

 Sponsored through Framework Programme Sixth (Call 5) by  Information Society Technologies		Document Information
		Version: 1.0 Date : 30 Nov, 2009 Dec 09 Pages : 13
		Owning Partner: URJC
		Author(s): Jesus M. Gonzalez-Barahona Daniel Izquierdo-Cortazar
		Reviewer(s): Paul Adams Jean Christophe Deprez Xavier Heymans
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)		To: CONSORTIUM
		Purpose of distribution:
Status: <input type="checkbox"/> Draft <input type="checkbox"/> To be reviewed <input type="checkbox"/> Proposal <input checked="" type="checkbox"/> Final/Released		Confidentiality: <input checked="" type="checkbox"/> Public - Intended for public use <input type="checkbox"/> Restricted - Intended for E.C. Program participants <input type="checkbox"/> Confidential - Intended for individual partner only
Deliverable ID: D6.5a		
Title: Collaboration Plan and Report		
License for distribution: This work is licensed under a Creative Commons Attribution-Share Alike 2.5 License . (The license can be found in http://creativecommons.org/licenses/by-sa/2.5/) 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.		

	<p>Collaboration plan and report</p> <p>Deliverable ID: D6.5b</p>	<p>Page : 2 of 13</p> <hr/> <p>Version: 1.0 Date: Dec 10, 09</p> <hr/> <p>Status : Proposal Confid : Public</p>
---	---	---


Deliverable: D6.5b

Title: Collaboration plan and report

Executive Summary:

This report explores the opportunities for collaboration and joint dissemination of the projects funded by the 6th Framework Programme of the European Commission in the area of libre (free, open source) software development and quality. Beside QUALOSS, other projects are FLOSSMETRICS, SQO-OSS, and QUALIPSO. Those projects, working in the same area, and sharing some partners, are interested in identifying synergies and avoiding duplication of efforts.

Several areas of potential collaboration are identified in this document, in particular, in relation to data exchange, tools and infrastructure, interactions between communities of researchers and of libre software developers, and dissemination effort.

	<p>Collaboration plan and report</p> <p>Deliverable ID: D6.5b</p>	<p>Page : 3 of 13</p> <hr/> <p>Version: 1.0 Date: Dec 10, 09</p> <hr/> <p>Status : Proposal Confid : Public</p>
---	---	---

CHANGE LOG

Ver.	Date	Author	Description
0.1	10.12.2007	Jesus M. Gonzalez-Barahona	First version, based on the slides used in the coordination meeting with other related projects.
0.2	13.12.2007	Jean-Christophe Deprez	review for spelling and phrasing.
1.0	13.12.2007	Jean-Christophe Deprez	second review for formatting and consistency.
1.1	30.11.2009	Daniel Izquierdo-Cortazar	Initial new version based on old document
1.2	8.12.2009	Paul Adams	First Review
1.21	8.12.2009	Daniel Izquierdo-Cortazar	Minor corrections

APPLICABLE DOCUMENT LIST

Ref.	Title, author, source, date, status	Deliverable Identification



	<p>Collaboration plan and report</p> <p>Deliverable ID: D6.5b</p>	<p>Page : 4 of 13</p> <hr/> <p>Version: 1.0 Date: Dec 10, 09</p> <hr/> <p>Status : Proposal Confid : Public</p>
---	---	---

TABLE OF CONTENTS

1. Introduction.....	5
1.1 Scope and status of this report	5
1.2 Projects involved.....	5
1.3 Resources involved.....	6
1.4 Main areas for collaboration.....	6
2. Data exchange.....	6
2.1 Description of the current scenario.....	6
2.2 Opportunities and advantages.....	7
2.3 Action plan.....	8
3. Tools and infrastructure.....	9
3.1 Description of the current scenario and opportunities for collaboration.....	9
3.2 Action plan.....	9
4. Relationship with community and industry.....	10
5. Dissemination activities.....	11
5.1 Conferences and workshops.....	11
5.2 Other dissemination actions.....	12
6. Specific actions for coordination.....	12
7. Final remarks and conclusions.....	12

	Collaboration plan and report Deliverable ID: D6.5b	Page : 5 of 13
		Version: 1.0 Date: Dec 10, 09
		Status : Proposal Confid : Public

1. INTRODUCTION

There are several projects in the area of libre (free, open source) software, quality and metrics funded under the 6th Framework Program of the European Commission, and some other projects (both in the European and worldwide context) with interests in the same area. Some coordination and collaboration between these projects seem to be interesting, and could enable a wider impact for all of them, also avoiding solving the same problems once and again. This report details the coordination actions to be explored and put into work among these projects.

1.1 SCOPE AND STATUS OF THIS REPORT

Although this report was initially intended for the QUALLOSS project, and is written in its context, it will be the base for a collaboration plan for all 6th Framework Program projects related by this common interest in libre software, metrics and quality.

1.2 PROJECTS INVOLVED

The list of 6th Framework Program projects involved in this area are:

- CALIBRE, <http://calibre.ie>
- EDOS, <http://edos-project.org>
- FLOSSWorld, <http://flossworld.org>
- FLOSSMETRICS, <http://flossmetrics.org>
- QUALLOSS, <http://qualoss.org>
- SQO-OSS, <http://sqo-oss.eu>
- QUALIPSO


Other projects and initiatives at the European level potentially interested in these issues are:

- Open Source Observatory and Repository (OSO-R), funded by IDABC, <http://ec.europa.eu/idabc/>

Although this document is focused on the collaboration between EU-funded projects, some other projects and initiatives at the international level could also be interested in some kind of collaboration:

- OpenBRR, <http://openbrr.org>
- FLOSSMole, <http://ossmole.sourceforge.net>
- TA-RE Corpus Working Group, <http://tare.dforge.cse.ucsc.edu>
- QSOS, <http://qsos.org>

Still some other unidentified initiatives, or running at other levels, could also be a subject of collaboration. They will be actively sought and contacted during the life of collaboration plan.

	Collaboration plan and report Deliverable ID: D6.5b	Page : 6 of 13
		Version: 1.0 Date: Dec 10, 09
		Status : Proposal Confid : Public

1.3 RESOURCES INVOLVED

There are no additional resources allocated to coordination tasks and activities besides those already allocated by each project involved. In QUALOSS tasks 6.5, which is dedicated to collaboration effort, was allocated 1 men-month per partner. No additional funds have been provided for this extended collaboration, either. However, all involved projects have the interest and the specific goal of collaborating with other projects in the area is present in their descriptions of work. Therefore, some funding will be available, when expenses can be charged to the budget of the corresponding projects. Other parties are welcome to also provide their own funding.

In some special cases, such as large conferences or dissemination activities, funding from sponsors could be actively considered, to complement own funding.

1.4 MAIN AREAS FOR COLLABORATION

Several areas have been identified for collaboration. Among them, the following are the most prominent:

- Data exchange
- Tools and infrastructure
- Relationship between community and industry
- Dissemination

These areas will be explored and described in some detail in this report. However, it is important to notice that the general motivation for collaboration is in fact wider than this list could suggest: the main target will be to exploit, in several ways, the large pool of talent available in the different projects in the referenced area.


2. DATA EXCHANGE

All the involved projects are either producers or consumers (in some cases potential consumers) of data about libre software development and quality of libre software projects. Therefore it can be very productive to find ways of sharing this data, and establishing a collaboration framework for validation, dissemination and experience of collection, analysis and use of data from libre software repositories.

2.1 DESCRIPTION OF THE CURRENT SCENARIO

Several projects (CALIBRE in the past, EDOS, FLOSSWorld, FLOSSMETRICS and SQO-OSS currently) are collecting and analysing data from libre software repositories. In particular, FLOSSMETRICS and SQO-OSS are interested (although with different goals) in massive collection of data from thousands of projects. However, each of these two projects is using different approaches, considering different objects of interest (both in the repositories and in the further analysis), different scopes, and is working in different contexts.

Some other projects, in addition to those already mentioned (specially QUALOSS and QUALIPSO), have also a component of data retrieval, but are more focused on data consumption. Some other initiatives in the European scenario (such as the Open Source Observatory and Repository from IDABC) could also benefit from data consumption. In the international scenario, several other projects (FLOSSMole, TA-RE working group, etc.) are also retrieving (or planning to retrieve) data about libre software development, and many others are consuming (or planning to consume) this kind of data.

	<p>Collaboration plan and report</p> <p>Deliverable ID: D6.5b</p>	<p>Page : 7 of 13</p> <hr/> <p>Version: 1.0 Date: Dec 10, 09</p> <hr/> <p>Status : Proposal Confid : Public</p>
--	---	---

In particular, there are two main communities in the area of consumption of data about software development:


- The community of researchers about libre software that is starting to understand the benefits of using publicly available datasets instead of having each group retrieving their own data directly from their repositories.
- The community of companies and institutions interested in ranking and assessing libre software products that is also starting to understand the benefits of using quantitative data available from reliable sources.

Of course, these cases are not unique, since many companies, public administrations and other institutions are demanding high quality data about libre software projects and products to take informed decisions, but these two types of communities are the most active currently.

2.2 OPPORTUNITIES AND ADVANTAGES

The described scenario already shows the opportunities for collaboration, and its advantages both for the actors involved and for the field in general. This is further detailed below:

- Specialization in retrieval of data. Fetching information from public software repositories, and filtering it in ways that make the retrieved data usable for research and for end users, is becoming a more and more complex task. The research community is interested in deeper analyses, correlation of data from different sources, follow-up of complex patterns, etc. End-users need comparative analysis of meaningful parameters, and detailed information about the evolution of libre projects. Decision-makers need deep but comprehensive views of the full landscape of libre software development. Getting and processing information with enough quality to satisfy these needs is a great challenge, which needs tools of great complexity, and techniques which are getting very complex. Therefore, it makes sense to promote the production of this information by specialized teams and let other groups more specialized in analysis use it.
- Validation by third parties. Validation of retrieved data is a difficult and time-consuming issue. Data in libre software repositories has all kinds of noise (ranging from automatic activities which have to be differentiated from human activities, to errors in time-stamping of records), which has to be removed or mitigated. In the process, several heuristics are used, which can lead to errors in the resulting data. In addition, the retrieval process itself can also introduce errors. When the quantity of retrieved data is huge, the chances of being hit by these errors increases. Therefore, applying some techniques for validation of data is a must. But these techniques are usually complex, time-consuming and in many cases linked to specific analysis. This means that the team producing the data cannot always devote the needed time to validation effort, which can quickly become a continuous, tedious process. Having contributions to validation by third parties does not only help to improve the quality and reliability of the data, but also to avoid the unavoidable skew of auto-validation.
- Sharing of ideas and approaches. Although scientific publications contribute to the spread of new ideas and approaches in the field, many details are not obvious when reading literature, and need of a more head-to-head relationship. In particular, early discussion of new approaches allows for a faster refinement, and helps to avoid duplication of efforts.
- Sharing of tools. Many tools in the field are developed by the teams themselves, and fortunately they are usually willing to share them (in fact, most of them are libre software, and thus shareable by nature). But willingness to share tools is usually not enough for experiencing actual sharing. In fact, these tools are usually complex, and built with more focus on functionality than usability. Therefore,

	Collaboration plan and report Deliverable ID: D6.5b	Page : 8 of 13
		Version: 1.0 Date: Dec 10, 09
		Status : Proposal Confid : Public

close contact between potential users and the creators of the tools is of great importance for their reuse.


- Creation of a critical mass. Several teams are working in the area, specially in the European Union. That opens plenty of opportunities for collaboration and creation of a true critical mass that permits to maintain the good position of these teams in the international landscape. This should also ease the transfer of results to the market, promoting the adoption libre software by companies and public administrations , and the creation of start-ups interested in marketing services based on those results.
- Contact with other international groups. The creation of the previous critical mass should also help to convert it as the focal point for other groups at the international level. Coordination with similar efforts, specially in the United States and Canada, should be possible on an equal basis.

2.3 ACTION PLAN

An action plan in this area is an evolving target, but in the current context, the directions proposed (and already being discussed with interested parties) are:

- Design of exchange formats and methods. Of course, for any kind of collaboration in this area, a pre-requisite is to agree on exchange formats. Therefore, one of the specific actions will be to give the needed steps to ensure that actual transfer of data from project to project and from team to team is possible. Fortunately, this is (in different contexts) in the description of work of many of the involved projects, and mainly coordination actions will be needed to ensure coordinated decisions (see below).
- Joint meetings. The current plan considers two coordination meetings per year. The first one is already scheduled to be co-located with FOSDEM 2007, in February 2007 in Brussels (BE), probably followed by another co-located with OSS 2007, in June 2007 in Limerick (IE). In general, the plan is to co-locate the meetings with events attended by persons involved in the E.C. projects and other interest initiatives.
- Global workshops. These will be sponsored by the involved projects. It will help in dissemination, high-quality contributions, reviewing process, and of course interesting discussion during the workshop. These workshops are expected to be not only a place for networking with other relevant international initiatives, but also a showcase for the main results of the projects involved.
- Common infrastructure. Since the data retrieval and data exchange needs of several projects are similar, the use of common pieces of infrastructure, and even the set-up of a common joint infrastructure is to be explored. However, some issues related to the specific description of work of the involved projects have already been identified, which will limit this action.
- Establishment of joint data-sets. Agreement on datasets that can be used for benchmarking and performing reference studies is a well identified target. The establishment of a joint archive of those datasets will be explored by the involved projects.
- Validation. The main action in this area will be to set up the means and procedures for allowing collaboration in the validation process, and validation by third parties, in particular, since some data will be provided by third parties, these validation procedures will help to improve the quality of the datasets.
- Involvement in global communities with the same interest. This includes not only research communities, but also industrial forums, and specific forums of libre software developers. This action should help to better understand the needs and issues of the potential users of the data produced by the involved projects, and get some feedback about the results.

This action plan is subject to discussion with all the projects involved.

	<p>Collaboration plan and report</p> <p>Deliverable ID: D6.5b</p>	Page : 9 of 13
		Version: 1.0 Date: Dec 10, 09
		Status : Proposal Confid : Public

3. TOOLS AND INFRASTRUCTURE

Some of the tools to be designed, implemented and used by some of the projects involved are of potential interest to others. Therefore, coordination actions in this area, which improve code reuse, avoidance of duplicate effort, and transfer of knowledge between projects, may improve meaningfully the aggregated output of all of them.

3.1 DESCRIPTION OF THE CURRENT SCENARIO AND OPPORTUNITIES FOR COLLABORATION

Most of the projects in the area are specially interested in tools distributed as libre software. In some cases they are produced by the projects themselves, in others, they are improved or customized. These tools can be found in several fields, from the already mentioned data retrieval to the specific analysis or website set-up. All of them are potentially shareable, and can, if convenient, be developed in coordination. That could have an impact not only on its quality and functionality, but also in their user base (which is important for their long-term sustainability).

In addition, at least two projects (SQO-OSS and FLOSSMETRICS) are planning to set-up a significant infrastructure for data retrieval and automated analysis, which will be a clear target for specific collaboration (as previously mentioned). These two projects will also have infrastructures with some resemblance in the area of data providing and web services. Other projects (such as QUALOSS) are clear candidates to be users of those web services.


Many of the involved projects are also actively researching the state of the art with similar targets, and in some cases, by related teams. Obviously, coordination in this area can be specially productive. Same can be said about documents designing the main systems to be built, although in this cases there are several constraints because of the description of work, and the work plan.

Another limit for coordination in this area are the different interests, targets and scopes of the projects involved.

3.2 ACTION PLAN

The actions proposed for fostering collaboration in this area are in many respects linked to those already identified for data exchange, and were already mentioned in that section. However, some specific actions can be identified:

- Common repository of tools. Given that the tool-set used in several projects could be similar, and that in any case, some tools used by one could be interesting for others, the establishment of a common repository of tools (with a clear focus on libre software tools) can be of interest. In addition, this repository could also be of interest to third parties, which would help in the dissemination and impact of the involved projects. This repository is not necessarily an actual storage point for the tools: it could also be a catalogue with information and clear links to the tools themselves (which, being libre software, are already available in the Internet).
- Collaboration in the design of new tools, and improvement of the existing ones. This a specially interesting area, but also a difficult one, since there are many implications for the functionality, maintainability and evolution of the tools themselves. Specific exploration of opportunities for collaboration will be identified in the joint meetings and other joint activities (see below).

	<p>Collaboration plan and report</p> <p>Deliverable ID: D6.5b</p>	Page : 10 of 13
		Version: 1.0
		Date: Dec 10, 09
		Status : Proposal Confid : Public

- Identification of common modules in infrastructures. Since infrastructures to be set up by several projects have resemblances, and similar modules, it will be explored how to share information at least in the design phase, and if possible, also in the implementation phase. If possible, modules that are directly shareable by several projects will also be identified, to avoid duplication of efforts. However, these activities will be limited by the scheduled work plan of each project, and their different focus and objectives.
- Joint meetings. As in the area of data exchange, joint meetings will be organized for discussion and identification of common interests. This will include virtual meetings using mailing lists and phone conferences.


This action plan should be revised carefully after the first studies on available tools, and after the design of the systems in all involved projects is ready.

4. RELATIONSHIP WITH COMMUNITY AND INDUSTRY

For maximizing the impact on society, and on the main targeted actors, which include the libre software community itself and industry at large as users and producers of software, a close relationship with these actors is needed. This relationship will be useful for obtaining the needed feedback by the group of projects, and will also help in the dissemination of the results of interest. In fact, the projects involved in our collaboration effort are recognizing these facts from their design phase thus means, procedures and plans to foster this involvement are built. These actions can be coordinated, and synergies between the different approaches can be explored.

Some specific ideas in this area are:

- Involvement of industrial and community partners. Take advantage of the involvement of industry and community partners in the involved projects, helping them to understand and communicate not only with the projects in which they participate directly, but also others. In addition, common events where these partners and other interested parties could participate and share impressions with all the involved projects will help to have a broader impact and feedback. This is already enacted with EDOS partners and AdaCore having article accepted for FOSDEM 2007 presentations workshop organized by Universidad Rey Juan Carlos.
- Combined reports. Since the results expected from several projects are related, combined reports (and other kinds of combined deliverables, such as combined dissemination events) could be produced. For these combined reports, establishing common means for delivery of the resulting documents, but also joint effort in producing comprehensive studies (by combining and coordination individual results) would make them more attractive for target users.
- Agreements with communities and libre software projects. Since the subject matter for most studies is libre software development, joint agreements with communities of developers, or specific libre software projects, could help to streamline the collection of results, and at the same time, to make those projects aware of the research actions on them. The promotion of a “research friendly” label for those projects (which could include special facilities for researchers, as well as special facilities and services by researchers to projects) could be one of the main lines of this collaboration between the research community and the libre software development community (including companies involved in libre software development).
- One-place information for projects, actors or artefacts. Different studies (performed by different research projects) will analyse libre software repositories from different points of view. Having all this information available in one place (probably organized by project, actor or artefact) will help to maximize the impact on potential users of that information. However, specific (and probably complex) technical means should be established to achieve this goal, which could be well above the

	<p>Collaboration plan and report</p> <p>Deliverable ID: D6.5b</p>	<p>Page : 11 of 13</p>
		<p>Version: 1.0</p> <p>Date: Dec 10, 09</p>
		<p>Status : Proposal</p> <p>Confid : Public</p>

descriptions of work of the involved projects. Maybe techniques such as federation (using RSS or similar means) or mashups could help in this direction.

5. DISSEMINATION ACTIVITIES

Joint dissemination activities are designed targeting three different populations: the libre software community, researchers in the field, and the industry with interests in libre software. Some activities, mainly presentations in conferences and workshops, will be specific to one group. Some others such as having a common website is for touching all three groups at once.

5.1 CONFERENCES AND WORKSHOPS


Presentations and conferences can be an interesting dissemination instrument. Organising workshops and special sessions co-located with those events will allow for a more focused dissemination. Of course, those workshops and sessions can be open to other related initiatives. Some of the activities which have been organized are the following, classified by target populations:

Libre software community:

- Research room @ FOSDEM 2007. Co-located with FOSDEM 2007, in Brussels (BE), 24-25 February 2007. This was the continuation of the CALIBRE room, organized by the CALIBRE project at FOSDEM in 2005 and 2006. It is an open workshop focused on the interaction of developers and researchers. FOSDEM, being the most important event for libre software developers in Europe, is a good event for this interaction. The research room includes a session with presentations of projects, some sessions with presentations of reviewed contributions (after an open call for contributions), and panels for discussing how research could help libre software development, and what is still missing in this area.
- Research room @ FOSDEM 2008. Co-located with FOSDEM 2008, in Brussels (BE), 23-24 February 2008. This was a continuation of the Research Room, sponsored by the QualOSS, FLOSSMetrics, SQO-OSS, QualipSO, MANCOOSI and FLOSSMole. This year the topic consisted of exploring the idea of "research friendly" where academia and libre software developers could participate together to achieve common goals as to avoid the overload of servers.
- Some presentations in internal seminars and regional meetings took place.

Research community:

- WoPDaSD 2007, co-located with OSS 2007, in Limerick (IE), 11-14 June 2007. This was the second edition of the workshop (Workshop on Public Data about Software Development). The focus is to discuss about data set with information about libre software development.
- WoPDaSD 2008, co-located with OSS 2008, in Milan (IT), 7-10 June 2008. This third edition focused on the main platform existing, such as FLOSSMole or FLOSSMetrics.
- WoPDaSD 2009, co-located with OSS 2009, in Skövde (SE), 3-6 June 2009. This fourth edition focused on the same topic than old editions, but adding new point of view related to the retrieval process and storage of data.
- Workshop on Emerging Topics on Open Source Software Research (ETOSSR), organized at the ICSE 2007, in Minneapolis (US), 20-26 May 2007. This new workshop deals with specific research on libre software.

	<p>Collaboration plan and report</p> <p>Deliverable ID: D6.5b</p>	<p>Page : 12 of 13</p>
		<p>Version: 1.0</p> <p>Date: Dec 10, 09</p>
		<p>Status : Proposal</p> <p>Confid : Public</p>

- ETOSSR 2009, organized at the ICSE 2009, in Vancouver (CA), 16-24 May 2009. This was the second edition.

Industry:

- Linux Tag 2008.

In all these cases, projects promoting the joint events or presentations participated by helping in the dissemination of calls, submitting contributions, helping in the organization, and participating in panels and other joint sessions.

5.2 OTHER DISSEMINATION ACTIONS

In addition to workshops and conferences, other planned joint dissemination actions were:

- Federation of websites, so that events and news announced in the website of one of the participating projects will be visible also in others. This is the case of the FLOSSQuality initiative (<http://flossquality.eu>, which aims to bring together the experience of QualOSS, FLOSSMetrics and QualipSO).

6. SPECIFIC ACTIONS FOR COORDINATION


Projects agreed on the general direction proposed in this document. Next steps were done in order to have some specific coordination in this field:

- Flyers and sheets with information targeted at specific audiences, and with information from the relevant projects for those audiences, were produced.
- Projects created a joint web-based dissemination portal (<http://www.flossquality.eu>), aggregating news content from each of the project's primary websites.
- Paul Adams (Sirius Corporation) coordinated dissemination for the SQO-OSS project. As that project finished before most of the others, ZEA (acting as dissemination partner in QUALOSS and FLOSSMetrics) employed Paul in order to ensure the transfer of skills and knowledge between the projects.

7. FINAL REMARKS AND CONCLUSIONS

Coordination of 6th Framework Programme projects in the area of libre software development and quality was an interesting opportunity. The pool of expertise, and the aggregated amount of effort and initiatives was significant to have an impact on the worldwide landscape of the RTD in the area of libre software. . The coordination of activities, plans, and a joint dissemination schedule between all E.C. projects had as a result several papers published between different projects, for instance, FLOSSMetrics and QualOSS.

Some of the common partners among the different projects helped to have a common agent to coordinate all these activities. This basically helped to avoid duplicating efforts in some initial tasks.

	<p>Collaboration plan and report</p> <p>Deliverable ID: D6.5b</p>	<p>Page : 13 of 13</p> <hr/> <p>Version: 1.0 Date: Dec 10, 09</p> <hr/> <p>Status : Proposal Confid : Public</p>
---	---	--