

# Conducting Qualitative Research in an International and Distributed Research Team: Challenges and Lessons Learned

Rafael Prikladnicki  
PUCRS, Brazil  
rafaelp@pucrs.br

Roberto Evaristo  
3M Center, U.S.  
jroberto.evaristo@gmail.com

Daniela Damian  
UVIC, Canada  
danielad@cs.uvic.ca

Jorge Luis N. Audy  
PUCRS, Brazil  
audy@pucrs.br

## Abstract

*In this paper, we discuss challenges for planning and executing qualitative research conducted by an international research project team. The study comprised an exploratory examination of strategies of offshoring and onshoring for software development. An important methodological challenge is that the members of the research team live in different countries, rely on different languages and originate from different cultures. These challenges are in many ways analogous to those inherent in the subject we are researching, distributed software development. To explore these issues, we present the difficulties we faced on collecting and analyzing the qualitative data. Our main contribution is the identification of challenges, strategies to overcome them, and a set of lessons learned.*

## 1. Introduction

More than a decade ago, many organizations began to experiment with Distributed Software Development (DSD), seeking reduction in costs, exploiting a broader skills base, and being closer to emerging markets [8]. For these reasons, DSD is a growing focus of attention in universities and organizations worldwide, and has attracted a large research effort in software engineering (SE) [2, 5, 6, 8, 13, 16, 23, 36, 39, 43]. Very much like in a DSD scenario, we found that sometimes a research team with interest in distribution is also distributed.

It has been common to find distributed research groups being formed exactly like software development project teams have been formed over the past ten years. A distributed and multicultural research team can collaborate in many ways, including data collection and analysis, for example, aiming at a better explanation of problems that most of the times can not be understood if a DSD research is conducted locally. On the other hand, the existence of such distributed team can create additional challenges for the research being conducted.

In this context, the main purpose of this paper is to discuss several challenges identified during the planning phase of an exploratory study conducted by an international and distributed research team. This study was part of a program of research, where the main

purpose is to develop a maturity model for wholly owned offshore subsidiaries for internal software development (internal offshoring).

As part of our research design, we engaged in a qualitative exploratory investigation initially designed to examine the practices of several companies involved in different DSD business models, offshoring and onshoring, and document their differences. This phase of the study involved a distributed research team located in Brazil, the U.S., and Canada. We found that the existence of a multicultural research team created methodological challenges to the qualitative research: different stakeholders collected and analyzed data, and results had to be discussed and validated with the other distributed researchers.

In this paper, we discuss the challenges identified, comparing them to what we faced in the execution phase. We also discuss the strategies developed to address such difficulties and lessons we learned. We will focus on how our experience can contribute to minimize the methodological issues in projects where researchers are geographically distributed, and therefore reliant in different cultures and languages.

To familiarize the reader with the subject of our program of research, in the next section we briefly present the concepts involved in onshoring and offshoring, and maturity models for distributed software development. In section 3 we present a discussion on qualitative research methods in SE and DSD. In section 4 we present the research design, the challenges identified in the planning phase, and the strategies develop to overcome them. In section 5 we discuss what ultimately happened during the execution of our study, as well as the main lessons learned. Section 6 concludes the paper.

## 2. Onshoring and offshoring

When organizations explore distributed software development, it is important to characterize DSD along several dimensions: organizational relationship structure, geographical distance, and other attributes such as temporal distance, stakeholders' location, etc.

When the relationship structure is analyzed along with the geographic location, for example, DSD can be characterized by business models [44]. Based on these

two dimensions, Prikladnicki et al [38] defined four of the most common business models for DSD, as follows:

- Internal domestic supply or onshore insourcing: there is a department in the company premises (or subsidiary) in the same country (onshore) that provides services throughout internal projects (insourcing);
- Onshore outsourcing or outsourcing: this model indicates that an external company (subcontract) is responsible for providing software development services or products for the client company (outsourcing). To characterize the onshore scenario, the subcontracting company is located in the client’s country (onshore).
- Offshore outsourcing: involves a relationship with an external company (outsourcing) not located in the client’s country (offshore).
- Offshore insourcing or captive / internal offshoring: a company creates its own software development center (subsidiary) to supply the internal demand (insourcing). This subsidiary is located in a different country than the company’s headquarters.

### 2.1. Prior studies on maturity models for DSD

To date, we have few studies that deal with a maturity perspective in distributed software development. One of the first studies was conducted by Carmel & Agarwal [7], where offshore IT sourcing was identified as a maturation process. Morstead & Blout [29] proposed a model for offshore organizations, called OMM (“Offsourcing” Maturity Model). According to the authors, the model is limited because it doesn’t have details of process, templates and history that drive specific task-oriented actions on a day-by-day level.

Hyder et al [18, 19] developed the eSourcing Capability Model for Service Providers (eSCM-SP). This model is a “best practices” capability model with the purpose to give service providers guidance that will help them to improve their capability across the sourcing life-cycle. The model was developed by a consortium led by CMU (Carnegie Mellon University), and the authors argue that the existing frameworks do not address all of the critical issues in eSourcing (IT-enabled Sourcing).

More recently, two other studies also pointed out the importance of studying the maturity evolution of DSD projects. The study conducted by Ramasubbu et al [43] identified 24 new key process areas for managing global projects. In the other study, conducted at IBM by Sengupta et al [46], the authors identified four areas where important research questions need to be addressed to make DSD more effective. One of the areas is related to the identification of key process areas to improve the current capability maturity models for software development. And finally, Meyer [27] also emphasizes that the area of process models could have significant improvements focusing the DSD environment.

### 3. Qualitative research methods: Motivation and Challenges

One of the major reasons for engaging in qualitative research is to more deeply understand a given phenomenon. Hence, we decided to investigate the maturity of the relationship that forms over time between an organization’s headquarters and its subsidiary, where internal offshoring approaches are used for software development. This investigation was part of a program of research (original study), which is outlined in detail in Evaristo et al [14].

The research methodology on the original study included first the development of a case study, represented in dark in figure 1. One of the main research questions for this case study was: “What are the main differences among DSD business models, and in particular in the internal offshoring when it is employed by way of a wholly-owned subsidiary?” Content analysis of the interview transcripts were planned to inform the second phase of the program of research, that is, to test the theory by developing a large scale web survey based on results of the exploratory phase.

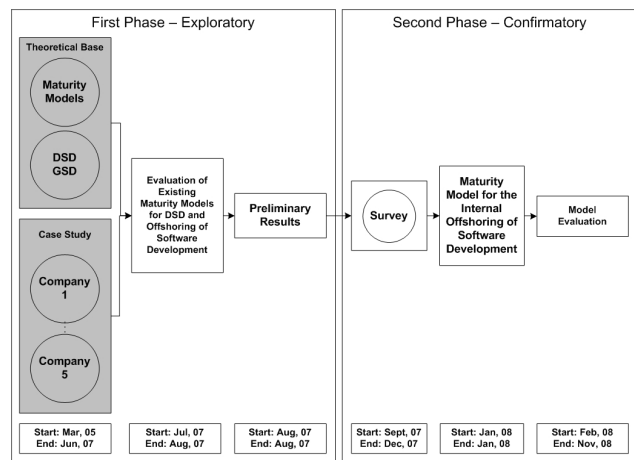


Figure 1. Research design of the original study

During the planning phase of the qualitative study, the geographical distribution of the research team was itself creating communication and cultural differences that needed to be taken into consideration during the execution phase. Geographical distance, disparity of language and differences in cultural origin increase the difficulty of collecting, analyzing, and achieving reliability and validity based on qualitative data. Interestingly, this is not well documented in the literature. Therefore, in parallel with the planning efforts, we identified the challenges that we were facing, and developed ways to overcome them. In the next section, we present a brief discussion of the importance of using qualitative methods in software engineering research, and specifically in this research.

### 3.1. Qualitative methods in SE and DSD

Hirschheim [17] points out that all research (whether quantitative or qualitative) is based on underlying assumptions about what constitutes “valid” research and which research methods are appropriate. In order to conduct qualitative research, it is therefore important, according to the author, to know what these assumptions are. The motivation for qualitative research according to Creswell [11], is to collect open-ended, emerging data in order to develop themes from the data. Qualitative methods can be used to explore substantive areas about which little is known or about which much is known to gain novel understandings (Strauss & Corbin [48]).

Qualitative research methods involve interviews, documents, and participant observation data, to understand and explain social and cultural phenomena [30]. The goal of understanding a phenomenon from the point of view of the participants is lost when textual data are quantified [22]. For this reason, Taylor & Bogdan [50] point out that qualitative research methods were designed mostly by educational researchers and other social scientists to study the complexities of human behavior (e.g., communication, difficulties in understanding). According to the author, much of human behavior cannot be adequately described and explained through statistics and other quantitative methods.

Examples of qualitative methods are action research, case study research, ethnography, and grounded theory. In Information Systems, according to Paré [34], there is a growing tradition of using qualitative research approaches, and case studies have had increasing acceptance over the past decades.

In Software Engineering, Seaman [45] notes that only recently empirical studies have achieved significant recognition in the software engineering research community. She also notes that only recently did empirical studies begin to address the human role in software development. And part of the reason for this interest among researchers comes from practitioners [45]. Industry players recognize that software development presents a number of unique management and organizational issues that need to be addressed and solved, leading to studies related not only to software engineering technical issues, but also to contextual and organizational issues, as well as to the intersection between the technical and contextual.

Another problem, according to Caver et al [9] is that complex software engineering issues can be difficult to study using a purely quantitative approach. The reason is that when we work with human subjects we typically have small sample sizes. In this case, controlled experiments can be very expensive to run. Alternatively, qualitative studies can generate well-grounded hypotheses and findings that incorporate the complexity of the phenomenon under study. They also offer richer

explanations and new areas for future study. They are also appropriate when variables are not defined or quantified and there is little prior theoretical or empirical work.

One of the benefits of using qualitative methods is that they force the researcher to delve into the complexity of the problem rather than abstract it away, and the results can be more informative [32; 51]. This is the intent of our research, since we wish to understand the challenges that internal offshoring environments hold for those involved software development activities, while comparing to other DSD business models. In this way we created a research design that combines qualitative and quantitative methods, in order to take advantage of the strengths of both. As Basili [4] states, software engineering requires the same high level approach for evolving the knowledge of the discipline like physics, medicine, and many other disciplines. We cannot rely solely on observation or qualitative data.

In the next section, we present the methodological approach for the exploratory study, involving an international and distributed research team. We also present the challenges identified during the planning of this exploratory phase.

## 4. Description of the qualitative research plan

According to the context described in the previous sections, the qualitative and exploratory phase of our study was proposed in order to understand the dynamics of management the onshoring and offshoring strategies, from a software development perspective. The objective was to identify the main differences among different distributed software development business models, to guide the next steps of the research. For this reason, we planned an exploratory case study in five multinational companies.

Our data collection methods included interviews and document reviews. We reviewed project plans, lessons learned, and the software development process. We conducted 20 hours of face to face individual interviews (20 interviews), with three levels of respondents: project manager, IT and Portfolio manager, and organizational management. We interviewed directors, center directors, project managers, and technical leaders. The data collection instrument was a semi-structured interview guide, with open-ended questions. The questions were divided in topics related to the software development process. Seven interviews were conducted in English (35%) and thirteen in Portuguese (65%).

The interview guide was developed through many interactions, involving six experts in the area (an offshore software development center director, a senior consultant in offshore software development, three senior researchers in offshore software development, and a PhD student). We created an initial interview guide having the constructs based on the literature review, and in the model proposed by Evaristo et al [14], which was adapted from

Evaristo et al [15]. We then conducted a review with experts and added some other subjects to develop the interview guide. The questions were related to the literature reviewed and the constructs approved by the experts. The interviews sought to uncover practices and challenges along three topics: Software Engineering, Software Quality and Software Project Management.

However, even before we started the data collection, we as the researchers realized that our own assumptions about the research methodology were colored by our intellectual heritages (as reflected by research training source) as well as by other intrinsic cultural differences either originating from national culture at birth or by extended residence abroad. In table 1 we present the background of each researcher.

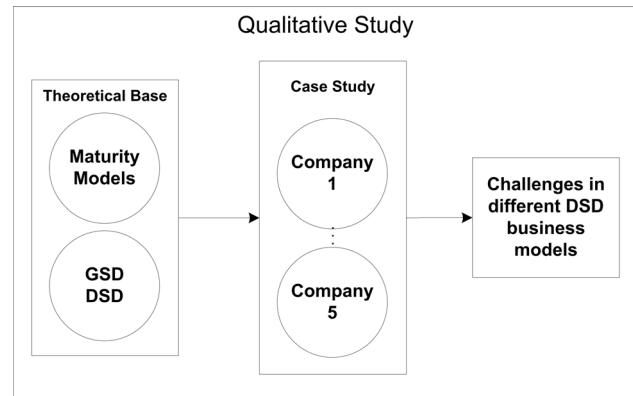
**Table 1. Researcher background**

R#	Residence (years)	Current Residence	Research Training	Language
R1	Brazil (25) U.S. (17) Europe (2)	U.S.	U.S. Europe	English Portuguese
R2	Brazil (40) U.S. (2) U.K. (2)	Brazil	Brazil U.S.	Portuguese English
R3	Romaine (24) Australia (1) Canada (11)	Canada	Europe Canada Australia	Romanian English
R4	U.S. (40)	U.S.	U.S.	English
R5	Brazil (27) Canada (1)	Brazil	Brazil Canada	Portuguese English

Ultimately, we identified challenges and developed strategies to propose a less biased approach to qualitative data methodology. We believe that our contribution relies on the fact that this situation is more likely to become the norm in software development research in the future.

#### 4.1. Research design and planning issues

The experience of conducting research in an international and distributed research team was based on the exploratory case study conducted [52]. Concurrently with a complete literature review, the case study was performed in the five multinational companies mentioned using software project as the unit of analysis. Figure 2 shows a subset of the research design presented previously (the research design of the original study in figure 1).



**Figure 2. Research design of the qualitative investigation**

We used several data collection methods, allowing for triangulation of data to increase reliability and validity of our findings. The interviews and document reviews were used according to Oates [31]. The convenience sample included large organizations that are heavily involved in DSD. The condition for participation was to commit to in-depth interviews by key informants, and also individuals' time in the future to enable survey responses at different points. With respect to data analysis, we run content analysis as proposed by Krippendorff [24].

First, categories were generated by an examination of a subset of data, until there was theoretical saturation, that is, no more categories were being generated. The remaining data was then coded according to these categories. Once all the data was examined, the concepts were organized by recurring theme that would become prime candidates for a set of stable and common categories. Significant differences should be resolved at this point.

#### 4.2. Challenges identified in the planning phase

This study also involved the collaboration with researchers from Brazil, the U.S., and Canada. For this reason, we have identified challenges on the research planning, data collection and data analysis, since it is part of a multi-university, multi-country research program [14]. The challenges were based on three main reasons: the physical and temporal distance among the researchers, the cultural diversity, and the language differences, all increasing the chance for difficulties in the data analysis due to the researcher bias.

Our initial concept was to involve all researchers in the development and validation of the interview guide, as well as the pilot and the data analysis. The idea was that all researchers could analyze the data to detect similarities and compare differences. But we faced a significant challenge in conducting the analysis concurrently involving the distributed researchers. Part of the problem faced was that much of data categorization could be influenced by culturally biases, distance and language.

Moreover, transcripts of interviews conducted in Brazil were originally prepared in Portuguese. The translation of such interviews to English could introduce cultural assumptions on meanings that may or not be caught by translators or the team. The issue that was raised during the category generation was therefore how stable such categories can be across cultures. At that time, we identified ten challenges, related to five of the main activities planned in our study (the data collection activity was not performed distributed): (1) development of the data collection instrument; (2) pilot execution; (3) data preparation; (4) data analysis; (5) and planning of the study itself. Table 2 presents the challenges, the reasons why they were identified, and in which activity each challenge was identified.

**Table 2. Challenges identified**

#	Challenge	Reason	Activity
1	The development of the interview guide could take longer than expected due to language, culture, and distance	Language Distance Culture	1
2	The validation of the interview guide by distributed researchers could be hard due to the distance	Distance	1
3	The planning, execution, and analysis of the pilot could delay the entire study	Language Distance	2
4	Category development of Portuguese based data and validation by English based researchers could lead to significant misunderstandings	Language Culture	4
5	Brazilian researchers could lose important context on the development of categories based on English data	Culture	4
6	Could be hard to transcribe and translate all the thirteen interviews in Portuguese	Language	3
7	Some important context could be lost on translating the interviews in Portuguese for analysis by English based researchers	Language Culture	3
8	The agreement on categories among researchers with different culture background could be hard	Culture	4
9	Transcription and coding of 21 interviews could be time consuming	Language	3, 4
10	The distance (both physical and temporal) among the distributed researchers could create unexpected challenges	Distance	5

Based on the challenges, we recognized the need to develop a new approach to category development, taking in account the physical distribution, language differences and cultural diversity of the research team. For each challenge, we describe the strategy planned.

*Challenge 1:* For the development of the interview guide, we considered the planning of face-to-face meetings among the researchers, together with teleconference and e-mails. This effort was planned to be managed by the PhD student involved in the research.

*Challenge 2:* To validate the interview guide by all researchers, we planned face-to-face meetings to better discuss possible changes and misunderstandings. While most of the meetings were arranged in advance, some of them were organized based on collaboration opportunities that appeared later.

*Challenge 3:* The challenge related to the pilot was identified because we were having difficulties in finding a set of people to run it. Besides, we also decided to run pilots both in Portuguese and in English, to avoid misunderstandings across languages. For this reason, we planned one pilot in Brazil as part of the study of a master student, and another pilot in the U.S., conducted by one of the senior researchers.

*Challenge 4:* In order to generate categories from the Portuguese based data, the Brazilian researchers decided to use the original transcripts in Portuguese. For the other researchers, we planned to translate the same transcripts to English. All researchers agreed to generate categories only in English. The validation of categories was decided to be conducted through teleconference and e-mail, in order to clarify possible misunderstandings.

*Challenge 5:* Based on the subset of interview transcripts in English, the Brazilian researchers were supposed to develop categories also in English. Separately, yet concurrently, the U.S. and Canada researchers were also supposed to create categories using the same subset of interview transcripts. Inter-category agreement was planned to be worked out via email and voice mail calls.

*Challenge 6:* The research team decided to do the analysis in English. For this reason, we decided to translate all interviews in Portuguese to English. To avoid delays in the analysis, we also proposed a plan to get funds and hire technical writers and translator professionals to help in this activity.

*Challenge 7:* To alleviate the risks associated to the translation of Portuguese interviews and the possible loss of context in this case, one of the researchers (that has proficiency in both the Portuguese and English languages) was supposed to play the liaison role, reviewing the translated transcriptions.

*Challenge 8:* To reach a common and stable set of categories by distributed and culturally different researchers, we planned many discussion sessions where all findings could be discussed.

*Challenge 9:* The effort necessary for the transcriptions and coding of all interviews was one of the limitations of our study. In this case, we planned to monitor the evolution of the analysis, and change to an alternative way if necessary (not planned at that time).

*Challenge 10:* In order to avoid possible challenges due to the physical distance, we set a proper infrastructure in the MuNDDoS [40] research lab in Brazil, including teleconference systems, communication and collaboration software tools. We also conducted trips to meet the other researchers face-to-face during the study planning and also during the analysis.

## 5. Discussion

The first set of interviews was conducted in February of 2006, in Canada. A preliminary analysis with the two Canadian companies was documented in Prikładnicki, Damian & Audy [41]. Then, the same interviews were conducted from April to August of 2006, in Brazil. The analysis of one organization in particular was documented in Szymanski & Prikładnicki [49]. A complete report can be found in Prikładnicki et al [38].

In this section, we discuss the challenges presented previously based on what happened during the data collection and analysis of our qualitative study. Together with this discussion, we identified important lessons learned from conducting research with distributed teams, and practical implications for future research on DSD.

*Challenge 1 and challenge 2:* Face-to-face meetings were planned to avoid possible misunderstandings and problems during the interview guide development and validation by all researchers. First, a meeting was conducted in Brazil, among three of the five researchers (R1, R2, R5). As the result of the meeting, a set of questions was developed and validated by the three researchers. Later, the researcher R1 validated the questions with the researcher R4 in a face-to-face meeting in the U.S., and shared with all team. Finally, the researcher R5 validated the interview guide with the researcher R3 in a face-to-face meeting in Canada. This process took longer than originally planned, but was important to ensure proper development of the interview guide (see table 3 for more details on face-to-face meetings).

*Challenge 3:* We conducted two pilot interviews, one in Portuguese and one in English. This was planned to avoid misunderstandings in both languages. The first pilot was executed in Brazil by the PhD student and a master student. The second pilot interview was executed in the U.S., conducted by both R1 and R4. Since researcher R1 had the same level of proficiency in both Portuguese and English, he was able to review all the questions. The pilots were recorded, and then analyzed by the Brazilian team.

*Challenge 4:* Originally, we had planned to translate the interviews in Portuguese into English. But due to the effort needed for this activity, we decided for an alternative plan. What was actually done was that researchers R1, R2 and R5 were responsible for the coding of the interviews in Portuguese. The validation of categories was then conducted through teleconference and e-mail, and agreed among all researchers.

*Challenge 5:* We also changed what was planned for the English based data. Because of the amount of resources and effort necessary, the Brazilian researchers first developed the coding. Then, instead of doing the same coding, a few of the English-speaking researchers reviewed the data coded, to guarantee the accuracy of the findings. Conflicts were resolved through teleconference and email. Frequently, a category previously generated by the Brazilian researchers was changed to better fit in English.

*Challenge 6:* We were not able to get funds as originally planned. Moreover, we didn't have enough resources to translate all interviews to English. Instead, we had external professionals helping in the interview transcriptions, but keeping the original language. For this reason, we applied the strategy explained in challenge 4, there is, researchers R1, R2 and R5 were responsible for the coding of the Portuguese interviews.

*Challenge 7:* The limited resource involved led us to change the plan related to the translations of the Portuguese interviews. Researcher R1 also played the reviewer role to avoid the possible loose of context during the coding. In some cases where a category was questioned by an English-speaking researcher, we followed two ways: explained what was said to clarify and reach a common understanding, or translated specific parts of the interviews to reach the consensus.

*Challenge 8:* To reach a common and stable set of categories, many interactions occurred. First, the categories were validated by teleconference and e-mail between researchers R3 and R5. After that, the categories were validated by R2, and finally by R1 and R4 together. We identified interesting discussions based on the categories, such as the way the category was written, and the amount of details necessary to describe them. At the end, researcher R5 was responsible to coordinate the research team into a common and stable set of categories, in English.

*Challenge 9:* The effort necessary to transcribe and code all interviews was indeed considerable. We were only able to transcribe all of them because we had external professionals working in parallel in a set of interviews. While we understand that this could cause some losses, since the data was read and transcribed by a person outside the research team, all transcriptions were carefully read by the PhD student responsible for the study.

*Challenge 10:* considering the challenges related to distance, we first define a communication protocol among the team. Researcher R5 was the main responsible for make things happen. A proper communication infrastructure was installed in all locations, including communication tools (MSN, Skype, e-mail). We also defined a common software tool for data analysis, and the same configuration was shared among the team. Besides that, we were able to scheduled at least ten face-to-face meetings, as shown in table 3. As said before, most of the meetings were scheduled in advance (indicated in bold).

**Table 3. Face-to-face meetings**

Where/ When	Who	Why (Purpose)	How	Result
<b>Brazil 2004</b>	R1 R2 R5	Overall research planning	R1 as visiting professor	[3]
<b>Brazil 2005</b>	R1 R2 R5	Interview guide development	R1 as visiting professor	[14]
U.S. 2005	R1, R2 R4, R5	Pilot discussion	ICIS [21] Participation	[42]
<b>Brazil 2005</b>	R1 R2 R3	Interview guide development and pilot discussion	R3 as visiting Professor	---
U.S. 2005	R1 R5	Discussion on data collection	R5 personal travel	---
<b>Canada 2006</b>	R3 R5	Interview guide discussion and data collection execution	R5 as visiting student	[41]
U.S. 2006	R1 R4	Discussion on data collection and analysis	R1 personal travel	---
<b>Brazil 2006</b>	R1 R2 R5	Discussion on data analysis	R1 as visiting professor	[35] [39]
<b>Brazil 2006</b>	R2 R3 R5	Discussion on data analysis	R3 as visiting professor	[38]
<b>Canada 2007</b>	R3 R5	Discussion on data analysis	R5 as visiting student	[37]

All meetings were important for clarification on many aspects of the study, such as questionnaire review, pilot planning, data analysis strategy, and general discussions. Due to the researchers' distribution, the meetings were helpful to discuss several issues of the research that could have taken longer if conducted through e-mail or teleconference. We also understand that we were only able to plan that amount of face-to-face meetings not only because of the opportunities we had as researchers, but also because we established some partnerships between universities and companies (interested in the research results) to support those interactions.

Additionally, we generated publications in books, journals, conferences, workshops, and technical reports (detailed in table 3). Those publications were instrumental in fostering discussion among the researchers.

### 5.1. Lessons learned

Based on the distribution of the research team, lessons were learned during the data collection and analysis of this qualitative study on DSD. The lessons are presented in the following table (Table 4).

**Table 4. Lessons learned**

#	Lesson
1	It is very important to plan and execute pilot to test the data collection instruments in all languages considered in the study (in our case English and Portuguese).
2	If it is not possible to have all researchers involved all the time due to distance and other assignments, one researcher should centralize all discussions and manage all actions.
3	Plan face-to-face meetings during the planning of data collection, when is possible.
4	If the research is to be conducted in more than one language, it is recommended to have at least one researcher proficient in all languages, to reduce the biases during translations and coding.
5	It is very important to define a common software tool to be used by the research team during data analysis, as well as proper infrastructure to communicate and collaborate over distances.
6	The research on DSD can benefit from a distributed research team if it involves data collection and analysis from multiple and multicultural sites.

On lesson one, the execution of two pilots (one in Portuguese and one in English) was very important to update the interview guide with significant changes based on the two languages. And in this case we believe that this should be done every time is possible.

To solve the problem of having five researchers geographically distant and with different agendas, as stated in lesson two, the PhD student involved in the research centralized all efforts from the planning phase until the execution and the analysis of all data. This helped to keep all the research team in the same page.

For lesson three, we understand that face-to-face meetings could be hard sometimes. But the meetings we had were very important to solve some misunderstandings and conflicts in all research phases. The PhD student (R5) participated in most of the meetings, centralizing most of the actions. It is important to say that in some meetings we also had a master student and a person from industry developing the interview guide. All meetings were important to minimize the effect of both physical and temporal distances among the researchers.

Lesson four is related to the better execution of a research when more than one language is involved. In our case the proficiency of researcher R1 in both English and Portuguese was very important to have more than one view of the same data being analyzed. This was very helpful to guide the discussion and conflicts solving due to the culture and language differences.

Lesson five was identified based on the results we achieved on using software tools to communicate and collaborate. Those tools were very important to optimize some activities that could take longer due to distance (some reviews or discussions). And the use of a software tool for data analysis kept the research team always on the same page. We used Sphinx 4.5 [47] for demographical data, Atlas TI [1] for coding and category development, and Microsoft Excel [28] for detailed analysis and reports.

Last, but not least, lesson six was identified because our experience indicated that when a research on DSD involves data collection from multiple and multicultural sites, the existence of a distributed research team can be an advantage. First, could be easier to have the initial contact at different sites (city or country, for example) made by local researchers. Moreover, data analysis could be improved with a less biased approach, as explained previously.

## 5.2. Practical implications

Based on the challenges and lessons learned presented, we can draw some practical implications for future investigations conducted by distributed research team and future research on DSD. First, it is important to say that the use of a distributed research team introduces additional challenges, and makes more sense if it can bring benefits to the research itself.

The implications presented in this section are based on the five activities observed and classified in table 2. With respect to the first activity, development of data collection instrument, the existence of researchers with different culture backgrounds can help, for example, to better understand which questions should be more appropriated to a specific culture, and how it should be written. In addition, having the research team distributed can facilitate the data collection from multiple sites, since respondents can be identified in the countries where each researcher is located, and the local researcher can act as a bridge (see lesson six), helping in the initial face-to-face contact and the trust acquisition.

Our experience also showed that in DSD research, pilot interviews have an additional purpose besides to test the research questions and to hone the interviewing skills of the researchers [10]. If data is collected from distributed sites, pilots must be planned to test the questions from the point of view of each site. The existence of a distributed research team can facilitate this activity. In the same way, preparation and analysis of data

gathered from different cultures and languages can benefit from a distributed research team (see lesson four).

Finally, while the planning might take longer in a study conducted by a distributed and multicultural research team instead of local teams, we believe that important risks related to DSD research can be minimized. To give an example, the influence of the bias of the researchers in the results can be reduced, if people from different cultures – from where the DSD teams come from – are planned to analyze and code the same set of data.

## 6. Conclusions

Seaman (p. 572) was eloquent in her call [45]: “Empiricists in software engineering often complain about the lack of opportunities to study software development and maintenance in real settings. This really implies that we must exploit to the fullest every opportunity we do have, by collecting and analyzing as much data of as many different types as possible. Qualitative data is richer than quantitative data, so using qualitative methods increases the amount of information contained in the data collected. It also increases the diversity of the data and thus can potentially increase confidence in the results through triangulation, multiple analyses, and interpretive ability.”

In this paper, we described additional issues that must be taken into consideration when researching distributed software development using qualitative methods. In particular, we identified some challenges and lessons learned when we have an international research team, with different language, different culture, and located in geographically and temporal distances. We presented practical strategies that we applied in our study to address the problem of the research biases.

The planning and execution of data collection and analysis with distributed researchers is not easy. But this experience was very rich in terms of results achieved. At the end, we realized that the research on distributed software development, in a scenario where the respondents are also distributed, and speak different languages, the unique background of each researcher was very important to understand and validate each phase. Our experience was also very helpful to find ways of collecting data when part of the interviews is conducted in a different language by researchers who originate from a different culture and are geographically distributed. Planned follow up studies include the sharing of similar experiences based on the next phases of our research, which include a detailed analysis of the internal offshoring business model, and the use of quantitative methods, both also planned to be conducted with distributed researchers.



Finally, in spite of the fact that many papers in the DSD literature describe qualitative studies with distributed teams [12, 15, 25, 26, 33], sharing a lot of experience (with distributed researchers or not), few papers document the existing difficulties and solutions to overcome the challenges. Our contribution relies on the fact that this situation is more likely to become the norm in software development research in the near future.

## 7. Acknowledgments

Study partially supported by the Research Group on DSD of the PDTI program, financed by Dell Computers of Brazil Ltd. (Law 8.248/91).

## 8. References

- [1] Atlas TI 5.0.66, retrieved on May of 2007, from <http://www.atlasti.com>.
- [2] Audy, R., Prikladnicki, R., "Distributed Software Development? Developing Software with Distributed Teams", Rio de Janeiro: Campus/Elsevier, 2007, in Portuguese.
- [3] Avritchir, J., Prikladnicki, R., Evaristo, R., Audy, J. L. N., "A Maturity Model for Offshore Insourcing: A Research Proposal", In the 10<sup>th</sup> Americas Conference on Information Systems, New York, pp. 3488-3492, 2004.
- [4] Basili, V., "The Role of Experimentation in Software Engineering? Past, Current, and Future", In the 18<sup>th</sup> International Conference on Software Engineering, Berlin, pp. 25-29, 1996.
- [5] Boehm, B., "A View of 20th and 21st Century Software Engineering", In the 28<sup>th</sup> International Conference on Software Engineering, Shanghai, China, pp. 12-29, 2006.
- [6] Carmel, E., Tija, P., "Offshoring Information Technology: Sourcing and Outsourcing to a Global Workforce", UK: Cambridge, 2005.
- [7] Carmel, E., Agarwal, R., "The Maturation of Offshore Outsourcing of Information Technology Work", MIS Quarterly Executive, 1(2), pp. 65-77, 2002.
- [8] Carmel, E., "Global Software Teams – Collaborating Across Borders and Time-Zones", Prentice Hall, 1999.
- [9] Carver, J., Seaman, C., Jeffery, R., "Using Qualitative Methods in Software Engineering", In International Advanced School of Empirical Software Engineering (IASESE04), Los Angeles, CA, USA, 2004.
- [10] Chrissis, M. B., Konrad, M., Shrum, S., "CMMI: Guidelines for Process Integration and Product Improvement", 2nd Edition, SEI Series on SE, US: Addison-Wesley, 2006.
- [11] Creswell, J. W., "Research Design: Qualitative, Quantitative, and Mixed Methods Approaches", USA: SAGE Publications, 2003.
- [12] Damian, D. Marczak, S., Kwan, I., "Collaboration Patterns and the Impact of Distance on Awareness in Requirements-Centred Social Networks", In the 15<sup>th</sup> IEEE International Requirements Engineering Conference, New Delhi, 2007.
- [13] Damian, D., Moitra, D., "Guest Editors' Introduction: Global Software Development: How far Have We Come?", IEEE Software, 23(5), pp.17-19, 2006.
- [14] Evaristo, R., Audy, J. L. N., Prikladnicki, R., Avritchir, J., "Wholly Owned Offshore Subsidiaries for IT Development: A Program of Research", In the 38<sup>th</sup> Hawaii International Conference on System Sciences (HICSS'05). IEEE Computer Society Press, Hawaii, USA, 2005.
- [15] Evaristo, R., Scudder, R., Desouza, K., "A Dimensional Analysis of Geographically Distributed Project Teams: A Case Study", Journal of Engineering and Technology Management. 21(3), pp. 75-189, 2004.
- [16] Herbsleb, J. D., Moitra, D., "Guest Editors' Introduction: Global Software Development", IEEE Software. 18(2), 16-20, 2001.
- [17] Hirschheim, R., "Information Systems Epistemology: An Historical Perspective", Information Systems Research: Issues, Methods and Practical Guidelines, R. Galliers (ed.), Blackwell Scientific Publications, Oxford, pp. 28-60, 1992.
- [18] Hyder, E. B., Heston, K. M., Paulk, M. C., "The sCM-SP v2.01: The eSourcing Capability Model for Service Providers (eSCM-SP) v2.01 – Model Overview", Technical Report CMU-ITSQC-06-006, Carnegie Mellon University, Available at <http://itsqc.cs.cmu.edu/>, Pittsburgh, 2006.
- [19] Hyder, E. B., Heston, K. M., Paulk, M. C., "The sCM-SP v2.01: The eSourcing Capability Model for Service Providers (eSCM-SP) v2.01 – Practice Details", Technical Report CMU-ITSQC-06-007, Carnegie Mellon University, Available at <http://itsqc.cs.cmu.edu/>, Pittsburgh, 2006.
- [20] Huen, W. H., "An Enterprise Perspective of Software Offshoring", In the 36<sup>th</sup> ASEE/IEEE Frontiers in Education Conference, pp. 17-22, San Diego, USA, 2006.
- [21] International Conference on Information Systems, retrieved on May of 2007, from <http://icisnet.aisnet.org/>.
- [22] Kaplan, B. and Maxwell, J.A., "Qualitative Research Methods for Evaluating Computer IS", Evaluating Health Care IS: Methods and Applications", J.G. Anderson, C.E. Aydin and S.J. Jay (eds.), Sage, Thousand Oaks, CA, pp. 45-68, 1994.
- [23] Karolak, D. W., "Global Software Development – Managing Virtual Teams and Environments", Los Alamitos: IEEE Computer Society, 1998.
- [24] Krippendorff, K., "Content Analysis: An Introduction to its Methodology", Sage Publications, 2004.

- [25] Komi-Sirviö, S., Tihinen, M., "Lessons Learned by Participants of Distributed Software Development", *Knowledge and Process Management*, 12(2), pp. 108-122, 2005.
- [26] Layman, L., Williams, L., Damian, D., Bures, H., "Essential communication practices for extreme programming in a global software development team", *Journal of Software and Technology*, 48(9), pp. 781-794, 2006.
- [27] Meyer, B., "The Unspoken Revolution in Software Engineering", *IEEE Computer*, 39(1), pp. 124, pp. 121-123, 2006.
- [28] Microsoft Excel, retrieved on May of 2007, from <http://office.microsoft.com/en-us/excel/FX100487621033.aspx>.
- [29] Morstead, S., Blout, G., "Offshore Ready: strategies to plan & profit from offshore IT enabled services", US: ISANI Press, 2003.
- [30] Myers, M. D., "Qualitative Research in IS", retrieved on May of 2007, from <http://www.qual.auckland.ac.nz/>.
- [31] Oates, B. J., "Researching Information Systems and Computing", London: Sage Publications, 2006.
- [32] Oza, N., Hall, T., Rainer, A., Grey, S., "Critical Factors in Software Outsourcing – A Pilot Study", In WISER, pp. 67-71, Newport Beach, California, USA, 2004.
- [33] Paasivara, M., Lassenius, C., "Communication needs, practices and supporting structures in global inter-organizational software development projects", *Journal of Software Process Improvement and Practice*, 8(4), pp. 183-199, 2003.
- [34] Paré, G., "Enhancing the rigor of qualitative research: Application of a case methodology to build theories of IT implementation", *The Qualitative Report*, 7(4), retrieved from <http://www.nova.edu/ssss/OR/OR7-4/pare.html>, April of 2007.
- [35] Prikladnicki, R., Evaristo, R., Audy, J. L. N., Yamaguti, M. H., "Minimizing the Challenges of Risk Management in Distributed IT Projects: The Importance of the Alignment of Strategic, Tactical, and Operational Levels", In Ned Kock (Org.), *E-Collaboration in Modern Organizations: Initiating and Managing Distributed Projects*, November of 2007.
- [36] Prikladnicki, R., Audy, J. L. N., "Distributed Software Development: Software Development with Distributed Teams", Rio de Janeiro: Campus/Elsevier, 2007, in Portuguese.
- [37] Prikladnicki, R., Audy, J. L. N. "How DSD Organizations Evolve in Internal Offshoring Environments? A Preliminary Qualitative Study", In the First Brazilian Workshop on DSD (collocated with the XXI Brazilian Symposium on Software Engineering), João Pessoa, 2007, in Portuguese.
- [38] Prikladnicki, R., Audy, J. L. N., Damian, D., Oliveira, T. C., "Distributed Software Development: Practices and Challenges in Different Business Strategies of Offshoring and Onshoring", In the IEEE 2<sup>nd</sup> International Conference on Global Software Engineering, Munich, Germany, pp. 262-271, 2007.
- [39] Prikladnicki, R., Evaristo, R., Audy, J. L. N., Yamaguti, M. H., "Risk Management in Distributed IT Projects: Integrating Strategic, Tactical, and Operational Levels", *International Journal of e-Collaboration*, special issue on Collaborative Project Management, Idea Group Inc., 2(4), 1-18, 2006.
- [40] Prikladnicki, R., Marczak, S., Audy, J. L. N., "MuNDDoS: A Research Group on Global Software Development", In the IEEE 1<sup>st</sup> International Conference on Global Software Engineering, pp. 251-252, Florianópolis, Brazil, 2006.
- [41] Prikladnicki, R., Damian, D., Audy, J. L. N., "Offshoring Distributed Software Development: Issues and Challenges", Tech Report DCS-317-IR, University of Victoria, Canada, 2006.
- [42] Prikladnicki, R., Evaristo, R., Gallagher, K., Lopes, L. T. & Audy, J. L. N., "The Role of Culture in Interpreting Qualitative Data: Methodological Issues in an Exploratory Study of Cross-Cultural Distributed Software Development", In the 13th Special Interest Group on Cross-Cultural Research in Information Systems (SIGCCRIS) at ICIS, Las Vegas, 2005.
- [43] Ramasubbu, N., Krishnan, M. S., Kompalli, P., "Leveraging Global Resources: A Process Maturity Framework for Managing Distributed Development", *IEEE Software*, 22(3), pp. 80-86, 2005.
- [44] Robinson, M., Kalakota, R., "Offshore Outsourcing: Business Models, ROI and Best Practices", Mivar Press, 2004.
- [45] Seaman, C. B., "Qualitative Methods in Empirical Studies of Software Engineering", *IEEE Transactions on Software Engineering*, 25(4), 557-572, 1999.
- [46] Sengupta, B., Chandra, S., Sinha, V., "A Research Agenda for Distributed Software Development", In the 28<sup>th</sup> International Conference on Software Engineering, Shanghai, China pp. 731-740, 2006.
- [47] Sphinx 4.5, retrieved on May of 2007, from <http://www.sphinxbrasil.com/po/>.
- [48] Strauss, A., Corbin, J., "Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory", 2<sup>nd</sup> Edition, USA: Sage Publications, 1998.
- [49] Szymanski, C. H., Prikladnicki, R., "The Evolution of the Internal Offshore Software Development Model at Dell Inc", In the IEEE 2<sup>nd</sup> International Conference on Global Software Engineering, Munich, Germany, pp. 40-47, 2007.
- [50] Taylor, S. J., Bogdan, R., "Introduction to Qualitative Research Methods", New York: John Wiley & Sons, 1984.
- [51] Wohlin, C., Höst, M., Henningson, K., "Empirical Research Methods in Software Engineering", In: R. Conradi and A.I. Wang (Eds.): *ESERNET 2001-2003*, LNCS 2765, pp. 7-23, Springer, 2003.
- [52] Yin, R. K., "Case Study Research: Design and Methods", 3rd edition, USA: Sage Publications, 2003.