


 <p>Sponsored through Framework Programme Sixth (Call 5) by</p>  		<b>Document Information</b>
		<b>Doc Version:</b> v1 <b>Date :</b> 01/12/10 <b>Pages :</b> 14
		<b>Person Responsible:</b> Ruediger Glott
		<b>Author(s):</b> Ruediger Glott (MERIT) Kirsten Haaland (MERIT) Rishab Ghosh (MERIT) Jean-Christophe Deprez (CETIC)
		<b>Reviewer(s):</b> Daniel Izquierdo Cortázar,
The QUALOSS Consortium consists of: CETIC (BE), Facultés Notre Dame de la Paix à Namur (BE), Universidad Rey Juan Carlos (ES), Fraunhofer IESE (DE), ZEA Partners (BE), MERIT (NL), AdaCore (FR), PEPITe (BE)		<b>To:</b> E.C. for review
<b>Status:</b> <input type="checkbox"/> Draft <input type="checkbox"/> To be reviewed <input type="checkbox"/> Proposal <input checked="" type="checkbox"/> Final/Released	<b>Confidentiality:</b> <input checked="" type="checkbox"/> Public - Intended for public use <input type="checkbox"/> Restricted - Intended for Evaluators only <input type="checkbox"/> Confidential - Intended for individual r only	
<b>Deliverable ID: D3.3</b>  <b>Title :</b> <b>Validation of Data and Measurements of Reference advanced F/OSS projects</b>  <b>Return from F/OSS Communities on the Standard QualOSS Assessment Method</b>		
Disclaimer: The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.		

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 2 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public


**Title: Validation of Data and Measurements of Advanced F/OSS Projects**

**Abstract:**

This document describes the validation of the QualOSS Standard Assessment from the FIOSS community point of view. For this purpose, in-depth interviews with representatives of two advanced FIOSS communities, CVSAAnaly and Evince, have been carried out.


Priority is thereby set on following aspects:

- appropriateness and understandability of the assessment results
- evaluation of data sources
- responsiveness of community
- community history and evolution in general
- evolution of the quality items examined in the QualOSS standard assessment

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 3 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

## TABLE OF CONTENTS

1. <a href="#">General overview of WP3</a> .....	<a href="#">5</a>
2. <a href="#">Assessment Objectives</a> .....	<a href="#">6</a>
2.1. <a href="#">Objectives</a> .....	<a href="#">6</a>
2.2. <a href="#">Approach</a> .....	<a href="#">6</a>
3. <a href="#">Structure Of The Interview</a> .....	<a href="#">8</a>
4. <a href="#">Evaluation Criteria and Purposes</a> .....	<a href="#">10</a>
5. <a href="#">Evaluation of the QualOSS Assessment by Community Members</a> .....	<a href="#">11</a>
5.1. <a href="#">Introduction</a> .....	<a href="#">11</a>
5.2. <a href="#">Understandability and acceptance</a> .....	<a href="#">11</a>
5.3. <a href="#">Dissent and its causes</a> .....	<a href="#">12</a>
5.4. <a href="#">Ideas for improvement</a> .....	<a href="#">12</a>
6. <a href="#">Conclusions</a> .....	<a href="#">13</a>

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 4 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

## 1. GENERAL OVERVIEW OF WP3

Work package 3 aims at identifying the FIOSS projects that will be measured (Task 3.1) and deploys the necessary instrumentations to execute the measurements (Task 3.2). Task 3.3, which is described in detail in this deliverable, consists of the validation of data and measurements that were collected and deployed in task 3.2.


The overall goal of Task 3.3. is to validate the data collected in the QUALOSS repository and to verify that the results of the QualOSS standard assessment are understandable and explainable from a community point of view. This also helps to check if the QualOSS assessment is capable to induce a software quality improvement in response to this assessment. Priority is thereby set on following aspects:

- appropriateness and understandability of the assessment results
- evaluation of data sources (did we use the correct data?)
- responsiveness of community (is there community interest in the QualOSS measurement?)
- community history and evolution in general (are there important breaks / milestones in the development of the community - e.g. a change of the core committers or a "take over" by a company - that must be considered with regard to the QualOSS standard assessment?)
- evolution of the quality items examined in the QualOSS standard assessment (have there been important changes with regard to these quality characteristics, such as a change in the release management?)

For this purpose, core community members (core committers) have been contacted and interviewed in order to review and validate measurements calculated by our measurement tools. This gives developers a chance to explain measurements that could sometimes be misconstrued.

In detail, the validation process consists of following steps:

- design interviews for core community members
- exercise interviews with core community members
- validate the results of the QualOSS standard measurement

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 5 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

## 2. ASSESSMENT OBJECTIVES

### 2.1. OBJECTIVES

While the validation of the QualOSS standard measurement from the business perspective (WP5) has to consider the measurement from the various viewpoints of different actors and different product and process aspects (usability, profitability), the validation of the QualOSS standard assessment from the community perspective focusses only on two aspects:

Do important community members understand the outcome of the QualOSS standard measurement?


In case of measurement results that are not understandable to important community members, can these community members explain why the QualOSS standard measurement may have turned out inappropriate results and how the measurement can be improved through alternative measurements?

To exemplify the latter point, one may imagine a FIOSS endeavour that has been assessed through the QualOSS standard assessment and has resulted in a negative assessment of, say, maintainability because it was observed that the number of bug reports and the number of bug fixes are decreasing. However, it might happen that the community uses other repositories and bug trackers than those that are publicly accessible. In that case it might be necessary to re-assess the maintainability measures with the correct data source and to document this in the measurement report.

### 2.2. APPROACH


The evaluation was executed by interviews in two stages: in the first stage, core committers were contacted by email and informed about the QualOSS standard measurement of their software and the outcome of the measurement. The main purpose of this initial contact was to check if there is interest of the core committers (i.e. the key actors of the FIOSS endeavour) in the measurement and willingness to respond to it, and to examine if the community understands the measurement results and how it interprets them. In total, 14 core committers of 3 FIOSS projects have been contacted: CVSSAnaly (2 committers), Eclipse (5 committers) and Evolution (7 committers). In case of interest and responsiveness, a core committer should be asked to be interviewed again by another email inquiry (stage 2) in order to talk about details of the measurement, as, for instance, the appropriateness of thresholds that were chosen by the QualOSS team in order to differentiate quality levels.

However, it turned out that FIOSS communities showed no interest in quality issues in the sense that they would respond to the outcomes of an exercise like the QualOSS Standard

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 6 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

Assessment. Therefore we had to change our approach and contact core community members that were known by the QualOSS consortium. This methodological constraint required to do in-depth interviews with the two core committers of Evince and CVSAnalY, Carlos Garcia Campos and Santiago Dueñas, instead of a broad survey of interviews capturing a larger number of projects.

Feedback from the two interviewees and previous experience with FLOSS developer communities suggest that the reason for the low community response is not a lack of awareness of quality issues within these communities but a strong focus on internal community matters. Requests for feedback from people outside the community are often rejected, probably due to the circumstance that many communities receive a lot of similar research requests. Another reason for the low response might be the fact that the community members are not aware of quality issues of their product since they have limited time to work on it, and most of the time they do not care about external studies, which do not really provide a big improvement to the source code or the community. Furthermore, quality issues might be regarded by the community members as something that should be discussed internally before external parties are allowed to receive community feedback.


	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 7 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

### 3. STRUCTURE OF THE INTERVIEW

The first interview, which was carried out by phone, included information about the results of the QualOSS Standard Assessment. It explained to the core committers the character of the QualOSS Standard Assessment and the meaning of the colours that are assigned to the different quality characteristics. In other words, it explained that the colours state the risk of the interviewees' software from a business point of view when using the data that is accessible, and that the overall purpose of QualOSS is to help companies in decision-making regarding the utilisation of a FLOSS endeavour.

This information included an explanation how these results have been generated and an overview of the background and purpose of the study. Questions that were asked in the first interview are:


- How do you interpret the results of the QualOSS assessment? Do you think it shows that the FLOSS endeavour features rather high or rather low quality?
- Did you expect that a measurement of factors that influence the quality of your FLOSS endeavour would have turned out this result?
- Where do you agree and where do you disagree with the results? Why?
- What colours would you assign to the different quality characteristics if you had to evaluate your FLOSS endeavour (using the same colour scheme as the QualOSS assessment).
- For each topic mentioned below, can you answer the following questions:
  - What colour would you assign to it (QualOSS assessment colour scheme)?
  - Do you have the feeling that the quality in this regard has been improving, regressing, or remaining stable over the last year?
- Code/product maintainability
- Code/product reliability
- Code/product security
- Documentation availability (i.e., is there a broad range of documentation of various aspects of the software?)
- Documentation content quality
- Test availability (unit and system tests)
- Test coverage (functional or code)
- Test repeatability (ease of performing testing in subsequent version of the software component)
- Community size and regeneration (are people who leave replaced by newcomers?)
- Community interactivity and liveliness

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 8 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

- Community composition - is it the right mix of
  - people with different expertise fields,
  - professionals and hobbyists
  - experts and laymen
- Requirement and Change management
- Release management
- Test management
- Community management
- Support management
- Compatibility with dependencies (license related but also simply in terms of compatibility between people in the two communities as well as compatibility between their code and also compatibility in release cycle, cooperation on bug reporting).

Besides these QualOSS related topics, the interview had also to cover aspects that provided the research team with some background information:

- Personal features of interviewee
  - since when are you member of this community?
  - since when do you hold committer rights?
  - are you a voluntary member of the community or are you contributing as an employee of a firm / as an owner of a firm (for firm's interests)
- Community evolution
  - since when does the community exist?
  - how did the number and composition of members develop since the community was established?
  - who plays the most important role for the development of the community (and the software)?
  - how many companies are involved in the community? - is it rather a volunteers or a business community?
  - what were the most important changes in the lifetime of the community, and when did these happen?
  - how did work products, software processes, and tools and FLOSS endeavour dependencies change during the lifetime of the community?

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 9 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

#### 4. EVALUATION CRITERIA AND PURPOSES

The evaluation provides a critical assessment, from a community point of view, of whether the QualOSS assessment has achieved its objectives. The main criterion in this regard is whether the community understands and interprets the results of the assessment largely in the same way as the QualOSS team would understand and interpret them. If there are strong discrepancies between the notions of the community and of the QualOSS team, or if the QualOSS assessment is refused by the community, the QualOSS measurement must be critically reviewed. Otherwise the QualOSS assessment would endorse the risks that

- a) the assessment is too rigid
- b) the assessment induces within the community rather resistance than willingness to improve quality


In both cases, the QualOSS assessment would have missed two important objectives.

However, for obvious reasons dissent between the QualOSS assessment and the assessment the community would exhibit alone can not suffice to draw fundamental conclusions on whether the QualOSS assessment has achieved or missed its goals. The fields in which dissent occurs, and the reasons community member give for their dissent, are also important to evaluate. For instance, there could be a systematic problem for the QualOSS assessment to capture the most appropriate data for a particular quality characteristic. Or a community could show reluctance towards such assessments in general although, from a neutral point of view, the assessment results in fact in a realistic depiction of the quality of the FLOSS endeavour. Thus, qualifying the response from the community is an important context factor to consider in order to evaluate whether the QualOSS assessment has achieved its objectives.

Finally, the quality of the QualOSS assessment must be evaluated with regard to whether or not it can be improved. If the community is able to provide feedback including ideas for improvements the QualOSS assessment should be critically reviewed. This review requires to check also the feasibility of these ideas. If there are no ideas for improvement or mainly ideas that are impossible to implement, the quality of the QualOSS assessment must be considered as high because it obviously cannot be improved. This might imply that it doesn't turn out results that would be considered as appropriate by the respective FLOSS community – however, it would, in this case, provide the best assessment one can get. In this case, it would be a task of future research and of FLOSS communities to improve the infrastructure that is necessary to run quality assessments like the QualOSS assessment.

Hence, the evaluation in the following section is structured as follows:

1. Understandability and acceptance
2. Dissent and its causes

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 10 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

### 3. Ideas for improvement

## 5. EVALUATION OF THE QUALOSS ASSESSMENT BY COMMUNITY MEMBERS

### 5.1. INTRODUCTION


One of the interviewees is an experienced member of the CVSAnalY and the Evince project. The other one is an experienced member of the CVSAnalY community. Both have committer rights and play a key role for these communities. Both FIOSS endeavours exist for a number of years and are well advanced with regard to community dynamics and community management. The commercial input to these FIOSS endeavours is limited but existent, as Evince has, for instance, been supported by Google. CVSAnalY is mainly used in academia. Both communities depend on a limited number of core committers, which are also largely responsible for all management tasks that must be performed. Both interviewees reported a stabilisation and professionalism of the communities over the past years.

### 5.2. UNDERSTANDABILITY AND ACCEPTANCE


Both interviewees confirmed that in principle they have understood the purpose of the QualOSS Standard Assessment and how it was conducted. However, they both expressed strong concerns regarding the composition of the indicators and, in particular, the number of measures that were covered by the assessment.

Given the restriction that detailed information about the measurement could only be retrieved from spreadsheets, both interviewees complained that the results are hard to understand when the details are considered. While it appeared traceable to them how the factors are composed, they had difficulties to understand how thresholds for the assignment of colours have been defined. Both demanded clear-cut information on this part of the methodology and largely refused to consider the results of the QualOSS Standard Assessment as acceptable without this additional information. They made clear, however, that this point was not due to distrust in the QualOSS methodology but to the fact that they were not able to fully understand how the measures were defined and generated.

The acceptance of such an approach in general was however disputed. While Santiago Dueñas fully supported the purpose and approach of QualOSS and saw especially possible benefits for FIOSS communities in terms of dissemination, Carlos Garcia Campos showed some reservation towards any kind of automated quality measurement and the business perspective. Both however strongly supported attempts that help to increase the quality of a FIOSS product and the quality awareness within FIOSS communities. In addition, both support attempts to raise interest of businesses into FIOSS. Mr. Garcia Campos' discontent was therefore not based on an anticommercial attitude but on the belief that quality issues must be looked at from different angles, and he considered the developer's angle to be more important than the business angle.

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 11 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

By and large, both interviewees emphasised that there is a strong sense of quality assurance within the communities, which they considered as a general feature of (and strong requirement from) and FIOSS community.

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 12 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

### 5.3. DISSENT AND ITS CAUSES

Regarding the acceptance of the results, the perception of the two interviewees was largely determined by the fact that they belong to the core producers of the software that was analysed by the QualOSS platform.

Mr. Dueñas largely agreed with the measurement results for CVSSAnaly but emphasised that some of these results are just due to the fact that some important aspects have been captured incorrectly. For instance, he maintained that bug reports and bug fixes in CVSSAnaly would have been better captured from the developers mailing list than from the bug tracker. He assumed that in future, when CVSSAnaly will use Bugzilla, the result would look more positive in this regard. He particularly disagreed with the QualOSS results regarding reliability of CVSSAnaly. To him, CVSSAnaly is extremely reliable and he could not understand how the QualOSS assessment could turn out such a negative result. Especially in this regard he demanded more explanation of the measurement's definitions.


Mr. Carlos Campos showed a slightly stronger refusal of the QualOSS assessment results than Mr. Dueñas, although Evince was evaluated much better than CVSSAnaly. He particularly disagreed with the "high risk" that was attributed to test availability and coverage. To Mr. Carlos Campos, this was just due to a measurement error, as the tests for Evince should be found within the GNOME repository.

An interesting point Mr. Dueñas and Mr. Garcia Campos explicitly made was that they, as community members, might have a different point of view when quality aspects are considered than a business representative might have. For instance, for them as core committers and community managers who know the code and the discussions within the community very well might not bother too much about a certain bug, or even a number of certain bugs, because they know these bugs are harmless or that it is not urgent to fix them. However, from a business perspective pending bugs might be a bad signal.


Finally, both interviewees pointed out that some of the measures examined through the QualOSS platform are heavily dependent on things the core developers and the community members cannot control. For instance, many quality issues could be fixed very effectively if more users would regularly give feedback to the community. In this regard, both interviewees expressed their hope that a tool like the QualOSS platform could add value, as it makes things visible to the developers that due to the lack of user feedback they did not see before.

### 5.4. IDEAS FOR IMPROVEMENT

Overall, the criticism of the two interviewees of the QualOSS measurement addressed not the measures directly but their understandability and traceability. They both confirmed that they cannot see any measure the QualOSS Standard Assessment might have missed. However, given the sheer number of measures they could not exclude that QualOSS might measure too much, though due to time restrictions and the lack of definitions they have perceived it was not possible for them to identify measures that possible could have

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 13 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

been ignored. At any rate, both interviewees made very clear that without a more clearcut description of how the measures and indicators are generated and composed they cannot consider the QualOSS platform as a reliable measurement tool.

	<b>Validation of Data and Measurements of Advanced F/OSS Projects</b>	Page : 14 of 14
		Doc Version: 3 Date: Jan 13, 10
		Status : Final Confid : Public

## 6. CONCLUSIONS

The interviews have shown that the QualOSS platform deems to be partly accepted and partly refused by the FIOSS community. This seems to some extent due to the limitations of information that can be provided to the user at a glance, but maybe also due to factors that are characteristic for FIOSS communities and can hardly be tackled by any (partly) automated measurement: The deep understanding and insight that core community members have into their software raises their requirements from a quality measurement tool to a level that can not be met, as it is impossible to reproduce this deep knowledge from outside the community. A factor that has surely not been taken into account is the time that is needed to do such a quality assessment. The strong expertise of core committers is built up over a long period of time, whereas a profitable quality assessment from the outside must perform its tasks and provide reliable results within a couple of days.

Thus, in order to tap the potential of the QualOSS platform for FIOSS communities, it would be worthwhile to communicate to the communities in a clear-cut way what the platform can achieve and where its limitations are. Both interviewees have clearly seen such possible benefits. However, the most important result of the interviews with regard to the purposes of QualOSS is that, apparently, the community members are aware of distinct quality perceptions of businesses and community members and that there is willingness to accept the business point of view - as long as the generation and composition of results is well documented.