Sponsored through Framework Programme Sixth (Call 5) by

Document Information

Version: 1.1
Date : Jan 13, 10
Pages : 18

Owning Partner:
AdaCore

Author(s):
José Ruiz (AdaCore)
Ruediger Glott (MERIT)

Reviewer(s):
José Ruiz (AdaCore)
Olivier Ramonat (AdaCore)
Jacques Flamand (CETIC)
Ruediger Glott (MERIT)

To:
European Commission

Purpose of distribution:
Final Submission

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)

Printed on 01/12/10 at 05:17:54 PM

Status:

[ ] Draft
[ ] To be reviewed
[ ] Proposal
[X] Final/Released

Confidentiality:

[X] Public - Intended for public use
[ ] Restricted - Intended for QUALOSS consortium only
[ ] Confidential - Intended for individual partner only

Deliverable ID: D5.3

Title:

Results of Case Studies

Deliverable: D5.3

Title: Results of Case Studies

Executive Summary:

One of the most important aspects in QualOSS is the validation of the QualOSS methodology and of the QualOSS assessment methods developed in the project, with a particular emphasis, in the initial case studies, on the validation of the standard QualOSS assessment method. To achieve such a validation, a set of case studies are devised to verify whether or not particular business goals are reached. A set of suitable pilot projects are identified to assess the applicability and utility of the QualOSS methodology and QualOSS methods.

People directly involved in the pilot projects will be interviewed to better understand the general context in which the QualOSS methodology and QualOSS assessment methods will be applied. These interviews will help to verify several hypotheses regarding user satisfaction and profitability. First, they will allow for comparing the results obtained from QualOSS assessments against human perception of the robustness and evolvability of the FlOSS endeavors assessed in each case study. Second, these interviews are also useful to study user satisfaction with the results obtained from the standard QualOSS assessment method and eventually of other more advanced QualOSS assessment methods. The focus in this document is set on the overall user satisfaction with the assessments.

TABLE OF CONTENTS

1. Introduction

1.1 Motivation

The strategic objective of the QualOSS project is to enhance the competitive position of the European software industry by providing methodologies and tools for improving their productivity and the quality of their software products. To achieve this objective, QualOSS notes that many organizations integrate Free *libre* Open Source Software (F/OSS) in their systems hence QualOSS goal is to facilitate the identification of the most robust and evolvable F/OSS development endeavors whose F/OSS components are worth integrating in industrial software products and systems. In the end, the QualOSS methodology and its QualOSS assessment methods will ease the selection of high quality open source components. The overall effect will be increased productivity and higher dependability for the industrial software products integrating F/OSS components.

To achieve this goal, QualOSS proposes to build a high level methodology to benchmark the quality of open source software. In particular, the QualOSS project delivers an assessment methodology for gauging the evolvability and robustness of open source software endeavors.

This fifth work package (WP5) verifies that the QualOSS methodology, its QualOSS assessment methods and the accompanying tools can be used to verify whether particular business goals set for the studied projects are reached or not. The first task of WP5 (T5.1) presents the broad context of the case studies and selects the pilot projects to be analyzed. Furthermore, T5.1 lists the hypotheses that will be checked by each case study. The second task (T5.2) of WP5 consists in a set of interviews and the application of QualOSS assessment methods on the selected FlOSS endeavors. In a first phase, T5.2 applies the standard QualOSS assessment method on the FlOSS endeavors. In a second phase, the adaptation of the standard QualOSS assessment method into more advanced methods and the application of these advanced methods will be studied. The results obtained from T5.2 are used in the final task (T5.3) to report the results of the case studies and to argue the validity of the hypotheses being tested.

1.2 Objectives

The goal of task 5.1 was to design the case studies and find the pilot projects on which these studies can be conducted. It included the identification of the hypotheses to test, the design of the general protocol to use in case studies, and the description of the pilot projects on which these studies are to be conducted.

Task 5.2 measured quality characteristics, indicators and metrics required to validate these hypotheses. People directly involved in the pilot projects are then interviewed, giving them access to the results of the QualOSS assessment.

Finally, Task 5.3 analyzes the human perception of the assessment and validates the set of hypotheses previously defined.

1.3 Approach

This document verifies the results obtained in the assessments related to the case studies defined in order to validate the set of hypotheses previously defined.

In order to determine whether the QualOSS platform and methodology are useful, the approach followed is based on a close interaction between the different actors (QualOSS experts, developer's community, and user's community). The information extracted is made available to all parties so that they can manipulate it and assess its validity and appropriateness based on their expertise and needs.

The results showed in this document are organized in tables containing the validation status of the different hypotheses for the different case studies, based on following scale:

1 = hypothesis falsified

2 = hypothesis largely refused
3 = hypothesis largely confirmed
4 = hypothesis verified

An explanation, when needed, is also added to the validation status to give the rationale for the score.

1.4 Strategy of WP5

T5.1 identifies the set of general hypotheses and a general protocol to follow in all our cases studies. The goal is that this general framework, made of hypotheses and general protocol, will steer our various hypotheses in a similar way so that their results can be aggregated at the end of WP5 to identify certain trends (e.g. all case studies were interested about code reliability and 80% of the users were satisfied with the results obtained from the standard QualOSS assessment method for code reliability). Such an aggregation from several case study results is only possible if the various studies were performed in a compatible way. By providing a set of general hypotheses and a general protocol, T5.1 proposes a first action to help conduct our case studies in a compatible fashion.

T5.2 uses the general hypotheses and the general protocol created during T5.1, and refines specific hypotheses and a specific protocol for each case study. It is important that T5.2 controls how the refinement takes place because this is how compatibility across case studies will be guaranteed. Once the refinement has produced the specific hypotheses and specific protocol, including the specific questionnaires for a particular case, then T5.2 also has the responsibility to collect the data for that case study.

T5.3 first analyzes the data for each case study individually and assesses whether the specific hypotheses verify or not. Then, it aggregates their results to determine the usefulness and validity of the QualOSS methodology.

1.5 Structure of the Deliverable

The rest of the deliverable is structured as follows: Section 2 provides the definition of terms used in the document. Sections 3, 4, 5, and 6 contain the evaluation of the different case studies, describing whether hypotheses were validated. Finally, section 7 contains the conclusions resulting from these case studies.

2. Terminology / Glossary

F/OSS Endeavor. F/OSS Endeavor is defined by 1) a set of work products, 2) the F/OSS community creating, updating and using these work products, 3) the tools used to act on these work products or to build or run the software product, and 4) the set of development processes executed by the community, these processes include rules and a division of labor accepted and followed by community members when interacting and creating work products.

3. AdaCore/GCC Version to Use as Back End for GNAT

3.1 Objectives

For AdaCore, quality metrics for the GCC back end (the one used by the GNAT compiler) related to its robustness are of paramount importance. AdaCore needs to upgrade to new versions of the GCC back end regularly (every one or two years) and the process to select the appropriate GCC version is critical for its business.

The QualOSS standard assessment has been used to analyze two different versions of the GCC back end (GCC 4.2 and 4.3), with the objective of help improving the decision process and increasing the confidence in the selected version.

The people interviewed were Olivier Hainque (AdaCore) whose role is that of project manager of the GCC integration (and active contributor to GCC), and Olivier Ramonat (AdaCore), whose role is quality assurance at AdaCore.

3.2 AdaCore/GCC Hypotheses

3.2.1 User Satisfaction

Regarding user satisfaction, the pre-assessment interviews confirmed largely that the QualOSS Standard Assessment was perceived to match very well the interviewees' requirements from a quality assessment.

Both interviewees agreed that the indicators chosen to measure quality related to the work product are relevant. Improvements for a more advanced measurement were suggested, in this regard, in order to capture license issues, the portability of code, and the functional impact of code changes made in a new version. To this end, it was suggested to include measures like the number of different licenses, license compatibility with existing products, the number of platforms supported, the time needed to support a new platform, and more qualitative measures that reflect the level of experience of community members that do changes on the code. For the latter point, it was also suggested to examine the community members' behavior in discussions. However, it was also clear to the interviewee who suggested this improvement that it is impossible to do this in an automated way and that it may consume a lot of time.

Regarding the quality measures related to community, software processes and dependencies, both interviewees said that the QualOSS standard measurement captures all relevant points they can think of, and both confirmed that these measures are very important.

After the assessment, the overall degree of satisfaction with the results of the measurement, based on a scale from 1 (not at all satisfied) to 5 (very satisfied) did not differ very much between the two interviewees. Both allocated mostly values between 3 and 4 to the measurement results. The overall quality of the QualOSS Standard Assessment was considered to be high by both interviewees.

Discontent was mainly caused by difficulties to understand the details of the measurement. For instance, Olivier Ramonat complained about the limited level of information the spreadsheet with the results provided on how the measurement of some reliability measures was conducted. With the provision of the online presentation of the measurement results at http://ingrid.cetic.be:33323/qualoss_assessment/index.php these problems should however be solved.

3.2.1.1 Quality Model Satisfaction

Hypothesis	*Validation Status*	*Comment*
Quality Model Completeness. All characteristics included.	4	
QMC-wp. All characteristics of work products included.	3	
QMC-cm. All characteristics of community included.	3	
QMC-sp. All characteristics of software processes included.	4	
Quality Model Minimality. Minimal set of characteristics included.	4	None of the measures was considered to be irrelevant
QMM-wp. Minimal set of characteristics of work products included.	4	
QMM-cm. Minimal set of characteristics of community included.	4	
QMM-sp. Minimal set of characteristics of software processes included.	4	

Table 1: Quality Model Satisfaction for GCC

3.2.1.2 Indicator Satisfaction

Hypothesis	*Validation Status*	*Comment*
Leaf Characteristic Convincingness. The value of leaf characteristics are convincing.	4	
Low-level Indicator Sufficiency. The indicators are sufficient to assess the characteristics.	4	
High-level Indicator Aggregation. Aggregation summarizes appropriately scores from leaf characteristics up to the root of the quality model tree.	3	Value is an estimation based on the observation of a few occurrences
Measure Aggregation Satisfaction. The rule for a given low-level indicator is a good way to assess risks related to the corresponding leaf characteristic.	4	
Indicator Drill-down Capability Results going to the level of detail of low-level indicator are satisfying.	4	
Measure Drill-down Capability. Results going to the level of detail of measures are satisfying.	4	

Table 2: Indicator Satisfaction for GCC

3.2.2 Profitability

Hypothesis	*Validation Status*	*Comment*
Characteristic Effort Profitability. Characteristics are profitable.	4	Question: Do you think that effort and time used for the assessment are reasonable?
Indicator Effort Profitability. Indicators are profitable.	4	Question: Do you think that effort and time used for the assessment are reasonable?
Specific Assessment Profitability. Adapting into a specific assessment method is profitable.	-	Could not be validated in interviews because of too high level of detail

Table 3: Profitability of GCC Assessment

3.3 AdaCore/GCC Conclusions

All hypotheses that could be tested in the interviews have either been fully or largely conformed. Improvements should be considered regarding QMC-wp and QMC-cm. The interviews turned out some ideas for such improvements.

As it was expected, the assessment did not find large differences between the two versions of the GCC back-end. This was not surprising because GCC is a very mature endeavor.

However, it provided some very interesting indicators that showed up as good to evaluate the quality of a version (mainly its robustness and reliability):

- the rapidity to implement bug fixes and enhancements, which was much better in GCC 4.3 than in 4.2 (roughly half of the time)
- the number of serious regressions, which was 30% smaller in GCC 4.3 than in 4.2
- the location of changes, indicating that in GCC 4.3 11.62% of the files changes from the first 4.3 release to the latest, while in the case of the GCC 4.2 branch the changes were much more pervasive (44.37% of the files changed)

AdaCore was satisfied with these results, and they will consider to measure at least these three indicators in the future to help making decisions when migrating to a new back-end version.

4. Freecode / Asterisk

4.1 Objectives

Freecode uses Asterisk for the provision and implementation of complete telecommunication infrastructures. This case study analyzes Asterisk 1.4.26, and compares the QualOSS results against an internal OpenBRR assessment that they performed on the same version.

4.2 Interview Results

The people interviewed were Tommy Jensen (Freecode), Anna Tannenberg (Freecode), and Arne-Kristian Groven (Norsk Regnesentral, Oslo), who are very knowledgeable about quality assessments and Asterisk.

4.3 Freecode/Asterisk Hypotheses

4.3.1 User Satisfaction

Regarding user satisfaction, the pre-assessment interview confirmed largely that the QualOSS Standard Assessment was perceived to match very well the interviewees' requirements from a quality assessment, though the reservation was made that a measurement that could be used at the client's site (less technical and more “marketing material”) to identify preferences and risks would be desirable. However, it was clear to the interviewee that any measurement of this kind would heavily rely on technical measures and terms that cannot be easily explained to an end user.

Both interviewees agreed that the measures and indicators of QualOSS appear comprehensive (rather too comprehensive than too limited) and are relevant. Identifying measures that could be ignored was difficult for the interviewees, as they pointed out that such a decision must take into account the links between different measures that are useful for the calculation of broader indicators or for other recombinations of information that might be useful with regard to specific objectives.

The post-assessment interview confirmed the impressions uttered in the pre-assessment interviews. However, despite the fact that some concern was raised by the sheer number of QualOSS measures, the interviewee made some suggestions for additional measures that an advanced version of the QualOSS Assessment could take into account: measures of the difficulty to become an active community member and measures on how good the community fulfills support functions.

The overall degree of satisfaction with the results of the measurement, based on a scale from 1 (not at all satisfied) to 5 (very satisfied) was allocated “somewhere close to 4”. The overall quality of the QualOSS Standard Assessment was considered to be high, though improvements in terms of understandability were demanded.

4.3.1.1 Quality Model Satisfaction

Hypothesis	***Validation Status***	***Comment***
Quality Model Completeness. All characteristics included.	4	
QMC-wp. All characteristics of work products included.	4	
QMC-cm. All characteristics of community included.	3	
QMC-sp. All characteristics of software processes included.	4	
Quality Model Minimality. Minimal set of characteristics included.	4	
QMM-wp. Minimal set of characteristics of work products included.	4	
QMM-cm. Minimal set of characteristics of community included.	4	
QMM-sp. Minimal set of characteristics of software processes included.	4	

Table 4: Quality Model Satisfaction for Asterisk

4.3.1.2 Indicator Satisfaction

Hypothesis	*Validation Status*	*Comment*
Leaf Characteristic Convincingness. The value of leaf characteristics are convincing.	3	Understandability and combination of leaves was questioned, but no better alternative could be developed
Low-level Indicator Sufficiency. The indicators are sufficient to assess the characteristics.	-	Could not be validated in interviews because of too high level of detail
High-level Indicator Aggregation. Aggregation summarizes appropriately scores from leaf characteristics up to the root of the quality model tree.	3	Value is an estimation based on the observation of a few occurrences
Measure Aggregation Satisfaction. The rule for a given low-level indicator is a good way to assess risks related to the corresponding leaf characteristic.	-	Could not be validated in interviews because of too high level of detail
Indicator Drill-down Capability Results going to the level of detail of low-level indicator are satisfying.	-	Could not be validated in interviews because of too high level of detail
Measure Drill-down Capability. Results going to the level of detail of measures are satisfying.	-	Could not be validated in interviews because of too high level of detail

Table 5: Indicator Satisfaction for Asterisk

4.3.2 Profitability

Hypothesis	*Validation Status*	*Comment*
Characteristic Effort Profitability. Characteristics are profitable.	4	Question: Do you think that effort and time used for the assessment are reasonable?
Indicator Effort Profitability. Indicators are profitable.	4	Question: Do you think that effort and time used for the assessment are reasonable?
Specific Assessment Profitability. Adapting into a specific assessment method is profitable.	-	Could not be validated in interviews because of too high level of detail

Table 6: Profitability of Asterisk Assessment

4.4 Freecode/Asterisk Conclusions

All hypotheses that could be tested in the interviews have either been fully or largely conformed. Improvements should be considered regarding QMC-cm, especially with respect to support functions.

The Freecode representatives and Arne-Kristian Groven felt that, overall, the QualOSS Standard Assessment provides a reliable and very detailed measurement of the quality of F/OSS products. However, the richness of measures and indicators also created a feeling of uncertainty and ambiguity with regard to

the clarity and composition of the indicators. They all emphasized that to understand it correctly and to tap its full potential requires some time to familiarize with the QualOSS measurement.

Since all interviewees had experience in assessing the quality of Asterisk with OpenBRR, they judged the QualOSS Standard Assessment in comparison to this tool. They commonly felt that the QualOSS Standard Assessment performs better than OpenBRR when used by a company with staff that has CS skills at its command. However, if the two assessment tools are compared from the final users' (e.g. a Freecode client) point of view, OpenBRR appeared to be better because it was not as technical and complicated (regarding the number of measures) as the QualOSS assessment.

On the other hand, the richness of the information provided by QualOSS was seen an an advantage especially for a firm with deep knowledge of the F/OSS endeavor and the knowledge, resources and interest in monitoring the advancement of this endeavor over time. In this sense, QualOSS appeared advantageous with regard to repeatability and verifiability of the measurement results.

5. Océ PRISMAspool to use the LPR client of the yanolc project

5.1 Objectives

Océ delivers the Océ PRISMAspool software product that provides an enhanced spooling system for Windows environments. A new LPR client should be delivered with a next version of PRISMAspool, and this case study analyzes the use of a subset of the Open Source project *yanolc* (Yet ANOther LPR Client). As PRISMAspool is used for professional printing, for transactional and mailing environment, a top most quality and reliability is required.

This assessment analyzes the most recent available version of yanolc, namely version 1.2.11, and it will study whether the QualOSS methodology will confirm the decision that has been taken to not use yanolc as a basis but well an internally developed tool.

The people interviewed were Jacques Flamand (Océ), whose role is that of project and product manager of the printer client, and François Thiry (Océ), who is one of the developers of the printer client.

5.2 Océ/Yanolc Hypotheses

5.2.1 User Satisfaction

In the pre-assessment interviews, both interviewees said that the QualOSS Standard Assessment was perceived to match very well the interviewees' requirements from a quality assessment. As a critical point, both interviewees emphasized that the QualOSS Standard Assessment contains too many measures for the F/OSS collaboration context of a fork. They also saw that the specific requirements of a company from a F/OSS product in such a collaboration context are highly company-specific, so that the QualOSS methodology must provide a broad range of measures in order to match all possible company specifics.

Both interviewees agreed that the indicators chosen to measure quality related to the work product are relevant. The overall degree of satisfaction with the results of the measurement, based on a scale from 1 (not at all satisfied) to 5 (very satisfied) was seen somewhere between 4 and 5. The overall quality of the QualOSS Standard Assessment was considered to be high.

As a recommendation for a more advanced assessment methodology, it was suggested to present the results of the assessment together with information about the (possible) meanings of these results in different collaboration contexts. However, it was also seen that this additional functionality might have the downside of confusing the users of the results.

5.2.1.1 Quality Model Satisfaction

Hypothesis	*Validation Status*	*Comment*
Quality Model Completeness. All characteristics included.	3	Though all characteristics needed for Océ have been included the measurement appeared to broad for the specifics of a fork
QMC-wp. All characteristics of work products included.	3	
QMC-cm. All characteristics of community included.	3	
QMC-sp. All characteristics of software processes included.	3	
Quality Model Minimality. Minimal set of characteristics included.	4	
QMM-wp. Minimal set of characteristics of work products included.	4	
QMM-cm. Minimal set of characteristics of community included.	4	
QMM-sp. Minimal set of characteristics of software processes included.	4	

Table 7: Quality Model Satisfaction for Yanolc

5.2.1.2 Indicator Satisfaction

Hypothesis	*Validation Status*	*Comment*
Leaf Characteristic Convincingness. The value of leaf characteristics are convincing.	4	
Low-level Indicator Sufficiency. The indicators are sufficient to assess the characteristics.	4	
High-level Indicator Aggregation. Aggregation summarizes appropriately scores from leaf characteristics up to the root of the quality model tree.	4	Value is an estimation based on the observation of a few occurrences
Measure Aggregation Satisfaction. The rule for a given low-level indicator is a good way to assess risks related to the corresponding leaf characteristic.	4	
Indicator Drill-down Capability Results going to the level of detail of low-level indicator are satisfying.	4	
Measure Drill-down Capability. Results going to the level of detail of measures are satisfying.	4	

Table 8: Indicator Satisfaction for Yanolc

5.2.2 Profitability

Hypothesis	*Validation Status*	*Comment*
Characteristic Effort Profitability. Characteristics are profitable.	4	Question: Do you think that effort and time used for the assessment are reasonable?
Indicator Effort Profitability. Indicators are profitable.	4	Question: Do you think that effort and time used for the assessment are reasonable?
Specific Assessment Profitability. Adapting into a specific assessment method is profitable.	-	Could not be validated in interviews because of too high level of detail

Table 9: Profitability of Yanolc Assessment

5.3 Océ/Yanolc Conclusions

All hypotheses that could be tested in the interviews have either been fully or largely conformed. Improvements should be considered regarding how to better adapt the measurement methodology to the specifics of the F/OSS collaboration context of a fork. The concrete shortcoming is that businesses, in this case, want to examine only a very specific and limited set of factors and measures. The QualOSS platform should provide opportunities to help firms in this respect.

The results of the QualOSS assessment of yanolc confirmed to Océ the mixed feelings they had about the robustness and evolvability of the yanolc project. Indeed, yanolc is a one-developer effort and hence the benefit of a F/OSS community would be very limited; moreover, the bug tracker is missing, so many indicators regarding reliability and maintainability could not be obtained, which was then viewed as a high risk by the QualOSS assessment. Additional factors that led Océ to the decision of not using yanolc were the fact that they would have to modify (to a rather large extent) the code to make yanolc fulfil Océ's functional requirements, and the dependency of yanolc on another (rather big) project.

6. AdaCore Couverture

6.1 Objectives

AdaCore, together with Open Wide, ENST and LIP6, is developing a Free Software coverage analysis toolset. In addition to the tools, the project aims to generate artifacts that allow the tools to be used for safety-critical software projects undergoing a DO-178B software audit process for all levels of criticality. The project is being developed following the open source philosophy, and it is being moved to an open forge to allow for external contributions.

AdaCore wants to use the QualOSS methodology to analyze and understand what could help to make the Couverture project a robust and evolvable F/OSS endeavor while moving it to an open forge. Hints about how to create and interact with the new community, documentation and procedures to create, and how to develop and organize the code would be very valuable.

The people interviewed were Tristan Gingold (AdaCore), whose role is that of main developer of Couverture, and José Ruiz, whose role is that of user.

6.2 AdaCore/Couverture Hypotheses

6.2.1 User Satisfaction

Both, José Ruiz and Tristan Gingold, said that the indicators used to measure quality are relevant. Suggestions for a more advanced measurement were to find ways that better take into account the specifics of young projects (i.e. the measurement should be less strict in such cases), to reduce the number of measures for the software process, and to put the results in a context that allows to better understand them (i.e. an explanatory framework) and to benchmark results of one measurement with another.

The overall degree of satisfaction with the results of the measurement, based on a scale from 1 (not at all satisfied) to 5 (very satisfied) did not differ between the two interviewees, as both allocated a value of 4 to the measurement results. The overall quality of the QualOSS Standard Assessment was considered to be high by both interviewees.

6.2.1.1 Quality Model Satisfaction

Hypothesis	*Validation Status*	*Comment*
Quality Model Completeness. All characteristics included.	3	Specifics of projects in early stages are nor captured properly
QMC-wp. All characteristics of work products included.	4	
QMC-cm. All characteristics of community included.	4	
QMC-sp. All characteristics of software processes included.	3	Too many measures
Quality Model Minimality. Minimal set of characteristics included.	4	
QMM-wp. Minimal set of characteristics of work products included.	4	
QMM-cm. Minimal set of characteristics of community included.	4	
QMM-sp. Minimal set of characteristics of software processes included.	4	

Table 10: Quality Model Satisfaction for Couverture

6.2.1.2 Indicator Satisfaction

Hypothesis	*Validation Status*	*Comment*
Leaf Characteristic Convincingness. The value of leaf characteristics are convincing.	4	
Low-level Indicator Sufficiency. The indicators are sufficient to assess the characteristics.	-	Could not be validated in interviews because not enough data
High-level Indicator Aggregation. Aggregation summarizes appropriately scores from leaf characteristics up to the root of the quality model tree.	4	Value is an estimation based on the observation of a few occurrences
Measure Aggregation Satisfaction. The rule for a given low-level indicator is a good way to assess risks related to the corresponding leaf characteristic.	-	Could not be validated in interviews because not enough data
Indicator Drill-down Capability Results going to the level of detail of low-level indicator are satisfying.	-	Could not be validated in interviews because not enough data
Measure Drill-down Capability. Results going to the level of detail of measures are satisfying.	-	Could not be validated in interviews because not enough data

Table 11: Indicator Satisfaction for Couverture

6.2.2 Profitability

Hypothesis	*Validation Status*	*Comment*
Characteristic Effort Profitability. Characteristics are profitable.	4	Question: Do you think that effort and time used for the assessment are reasonable?
Indicator Effort Profitability. Indicators are profitable.	4	Question: Do you think that effort and time used for the assessment are reasonable?
Specific Assessment Profitability. Adapting into a specific assessment method is profitable.	-	Could not be validated in interviews because of too high level of detail

Table 12: Profitability of Couverture Assessment

6.3 AdaCore/Couverture Conclusions

All hypotheses that could be tested in the interviews have either been fully or largely conformed. Improvements should be considered regarding how the QualOSS platform could be adapted to the specifics of F/OSS projects that are at an early stage. The problem here is that a lot of data might not be available for the measurement. A recommendation should be given if and to what degree the QualOSS Assessment can be applied in such a case, and how the results can be interpreted.

The QualOSS assessment helped AdaCore understanding some hints about how to facilitate external contributions to Couverture. Test indicators were very useful, helping to reinforce the view that testing is still weak and not well documented. This is one of the aspects on which AdaCore would need to work to have reliable external contributions. Documentation results were useful, showing a lack of documentation (with was expected) but it helped to identify specific important items that are missing (AdaCore considered important to work on testing documentation and FAQ). Finally, indicators related to software processes were also very useful, showing that the community knowledge is not formalized, and it relies on direct interactions among the developers (most of them work for AdaCore), which is a risk when opening the development to external people.

In a few words, the QualOSS assessment helped to determine that there are two critical aspects on which AdaCore needs to work to be successful in having external contributions (Tests and Software Processes), with Documentation being also important but with lower criticality.

7. Overall Conclusions

Overall, the case studies have validated all testable hypotheses, either fully or largely. None of the interviews has turned out that the QualOSS Standard Assessment is incomplete in some respect or misses out an important measure, though suggestions have been made how a more advanced QualOSS platform could be improved as compared to the Standard Assessment. In all cases, the QualOSS Standard Assessment turned out to be profitable. However, a shortcoming of the level of detail captured by the QualOSS Standards Assessment is that hypotheses that required to dig deep down into the measurement details could only be validated in interviews with QualOSS experts.

People found the QualOSS assessment useful and meaningful. Sometimes, it showed important or interesting quality factors or indicators that were not taken into account before. The large number of measures and indicators was often seen as an advantage because it provides a flexible way to look at quality issues from many different angles.

Most of the times, the development of an Advanced QualOSS Assessment would be very useful. The idea is to take the Standard QualOSS Assessment as a base, and tailor it removing the characteristics or indicators that are not of interest and adding those specific to the endeavor under analysis. This way, interesting indicators are extracted at a reduced cost.

The mix of both internal case studies (on companies belonging to the QualOSS consortium) and external ones helped validating the QualOSS methodology. It allowed both in-depth analyses on internal case studies and impartial feedback from external ones.

Additionally, during the Asterisk case study the results of the QualOSS assessment were informally compared against a previous OpenBRR evaluation of the same endeavor. The Standard QualOSS Assessment showed up as a richer evaluation, although targeted to people technically more knowledgeable.