2009 Fourth IEEE International Conference on Global Software Engineering

Risks and Safeguards for the Requirements Engineering Process in Global Software Development

Alejandro López, Joaquín Nicolás, Ambrosio Toval
Software Engineering Research Group
Universidad de Murcia
Campus de Espinardo 30071 Murcia, Spain
Email: {alexlopez,jnr,atoval}@um.es

Abstract

The rise of new development paradigms such as Global Software Development (GSD) forces Requirements Engineering (RE) to face up to new challenges and risks not common in traditional development models. When an organization first embarks upon a GSD project it exposes itself to plenty of risks. Many of these risks appear as a result of the lack of experience of the development team on GSD projects. A Systematic Literature Review (SLR) has led to the compilation of a repository which gathers the risks that concern RE when developed in a distributed software development environment, as well as a set of safeguards, which help overcoming such risks. This repository, open to progressive enrichment, provides a novel basis to assist in RE activities for GSD projects. It could be especially useful when the developers have little experience in distributed projects and cannot prevent communication, distance, cultural, or organizational risks. Although the repository collects the main contributions from published work in the field, it has not been validated on real GSD projects yet.

1. Introduction

It is undeniable that the globalization process, which has affected every industry, has also reached Software Engineering (SE) in the form of Global Software Development (GSD) [6]. One of the most relevant features of this paradigm is the dispersion of the stakeholders across different countries, continents and time zones. Cheng and Atlee [3] recently identified globalization as a hot research topic regarding Requirements Engineering (RE). The rise of this trend is based on the advantages that GSD brings in comparison to the traditional, non-distributed software development process: low cost of work forces in certain countries, 24 hour-workday (*around-the-clock* work model), access to a global pool of highly-qualified resources, and proximity to the client. This new paradigm implies that many of the RE activities have to undergo changes, as it is no longer possible to carry them out as if the participants were co-located. Furthermore, the novelty of the paradigm increases the risk of project failure, since there is a lack of experience in managing such projects. A key issue to achieve success in GSD is overcoming the pitfalls which GSD poses to RE.

The goal of this paper is to present a novel proposal of a risks and safeguards repository, which compiles a collection of good practices to help minimizing a set of risks regarding different aspects which may provoke the failure of the project, among them it is worth mentioning knowledge management and awareness issues. Concerning our purposes, *awareness* can be defined as the mechanism that allows developers to know who is working on certain requirements, or who has certain skills. The knowledge used to fulfill this purpose was mainly extracted from literature through the execution of a Systematic Literature Review (SLR) [11]. Depicting the full process of the SLR is out of the scope of this paper. The true purpose of this writing is not to show just a SLR, but a repository which would help practitioners on their GSD projects, and would continually evolve to show not only the state of research, but also proposed safeguards to be validated. The SLR is only the means to achieve the results here presented; it is the basis of the repository.

The rest of this paper is structured as follows: Section 2 reviews the methodology which was used to obtain the risks and safeguards repository. Section 3 shows the structure of the risks and safeguards repository along with an excerpt from it and a brief analysis. Finally Section 4 presents the conclusions and further work.

2. Research Methodology: A Systematic Literature Review

A SLR is a rigorous literature reviewing technique which has been used in this research to pursue two goals: (1) to find out the risks and challenges identified in the application

Partially financed by the CICYT (Science and Technology Joint Committee), Spanish Ministry of Science and Technology (DEDALO project, TIN2006-15175-C05-03) and the Science and Technology regional ministry of the Junta de Comunidades de Castilla-La Mancha (MELISA project, PAC08-0142-335).

978-0-7695-3710-8/09 $25.00 © 2009 IEEE
DOI 10.1109/ICGSE.2009.62

of RE on GSD projects, which were not common in the traditional non-distributed development process, and (2) to find out the solutions proposed in the literature to solve such risks and challenges.

Following the SLR process, two questions have been formulated: $\mu 0$: *Which new risks and challenges are identified in the literature about the application of RE on GSD environments?*, and $\mu 1$: *Which solutions have been proposed in the literature to face the risks identified from question $\mu 0$?*

The following search string has been formulated to address both $\mu 0$ and $\mu 1$: *Global Software Development AND Requirements Engineering*. This string has been used in the search engines of some of the most important scientific publications, namely: IEEE Digital Library, ACM Digital Library, MetaPress (Kluwer + Springer), Science@Direct, Wiley InterScience, and DBLP. Moreover Google Scholar has been taken into account, as it could be an important source of grey literature.

The inclusion criteria which have been applied to mark the found papers as candidates are the following: only studies regarding the risks and challenges which are posed by the application of RE on GSD projects, or studies which propose solutions to these risks and challenges are included. These studies should directly address the relationship between RE and GSD, not just tangentially mention it, and should not deal with GSD from an educational point of view, i.e. present teaching methodologies for GSD. From this set of candidate studies, those duplicated in different search engines, published before year 2000 or have less than 4 pages are excluded.

During the SLR execution process, an initial set of 564 documents was found by applying the search strings on the search engines. Then, the inclusion criteria defined above were applied to get a subset of 175 candidate articles. Then, exclusion criteria were applied to get the final subset of 36 selected documents. In order to do this, the abstract of the articles, and, in some cases, selected sections of them, were read. Table 1 and Fig. 1 show the set of studies broken down by sources (identical studies found in several sources have not been removed). The search engines without selected results are not displayed. In both Table 1 and Fig. 1, it can be seen that Google Scholar is the search engine which found more documents. This can be explained not only because it returned papers published in other sources than IEEE, ACM or MetaPress, but also because it found other kind of documents, namely book chapters, technical reports, theses, and articles located in their authors' personal websites. Therefore, we believe that for further SLR, Google Scholar should be included in the list of search engines, not to replace the other digital libraries, but to complement them.

The SLR protocol has been designed and executed by the first author of this paper as part of his final project, and his advisor (the second author of this paper) has revised the protocol, the included and excluded papers, and has discussed the results of the review with him. From the whole set of 36 articles, 16 of them come from the industrial area. These are either experience papers or have used real GSD projects to validate the methods they propose (e.g. [2], [7]).

Table 1: Number of Results per Source

Source	Total found	Candidates	Selected
IEEE	100	61	6
ACM	83	9	7
MetaPress	54	29	1
Google Scholar	327	76	22
Total	564	175	36

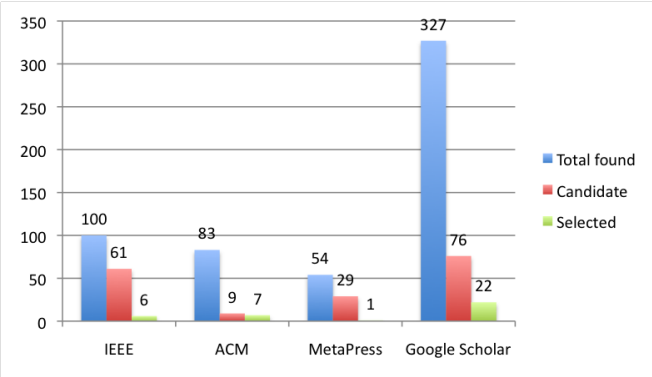

Figure 1: Number of Results per Source

After being thoroughly read, the articles have been classified according to their contribution into the following disjoint categories:

- *Only identify risks*: Studies in which risks in the application of RE to GSD environments are identified but no solution is proposed. Some of them are based on real GSD experiences or controlled experiments.
- *Make theoretical proposals*: Studies in which proposals with no direct application are made. They are based on different theories, but no guide to put them in practice is given.
- *Propose methods*: Studies in which methods to overcome certain risks are proposed. Some of them are backed by controlled experiments or real-world cases.
- *Propose tools*: Studies in which tools that help alleviate some problems or risks are presented. Some of them have been tested in controlled or real world environments.
- *Offer good practices*: Studies in which good practices to successfully face GSD projects are offered. Most of them are experience papers about real GSD projects.
- *Do not fall in any of the above categories*: Studies whose content cannot be classified in any of the above categories.

Figure 2 shows the set of selected documents broken down into the different types of contributions.

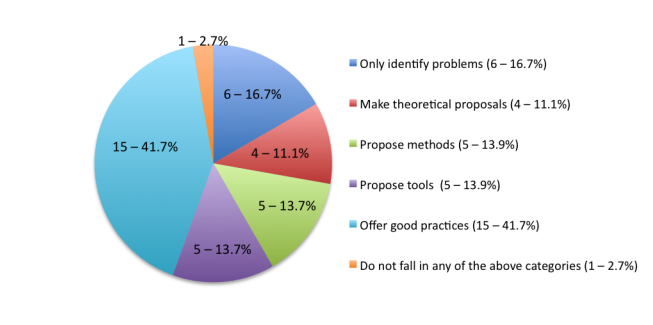

Figure 2: Classification of Studies According to their Contribution

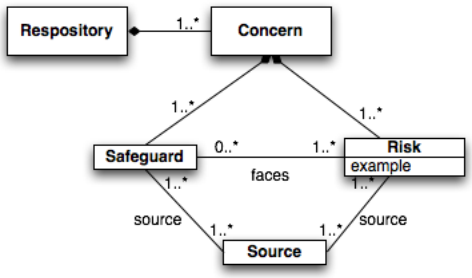

Figure 3: UML Class Diagram Showing the Structure of the Repository

3. A Repository of Risks and Safeguards for the Application of RE on GSD Projects

In this section the risk and safeguards repository is reviewed. First, the structure of the repository is explained and some examples of risks and safeguards are given in Section 3.1. Then the risks and safeguards pertaining to one of the concerns, *Knowledge Management and Awareness* are shown in Section 3.2. Finally some significant issues regarding the repository are addressed in Section 3.3.

3.1. Structure of the Repository

In order to better understand the structure of the repository the terms *risk* and *safeguard* should be now presented. In this context we define a *risk* as an aspect of the application of RE in a GSD environment which could endanger the RE process if overlooked and therefore it should be taken into account. A *safeguard* is defined as a good practice, some kind of advice which helps alleviate the corresponding risk(s).

Based on the SLR, a taxonomy composed by seven different types of concerns, i.e. areas of interest, was compiled. Each of these classes of concerns contains a set of risks and their corresponding safeguards regarding the specific aspect of GSD which motivates the concern. These classes are not fully disjointed, but their boundaries are fuzzy, and in some cases a risk or safeguard may fit in more than one of them. Each of the classes of this taxonomy is briefly explained next:

- *Comunication and distance*: Risks derived from the dispersion of the stakeholders across countries and time zones.
- *Knowledge management and awareness*: Risks derived from the difficulties of keeping awareness, cohesion, and coherence of knowledge when different working groups try to access it concurrently.
- *Cultural differences*: Risks derived from the interaction among groups where people have cultural backgrounds which vary greatly.
- *Management and project coordination*: Risks derived from the establishment of organizational structures, role definition, and coordination procedures.
- *Tools which support the processes*: Risks derived from the lack of suitable tools which fully support the RE process.
- *Clients*: Risks derived from the interaction with distributed clients.
- *Miscellany*: Risks not related to any of the categories described above.

Both risks and safeguards have an attribute named *Source*, which identifies the source document in the SLR from where they were elicitated (in case they do not have their origin in a document, i.e. they are our own proposal, labelled as *Proposed*). Each safeguard has an additional attribute, *Faces*, which refers to the risk(s) it faces, and it is used to keep traceability between risks and safeguards. Some risks also have an attribute *Example* when further explanation would be useful to understand the problem posed by the risk. Moreover, risks and safeguards are hierarchically structured: the most general ones (parents) are refined by other more specific ones (children). The practitioner should notice that the safeguards labelled as *Proposed* are not validated and, although they make sense to the authors of this work, cannot be taken as universal truths. If the practitioner decides to follow their advice and finds out that they actually help alleviate their associated risks, they will no longer be proposed safeguards but validated ones. This structure is shown in Fig. 3 by means of a UML class diagram.

To illustrate the kind of risks and safeguards covered by the previously defined categories, an example of some of them is provided:

- *Cultural differences*
  - *Risk*: **R19.3** In GSD, barriers to work ethics may appear: some groups find difficulties in expressing their dissatisfaction to clients, and key features and look-and-feel issues can be equally treated without taking into account priorities.
    - Source: Illes-Seifert *et al.* [10]
  - *Safeguard*: **S12.1** Hofstede's national and organi-

zational dimensions shall be studied in order to compare cultures.
    - Faces: R19, R19.2, R19.3, R19.4
    - Source: Aranda *et al.* [1]
- *Management and project coordination*
  - *Risk*: **R25.2** In GSD, the lack of a centralized, skilled, and firm analyst role might seriously affect the project development.
    - Source: Berenbach [2]
  - *Safeguard*: **S22** It shall be established a well-defined agile leadership hierarchy, with an analyst playing the role of coordinator. Their responsibilities shall be well defined and shall play their role in a skilled and firm way. Below them in the hierarchy shall be team heads, which shall be responsible for the analysts belonging their work group.
    - Faces: R25, R25.1, R25.2, R26, R29, R29.1, R30, R33, R35
    - Source: Berenbach [2]
- *Clients*
  - *Risk*: **R49** In GSD, it might be difficult to indentify key users if analysts do not have access to the majority of personnel in their work place, and thus, cannot see the tasks they carry out.
    - Source: Hanisch and Corbitt [7]
  - *Safeguard*: **S29** Before the beginning of the project, client and vendor shall sign a set of conditions which grant the involvement of the client in the product development, in particular in the requirements analysis process.
    - Faces: R41, R41.1, R41.2, R41.3, R43.1, R45, R46, R46.1, R47, R48, R49, R49.1, R49.2
    - Source: Proposed
- *Miscellany*
  - *Risk*: **R58** In GSD, developers often realize that the rationale of requirements is not clear, and they do not know why the requirements were formulated as they were, so it is quite common to misunderstand them, causing delays and needing additional explanation.
    - Source: Heindl and Biffl [8]
  - *Safeguard*: **S32** An attribute shall be included within the requirement to explain its rationale, and to justify the way they were formulated.
    - Faces: R58
    - Source: Heindl and Biffl [8]

In the previously shown set of risks and safeguards which exemplify the ones contained in the repository, some roles which have important functions can be found. This set of roles played by the stakeholders is defined as follows:

- *Coordinator*, who coordinates the work of each participant in the project.
- *Moderator*, who chairs requirements negotiation meetings.
- *Team Head*, who represents a work group in requirements negotiation meetings and speaks for it with the Coordinator and other Team Heads.
- *Analyst*, who develops similar work as any analyst taking part on a non-distributed software development project.
- *Key User*, who knows the whole system and contributes with essential knowledge to build the requirements documents.
- *User*, who knows part of the system and contributes with partial knowledge to build the requirements documents.

3.2. Risks and Safeguards: an Excerpt

The repository is too large to be fully included in this paper, and thus only a few examples were presented in Section 3.1. In this section the set of risks and safeguards associated with the concern *Knowledge Management and Awareness* will be further detailed. The full version of the repository (see [12]) includes the whole set of risks and safeguards.

The following risks have been identified regarding *Knowledge Management and Awareness*:

R14 In GSD, work groups do not get the benefits of social mechanisms and natural processes present in co-located workspaces, which reduces the need to define explicit support for workspace awareness.
Source: Damian et al. [5]

R14.1 In GSD, geographically distributed work teams cannot exchange information through social and informal mechanisms.
Source: Damian et al. [5]

R14.2 In GSD, awareness is not maintained.
Source: Damian et al. [5]

R15 In GSD, the lack of trust and inability to share knowledge makes it difficult to reach a common methodology, which fosters the exchange of critical information during requirements management.
Source: Damian [4]

R16 In GSD, changes on requirements are normally reported with delay and then to others than relevant stakeholders.
Source: Heindl et al. [9]

The set of safeguards related to the above-described risks are:

S8 The system shall keep information about analysts and work groups working on each requirement through attributes stored along with them.
Faces: R14, R16
Source: Proposed

Table 2: Traceability matrix

	S8	S9	S11
R14	X	X	
R14.1			
R14.2			X
R15			X
R16	X	X	

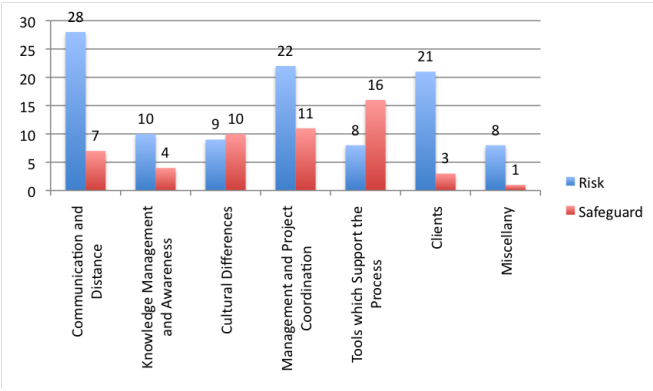

Figure 4: Structure of the repository

S9 Each change on a requirement shall be notified to the analysts who belong to the work groups responsible for it. This association shall be stored in an attribute along with the requirement.
Faces: R14, R16
Source: Proposed

S11 Requirements shall use attributes to store discussion threads and votes.
Faces: R14.2, RT15
Source: Proposed

The traceability matrix that relates the risks and safeguards shown above is displayed in Table 2.

3.3. Analysis

The repository is composed of 106 risks and 52 safeguards, divided into seven different classes of concerns as depicted in Fig. 4. Two group of concerns can be identified based on the number of risks they contain: (1) The first group would include the concerns *Communication and Distance*, *Management and Project Coordination*, and *Clients*. These are the most prone to provoke the GSD project to fail, as the greatest part of the risks belongs to these categories. (2) The rest of concerns, i.e. *Knowledge Management and Awareness*, *Cultural Differences*, *Tools which Support the Process*, and *Miscellany*, compose a second group which seems to be less prone to risks. We think of two explanations for this difference in the population of the risks in the two groups:

The first hypothesis considered is that (1) is more prone to risks because of the nature of GSD. The areas composing this group are critically affected if its related tasks are performed in a distributed way or even in different time zones. This seems strange, however, as concerns of (2) are very important, therefore this first hypothesis was left aside, and another one was formulated. Risks of (1) are more obvious and arise easily when GSD projects are undertaken. Because risks regarding (2) are somehow more subtle and not so easily found, further research effort should be done in order to find more hidden risks regarding areas included in (2). Intuitively we believe this second explanation is more convincing.

Table 3: Uncovered Risks and Proposed Safeguards

Concern	Risks	Uncovered	Safeguards	Proposed
Communication	28	3	7	2
Management	22	0	11	7
Clients	21	3	3	3
Knowledge Management	10	2	4	4
Cultural	9	1	10	6
Tools	8	0	16	3
Miscellany	8	7	1	0

In Table 3 (names of concerns have been abridged to improve readability) the safeguards are reviewed in numbers. For each category not only the number of risks and safeguards is shown, but also the amount of uncovered risks, i.e. risks which do not have associated safeguards, and safeguards which were not extracted from the literature, but proposed by the authors of this paper. Although the areas of *Management and Project Coordination* and *Clients* are relatively well studied, we could not find a significant number of safeguards to the many risks detected in these areas, but most of the safeguards are our own proposals (64% and 100% respectively). Not only these, but every proposed safeguard should be validated with further research, ideally in industrial cases, as they are 48% of the total amount of proposed safeguards. The *Miscellany* group is too heterogeneous. If new risks belonging to this group are identified as more GSD projects are undertaken, it could be refined into more fine-grained sub-groups, and then maybe new safeguards could be proposed to address the concrete concerns. In the *Knowledge Management and Awareness* area more research should be done, as 100% of its safeguards are proposed. Regarding this group it might be of interest to explore the possibility of using ontologies as it was pointed out in [1]. In that paper they propose ontologies as a common frame to define and formally structure the domain concepts, as they are easily extensible and non-dependant of languages. Further research in this direction might prove many of the other features of ontologies useful to the issue of RE application in GSD projects. Although there is a significant number of authors which have published articles identifying risks or proposing safeguards, some of the most active researchers (or those who have made more

contributions by means of their papers) are the SEGAL group of Daniela Damian, Timea Illes-Seifert, Jyoti M. Bhat and colleagues, Brian Berenbach, and Matthias Heindl and colleagues. The full list of contributors can be found in [12].

4. Conclusions and Further Work

Many partial solutions to the challenges posed by the application of RE in a GSD environment have been proposed, but none is complete enough to cover the whole RE process or every aspect of it. Hence the contribution of this work is a public updated repository, structured in the form of risks and safeguards, which is intended to help requirements engineers with little experience on GSD projects. This repository will make them aware of risks that they might otherwise overlook, and provide them with safeguards that have already been validated in several real GSD projects. Furthermore it will help other more experienced requirement engineers who want to adapt their current RE methods to make them applicable to the emerging GSD paradigm.

Developing the repository presented in this paper through of a flexible means such as a Wiki, together with the fact that it has been structured in a clear, modular manner, with different sections which group similar risks/safeguards, would help make it extensible, scalable, and would foster collaborative work, as it could be modified by analysts no matter where they are located. Moreover the Wiki would also permit taking the repository a step closer to the Web 2.0 world. Changing the taxonomy here presented in a more flexible tagging system, a *Folksonomy*, would allow risks and safeguards to have their own category depending on the tags the practitioners give them.

The repository also has to be validated by GSD experts and tested on a real GSD project. As it is applied in different projects, new risks will probably arise, and new safeguards will have to be consequently created. Practitioners will be, therefore, engaged on a continuous repository-improving process as they will validate the proposed safeguards and propose some new ones to face some risks. An RE method, based on a previous proposal of the authors ([13], [14]), is being developed to implement the application of the safeguards in the repository.

References

[1] G. Aranda, A. Vizcaíno, A. Cechich, M. Piattini, and J. Soto. Una metodología para elicitación de requisitos en proyectos gsd. *XII Jornadas de la Ingeniería del Software y bases de datos (JISBD'2007)*, 2007.

[2] B. Berenbach. Impact of organizational structure on distributed requirements engineering processes: Lessons learned. *International Workshop on Global Software Development 2006. GSD'06*, 2006.

[3] B. H. Cheng and J. M. Atlee. Research directions in requirements engineering. In *Future of Software Engineering, 2007. FOSE '07*, pages 285–303, 2007.

[4] D. Damian. Stakeholders in global requirements engineering: Lessons learned from practice. *IEEE Software*, 24(2):21–27, 2007.

[5] D. Damian, J. Chisan, P. Allen, and B. Corrie. Awareness meets requirements management: Awareness needs in global software development. In *International Workshop on Global Software Development at the 25th International Conference on Software Engineering*, pages 7–12, May 2003.

[6] K. Fryer and M. Goethe. Global software development and delivery: Trends and challenges. http://www.ibm.com/developerworks/rational/library/edge/08/jan08/fryer_gothe/index.html, 2008. Last accessed on March 13, 2009.

[7] J. Hanisch and B. Corbitt. Requirements engineering during global software development: Some impediments to the requirements engineering process – a case study. In *European Conference on Information Systems '04*, 2004.

[8] M. Heindl and S. Biffl. Risk management with enhanced tracing of requirement rationale in highly distributed projects. *International Workshop on Global Software Development 2006. GSD'06*, 2006.

[9] M. Heindl, F. Reinisch, and S. Biffl. Requirements management infrastructures in global software development - towards application lifecycle management with role-based in-time notification. In *Proceedings for International Workshop on Tool Support and Requirements Management in Distributed Project (REMIDI 2007)*, pages 43–48, 2007.

[10] T. Illes-Seifert, A. Herrmann, M. Geisser, and T. Hildenbrand. The challenges of distributed software engineering and requirements engineering: Results of an online survey. In *Proceedings for 1st International Global Requirements Engineering Workshop (GREW'07)*, pages 55–66, 2007.

[11] B. Kitchenham, O. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman. Systematic literature reviews in software engineering – a systematic literature review. *Information and Software Technology*, 51(1):7–15, 2007.

[12] A. López, J. Nicolás, and A. Toval. A repository of risks and safeguards for the requirements engineering process in global software development. http://www.um.es/giisw/GSD/repository.html, December 2008. Last accessed on April 9, 2009.

[13] A. Toval, B. Moros, J. Nicolás, and J. Lasheras. Eight key issues for an effective reuse-based requirements process. *International Journal of Computer Science and Engineering*, 23(6), November 2008.

[14] A. Toval, J. Nicolás, B. Moros, and F. García. Requirements reuse for improving information systems security: A practitioner's approach. *Requirements Engineering*, 6(4):205–219, 2002.