 500713781, (tel)+31 6 17506168
Amsterdam, 2nd of March 2018
Amsterdam University of Applied Sciences
HBO-ICT, Game Development
C.J. Rijsenbrij
Software for Science
Marten Teitsma
February Semester, 2017-2018

June 27, 2018

F.P. van der Meulen, 500713781, (tel)+31 6 17506168
Amsterdam, 2nd of March 2018
Amsterdam University of Applied Sciences
HBO-ICT, Game Development
C.J. Rijsenbrij
Software for Science
Marten Teitsma
February Semester, 2017-2018

Contents

Preface

I would like to thank the following persons:

Abstraction

The abstraction of the report.

1 Introduction

This is a student thesis from the Amsterdam University of Applied Sciences from the HBO-ICT-Game Development study course.
Software for Science is an organization, lead by Dr. Marten Teitsma, that combines software with science. This organization is created as an organization from the Amsterdam University of Applied Sciences. The organization works with Astron, eScience center and CERN to create software for their scientific experiments.
Conseil Europen pour la Recherche Nuclaire, CERN, is a scientific research center in Geneva, Switzerland, focused on researching nuclear energy. CERN works mostly with particle colliders, e.g., shooting particles against each other in order to find out what energy comes free from one of those runs. CERN works with multiple particle colliders to create new particles, or to research events that has happened in the past. One of those particle colliders is ALICE.
ALICE, A Large Ion Collider Experiment, has its own bookkeeping system. This system makes sure that the runs ALICE makes are recorded into the bookkeeping system so that researchers, collaborators with CERN and other people can look back upon the runs. This also includes reports whenever a shifter, a scientist or student that works with ALICE, has completed their shift and need to write down what the shifter did during their shift. Other kind of reports are the reports about incidents with ALICE, reports about filling the gas etcetera.

1.1 The assignment

In December of 2018, ALICE is going under maintenance. CERN calls it a long shutdown. During the maintenance, subsystems and parts of ALICE are being renewed. Before the maintenance started, the request for a new bookkeeping system was made. There are three reasons for a new bookkeeping system. The first reason is that the new system must combine the bookkeeping system with AliMonitor, a monitoring system. The second reason is that there are multiple desired functions that are not added to the current system, such as automated report making with the help of a template(at the moment, the shifter that has to make a report copies the previous report. The third reason is that most of the technologies and frameworks used to develop the system is outdated due to the fact that development started around 2009. From 2009 until now, new features were added to the system and new tables were created in the database. This caused that the structure of the code is not organized, the database structure was not efficient in handling requests from the client.
Software for Science is planning to deliver the new bookkeeping system in January. At first, they plan to show a prototype of the bookkeeping system in June to CERN software developers and users that are involved with the current bookkeeping system. The reason why a prototype needs to be created is that Software for Science wants to learn more about creating a bookkeeping system. With the prototype, it will be easier to detect possible flaws when creating the actual bookkeeping system.

After the prototype, Software for Science hosts a summer school, a special program for developers and students that are interested in developing software for scientific research. At the summer school, new features are added to the prototype. Finally, at September of 2018, a semester at the University of Applied Sciences will be held in order to create the final product.

Due to the size of the project, the little amount of time available for the development, and the many requirements that CERN has given for the actual product, not every requirement and feature can be implemented into the prototype. Furthermore, it is important that possible flaws with the chosen requirements can be detected as early as possible. Based upon the current situation, the following research question can be formulated:

Which requirements can be implemented into the logbook system for ALICE?

In order to make a prototype for the new bookkeeping system, the back end and the front end are separated from each other. This was done due to the fact that the scope of the project would become too big and thus there would be less requirements implemented into the prototype. Another reason was that the front end and back end of the application will run differently. While the back end will run on a server, the front end must run on every computer at CERN in June. This thesis handles the back end of the new bookkeeping system. The front end will be developed by Naomi Nazar.
This thesis is split into four parts. Every part resembles a sub research question. An explanation why these sub research questions are important and how they could help solving the main research question can be found later in the thesis. The sub research questions are as followed:

1. Which requirements are important for the logbook system for ALICE?
2. What are the consequences for the development process of the prototype based upon the important requirements and the database choices?
3. How can the prototype be developed?
4. How to test the prototype?

The second research question will be completed with the help of the front-end since the front-end needs to design and create the look of the prototype.
Based upon the research questions, a bookkeeping prototype can be created. The prototype will be demonstrated to the developers at CERN. The prototype will be delivered with the thesis.

2 Methods

This chapter discusses the methods used to solve the research questions. The four research questions that were earlier created will be discussed and further explored. At first, the method on how to complete the research questions will be explained, then the outcome of the use of those methods will be discussed and finally a conclusion will be given to the research questions.

2.1 Req analysis

This section of the thesis will discuss the requirements analysis. There are over 120 requirements created by CERN, these can be found as an attachment. It is impossible to use every requirement in the prototype, due to scale of the requirements and due to time constraints that are set. An analysis is needed in order to check which requirements are necessary to add to the prototype and which requirements can be left out.
Inside the requirement document were some vague terms that weren't explained. The terms were follow-ups, on call intervention, announcements, and an EOS report. These vague terms would create confusion when prioritizing the requirements. And that could cause that important features could be skipped.
To solve the vague requirements and unknown wishes, a former shifter by the name of Pascal Buschoten has been interviewed. The full transcript can be found in the attachments. The interview with Pascal clarified terms like the on call intervention and the end of shift reports with templates. On call Interventions are emergency issues with a detector, a particle collider or another system that need to be solved as soon as possible. These issues need to be fixed within a very short amount of time. Currently, there is not a formal template available for the E.O.S., End Of Shift, reports and because of that, all the reports do not follow the same structure. With a template, it is better to evaluate all the E.O.S. reports).
Software for Science has given a criteria for the prototype. This criteria was that all the requirements related to the Subsystem Run Coordinator must be added to the prototype. This criteria was crucial for the requirement analysis.
During a meeting at the ninth of may, one of the developers from CERN, Vasco Chibante Barroso, stated that he prefers that the prototype would focus more on the features instead of the roles and the authentication. This complicates the planning of the new features for the prototype. Therefore, there will be a focus on both the features that CERN requested and a focus on the Sub System Run Coordinator requirements that Software for Science has asked for.
The technique that will be used is the Analytic Hierarchy Process technique.[9] "In A.H.P., initially whole requirements are recognized and then criteria under which these requirements will be preferred. In A.H.P. we pair wise analysing between the probable pairs of the hierarchy. Now users can recognize the possible relationship between the hierarchies. We then pair wise analyse them and users can select its preferences from the scale which ranges from one to nine."(Javed Ali Khan, Izaz Ur Rehman, Yawar Hayat Khan, Iftikhar Javed Khan, Salman Rash, 2015). One of the main advantages of using this technique is that [10]. The second software requirements analysis technique that will be used for the software is the Hierachy Analytic Hierachy Process technique[9]. H.A.H.P. is a technique that creates so called planes of requirements in order to simplify the requirement analysis process. A plane is a group of requirements that are grouped together if they share a similarity. These planes can, for example, align to a user of the system or a feature that will be implemented into the final product. For this requirement analysis, the planes will consists of features since this will be more important for the prototype than the users themself.

Lastly, time will be an important factor for the requirement analysis. The time will be estimated using the story points technique. This technique consists of giving points to requirements in order to organize the effort and scale of a requirement. The range of these points are: zero, zero point five, one, two, three, five, eight, 13, 20, 40 and 100.

Applying the chosen techniques to the requirements document results in tables with prioritized requirements with scrum points given to the requirements. The requirements are grouped by feature so that it is easier to read. There were numerous requirements that were double. The cause of this was that different users have the same requirement. These requirements are grouped together to prevent duplicity.

2.2 Consequences

This section of the thesis will be about choosing which requirements can be used to develop the prototype. Now that the requirements are prioritized according to the criteria that has been set up earlier(the criteria were from CERN(focus on multiple roles), S.F.S(focus on the subsystem run coordinator) and estimated amount of time to complete the requirement), it is time to choose which requirements can be added to the prototype. At first, the list of prioritized requirements will be used to choose requirements that would fit the prototype. Once the requirements are removed from the list, a selection of the requirements will be made with the help of the front-end developer, Naomi Nazar. The front-end needs to know which requirements will be implemented into the prototype so that the front-end can start working on these requirements, therefore, a cooperation with the front-end is mandatory. The results from this selection will be a list of requirements that will be implemented into the prototype. With the list will be an explanation for each requirement why this requirement was picked by both the front-end and the back-end.

The requirements that have a priority of eight or nine will be instantly picked from the requirements list. Since these requirements have the closets match to the criteria set by CERN, S.F.S and are achievable within the available amount of time, these requirements are the most likeable to implement them in the prototype. Since the rest of the requirements either don't match or could match one criteria, but the other criteria not, they will not be implemented into the prototype. After applying this method on the list of prioritized requirements, over 50 requirements could be left out of the prioritized requirements list. With the smaller list, it will be more manageable to choose which requirements from the priority list can be implemented into the prototype.

2.3 Prototype

After identifying which requirements can be implemented into the prototype, the development of the prototype can start. The requirements need special kinds of technologies, e.g., uploading files to the server need special technologies in order to make it work. Including with the use of the technologies are frameworks. The frameworks can make the development process easier in the long term, due to the fact that it will not be necessary to write additional code that has been written before. The code itself will be developed according to a Representational State Transfer, REST in short, API. The REST api was required by SFS to be used for the prototype. It is important to think of the use of design patterns, a reusable piece of code to a given problem within software engineering. Design patterns could harm the performance of a system, therefore, design patterns should be used only when there is either a performance gain possible, the design patterns can improve the reuse ability of the code, or the design pattern can improve the security of the code. Once these important factors are thought out, it will be easier to create a class diagram of the prototype and a Entity Relation Diagram to have a guideline for the development process.

The main programming language for this research is Node JavaScript. Node JavaScript is a hard set requirement, e.g. the project must use JavaScript on the back and the front end. The version of Node that is used is version 9.4.0. At the start of the development, this was the most recent version of node JavaScript that was used. The preference of CERN is to use CERN's own developed frameworks as much as it is possible to do so. The WebUi framework is a framework to handle HTTP requests made by the client, in this case, the front end. The base of this framework is the ExpressJs framework. The ExpressJs[1] framework is a lightweight framework for handling HTTP calls. CERN has expanded this framework with features such as Json Web token support, debug logger systems and support for CERN's own authentication system, SAMS. Since Software for Science did not had access to SAMS at the time, it was decided to create a custom login system. The testing framework that has been chosen is the Mocha[4] testing framework. It is important that the entire application is tested. For instance, whenever a new feature is added to the prototype, it is important that the older features are still working as intended. With using tests, there is a better check if a feature still works. Mocha was one of the requirements set by CERN to use as the testing framework.
Scrum[7] is the work management method that is used for the development of the prototype. Both the front end and the back end make use of this work management method. With scrum, sprints are created to develop features on the according time. The final four sprints of the prototype will be discussed later in the thesis. Git has been used as the version control system for the prototype and for the thesis. With Git, it was possible to store the code online, making sure it was always available for use on any computer. This is the same case for the thesis.

Before choosing which design patterns can be implemented into the prototype, it is important to identify what possible problems could occur during the development of the prototype. The first problem could be the readability of the code. Since the backend of the prototype will be written like a REST API, it will feature http-requests, requests from the frontend. These requests could look like this:

```
server.get('/test/single/entry', handler function());
server.post('/test/post/entry', handler function());
server.get('/test/user/info', handler function());
```

The REST API could handle way more http requests than these three. The code could be harder to read. Another problem could be that it will be difficult to modify the code in later stages or add end points that could use the same handler function. To solve these problems, the Model View Controller design pattern can be used. With the MVC pattern, the data that the request wants(Model) can be isolated from the user interfaces(Views) and this data can be managed with the help of the Controller(BRON BOEK). This design pattern can help with organizing the code. For example, all requests that feature the log entry could go into the controller file. To obtain the required data, the specific log entry model could be called and then this data can be rendered according to the required view. The rendering of the view will be different for the prototype, due to the fact that the front-end, Naomi, will handle all the views. Therefore, the view will be more of a way that renders the data in a specific way that the front-end can work with. With using the REST API, another possible problem could occur. The data will be stored into the database. The REST API will likely want to retrieve or to write data to the database. This would mean that the database could be called multiple times within a second. Starting up the database and closing it down every time a call takes place would reduce the performance of the database significantly. The database will be also called in multiple places within the prototype. This could be the case with the configuration settings of the prototype as well, since this will be also called in multiple places. The pattern Singleton would solve these problems. The singleton design pattern creates one instance of an object that is instantiated once during initialization. This is useful for the database, since this would mean that the database is instantiated once during the start up of the application and that the database doesn't need to be turned off during run time.

2.4 Test cases

This section of the report will be about the creation of test cases for the prototype. These test cases will be important for later development of the bookkeeping system, because with setting up test cases this early will identify possible flaws with the actual product. Furthermore, it will be easier to create test cases for the final product when there are test cases already available. At first, it is important to identify which requirements of the prototype should be tested and in what ways it should be tested. After the identification of what to test is done, a test case template will be created. This is done to organize the way the software should be tested. Finally, to execute the test cases, pre programmed tests will be made so that the software stays tested. The results will be test cases that cover most of the software accompanied with programmed tests.

One of the mandatory things to test are that the requirements are correctly implemented into the prototype. These requirements could be the standard of a test. With every requirement, multiple tests can be thought of. For example, with creating a log entry, test could be about that creating the log entry is successful. It is also possible that a user did not fill in all the required fields of a log entry or that the user tries to add code that enables the user to hack the server and so on. Therefore, it is important that a list of possible scenario's is made. These scenario's consists of a single sentence that gives enough information about the scenario and the goal of the test. These scenario's can be found as an attachment. There double scenario's within the document. For instance, one of the recurring scenario's was with the token not being provided or that the token is not from the server itself. These scenario's are removed and written only once.

3 Results

This section of the report will discuss the results, which are in this case the prototype. An U.M.L. diagram of the final prototype will be shown, and code snippets that are in the prototype will be shown, there will also be a section related to the requirements that are implemented.

3.1 Requirements analysis

Applying the analysis methods on the requirements results into requirements that have gained a priority rating and an estimated time to complete. Due to the size of the requirements, four sections of the analysed requirements will be shown in the results. The rest of the analysed requirements can be found as an attachment.

3.1.1 Entries

These requirements are about the different kind of entries into the system(log entries, on call interventions, run entries etc.). The rest of these requirements can be found as an attachment. This feature sits at the core of the to be delivered product and therefore has the most requirements attached to it. Due to the amount of requirements that are related to the entries, the section in the requirements document is divided into different sections for readability.

Role	User Story	Priority	Estimated time
Shifter	A shifter makes an entry into the database consisting of several items. Each entry records the following items: time of creation,which class the creator originates: human, type of entry, general, EOS, DCS, number of run, author of the entry, title of the entry, log entry, follow ups, files and actions	9	8
User	As a user, I want to search log entries by different criteria (e.g. title, content, author, creation date,(...) and have the results listed.	9	8
Run coordinator, S.R.C. , S.T.C. and Shifter	As run coordinator, S.R.C. , S.T.C. or as Shifter, I may need to attach files to log entries. These files may contain text or binary information (PNGs, JPGs etc.) (Roberto Divia).	8	8
Subsystem expert	As a subsystem expert, I want to attach quality flags to runs so that physicists can use them while searching for good data sets for their analysis (Vasco Chibante Barroso).	8	8

3.1.2 Reports

A requested feature in the new final product was the automatic creation of reports and templates. This is not available in the current product, therefore, there are requirements referencing towards the creation of templates and reports.

Role	User Story	Priority	Time
Subsystem Run Coordinator	As a SRC I would like to be able to create my own detector specific templates for example On-Call interventions. In this case I can specify the relevant information which are required from the OnCall shifter for different kind of standard events (Robert Munzer).	8	20
Shifter	As a shifter, I want to have templates that prefill most of my end-of-shift reports from the available metadata so that I dont need to fill inmyself what the system already knows (Vasco Chibante Barroso).	8	13

3.1.3 Email

Sending emails is available in the current system, however, the current way the emails are send are sent in a spam related way, e.g., a shifter receives an email if another shifter makes an entry from another subsystem. Therefore, the email system needs to be optimized and it should be designed in a way that will not be annoying to the user. The reason that this is of more importance, is that it is easy to adjust the way the emails are sent, how the emails look like and who receives an email. That why, with adding it to the prototype, it will be easier to adjust it to the wishes of the users.

Role	User Story	Priority	Time
ALICE member	As an ALICE member, I would like to receive via email a global summary of each LHC Fill in order to follow ALICE operations without visiting the bookkeeping tools. Currently in the ALICE logbook, I like that I receive via email a document with info on efficiency and EOR Reasons and that on the body of the email there is a summary for each fill (Vasco Chibante Barroso).	8	20
Subsystem Run Coordinator	As a subsystem responsible, I want to be notified by email (or other channels) of log entries which are related with my subsystem so that I can better follow-up activities without having to constantly visit the product, e.g. EOS report (Robert Munzer) (Vasco Chibante Barroso).	8	13

Run coordinator	As run coordinator I may request to receive automatic e-mails concerning all Logbook entries that include all systems (either without distinction or using special selection criterias). The e-mail delivery address will probably be an e-group (single e-mail address ¡...¿@cern.ch)(Roberto Divia).	6	13

3.1.4 Roles and Authentication

There are different kinds of roles and users in the bookkeeping system. It is possible that a user is assigned to multiple subsystems with different roles. Only certain roles can move in and out of subsystems. Since the different roles play an important part in the bookkeeping system, it would be important to have it.

Role	User Story	Piority	Time
User	As a user, I want to be able to login with my CERN credentials to avoid having to remember a new set of credentials. This should be done by using the CERN authentication method.	9	13
Run cordinator, SRC and Administrator	As run cordinator, SRC or Admin, I must be able to move collaborators to and out of subsystem teams. These action may be conflict the information stored in SAMS (Roberto Divia).	9	13
Run cordinator and SRC	As run cordinator or SRC I need to give ALICE collaborators write or read-only access to the logbook. These rights will be superseeded by equivalent rights given according to the function of the user (e.g. a ALICE collaborator with read-only access will be given write access during the time of his/her duties as a shifter, subsystem run cordinator or system team member) (Roberto Divia).	6	13

After the terms that were vaguely written or had terms in that were not explained elsewhere were made clear thanks to the interview with Pascal, the prioritization of the requirements could start based upon the wishes from CERN and Software for Science, the amount of time estimated to complete a requirement and feature. This was done with the help of the AHP technique. The results were prioritized requirements, ordered by feature.

3.2 Consequences

These are the requirements that have chosen to be added to the prototype.

Role	User Story	Priority	Time
Run coordinator, SRC, STC and Shifter	As run coordinator, SRC, STC or as Shifter, I may need to attach files to log entries. These files may contain text or binary information (PNGs, JPGs etc...) (Roberto Divia).	8	8
Shifter	As a shifter, I want to have templates that prefill most of my end-of-shift reports from the available metadata so that I dont need to fill in myself what the system already knows (Vasco Chibante Barroso).	8	13
Subsystem expert	As a subsystem expert, I want to attach quality flags to runs so that physicists can use them while searching for good data sets for their analysis (Vasco Chibante Barroso).	8	8
User	As a user, I want to search log entries by different criteria (e.g. title, content, author, creation date,(...) and have the results listed.	9	8
User	As a user, I want to list all runs that match a given criteria to create my own run set.	8	8
Subsystem Run Coordinator	As a subsystem responsible, I want to be notified by email (or other channels) of log entries which are related with my subsystem so that I can better follow-up activities without having to constantly visit the product, e.g. EOS report (Robert Munzer) (Vasco Chibante Barroso).	8	13
SRC	As subsystem run coordinator I may request to receive automatic emails concerning all Logbook entries that include the System I am working for (either without distinction or using special selection criterias). The e-mail delivery address will probably be an e-group (single e-mail address ¡...¿@cern.ch) (Roberto Divia).	8	13
Run coordinator, SRC and Administrator	As run coordinator, S.R.C. or Admin, I must be able to move collaborators to and out of subsystem teams. These action may be conflict the information stored in SAMS (Roberto Divia).	9	13

There is an additional requirement, however, this requirement is dependent on access of the SAMS API used by CERN. This has not been given at the moment of writing, therefore, this requirement will be added to the list of the other requirements once access to SAMS is available.

Role	User Story	Piority	Time
Run coordinator, S.R.C. and Administrator	As run coordinator, S.R.C. or Admin, I must be able to move collaborators to and out of subsystem teams. These action may be conflict the information stored in SAMS (Roberto Divia).	9	13

With the help from the prioritized requirements, some requirements were chosen(with the front end) to be featured in the prototype.

3.3 Prototype

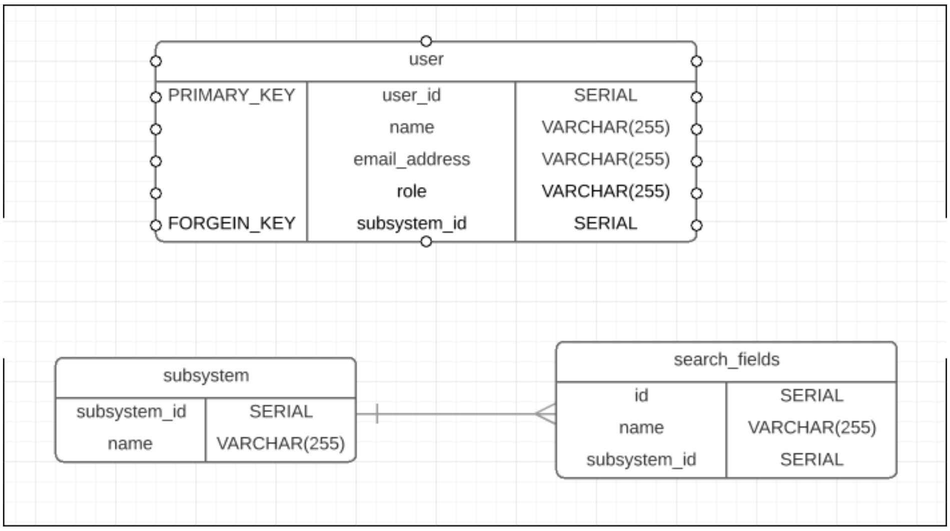

Figure 1: The Entity Relationship Diagram of the user schema

The user schema was build with simplicity in mind. The user authentication process is custom made. For the real product, CERN's own authentication process called OAuth will be used. Since this is temporary, it was decided to make this as simple as it was possible, since this is something that will disappear once the OAuth is working.

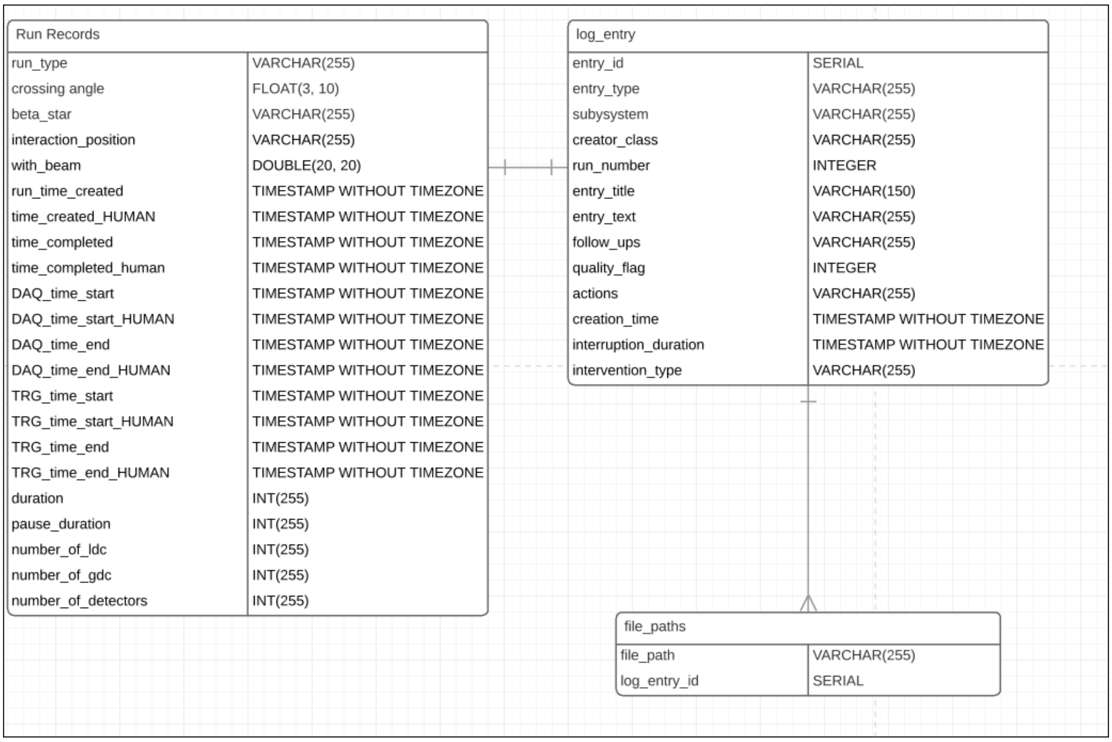

Figure 2: The Entity Relationship Diagram of the log entry schema

The goal of designing the log entry schema was to keep it as simple as possible. One of the first iterations was a table for each kind of log entry(End of Shift report, On Call Intervention etc.), however, after receiving the feedback that the database design could combine the different types of log entries into one table, this iteration was chosen. With this design, it will be much easier to extend the log entry table where needed. The file paths are separated from the log entries, since with this design, it will be easier to add multiple files to one log entry. The run records can be easily attached to one log entry. These records are not with every log entry type, so separating this makes the log entry table much easier to read and expand upon.

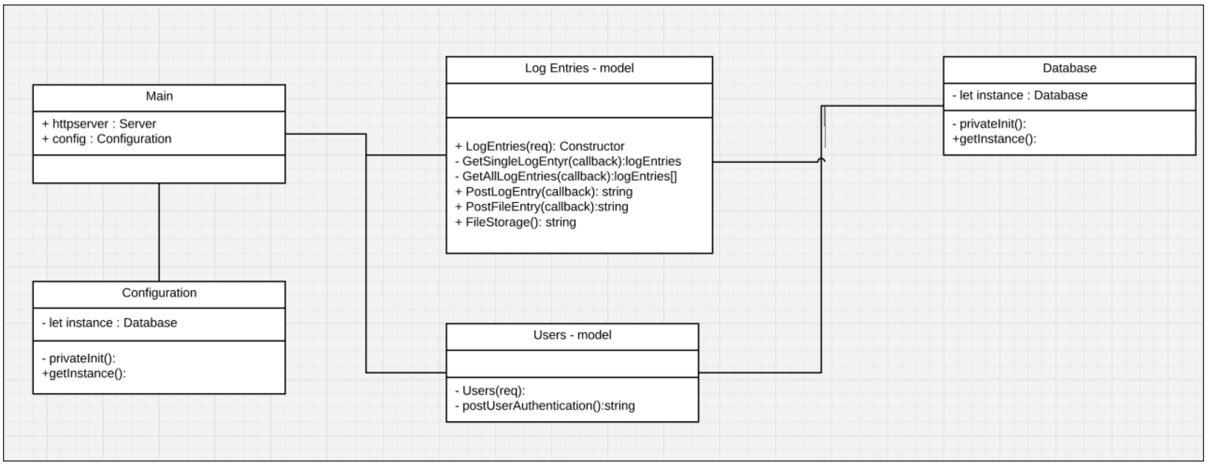

Figure 3: The class diagram of the prototype

The class diagram contains the two chosen design patterns and the requirements that were chosen to be implemented into the prototype.
With the two ERD's and the Class diagram, the prototype could be created.

3.4 Test cases

4 Conclusion

The conclusions of the report will go into this section of the thesis. This will be completed once there is a definitive prototype that will be shown at CERN.

5 Recommendations

Recommendations for future use of the electronic bookkeeping system.

6 Glossary

This section of the report will explain terms that will be used during the report.

1. Http-request = Hypertext Transfer Protocol is an application protocol used to send and retrieve information over the internet. An request is whenever the client asks to the website to do an task.

2. Client = The computer that the users uses to e.g. look at websites.

3. Server = A computer or device that provides functionality to other clients

4. JsonWebToken = JSON Web Token is an open standard (RFC 7519) that defines a compact and self-contained way for securely transmitting information between parties as a JSON object.

7 Bibliography

(1): Fast, unopinionated, minimalist web framework for Node.js. (n.d.). Retrieved May 16, 2018, from https://expressjs.com
(2): The World's Most Advanced open source relational database. (n.d.). Retrieved May 16, 2018, from https://www.postgresql.org/
(3): MariaDB.org. (n.d.). Retrieved May 16, 2018, from https://mariadb.org/
(4): The fun, simple, flexible JavaScript test framework. (n.d.). Retrieved May 16, 2018, from https://mochajs.org/
(5): Sublime text 3. (n.d.). Retrieved May 16, 2018, from https://www.sublimetext.com/3
(6): Schwaber, K. (1997). SCRUM Development Process. Business Object Design and Implementation, 117-134. doi:10.1007/978-1-4471-0947-1 11
(7): Git. (n.d.). Retrieved May 16, 2018, from https://git-scm.com/
(8): Buede, Dennis M.; Miller, William D. (2016). The Engineering Design of Systems. Wiley.
(9): Achimugu, P., Selamat, A., Ibrahim, R., and Mahrin, M. N. (2014). A systematic literature review of software requirements prioritization research. Information and Software Technology, 56(6), 568-585. doi:10.1016/j.infsof.2014.02.001
(10): Karlsson, J., Wohlin, C., and Regnell, B. (1998). An evaluation of methods for prioritizing software requirements. Information and Software Technology, 39(14-15), 939-947. doi:10.1016/s0950-5849(97)00053-0
(11): Teitsma, M., Dr. (2018, March 29). Requirements Bookkeeping application ALICE [Pdf].
(12): Online Diagram Software and Visual Solution. (2018, May 01). Retrieved May 16, 2018, from https://www.lucidchart.com/

8 Attachments

8.1 Interview transscript

Transscript Interview

Frederick van der Meulen

April 21, 2018

1 Overview

To gain more in-dept information regarding some requirements, Pascal Buschoten, a former shifter, has been interviewed.

2 Interview questions + answers

The interview is split into three parts; common, front-end and back-end related questions. The common questions are about features or rquiremetns that aren't clear for the front-end and the back-end. The front-end questions are about the front-end and finally the back-end questions are about the back-end of the new logbooking system.
Pascal is answering all the questions. Sometimes during the interview, Heiko and Dr. Marten stepped in to expand the answer. They're answers and comments are given into the answers of the research questions.

2.1 Common questions

Do you have any objections if this interview is recorded? The record will only be used to write the full answers down as s source for the thesis? I do not have any objections that this interview is recorded.

Can you introduce yourself and tell what you have done for CERN? I'm Pascal Buschoten and, for my graduation project, I've spend one year working at CERN. I've worked at a prototype for data process algorithmes for the upgrade of OSquare. That was mainly aimed at the TPC, Time Projection Chamber, that is the biggest detector of CERN. The data that comes free from that detector contains a lot of noise, for example, particles that make signals that are not usefull. I've tried to remove this noise from the data. After that, I've stayed there for two years as a trainee and I've worked a lot with drivers for the PCI-Express Reroute cards, cards that read out the data from the cables that are connected to the detectors. That was a big task, because the drivers needed to be developed, maintained and a lot of new requirements we're added to the system. My second task was evaluating a database for the process-configuration, that is all the processes that are running on the OSquare farm. At the start-up of the processes, they need to gain parameters, and it happends often in one big burst at the start of a run. I've looked at a couple of back-end systems to look at which back-end fitted the most for this process. I've also written an interface for this system. And that's what I've done during my time at CERN.

This interview is divided into three parts: some common questions regarding concepts that are not clear to us(Naomi and Frederick red.) so that we gain insight into them. Is this clear for you? Yes.

During reading the requirements for the new logbooking system, the term

follow-ups are reguarly mentioned. Do you know what they mean with this? I've got a vague guess, but I'm not sure. Perhaps they mean that, when they have an on call intervention, then they write down what's next to do or what the other problems could be? I'm not sure.

One of the posibillities for the new system is the functionallity to add files to log entries. What kind of files are added? Are these text based files or special files related to ALICE? I think that these are text based files, such as logs, and maybe crash-dumps, but generally, if you want to give more information, you write it down into a bug report.

There are announcements in the current system. What are these announcements and who creates these announcements? Can you show these announcements(asked to Heiko van der Heijden)? I don't know exacly what this means. Is this into the current system? (Heiko shows the announcements). I have no idea what this means.

What are on call interventions? Who can make these and how many times do they appear I've done on calls when I was in the DAC(Data Aquisition Chain) team, and everone in the team is responisble or maintains a piece of software in the DAC. On every moment, someone need to be accesible when problems arise that the shifters cannot solve in the control room. So the team has divided that every member must take a sum of days in a year to be on call duty. When your are on call duty, you have a CERN phone with you and if there's a problem, you will be called. Once your called, you must run to a terminal and log in remotely and look if you can solve the problem. Usually, you look at logs, and if you can't solve this, then you call the specialist on that area. For example, if there's something wrong with the logbook, you call Vasco, the specialist on that area. There are levels of perparedness. So when there is an important period of runs, then you must be reachable in a very small amount of time and being able to perform an intervention. (Frederick): To sum it up, On call interventions are interventions that cannot be created by shifters, and they only happen when there are problems with the system, then you need to solve the problems, and if you cannot solve them, then you need to call a specialist. Did I understood it correctly? (Pascal): yes, a shift leader can say, for instance that you need to call that person. (Naomi): does a on call intervention get added to a log? (Pascal): Every 24 hours, an on caller must make a log entry, and there you give a summary on what you have done. But i'ts not attached to one specific run. It's usually on multiple runs. (Dr. Marten): Every run can take 2 minutes, 2 hours or 0 seconds. (Pascal): That happends sometime (Dr. Marten): Why does that happen actually? (Pascal): Wrong configuration usually. (Dr. Marten): But how is it possible that that can start? (Pascal): It's not litteraly zero seconds, but it is possible that something can go wrong in a short amount of time. It doesn't happen that often though.

The requirements refer to a template for making an EOS report. How does a template look like and what is a EOS report? I've been told to look at the previous report and that's your template. I don't know if there is a formal template. More like this is the structure of the report you need to use. (Heiko): It is a wish from Roberto(one of the product owners at CERN red.), becaouse everyone has their own way of writing down things. (Pascal): I can imagine that. (Dr. Marten): It is also Robert's wish. Once every while they look at the logs to look why their machine is not working well. When everyone is just writing down stuff and use different terms for the same thing, then it's very difficult to check the reports. You can't do search and find due to the different terms. So that's why CERN says that it maybe be usefull to use a template. The other problem is that, if everything becomes a template, then its click click and there issues could arise aswell. So, we need to look at how we can create this.(Frederick): But there are not standard thing you need to fill in such as your name, to keep things simple or a description of your shift? (Pascal): Not in the way of a form, but you write down the time and a summary on what you did and how you fixed it?

One of the wishes for the new system is a forum like enviroment. This forum idea consists of leaving comments after entries. How do you envision this? I think that you need to keep it as a simple forum. You can place a comment under every log. You don't need threads and other difficult stuff as if people want to have a discussion, they can do that either in e-mail, Slack or personal. (Frederick): So the forum is just to say something small? (Pascal): Yes

Are logs only created on location at CERN or can this be made at a different location? These logs can be made everywhere. When you do an on call, you do not sit ever time at the headquarters of CERN. It 24 hours, so there is not a formal time you need to make the report. Not anymore though. In the past, there was not a formal time, but now there is every day a meeting with Roberto, so the report needs to be made before the meeting.

2.2 Front-end related questions

To gain some insight into the front-end, I(Naomi red.) was wondering what you think about the current front-end of the logbooking system and what you liked and disliked about them What I liked about the system is that it works, there are a lot of search options, you can search very specifically on issues
What I didn't like about the system is that it's a bit old fashioned and a bit slow. (Frederick): What do you mean about old-fashioned? Does it feel old-fashioned, does it look old-fashioned? (Pascal): Yes, those two options that you mentioned and to be specific it doesn't work very well on smart phones. It would be fine if it was possible on a mobile phone.

What kind of colour scheme would you like to see on the new logbooking

system and what did you think about the current white system? The white looks a bit boring, and there could be some improvements. The people at CERN are toleralble with thing that don't look posh, but they prefer it. (Frederick): Can you give us some colour examples? For example no use of yellow? (Pascal): I would prefer it if you could add the colours of ALICE into the design. Just the highlights would be enough.

How does a big screen view look like and what does it mean? That's about the control room! I'm not sure if you saw some pictures about the control room? Everyone sits in a glass box and people sit in front of a couple of computer screens and at the top of those screens there are big televisions that shows the important information. (Heiko shows a Googled pictured of the control room)(Naomi): What kind of information is shown on those big screens? (Pascal): For example, the status of the detectors, it depends on the piorities. Some graphs. (Heiko): The information is more like Teletekst. (Naomi): And everyone sees that? (Pascal): yes. (Frederick): To come back to the graphs, what kind of information do they show? (Pascal): I don't know but we need to talk specifically about every graph.

2.3 Back-end related questions

Can you tells us more about the OAuthentication used to log into CERN's products and how it exacly works? Are you talking on a technical level or al global level? (Frederick): On a technical level and on a global level? (Pascal): From the users side you can log into CERN's products, thats very nice. On a technical level, I don't know that much. (Frederick): So it's more like a token you receive to use CERN's products? (Pascal): yes

Can you tell us more about SAMS? My experience with SAMS as a on caller and as a shifter is that I use it to book my shifts. You need to do this early since the other shifters want to book their shifts depending on their vacation and the free days they have. At sams you can look at a kind of agenda and with that you can book your shifts. Sams can also be used to search for phonenumbers from shifters to reach somebody. Sams is also used for a big screen view, that sits into the coner of the control room. There, all the cellphone numbers are shown for the on callers.

8.2 Tables with pioritized requirements

8.3 Entries

Role	User Story	Piority	Time
Administrator	Only administrator may be given the possibility to remove log entries (and I am not even sure about this) (Roberto Divia).	1	8

8.3.1 Forum

Role	User Story	Piority	Time
User	As a user, I want to reply to existing log messages so that a conversation stays in a well-defined thread	4	20
User	People can create issues (Pierre vanden Vyvre)	1	13

8.3.2 View

Role	User Story	Piority	Time
User	As a user, I want to list log entries in a summary view so that I can get an overview of what happened in a given period.	8	8
User	As a user, I want to list log entries in a detailed view so that I can read them one after the other.	8	2
User	As a user, I want to browse through all the available metadata associated with a given run to understand on which conditions the run was made.	4	5
Shifter	. As a shifter I want to view log entries.	9	5
Shifter	As a shifter I want to view on call interventions.	4	5
Run cordinator	As run coordinator, I want to specify acquisition targets for certain time periods and check how far we are in achieving them so that I can keep track of progress (Vasco Chibante Barroso).	5	8
SRC	As System Run Coordinator I need ways to interrogate all the runs where the System I am responsible for participated and to get access to individual run entries and to summary statistics (Roberto Divia).	2	13

8.3.3 Search

Role	User Story	Piority	Time
User	As a user, I want to search log entries by different criteria (e.g. title, content, author, creation date,(...) and have the results listed.	9	8
User	As a user, I want to list all runs that match a given criteria to create my own run set.	8	8
User	As ALICE collaborator I need to check the details of any run: EOR reason, statistics, log entries (Roberto Divia).	6	5
Run cordinator	As run cordinator I may have to cross-reference log entries (e.g. by URL, by unique Reference ID, or by run number) (Roberto Divia).	4	8
Run cordinator, Shifter, SRC and STC	As run cordinator, Shifter, SRC and STC I may need to cross reference log entries or other logbook fields (e.g. run numbers, fill numbers etc...) with whatever issue tracking system will be used by the ALICE collaboration (today: Jira). This association may also be done automatically by daemons(e.g. what is done today for EOR reasons and Jira tickets) (Roberto Divia).	2	8
Subsystem expert	As a subsystem expert, I want to store custom fields that are only relevant to my subsystem so that I can correlate them with the rest of the metadata repository (e.g. fetch all runs with configuration X where this happened to my detector) (Vasco Chibante Barroso).	1	4
ECS / DAQ and SRC	As ECS/DAQ System Run Coordinator I need a way to access information of runs matching a selection criteria I specify (timestamps, run numbers, run types, included detectors etc...). Navigation between runs must be easy and quick. The target is to check the global runs (production and tests) for quality and errors (Roberto Divia).	2	8
SRC and System Team Member	As subsystem run cordinator I may have to cross-reference log entries (e.g. by URL, by unique Reference ID, or by run number) (Roberto Divia).	4	5
CERN Administrator Officer	As CERN administration officer I need to check all the on-call intervention records issued by CERN personnel (use case to be cross-checked with EP-AID-DA management) (Roberto Divia).	1	13

Developer	As a developer, I want to programmatically fetch log entries that match a given criteria so that I can build custom logic or applications based on existing data (Vasco Chibante Barroso).	1	8

8.3.4 Creation

Role	User Story	Piority	Time
User	As a user, I want to have a smart editor to create my log entries (WYSIWYG or Markup) and be able to use smart text so that messages look nice (e.g. links, code, . ..)	4	8
User	As a user, I want to be able to save search criteria for later use so that I dont lose time defining them at each visit.	1	8
Shifter	A shifter makes an entry into the database consisting of several items. Each entry records the following items: time of creation,which class the creator originates: human, type of entry, general, EOS, DCS, number of run, author of the entry, title of the entry, log entry, follow ups, files and actions	9	8
Shifter	As a shifter I want to be able to create log entries.	9	8
Shifter	As shifter I have to create log entries concerning any system (alone or in combination) (Roberto Divia).	8	5
Run cordinator	As run cordinator I have to create Logbook entries that cover almost all the Systems (e.g. global announcements or minutes) (Roberto Divia).	1	8
Run cordinator, SRC and STC	As run cordinator, subsystem run cordinator, system team member I have to create log entries concerning any system (alone or in combination) (Roberto Divia).	4	8

On Call Expert	A person who is called for a specific intervention makes an entry into the log system consisting of the following items; time of creation, author, type of intervention;remote, onsite, title of entry, log entry	2	8
Gas Technician	As a gas technician I want to create log entries when I delivered gas and other substances at Point 2.	4	3
Observer	As an observer I want to be able to look at the bookkeeping without the chance of adding or manipulating data.	1	5

8.3.5 Files

Role	User Story	Piority	Time
User	As a user, I want to attach files to log entries so that I can add additional non-textual information	9	8
Run cordinator, SRC, STC and Shifter	As run cordinator, SRC, STC or as Shifter, I may need to attach files to log entries. These files may contain text or binary information (PNGs, JPGs etc...) (Roberto Divia).	8	8

8.3.6 Flags

Role	User Story	Piority	Time
Run cordinator and Administrator	As run cordinator or administrator, I need to be able to update the logbook information for what concerns subsystems, in particular the run quality flag and the EOR reason(s). The question arises if subsystem run cordinators can update information associated to other systems (e.g. EOR reasons) as it is the case today (Roberto Divia).	2	8
Subsystem expert	As a subsystem expert, I want to attach quality flags to runs so that physicists can use them while searching for good data sets for their analysis (Vasco Chibante Barroso).	8	8

8.3.7 Data Extraction

Role	User Story	Piority	Time
Detector Expert	As a detector expert I would like be able to extract run/fill information in a format, which allows easier analysis than txt files, e.g. root-files to be able to do specific statistical analysis (Robert Munzer).	1	13
SRC	As subsystem run cordinator I need to be able to update the logbook information for what concerns my system and other systems, in particular the run quality flag and the EOR reason(s). The question arises if subsystem run cordinators can update information associated to other systems (e.g. EOR reasons) as it is the case today (Roberto Divia).	1	8

8.4 Reports

Role	User Story	Piority	Time
ALICE collaborator	As ALICE collaborator I have to create statistics reports such as number of runs, quantity of data, number of events, summaries by trigger classes etc... These reports will use selection criterias I will specify such as time spans, active systems (e.g. only the runs including my particular system), run type etc...	1	20
Shifter	As a shifter, I want to have templates that prefill most of my end-of-shift reports from the available metadata so that I dont need to fill inmyself what the system already knows (Vasco Chibante Barroso).	2	13
Shifter	As a Shifter, I would like to have templates that automatically compile and format the data available in the system in order to write my end of shift report in a fast and uniform way. Currently in the ALICE logbook, I dont like that I need to compile all the information myself and that not all shifters use the same structure (Vasco Chibante Barroso).	2	13
Run cordinator	As run cordinator, I want shifters to use templates so that it is easier and faster to read them (Vasco Chibante Barroso).	1	13
Subsystem Run Cordinator	As a SRC I would like to be able to create my own detector specific templates for example OnCall interventions. In this case I can specify the relevant information which are required from the OnCall shifter for different kind of standard events (Robert Munzer).	8	20

8.5 Email

Role	User Story	Piority	Time
ALICE member	As an ALICE member, I would like to receive via email a global summary of each LHC Fill in order to follow ALICE operations without visiting the bookkeeping tools. Currently in the ALICE logbook, I like that I receive via email a document with info on efficiency and EOR Reasons and that on the body of the email there is a summary for each fill (Vasco Chibante Barroso).	8	20
Run cordinator	As run cordinator I may request to receive automatic e-mails concerning all Logbook entries that include all systems (either without distinction or using special selection criterias). The e-mail delivery address will probably be an e-group (single e-mail address ¡...¿@cern.ch)(Roberto Divia).	6	13
Subsystem Run Cordinator	As a subsystem responsible, I want to be notified by email (or other channels) of log entries which are related with my subsystem so that I can better follow-up activities without having to constantly visit the product, e.g. EOS report (Robert Munzer) (Vasco Chibante Barroso).	8	13
SRC	As subsystem run cordinator I may request to receive automatic emails concerning all Logbook entries that include the System I am working for (either without distinction or using special selection criterias). The e-mail delivery address will probably be an e-group (single e-mail address ¡...¿@cern.ch) (Roberto Divia).	8	13
SRC	. The subsystem cordinator wants to be reported when something is going on with his system. He should not have to take action for himself to find out things (Robert Helmut Munzer).	6	13

8.6 Roles and Authentication

Role	User Story	Piority	Time
User	As a user, I want to be able to login with my CERN credentials to avoid having to remember a new set of credentials. This should be done by using the CERN authentication method.	9	13
Run cordinator, SRC and Administrator	As run cordinator, SRC or Admin, I must be able to move collaborators to and out of subsystem teams. These action may be conflict the information stored in SAMS (Roberto Divia).	9	13
Run cordinator, SRC, Administrator	As run cordinator, SRC or Administrator, access to Logbook actions restricted to my role should be granted without external interventions and for the time span of my duties (e.g. for shifters the shifts before and after mine, plus my own shift) (Roberto Divia).	4	8
Run cordinator and SRC	As run cordinator or SRC I need to give ALICE collaborators write or read-only access to the logbook. These rights will be superseeded by equivalent rights given according to the function of the user (e.g. a ALICE collaborator with read-only access will be given write access during the time of his/her duties as a shifter, subsystem run cordinator or system team member) (Roberto Divia).	6	13

8.7 View with Dash boards

Role	User Story	Piority	Time
User	As a user, I want to see in a dashboard the metadata associated with an LHC Fill so that I can have a global image of what happened during that LHC Fill.	2	13
User	As a user, I want to be able to customize dashboards so that I only see the fields relevant to me.	1	13
User	As ALICE collaborator I may have to open multiple GUIs with independent selection criterias (e.g. one browser window for day-to-day work and a second browser window for statistics) (Roberto Divia).	1	20
ALICE collaborator	As ALICE collaborator I need to be able to access the Logbook on a run-per-run summary view (possibly using a selection criteria I specify) and on a log entry by log entry view (possibly using a selection criteria I specify) (Roberto Divia).	1	13
Shifter	As a shifter I want to view data about calibration of the detector.	4	5
Shifter	As a shifter I want to be able to have a big screen view.	2	3
Shifter	As a shifter I want to view data about the fill.	1	5
Physics Community	The Physics Board has several needs or questions: 1. To make the planning possible an overview of storage and processing power (CPU) is needed. 2. The use of resources per user to run jobs could be more detailed. 3. How much PB is available on disk for storage. 4. For MC-storage a fine grained but lacks an overview. 5. When I want to clean up, where do I have to look? 6. MC production requests. 7. Usage statistics (which data is popular?). 8. Sort out why a train takes a specific time to process.	1	13
Physics Community	Most data is replicated because a lot of people use the data. There are two views from the Physics Board: clean up, to know what could be cleaned up planning, when can this MC be run?	1	13

Administrator	As an administrator, I want to have a dashboard that gives me log- entry related analytics so that I follow the evolution of the repository (Vasco Chibante Barroso).	1	13

8.8 Statistics

Role	User Story	Piority	Time
Shifter	As a shifter I want to view some statistics of runs and other stuff.	1	8
Run cordinator	As run cordinator I need to gather statistics on the runs selected by using custom rules (timestamps, run numbers, run types, included detectors etc...). These statistics will include EOR reasons, per-detector and per-system summaries, error recovery (PARs) rates etc.(Roberto Divia)	1	13
Physics Community	Each week global and specific statistics about the system are needed CPU usage data storage etc.	1	13

8.9 Run

Role	User Story	Piority	Time
Run cordinator	As run cordinator, I want to attach tags to runs so that I can then use them while searching (Vasco Chibante Barroso).	4	8
Run cordinator	As run cordinator, I want to edit certain specialized fields associated to a run (e.g. EOR Reason) so that I correct wrong information inserted by the O 2 software (Vasco Chibante Barroso).	1	5
Developer	As a developer, I want to programmatically fetch runs that match a given criteria so that I can build custom logic or applications based on existing data (Vasco Chibante Barroso).	1	8

8.10 Announcements

Role	User Story	Piority	Time
Shifter	As a shifter I want to view announcements.	4	5
Adminis- trator	System administrators can create an announce- ment. This announce- ment consists of the following items: time of creation, validity, duration of interruption of the system, author, title of the entry, log entry	2	8

8.11 Other

Role	User Story	Piority	Time
Managet	As a manager I want to know whether all the relevant people are involved with respect to an issue (Pierre vanden Vyvre).	1	13
Adminis- trator	As administrator I may request to replicate either selected portions or all of the Logbook data to external sites and to provide adequate access tools to it (to facilitate read-only accesses) (Roberto Divia).	1	20
Adminis- trator	As an administrator I must be able to configure the system.	1	13