

AN OPEN SOURCE COLLABORATION INFRASTRUCTURE FOR CALIBRE

David Nutter, Stephen Rank, and Cornelia Boldyreff

Abstract:

The study of Free and Open Source (Libre) software and the benefits provided by its processes and products to collaborative software development has been somewhat ad hoc. Each project wishing to use tools and techniques drawn from Libre software conducts its own research, thus duplicating effort, consequently there is a lack of established community practice on which new projects can draw. Long-standing intuitive theories of Libre development lack empirical validation. The long-term goal is to provide a resource to guide the evolution of Libre-software projects, from inception to maturity. The CALIBRE project is a co-ordination action aiming to address these issues through its research, its wider educational goals, and with an open invitation to the community to contribute. To succeed, the CALIBRE project needs an effective technological infrastructure which must support internal and external collaboration, communication and contribution to the project. The requirements of CALIBRE are similar to those of a Libre software project; this suggests that adopting a SourceForge-style environment which will be incrementally enhanced with further specialised tools as the requirements become better understood will be a sensible strategy.

1 Introduction

The importance of Free and Open Source (henceforth collectively referred to as 'Libre') software, its techniques and philosophy has been growing of late in areas as diverse as government information systems[12], public library systems[13], and education[2,3]. The European Union IST programme has encouraged projects such as GENESIS[9] and OPHELIA[8,10] to release their software under Libre licences, with mixed results.

Though the approaches taken by Libre developers are successful in many instances, the study of such techniques, much less attempts to categorise and disseminate the knowledge gained, has been confined to single interested researchers or projects with little recording of findings as a cohesive body of knowledge or expressed as well-founded theories. To address this deficiency CALIBRE [4]--*Coordination Action for LIBRE software engineering for open development platforms for software and services*--has been established under the European Community's Sixth Framework as a co-ordination action.

Libre software projects are necessarily distributed, consequently CALIBRE's investigation into the application of these techniques has import for the distributed software engineering field alongside assisting the development of Libre software. In particular we believe that distributed software engineering has much to learn from Libre software, for example user participation in development and release management.

2 CALIBRE Project

The CALIBRE project aims to study and improve interaction between the academic, industrial, and Libre communities; to develop an improved theoretical basis for Libre collaborative development; and to establish an online knowledge base for projects wishing to engage in Libre development[4]. It concentrates on knowledge dissemination and co-ordination of research related to Libre software

in order to achieve three goals. Firstly, integration of Libre Software research with practice, especially in the secondary/embedded software domain; secondly facilitating transfer of lessons learned from Libre software to everyday software engineering practice and finally establishing a European forum for Libre software research policy[4].

To realise these goals, it is necessary to study Libre projects, both successful and unsuccessful, as the characteristics that make projects suitable for the application of Libre development strategies are not yet well known. Existing research in this area has usually confined itself to studies of successful Libre projects which may have biased the findings somewhat, given that many Libre software projects are unsuccessful[7]. There are some notable exceptions, such as the FLOSS survey[11].

We believe that much interaction between large academic/industrial institutions and the Libre community is limited to single projects without a cohesive strategy behind them: strategies such as those employed by Sun and IBM are the exception not the rule. By a review of the available literature, CALIBRE intends first to establish the truth of this belief and, if true, to find out why. We wish to determine whether such collaborations can be better managed with higher-level managerial involvement or whether interference with current grass-roots support would be harmful. Using CALIBRE's industrial members, we initially intend to adopt a case study approach including both successful and unsuccessful projects to determine the collaborative procedures in use. The distributed nature of Libre software projects make empirical studies difficult; consequently exploratory case studies will provide an overview of the subject area and subsequent empirical studies will examine selected projects in more detail. These studies will require significant tool support for recording data and commentary.

These two studies naturally suggest a third: develop the theoretical basis of Libre collaboration for both pure (*ab initio* development) and hybrid (component re-use and evolutionary) Libre projects. In order to refine our understanding of Libre development, new theories based on data from the above empirical studies can be compared with current theories. Though eloquent and often insightful, many of the current theories of Libre development are based on intuition; establishing a sound empirical base for them will enhance the credibility of Libre development methodologies. Armed with improved theories and associated studies, the Libre community can begin to organise projects to best take advantage of Libre techniques.

Finally, this knowledge needs to be disseminated, primarily by establishing an online body of knowledge (BoK), through a series of workshops, and by developing educational materials suitable for training potential developers and managers who need to engage with the Libre communities.

2.1 Consortium Description

The CALIBRE consortium consists of eight academic and four industrial partners across nine countries. The action is coordinated by the University of Limerick and the remaining academic partners are the Universities of Maastricht/MERIT, Lincoln, Rey Juan Carlos, Catalunya, Cork, IMRI Paris IV, Poznan and Skovde. The industrial members of the consortium are the Business Innovation Centre of Alto Adige-Südtirol, Groupe des Ecoles Telecommunications, National Microelectronics Applications Ltd and Chinasoft. The geographical distribution of the partners suggests the use of CSCW systems to ensure effective collaboration across the partner sites.

3 Comparative Libre Research and Development

Between 2000 and 2003, GENESIS and OPHELIA, two European Commission Framework 5 projects with similar goals to produce Libre software engineering environments, ran in parallel.

GENESIS approached the problem by developing a tightly integrated workflow, resource and artefact management system while OPHELIA designed a tool integration framework and produced a prototype (ORPHEUS) integrating several common development tools. A comparative study[5] of the two projects was undertaken by researchers from both projects, detailing the difference between the development strategies, architectural design decisions, and consequently, the software components that each project produced.

One of these differences--not explicitly discussed in[5]-- was each project's attitude to collaboration in Open Source development. Though both utilised Libre software such as databases and configuration management tools both as components in their systems and as development aids, their release strategies were quite different. OPHELIA incrementally released their whole system via a public CVS repository and builds whereas GENESIS preferred to release finished configurations via SourceForge. Neither project directly developed using SourceForge.

Moreover, neither project received significant contributions from the Libre community despite publicity through academic channels and expressions of interest from external developers. There are several potential explanations for this: firstly the lack of project maturity, secondly the possibility that the requirements the projects addressed were too industrially oriented to be of interest to the wider Libre community and finally perhaps neither project did enough to foster wider participation, so promised contributions did not materialise.

Whilst CALIBRE is not aiming to address the first and second problems, it can potentially provide a solution to the third by determining the requirements for external communication support to ensure participation from the wider Libre community and requirements for internal support for communication and collaboration within the CALIBRE consortium.

4 CALIBRE Requirements

CALIBRE itself will function as an Libre organisation though, as a Coordination Action rather than an Integrated Project, it will not be directly engaged in development. Nevertheless, CALIBRE members will contribute their research and may be engaged in Libre development. External contributions and participation by the wider Libre community in CALIBRE is key to achieving its goals.

To accurately reflect practices in most Open Source projects and encourage contributions, CALIBRE has chosen technologies familiar to many Libre developers. Though many of the partners are familiar with "industrial-strength" tools and techniques, these are too complex to permit simple integration of new participants.

4.1 Communication and Collaboration Infrastructure

Mailing lists are the main form of communication within the consortium. Core management lists are private and where necessary public lists are created. Phone calls, private e-mails, face-to-face meetings and the like are also used but the intention is to keep much interaction on-line and therefore archived. Consequently, those interested in an aspect of CALIBRE merely have to read the relevant mailing lists and statements of intent to gain an overview of the discussions to date before contributing. Several concurrent lists will exist, each covering an aspect of the project's activities (theory development, organising workshops, etc) with some overlap between them.

CALIBRE will require a sophisticated, flexible and robust collaboration, communication and dissemination infrastructure to support its activities. To address these needs, a web portal will be created using Libre software components. The portal and its underlying infrastructure will be

engineered to address needs common to academic, commercial and public sector organisations and will be distributed under a Libre software license as part of CALIBRE's dissemination efforts. It will be evolved as a Libre project during the lifetime of the action. The portal will be an important focal point for access to CALIBRE and its research, providing access to three major sub-systems: the CALIBRE Work Environment, the CALIBRE Research Repository, and the CALIBRE Artefact Repository. Routine updates will ensure the portal provides accurate information including news, software and documentation about Libre software throughout the lifetime of the project, helping to transfer knowledge to the wider community.

The CALIBRE Work Environment (CWE) will consist of a distributed workspace, a virtual network collaboratory, comprising a number of workspaces for the research and software development projects within CALIBRE. Consequently researchers from all disciplinary backgrounds will be able to navigate between projects easily.

The CALIBRE Research Repository (CRR) will consist of a persistent, publicly available library of completed research artefacts and documents, while the CALIBRE Artefact Repository will contain material under active development by the project, both documents and software. A configuration management system, e.g. CVS or Subversion, will be used with a content management system to present the information.

Collaboration within the project will use the CWE system, probably based on one of derivatives of the popular SourceForge environment, for example, GForge. Documents will be prepared using Libre tools, such as OpenOffice, and managed using a Libre Content Management System. The use of a SourceForge-like environment would allow both easy separation of the subprojects for each work-package rather than one unwieldy resource tree and easy addition of any new resource under development as a separate ``project". As with mailing lists, those outside the project will be able to access the documents under development and thus contribute to the research. The choice of an environment and tools already familiar to members of the Libre community as the basis of the CALIBRE collaboration infrastructure will facilitate their participation and common tools across work packages will ensure cross-consortium working is made easier by allowing integration of tools such as an instant messenger system, wiki, and so forth.

There are various potential Libre software solutions for CALIBRE's needs, based on an initial evaluation of their claimed capabilities. These include, but are not limited to, content management systems such as Plone, Pathos, MaxDev, Mambo, eZ publish, and Xaraya. SourceForge-like systems provide development-specific capabilities, while content-management solutions are aimed at community-building and content creation. It will be necessary to take into account the specific requirements of a community of interacting researchers (and research groups) before making a final decision.

Explicit support for research activity is missing from these collaborative environments though support tools for tasks such as conference organisation and publications management are available elsewhere. Consequently, the consortium will have to decide which of these tools, if any, need integrating with the base development environment and what new tools need to be written. However, large-scale development must be avoided where possible, even at the expense of sacrificing desirable but inessential functionality.

4.2 Configuration Management

Since a SourceForge-type environment provides integrated configuration management, we initially propose to use that to manage all data including the website and other promotional materials. This provides traceability, gathers data for the CALIBRE study goals and allows external contributors to

provide precise patches to specific works rather than nebulous criticism. A CMS may then be introduced to meet requirements for more sophisticated content management.

Though version control is necessary as a baseline for collaboration, it does not support some research activities. However, attempting to introduce a complex, integrated system to perform both configuration and content management at the start of the project may be problematic. Consequently, a strategy of incremental adoption will be pursued; at first very simple tools will be deployed project-wide (email lists, version control) and other tools introduced when necessary.

4.3 Outreach and Body of Knowledge contributions

Outreach and dissemination are key activities within CALIBRE and both require tool support. Such support is separate from that required for mundane day-to-day communications. The highest profile item in this aspect of CALIBRE's work plan is a series of workshops which aim to bring practitioners and academics together to foster collaboration. The results of these workshops will be recorded and made available on-line to ensure the widest possible dissemination. Collaboration beyond the workshops is obviously necessary to build the body of knowledge and other deliverables; the workshops themselves merely provide a public face for the project and a forum for participants to discuss the ongoing activities.

Secondly, best practice guides for collaboration based on information derived from ongoing research will be developed. Additionally the project's raw data and sources for the technical paper series will be available to benefit future research. Finally, a searchable repository of all this material will be prepared, as the Libre community is now recognisable as a community of practice (CoP)[6], and this approach has been used successfully elsewhere to support CoPs.

Though the success of SourceForge and Savannah has provided Libre software projects with development resources, there seems little point in disseminating this success story further as they already have a high degree of recognition in the Libre software community. Instead, CALIBRE intends to promote effective ways of using available resources (including SourceForge and other environments) to achieve the "next level" beyond mere provision of support tools.

4.4 MasterLIBRE

Building on the outreach program, the MasterLIBRE European Masters Degree in Libre Software initiative has been proposed. If successful, this initiative will complement CALIBRE's wider activities and will run at the MasterLIBRE partners' universities, some of which (such as the Universitat Oberta de Catalunya) are already running courses or modules covering Libre Software Engineering. As well as training future software engineers in the philosophy and methodology of Libre activities, the MasterLIBRE partners will develop course materials used by each of the teaching partners and make them available for wider use. Materials, which are developed as Libre projects, include textbooks, case studies, such examination material that can be released prudently and guidelines for their use. This permits other educational institutions to either join the MasterLIBRE programme or derive their own programmes from the materials.

Obviously, material from the MasterLIBRE programme will be useful in the context of the BoK and vice-versa. However, learning materials have requirements beyond those of other CALIBRE dissemination activities, such as adherence to educational regulations and reflection of the existing organisational structure of higher degrees.

5 Contributing to CALIBRE

Though the CALIBRE consortium is large, it requires external contributions and candidates for study. Throughout the project, such participation will lead to incremental improvements to the Body of Knowledge. A number of ways in which people can participate in the CALIBRE project are outlined below.

5.1 Workshops

Firstly, the workshops will provide a forum to publish technical papers and meet others interested in Libre development. No other venue comprehensively addresses this need, so a wide cross-section of the technical spectrum are likely publish work of interest to the Libre community at these workshops. Moreover, the workshops will naturally host meetings of the CALIBRE project core, allowing attendees to learn of the latest research developments, give their feedback, and suggest future research directions.

5.2 Online Material

Once mature, the Online BoK will be a comprehensive guide to those planning to engage with the Libre community, hopefully removing some of the barriers for newcomers. From the authors' own experience, engaging with the Libre community is a difficult but exciting experience with potentially great benefits in software improvement, research activity, and adoption terms.

The areas covered by the BoK include techniques, tools and philosophy; attracting and retaining developers; managing culture clash; case studies of successful strategies that may be adapted to specific project needs and characteristic signs of failure. For educators, the BoK and associated course material should help to reduce the difficulty of integrating this knowledge into degree-level and other courses.

5.3 Libre Software and the EU

Much as the IST Results web site^[1] covers the wider EU Information Society Technologies programme, CALIBRE aims to foster collaboration between complementary Libre projects just as GENESIS and OPHELIA attempted to do. To this end, the CALIBRE site will provide a resource for finding collaborations and via the workshops and other publication media a forum for disseminating results. As a complement to IST Results for Libre projects, CALIBRE will attempt to meet the needs of the Libre software research community with necessarily incremental results.

6 Conclusion

Some aspects of the CALIBRE project have been presented in an attempt to identify some of its novel collaboration and communication requirements. Many of these are similar to those found in Libre development projects. CALIBRE's combination of research and dissemination activities ensures that it is a wonderful opportunity for practitioners wishing to contribute their experience or gain knowledge about Libre research and development and to produce a living contribution to Software Engineering knowledge. In the future, the proceedings from the first CALIBRE workshop (planned for November 2004) will be made available through the CALIBRE web portal. CALIBRE's focus on identifying and documenting successful Libre software techniques will provide a useful contribution to the field of distributed software engineering by making techniques that are currently poorly understood accessible to practitioners.

Bibliography

- [1] IST results - promoting innovation for the information society.
<http://istresults.cordis.lu/index.cfm>.
- [2] MIT open courseware. <http://ocw.mit.edu/OcwWeb/Global/AboutOCW/about-ocw.htm>.
- [3] Schoolforge: Working together to unleash the power of Open Source tools in education.
Web. <http://www.schoolforge.net/what.php>.
- [4] Coordination Action for Libre Software Engineering for Open Development Platforms for Software and Services -- Annex 1, May 2004.
- [5] BOLDYREFF, C., SMITH, M., WEISS, D., NUTTER, D., WILCOX, P., RANK, S., DEWAR, R., RITROVATO, P., Environments to Support Collaborative Software Engineering. In Proc. of the 2nd Workshop of Cooperative Supports for Distributed Software Engineering Processes, pages 23-29. IEEE, 2003.
- [6] CANFORA, G., CIMITILE, A., VISAGGIO, C.A.,
Dynamic software communities of practice through three-layered competence model.
In Cimitile, A., Lucia, A., and Gall, H., editors, Cooperative Methods and Tools for Distributed Software Processes, pages 135-149, Milan, Italy, 2003. FrancoAngeli.
- [7] CAPILUPPI, A., LAGO, P., MORISIO, M., Evidences in the evolution of OS projects through changelog analyses. In Feller, J., Fitzgerald, B., Hissam, S., and Lakhani, K., editors, Proceedings of the 3rd Workshop on Open Source Engineering, ICSE 2003, pages 19-24, Portland, Oregon, May 2003.
- [8] DEWAR, R., MACKINNON, L., SMITH, M., WILCOX, P., SMITH, R., POOLEY, L., The OPHELIA project: Supporting software development in a distributed environment. In Proceedings of WWW/Internet 2002. IADIS, 2002.
- [9] GAETA, M., RITROVATO, P., Generalised Environment for Process Management in Cooperative Software Engineering. In International Computer Software and Applications Conference, volume 26, pages 1049-1059, Oxford, England, August 2002. IEEE.
- [10] HAPKE, M., JASKIEWICZ, A., PERANI, S., OPHELIA - Open Platform and methodologies for development tools Integration in a distributed environment.
In Proceedings of 3rd National Conference on Software Engineering, pages 189-198, Otwock/Warsaw, 2001.
- [11] International Institute of Infonomics, University of Maastricht and Berlecon Research GmbH.
Free/Libre and Open Source Software: Survey and study. <http://www.infonomics.nl/FLOSS/report/>, June 2002.
- [12] SCHMITZ, P.E., Study into the use of open source software in the public sector. part 2: Use of open source in Europe. <http://europa.eu.int/ISPO/ida/export/files/en/1973.pdf>.
- [13] STEWART, G., One man's crusade to push open source software.
article in The Guardian.