

Reporting Empirical Evidence in Distributed Software Development: An Extended Taxonomy

Antônio Rafael da Rosa Techio, Rafael Prikladnicki, Sabrina Marczak
MuNDDoS Research Group - Computer Science School
Pontifícia Universidade Católica do Rio Grande do Sul
Porto Alegre, Brazil
antonio.techio@gmail.com, {rafaelp, sabrina.marczak}@puccrs.br

Abstract—Distributed Software Development (DSD) has been discussed by industry and academia for almost two decades now, and, as consequence, there is a large number of empirical scientific papers and industrial reports on it. However, the description of the context in which the empirical study was conducted is not always clear or complete, making the process of searching for empirical evidence burdensome. It becomes difficult to understand or to judge the relevance of study given that DSD scenarios are diverse. What works in one context might not apply to another. To reduce such difficulty, we need, as a research community, to have means to standardize how we report empirical studies and their findings aiming to make them more readily available to practitioners and researchers. In this paper we present an extended taxonomy to classify empirical DSD evidence. We conducted an expert opinion survey with researchers and practitioners to identify elements to compose the taxonomy. Preliminary evaluation of the proposed taxonomy suggests that it can be used to synthesize existing knowledge, to identify gaps in literature, to identify related work and to help researchers who will publish or review further empirical work, as well as practitioners who are interested in published empirical studies.

Keywords— *Systematization of Knowledge; Expert Opinion Survey; Taxonomy; Empirical Evidence; Distributed Software Development.*

I. INTRODUCTION

For almost two decades now the Software Engineering community has witnessed considerable developments in a relatively recent research field, named by some as Distributed Software Development (DSD) and by others as Global Software Engineering [1][2][3][4]. Nowadays, companies are increasingly distributing their software development processes around the world aiming to take advantage of technological improvements and global trade regimes [1]. In this scenario, DSD use has been expanding among large organizations that are converting local markets into global ones, and creating new forms of competition and collaboration with software development players [5].

Because global software projects are so widely varied, many new terms have been created to describe their attributes. The description of empirical studies context is sometimes unclear or incomplete, making it difficult to

understand the study context even after one has read the entire report. Adding to this new and conflicting terminology, the process of searching for empirical evidence is today a complex task. Databases index papers and articles in different ways, using different criteria, making it hard for those using the databases to know how to search for information. These factors taken together make it hard for researchers and practitioners to identify cases that may be of their interest [6].

In addition, due to the many ways global collaborative work can be conducted, lessons learned in one context may not directly apply to another. This means that it may require considerable effort for practitioners to understand the applicability of a study's findings, and thus put increased demand on researchers to clearly describe a study's context when reporting empirical results [7]. Thus, it is necessary to understand how empirical evidence could best be reported in order to be accessible to practitioners and researchers. Empirical research in DSD could be improved if there was a systematic way to identify the various dimensions and characteristics to which empirical evidence applies, as well as complete descriptions of the contexts in which studies are conducted.

This paper answers this need by identifying the descriptors of empirical studies context in DSD that are relevant to researchers and practitioners through an expert opinion survey. Our contribution is an extended taxonomy based on Smite et al. [6] initial work focused on classifying DSD empirical studies. We aim to extend their taxonomy in order to classify empirical evidence in order to contribute to the development of future research in the field.

The remainder of this paper is organized as follows. Section 2 presents related work to ours. Section 3 introduces our research methodology. Section 4 presents the results of the expert opinion survey. Section 5 describes our proposed extended taxonomy and its preliminary evaluation. Section 6 concludes the paper.

II. RELATED WORK

Previous studies have proposed strategies to systematize empirical DSD evidence. Gumm [8] proposed a taxonomy

of distribution used to describe the ways in which people or artifacts are distributed. The dimensions are: physical, organizational, and temporal and geographic distribution among stakeholder groups measured on a scale of High-Medium-Low.

Smite et al. [7] conducted a study focusing specifically on empirical DSD evidence. The study proposed a classification scheme to extract data from empirical studies and systematize the existing empirical DSD studies. The proposed classification scheme assisted in the categorization of extracted data with respect to study population, empirical background and findings. The study revealed that gathering papers with empirical data related to the topic of the desired systematic review was one of the main challenges since globally distributed work is at the cutting edge of cross-disciplinary research. Another difficulty reported was related to the process of deducing and collecting information about empirical background work. Taken altogether, these underscore the necessity for thorough descriptions of contexts in which empirical studies are carried out.

More recently, Smite et al. [6] conducted a study that proposed a terminology and taxonomy for categorizing DSD terms and showed how the taxonomy could be used to classify and to map existing knowledge. The study investigated the state of the use of terminology that characterizes DSD sourcing strategies (e.g., Offshore outsourcing, offshore insourcing) in relation to consistency in spelling and meaning. The result provided a systematically accumulated set of terms categorized in the form of a taxonomy.

The taxonomy of sourcing strategies developed by Smite et al. [6] is the most closely related work to ours, hence our decision to use their taxonomy as an starting point and extend it. Therefore, our goal is to propose an extended taxonomy to classify empirical evidence more broadly, and includes other categories related to empirical background, empirical focus, subjects of investigation and sources of data collection, among other aspects.

III. RESEARCH METHODOLOGY

The work reported in this paper targets researchers who publish or synthesize empirical DSD work, and also practitioners who are interested in published empirical cases. The main goal of our research is to propose a way to systematize the empirical evidence generated from scientific research conducted in the DSD field. The research questions that drove this empirical study are:

RQ1: Which context descriptors of empirical studies are relevant to professionals who work and research in the DSD field?

RQ2: How should empirical evidence generated from scientific research in the DSD field be reported?

To address these questions, this research was designed as an exploratory study. We have conducted an empirical expert opinion survey with experienced DSD researchers and practitioners.

We conducted a survey based on expert opinion [9] to identify the elements of our extended taxonomy. A survey is not just a questionnaire or a checklist for gathering information, it is a research method used to collect information to describe, compare or explain knowledge, attitude and behavior [10]. It allows generalizing about beliefs and opinions of a population by studying a representative subset of people, named a *sample* [11].

The study was carried out with 15 experienced professionals, where 7 respondents were from academia and 8 from industry. Table 1 summarizes participants' information regarding their location, job function, work experience, and education level.

TABLE I. PARTICIPANTS CHARACTERISTICS

ID	Country	Occupation	Function	Industry experience	Research experience	Education
P1	Sweden	Academia	Researcher	NA	From 6 to 10 years	PhD
P2	Italy	Academia	Researcher	NA	More than 10 years	PhD
P3	Italy	Academia	Researcher	From 2 to 4 years	From 6 to 10 years	PhD
P4	Germany	Academia	Researcher	From 4 to 6 years	From 4 to 6 years	PhD
P5	Brazil	Academia	Researcher	From 2 to 4 years	More than 10 years	PhD
P6	Brazil	Academia	Researcher	From 2 to 4 years	From 2 to 4 years	PhD
P7	Brazil	Academia	Researcher	NA	More than 10 years	PhD
P8	Brazil	Industry	IT Manager	More than 10 years	From 2 to 4 years	MSc
P9	Brazil	Industry	IT Manager	More than 10 years	More than 10 years	MSc
P10	Brazil	Industry	IT Manager	From 6 to 10 years	From 2 to 4 years	MSc
P11	Brazil	Industry	Project Leader	More than 10 years	From 2 to 4 years	MSc
P12	Brazil	Industry	Project Leader	From 6 to 10 years	From 2 to 4 years	MSc
P13	India	Industry	IT Executive	More than 10 years	More than 10 years	MBA
P14	USA	Industry	Researcher	From 6 to 10 years	More than 10 years	PhD
P15	India	Industry	Researcher	From 2 to 4 years	From 2 to 4 years	MBA

A. Survey purpose and design

This study is intended to identify context descriptions that are relevant for DSD practitioners and researchers. Thus, for convenience, we chose for interview professionals who have large DSD experience or who have conducted research in the field.

Our survey is categorized as descriptive as per Fink's definition [10]. A descriptive survey serves the purpose of producing information regarding existing groups or a phenomenon. We conducted interviews. A survey interview is one in which the interviewer asks questions from a prepared questionnaire and records the information [9]. Interviews can be conducted by the researcher working directly with the respondent either face-to-face or by telephone [11]. In our study, face-to-face interviews were used with local participants, and with those located in other cities we used Skype—a two-way Internet audio/video communication in lieu of telephone. We also used e-mail when the subject was not available over Skype.

B. Survey questionnaire and data analysis

Once the research goal was defined and the target audience was identified and characterized, the survey

instrument was designed. A questionnaire, which contains the survey, was defined based on the steps outlined in Kitchenham and Pfleeger [9] and Fink [10], which are: to search the relevant literature; to construct an instrument; to evaluate the instrument; and to document the instrument.

For this research, our interview instrument was a semi-structured questionnaire with open and closed questions. The one-on-one semi-structured interview instrument consisted of two main sections. In the first section, we focused on gaining an understanding of the participant's background as indicated in Table 1. In the second section, we focused primarily on gaining an understanding of empirical context descriptions in DSD studies from the subjects' perspectives. The questions were classified according to four categories: motivations, expectations, challenges, and study context descriptions.

The first category, *motivations*, sought a broad understanding of the interviewee's goals for searching through empirical DSD studies. In the second category, *expectations*, the participants were to report what they expected to find in published empirical studies. These two categories used open questions and were made to contextualize the participant in relation to the survey's objectives.

In the *challenges* category there were two questions intended to shed light on the challenges related to searching empirical DSD papers. We created one closed question to map how practitioners and researchers classify the complexity of the search process. Response options were limited to: very low complexity, low complexity, moderate complexity, high complexity and very high complexity. The open question that followed explored reasons for their response to the previous question.

The *context* category was intended to identify what context descriptors in DSD papers were relevant for practitioners and researchers. Here the context descriptions were related to the characteristics of the study, such as research method, data collection, aspects of the study population, characterization of the organization, projects and teams, geographic distribution (being either shared location, or within the same country, the same continent, or across the globe), among others.

In order to avoid interpretation difficulties, a senior researcher reviewed the data collection instrument and a pilot test was performed to simulate the survey application. It should be noted that the authors were not subjects of the pilot test. After transcribing the interviews, we conducted a qualitative analysis of the data collected and preliminary categories were identified through content analysis [12].

IV. FINDINGS

We present in this section the results of the analysis of the data collected. We organize the findings per target audience: first those from academia, next those from

industry. The results from both analyses were then consolidated.

To answer RQ1, we analyzed data collected in our expert opinion survey based on the four categories described in Section 3—motivations, expectations, challenges and context descriptions.

A. Results from academia

The interviews were conducted with 7 researchers. Five were conducted through Skype, one face-to-face, and one by e-mail. Three interviewees were from Brazil and the remaining from Europe as presented in Table 1.

Some participants emphasized the reasons why they would look for empirical papers describing scientific studies or industrial experience reports: to answer questions posed by industry partners, to obtain the sources of related studies, to find solutions for actual issues in practice, and to elaborate theoretical foundations for a new research topic. The researchers also reported other motivations, such as: to identify inherent difficulties to DSD, to find how they are resolved, and to familiarize themselves with the field's state-of-the-art. Some respondents reported specific research topics such as communication, trust, and DSD management and cost reduction as the ones that most interest them.

In the expectations category, items related to the subject company's context (company size, number of sites and locations, etc.), research methodology, source of data collections, research questions, clear descriptions of the contributions to academia and industry, characteristics of the studied populations (including distinctions between practitioners and students as subjects of study), and lessons learned were cited. Some respondents reported specific expectations, such as to find reviews of tools or tools prototypes and case studies addressing aspects of DSD communication.

Five out of 7 participants reported that searching for empirical studies in the literature is an activity with high complexity. The interviewees described some challenges, including that studies' context descriptions are sometimes unclear, empirical data from the scientific perspective is usually incomplete, there is no unified terminology in the field, and the abstracts of scientific papers are not structured, i.e., they do not present important information such as goals, how the research was conducted, results or contributions. They also mentioned difficulty in judging the relevance for, and applicability of, the evidence in their own context of interest.

Characterization of the companies involved and the projects analyzed in studies and their data sources were the main reasons for systematizing evidence reported by the respondents. They emphasized that information about the project team, such as size, helped them compare a study with other contexts. The geographic distribution of teams and information about temporal distance also help them

understand a study’s context. Some participants also mentioned that it was important for them to read whether the results of the study were based on empirical evidence from industry or from academia.

B. Results from industry

These interviews were conducted with eight practitioners, including executives (1), managers (3), project leaders (2), and researchers working in industry (2) as presented in Table 1. Five interviews were face-to-face and 3 by e-mail. Five interviewees were from Brazil, 2 from India, and 1 from the United States.

Regarding practitioners’ motivations to look for empirical DSD studies they reported several reasons, such as seeking best practices and benchmarking reports and studies that describe difficulties associated with DSD, what these are, and what are their solutions.

Practitioners respondents mentioned that they would like to find studies with stories of success and failure, with details related to costs and benefits of a particular practice or solution and industry reports of challenges, how they were resolved, lessons learned and the research methodology used in the studies.

The task of searching for empirical studies in the literature was considered an activity with high complexity by 50% of the industry participants. The respondents cited challenges with the field’s terminology, and difficulty with judging the relevance and quality of the published studies.

The main points sought by the practitioners regarding study context included the goal of the study, characterization of the organizations and projects studied, and data sources. The respondents also cited that the size and distribution of the project teams were important when such details could facilitate the comparison of studies with other contexts.

C. Consolidated results

The consolidated results based on the individual results from academia and industry is presented in Table 2. The results are mapped to the four categories analyzed in the opinion survey: motivations, expectations, challenges and context descriptions of empirical studies.

Questions in the motivations category explored reasons why researchers and practitioners would look for DSD papers with empirical evidence. The results involve identifying best practices, to answer a specific industry question, among others.

The second category, expectations, explored what researchers and practitioners expected to find in empirical DSD studies. The results showed that the description of the organizations and the projects’ context, research methodology and lessons learned were important for professionals working and researching DSD.

TABLE II. CONSOLIDATED RESULTS FROM THE EXPERT OPINION SURVEY

Category	Result	Source
Motivations	Theoretical foundation	Academia
	To answer industry’s questions	Academia
	Best practices	Academia & Industry
	To identify state-of-the-art	Academia
	Benchmarking	Industry
Expectations	Organization and project context	Academia & Industry
	Research method and data collection method	Academia & Industry
	Characterization of study population	Academia
	Contribution of the study	Academia
	Lessons learned	Academia & Industry
	Results of success or failure	Industry
	Costs and benefits of DSD practices	Industry
	Challenges and solutions of the DSD field.	Industry
Challenges	Empirical or non-empirical study classification	Academia
	Field terminology	Academia & Industry
	Research field maturity	Academia
	Non-structured abstract	Academia
	Judge the relevance of the study	Academia & Industry
	Quality of the study	Industry
Context	Empirical background (industry or academia)	Academia & Industry
	Characterization of the organizations studied	Academia & Industry
	Characterization of the projects studied	Academia & Industry
	Characterization of study population	Academia
	Research method and data collection method	Academia & Industry

The results showed that both academia and industry participants reported that the field’s terminology and difficulty in judging the relevance of studies as the main challenges to searching for empirical studies. The studies not being noted as either empirical or non-empirical was also cited as a challenge to finding empirical evidence.

Finally, the category of study context description explored what kind of description was relevant to practitioners and researchers. The empirical context of the study (either academic or industrial), the categorization of organizations and projects studied, and research methodology were mentioned by respondents as being important elements for describing the context of an empirical study.

D. Lessons learned with the survey

We summarize here the answers to RQ1 by presenting the lessons learned from the empirical findings of the expert opinion survey.

Lesson 1: There is evidence that those in industry are looking for solutions and best practices in DSD studies to compare with their own. Also, academics are focused on finding answers there to questions posed by their industry partners.

Lesson 2: There is a need to characterize studies as either empirical or non-empirical. Results showed that researchers do not always mention when their studies contain empirical evidence. Such information would facilitate searches in the literature, as researchers and practitioners could filter by keywords.

Lesson 3: It is important to identify the empirical background in which evidence is based. Results indicated that it is important to report whether empirical data is derived from industry or a research laboratory so that practitioners and researchers can better understand study results.

Lesson 4: Research methods and empirical data sources are important to include in context descriptions for both industry and academia. The most relevant context description is the source of empirical data, such as interviews, questionnaires, results of post-mortem meetings or archive analysis, among others.

Lesson 5: Characterization of a study's population was identified as an important factor in understanding results since they involve cultural aspects and because they can include other details such as the size and distribution of the project team. In addition, participants reported that it is necessary to identify whether a study population consists of students or practitioners.

Lesson 6: Characterization of the analyzed organizations and projects help readers to evaluate the applicability of study findings in other contexts and to understand the empirical evidence itself. Participants reported that details like team size, geographic distribution and temporal distance are important details in helping readers understand a study's context.

Lesson 7: There is need for a unified terminology to facilitate researchers and practitioners in searching for empirical evidence in available databases. Results indicated that terminology is a limiting factor in the search of empirical studies, especially when a researcher is new to the field. Participants reported that there are several terms for explaining the same phenomenon and that this makes searches for empirical evidence a complex task.

Lesson 8: The clear description of a study's results as being either successful or failing is important for practitioners. This kind of information can help in classifying empirical studies and thus facilitate the process of searching for empirical evidence in the literature.

E. Limitations

There are three tests for establishing the quality of descriptive empirical social research: construct validity, external validity and reliability [17].

Construct validity can be ensured through using multiple sources of evidence. This has been achieved here through conducting interviews with researchers and practitioners with diverse backgrounds and experience in DSD. To this end, the study has included researchers, managers, executives and project team leaders.

External validity is the extent to which the results of the study can be generalized to other situations. Since our study was done based on a limited number of respondents, it cannot be generalized this time.

Reliability was developed here through use of an interview instrument. This ensured consistency in data collection. Another senior researcher reviewed the data collection instrument to provide feedback on the clarity and understandability of the questions.

V. THE PROPOSED EXTENDED TAXONOMY

One of the challenges faced by researchers and practitioners described in Section 4 is related to the process of analyzing and understanding the empirical evidence reported in DSD papers and reports. Thus there is a need for the better characterization of the empirical context in this field. This section presents an extended taxonomy for the systematization of empirical DSD evidence.

A taxonomy is a hierarchical classification of a topic or area [6]. It is the grouping of similar items based on established criteria [13]. Forward and Lethbridge [14] add that a taxonomy can be used to help categorize empirical studies and other evidence obtained by researchers so that the applicability of the evidence can be systematically understood. According to Smite et al. [6], a taxonomy in the DSD field could help researchers apply their research systematically, and facilitate comparison and the appropriate application of research in the field.

This taxonomy can be used in the practical application of research results. For example, research results could be categorized in the taxonomy, thereby facilitating the systematization of knowledge. Thus, the proposed taxonomy is intended to help researchers and practitioners understand how empirical evidence could be reported. The taxonomy can be used both theoretically and in practice, thus improving empirical study publication as well as experience reports, and their searches.

First, we provide a description of the taxonomy's categories and its elements. Next, we evaluate the taxonomy by classifying selected empirical studies according to the requirements defined in the taxonomy.

A. Categories and elements of the taxonomy

In response to RQ2, this study proposes an extended taxonomy to help researchers and practitioners in understanding how empirical evidence could be reported more productively. The proposed taxonomy illustrated in Appendix III has 12 categories, each indicated in a column. Each category contains elements used to classify empirical evidence, each indicated as a labeled square. The summary of the categories and elements of the taxonomy as well as their sources (related work or the literature on empirical research methods) are presented in Table 3. In summary, the taxonomy has 40 distinct elements divided into 12 categories. The first eight categories and its elements were identified through analysis of the theoretical foundation described in Section 2 and the results of the expert opinion survey presented in Section 4. The last four categories were extracted from Smite et al. [6] study and added as an extension to the proposed taxonomy.

The first category, *DSD*, represents the starting point of the taxonomy and contains only one element: empirical evidence. The taxonomy does not include any non-empirical papers.

Empirical focus defines the second category. Smite et al. [7] have written that since research within immature disciplines tends to be more exploratory in nature than research in mature fields that focuses more on testing hypotheses, methods or tools, there is a need to differentiate between empirically-based and empirically-evaluated research in the DSD field. *Empirically-based* refers to studies that base their conclusions on empirical data, but do not perform any true empirical evaluation. If a study evaluates a practice, a method, a framework or a tool, it is referred to as *empirically-evaluated* research.

The third category, *empirical background*, defines whether a study's empirical evidence is derived from *industry* or a *research laboratory*. The need to specify the empirical background was found in the opinion survey as a required context description to aid in the understanding of a study's results.

The fourth category defines the *subjects of investigation*, which can be classified as *students* or *practitioners*. As discovered in the opinion survey, the identification of the subjects can help researchers and practitioners understand study results and compare them with their own context.

The fifth category, *study results*, classifies the results of the studies. The opinion survey showed that this information is important for practitioners since it facilitates the search process. The elements of this category were drawn from the systematic review conducted by Smite et al. [7]. This review mapped the results to five categories we have used here: *success story*, *successful practices*, *problem report*, *failed practices* and *failure story*.

TABLE III. SUMMARY OF CATEGORIES AND ELEMENTS OF THE TAXONOMY

Category	Element	Source
DSD	Empirical Evidence	Survey
Empirical Focus	Empirically based or Empirically evaluated	[7]
Empirical Background	Industry or Laboratory	Survey, [7]
Subjects of Investigation	Students or Practitioners	Survey, [7]
Study results	Success story, Successful practices, Problem report, Failed practices or Failure story.	Survey, [7]
Empirical Research Method	Survey, Controlled experiment, Case study, Action research, Ethnography or Grounded Theory.	Survey, [15], [16]
Source of Empirical Evidence	Observation, Archive analysis, Interview, Questionnaire or Postmortem.	Survey, [7]
Number of Sites	More than four sites, four sites, three sites or two sites.	Survey, [7]
Location	Onshore or Offshore	[6]
Legal Entity	Insourcing or Outsourcing	[6]
Geographic Distance	Close, Distant, Near or Far	[6]
Temporal Distance	Similar, Different, Small or Large	[6]

The description of the *empirical research method* was reported by most of the professionals as important information for understanding the context of the results, thus defining the sixth category of the taxonomy. Easterbrook et al. [15] describe five classes of research methods that are most relevant for empirical research in software engineering: *controlled experiment*, *case study*, *survey*, *action research* and *ethnography*. Glaser and Strauss [16] note in their book that it is common to find empirical studies presenting results based on *grounded theory*, so this method as also included in the taxonomy.

The results of the opinion survey indicated that the source of empirical evidence is necessary to understand the results and reproduce the study in another context. Therefore, the seventh category of the taxonomy defines the *source of empirical evidence*. The elements were drawn from the systematic review performed by Smite et al. [7], which mapped the sources of empirical evidence in studies in the DSD field, such as *observation*, *archive analysis*, *interviews*, *questionnaire* and *post-mortem*.

Software development became global in the 1990s as a consequence of the broad use of personal computers and the World Wide Web, and with the subsequent challenges of tight budgets, many companies were motivated to seek partners or to set up development sites in other countries [4]. As an outcome of this evolution, many companies built joint ventures and relocated their development centers to countries with low labor costs. Organization characteristics were thus found in the opinion survey results to be an important context descriptor for DSD studies. The eighth

category, *number of sites*, classifies studies in regard to the number of sites involved in collaboration. The studies can be classified in four ways: *more than four sites*, *four sites*, *three sites* or *two sites*.

Despite the number of sites, the characteristics of the organization can be further classified with sourcing details. The taxonomy proposed by Smite et al. [6] defines 4 categories which are sourcing *location* (onshore or offshore), *legal entity* (insourcing or outsourcing), *geographic distance* (close, distant, near or far) and *temporal distance* (similar, different, small or large). These 4 categories and its elements were incorporated from Smite et al [6] taxonomy and as such are indicated in grey in Appendix III to indicate what is not our original contribution.

It is important to mention that the research and data collection methods (sources of empirical evidence) employ generic categories and thus can be used in any area of research. Also, the study presented here is not intended to overlap with experimental software engineering studies, the goals of which are, among others, to define and to explain the existing research methods and data collection methods.

B. Taxonomy evaluation

In this section we illustrate the use of the taxonomy and explain how it can help synthesize existing knowledge. We have classified 46 articles in our study of the field’s literature according to the dimensions defined in the taxonomy.

The taxonomy was evaluated according to three scenarios (see Table 4). The first scenario (C1) includes empirical papers that can be easily classified because they contain all required context descriptors. Empirical papers with clear context descriptions that cannot be classified in the taxonomy (C2) result in failure of the evaluation process. This happens because the taxonomy does not provide the required elements to classify the study. The studies with unclear context description (C3) do not represent a failure in the process of evaluation because a study’s context description is unclear or not available, and thus, from the perspective of process evaluation, this scenario indicates success through exclusion.

TABLE IV. TAXONOMY EVALUATION SCENARIOS

Scenario	Description	Result	Evaluation
C1	Empirical study with clear context description.	Paper can be classified.	Success
C2	Empirical study with clear context description.	Paper cannot be classified.	Failure
C3	Empirical study with unclear context description.	Paper cannot be classified.	Success

The strategy used to evaluate the taxonomy involved selecting papers published in the International Conference on Global Software Engineering (ICGSE) given that this conference topic is the same our taxonomy tackles. We selected all papers published in the first and eighth edition of the conference, which occurred in 2006 and 2013, respectively (Appendix I). These two issues were chosen because they represent different periods of publications in this field, the first and the most recent one available at the time of the study. Tutorials, short papers, panels and editorials were excluded from the selection. A total of 46 papers were selected: 21 came from the first edition (2006) and 25 from the eighth edition (2013). To classify existing knowledge from the 46 articles selected in the ICGSE, we started with identifying the papers containing empirical evidence and excluded papers that were non-empirical. This resulted in 26 papers after the first screening to be used to evaluate the taxonomy. Data necessary to classify the papers was then drawn from the papers themselves. More specifically, we extracted the information necessary to map the context description to the taxonomy, and to evaluate the clarity of the context description, i.e., whether the paper’s authors provided details clear enough to be used in deducing the necessary information, or whether they had left the context unclear.

Scenario C1: Out of the 26 empirical studies, 15 studies contained clear context descriptions. Papers A10 and A24 used students as subjects of the study, and while the taxonomy proposed by Smite et al. [6] do not apply in this case, they do in our own taxonomy. The same applies to paper A25, which used professionals as subjects of study, but was performed as a controlled experiment with an academic and empirical context.

Scenario C2: We have not found empirical papers with clear context descriptions that could not be mapped in the taxonomy. However, we believe that this result may be due to the low number of papers analyzed, and also to the fact that we used only one source, the ICGSE.

Scenario C3: From a total of 26 papers, 9 papers were not classified in the taxonomy due to unclear context descriptions. As noted earlier, this scenario does not represent failure in the validation process since it was not possible to determine the classification elements from the context descriptions of these studies. In Table 5 we distribute the 26 papers among the taxonomy evaluation scenarios. We also present the list of papers in Appendix II.

TABLE V. RESULTS OF THE TAXONOMY EVALUATION SCENARIOS

Scenario	Papers	Evaluation
C1	A1, A2, A3, A5, A6, A8, A10, A17, A18, A20, A22, A23, A24, A25, A26	Success
C2	No papers found in this scenario.	Failure
C3	A4, A7, A11, A12, A13, A14, A15, A19, A21	Success

The taxonomy can also be used when a study has more than one possible classification. For example, a study may have used more than one source of data collection or the practitioners who participated as subjects of the study can be distributed in different ways, e.g., both onshore and offshore. In Table 6 we illustrate a case where article A3 has three sources of empirical evidence. The study used the interview method as a primary source, and observation and analysis of documents as secondary methods.

TABLE VI. A STUDY WITH THREE SOURCES OF EMPIRICAL EVIDENCE

ID	Research method	Source of empirical evidence
A3	Case Study	Interview
A3	Case Study	Observation
A3	Case Study	Archive analysis (analysis of documents)

Likewise, paper A6, for example, can be classified twice according to the location of practitioners, as illustrated in Table 7.

TABLE VII. A STUDY WITH DIFFERENT SOURCING LOCATIONS

ID	Location	Legal entity	Geographical distance	Temporal distance
A6	Offshore	Outsourcing	Far	Large
A6	Onshore	Outsourcing	Distant	Similar

In summary, having a clear classification for each paper simplifies the search for specific study contexts. This is particularly important for researchers looking for related work and for practitioners attempting to learn from existing research related to their own contexts or topics of interest.

VI. CONCLUSIONS

In order to identify context descriptions that are relevant for researchers and practitioners, this study conducted an expert opinion survey with academia and industry experienced professionals in DSD. The process of searching for empirical studies is complex, DSD terminology is not standardized, and the context descriptions of empirical studies are often unclear, therefore, the need for a taxonomy to guide how empirical work is reported and shared with the community.

The extended taxonomy has two primary usages. First, it provides a basis for researchers to classify their own studies and related studies. Second, once studies are classified using the taxonomy it is possible to identify a set of studies describing a particular situation. Researchers can use it to synthesize existing knowledge, to identify gaps in literature and to search for related work. Practitioners can use the taxonomy to find answers to a specific question in a context that is similar to their own.

ACKNOWLEDGMENTS

We would like to thank all expert opinion survey participants who made it possible for us to conduct this empirical study. This research is also partially funded by FAPERGS, project 002061-2551/13, and CNPq (project 309000/2012-2).

REFERENCES

- [1] Prikladnicki, R., A. Boden, G. Avram, C. B. Souza and V. Wulf. "Data collection in global software engineering research: learning from past experience". Empirical Soft Engineering, 2013, pp. 1-35.
- [2] Beecham, S.; OLeary, P.; Richardson, I.; Baker, S.; Noll, J., "Who Are We Doing Global Software Engineering Research For?," Global Software Engineering (ICGSE), 2013 IEEE 8th International Conference on , vol., no., pp.41,50, 26-29 Aug. 2013.
- [3] Herbsleb, James. 2007. "Global Software Engineering: The Future of Socio-technical Coordination". 29th ICSE, 188-198.
- [4] Carmel, Erran. 1999. "Global Software Teams - Collaborating Across Borders and Time Zones". USA: Prentice Hall, 269 p.
- [5] Herbsleb, James; Moitra, Deependra. Global Software Development. IEEE Software, California, v. 16, n.2, p. 16-20, Mar. /Abr. 2001.
- [6] Smite, D., Wohlin, C., Galvina, Z., and Prikladnicki, R. 2012. "An Empirically Based Terminology and Taxonomy for Global Software Engineering." Empirical Soft Engineering: An International Journal, Published online 18 July 2012. DOI: 10.1007/s10664-012-9217-9.
- [7] Smite, D, Wohlin C, Feldt R, Gorschek T. 2010. "Empirical Evidence in Global Software Engineering: A Systematic Review". Empirical Software Engineering Journal 15(1):91-118.
- [8] Gumm, D. C. 2006. "Distribution Dimensions in Software Development Projects: A Taxonomy." IEEE Software 23 (5): 45-51.
- [9] Kitchenham, B.A., Pfleeger, S.L., 2008. "Personal Opinion Surveys". Springer, London.
- [10] Fink, A., 2003. The Survey Handbook. Sage Publications, Los Angeles, USA.
- [11] Kasunic, M., 2005. Designing an effective survey. Technical report, handbook CMU/SEI-2005-HB-004. Software Engineering Institute, Carnegie Mellon University.
- [12] Stemler, S., "An Overview of Content Analysis", Practical Assessment, Research & Evaluation, vol. 7, no. 17, 2001.
- [13] Prieto-Diaz, R. "A Faceted Approach to Building Ontologies", 21st. International Conference on Conceptual Modeling-ER, Tampere, Finland. 7-11, 2002.
- [14] Forward, A. and Lethbridge, T. C. 2008. "A Taxonomy of Software Types to Facilitate Search and Evidence-Based Software Engineering". Proceedings of the 2008 Conference of the Center for Advanced Studies, CASCON'08
- [15] Easterbrook, S. M., Singer, J., Storey, M., Damian, D., "Selecting Empirical Methods for Software Engineering Research". In: Guide to Advanced Empirical Software Engineering, pp. 285-311. Springer, Heidelberg (2008).
- [16] Glaser, B.G. and Strauss, AL. 2009. "The Discovery of Grounded Theory: Strategies for Qualitative Research." Chicago, Aldine Transaction, 271 pages.
- [17] R. K. Yin, Case Study Research: Design and Methods, 4th ed. Thousand Oaks, California: Sage Publications, Inc, 2009.

APPENDIX I – TAXONOMY EVALUATION PAPERS

- A1 - P Al-Ani, B.; Marczak, S.; Prikladnicki, R.; Redmiles, D., "Revisiting the Factors that Engender Trust of Global Systems Engineers," ICGSE, pp.31-40, 26-29 Aug. 2013.
- A2 - Andaloussi, B.S.; Braun, A., "A Test Specification Method for Software Interoperability Tests in Offshore Scenarios: A Case Study," ICGSE, pp.169-178, Oct. 2006.

- A3 - Anh Nguyen-Duc; Cruzes, D.S., "Coordination of Software Development Teams across Organizational Boundary -- An Exploratory Study," ICGSE, pp.216-225, 26-29 Aug. 2013.
- A4 - Aranda, G.N.; Vizcaino, A.; Cechich, A.; Piattini, M., "Technology Selection to Improve Global Collaboration," ICGSE, pp.223-232, Oct. 2006.
- A5 - Avritzer, A.; Ostrand, T.; Weyuker, E.J., "Experience Developing Software Using a Globally Distributed Workforce," ICGSE, pp.179-186, Oct. 2006.
- A6 - Bass, J.M., "Agile Method Tailoring in Distributed Enterprises: Product Owner Teams," ICGSE, pp.154,163, 26-29 Aug. 2013.
- A7 - Beecham, S.; OLeary, P.; Richardson, I.; Baker, S.; Noll, J., "Who Are We Doing Global Software Engineering Research For?," ICGSE, pp.41-50, 26-29 Aug. 2013.
- A8 - Bruegge, B.; De Lucia, A.; Fasano, F.; Tortora, G., "Supporting Distributed Software Development with fine-grained Artefact Management," ICGSE, pp.213-222, Oct. 2006.
- A9 - Bruegge, B.; Dutoit, Allen H.; Wolf, Timo, "Sisyphus: Enabling informal collaboration in global software development," ICGSE, pp.139-148, Oct. 2006.
- A10 - Casey, V.; Richardson, I., "Project Management within Virtual Software Teams," ICGSE, pp.33-42, Oct. 2006.
- A11 - Deshpande, S.; Beecham, S.; Richardson, I., "Using the PMBOK Guide to Frame GSD Coordination Strategies," ICGSE, pp.188-196, 26-29 Aug. 2013.
- A12 - Deuter, A., "Slicing the V-Model -- Reduced Effort, Higher Flexibility," ICGSE, pp.1-10, 26-29 Aug. 2013.
- A13 - Dorairaj, S.; Noble, J.; Allan, G., "Agile Software Development with Distributed Teams: Senior Management Support," ICGSE, pp.197-205, 26-29 Aug. 2013.
- A14 - Dullemond, K.; van Gameren, B., "What Distributed Software Teams Need to Know and When: An Empirical Study," ICGSE, pp.61-70, 26-29 Aug. 2013.
- A15 - Ehrlich, Kate; Chang, K., "Leveraging expertise in global software teams: Going outside boundaries," ICGSE, pp.149-158, Oct. 2006.
- A16 - Estler, H.C.; Nordio, M.; Furia, C.A.; Meyer, B., "Collaborative Debugging," ICGSE, pp.110-119, 26-29 Aug. 2013.
- A17 - Holmstrom, H.; Conchuir, E.O.; Agerfalk, P.J.; Fitzgerald, B., "Global Software Development Challenges: A Case Study on Temporal, Geographical and Socio-Cultural Distance," ICGSE, pp.3-11, Oct. 2006.
- A18 - Kiani, Z.U.R.; Mite, D.; Riaz, A., "Measuring Awareness in Cross-Team Collaborations -- Distance Matters," ICGSE, pp.71-79, 26-29 Aug. 2013.
- A19 - Marques, A.B.; Carvalho, J.R.; Rodrigues, R.; Conte, T.; Prikladnicki, R.; Marczak, S., "An Ontology for Task Allocation to Teams in Distributed Software Development," ICGSE, pp.21-30, 26-29 Aug. 2013.
- A20 - Modi, S.; Abbott, P.; Counsell, S., "Negotiating Common Ground in Distributed Agile Development: A Case Study Perspective," ICGSE, pp.80-89, 26-29 Aug. 2013.
- A21 - O'Conchuir, E.; Holmstrom, H.; Agerfalk, P.J.; Fitzgerald, B., "Exploring the Assumed Benefits of Global Software Development," ICGSE, pp.159-168, Oct. 2006.
- A22 - Paasivaara, M.; Lassenius, C.; Heikkila, V.T.; Dikert, K.; Engblom, C., "Integrating Global Sites into the Lean and Agile Transformation at Ericsson," ICGSE, pp.134-143, 26-29 Aug. 2013.
- A23 - Shah, H.; Harold, M.J., "Culture and Testing: What is the Relationship?," ICGSE, pp.51-60, 26-29 Aug. 2013.
- A24 - Tell, P.; Babar, M.A.; Grundy, J., "A Preliminary User Evaluation of an Infrastructure to Support Activity-Based Computing in Global Software Development (ABC4GSD)," ICGSE, pp.100-109, 26-29 Aug. 2013.
- A25 - van Gameren, B.; van Solingen, R.; Dullemond, K., "Auto-erecting Virtual Office Walls: A Controlled Experiment," ICGSE, pp.206-215, 26-29 Aug. 2013.
- A26 - Zieris, F.; Salinger, S., "Doing Scrum Rather Than Being Agile: A Case Study on Actual Nearshoring Practices," ICGSE, pp.144-153, 26-29 Aug. 2013.

APPENDIX II – TAXONOMY EVALUATION RESULTS

Scenario C1:

ID	Year	Empirical evidence	Empirical focus	Empirical background	Subject of investigation	Study results	Empirical research method	Source of empirical evidence	Number of sites	Location	Legal entity	Geographical distance	Temporal distance
A1	2013	Yes	Empirically-based	Industry	Practitioners	Success story	Case Study	Interview	Irrelevant	Offshore	Insourcing	Far	Large
A2	2006	Yes	Empirically-evaluated	Industry	Practitioners	Success of practices	Case Study	Document analysis	2	Offshore	Outsourcing	Far	Large
A3	2013	Yes	Empirically-based	Industry	Practitioners	Success of practices	Case Study	Interview	3	Offshore	Insourcing	Far	Large
A5	2006	Yes	Empirically-based	Industry	Practitioners	Problems report	Case Study	Observation	2	Offshore	Outsourcing	Far	Large
A6	2013	Yes	Empirically-based	Industry	Practitioners	Success story	Grounded Theory	Interview	2	Offshore	Outsourcing	Far	Large
A8	2006	Yes	Empirically-evaluated	Laboratory	Students	Success story	Controlled experiment	Survey	Irrelevant	Irrelevant	Irrelevant	Irrelevant	Irrelevant
A10	2006	Yes	Empirically-based	Industry	Practitioners	Problems report	Case Study	Observation	2	Offshore	Insourcing	Far	Large
A17	2006	Yes	Empirically-based	Industry	Practitioners	Problems report	Case Study	Interview	2	Offshore	Insourcing	Far	Large
A18	2013	Yes	Empirically-based	Industry	Practitioners	Problems report	Survey	Interview	2	Offshore	Insourcing	Far	Large
A20	2013	Yes	Empirically-based	Industry	Practitioners	Problems report	Case Study	Interview	2	Offshore	Insourcing	Far	Large
A22	2013	Yes	Empirically-based	Industry	Practitioners	Problems report	Case Study	Interview	3	Offshore	Insourcing	Far	Large
A23	2013	Yes	Empirically-based	Industry	Practitioners	Problems report	Ethnography	Observation	2	Offshore	Outsourcing	Far	Large
A24	2013	Yes	Empirically-evaluated	Laboratory	Students	Success story	Controlled experiment	Survey	Irrelevant	Irrelevant	Irrelevant	Irrelevant	Irrelevant
A25	2013	Yes	Empirically-based	Laboratory	Practitioners	Success story	Controlled experiment	Survey	Irrelevant	Irrelevant	Irrelevant	Irrelevant	Irrelevant
A26	2013	Yes	Empirically-based	Industry	Practitioners	Success story	Case Study	Interview	2	Offshore	Outsourcing	Near	Small

Scenario C3:

ID	Year	Empirical evidence	Empirical focus	Empirical background	Subject of investigation	Study results	Empirical research method	Source of empirical evidence	Number of sites	Location	Legal entity	Geographical distance	Temporal distance
A4	2006	Yes	Empirically-evaluated	Industry	Practitioners		Case Study	Survey					
A7	2013	Yes	Empirically-based	Industry	Practitioners	Problems report	Survey	Interview	More than 4				
A11	2013	Yes	Empirically-based	Industry	Practitioners	Success of practices	Case Study	Survey					
A12	2013	Yes	Empirically-evaluated	Industry	Practitioners	Success of practices			3	Offshore	Outsourcing	Far	Large
A13	2013	Yes	Empirically-based	Industry	Practitioners	Problems report	Grounded Theory	Interview		Offshore			
A14	2013	Yes	Empirically-based	Industry	Practitioners			Survey					
A15	2006	Yes	Empirically-based	Industry	Practitioners	Problems report	Case Study	Survey		Offshore			
A19	2013	Yes	Empirically-evaluated	Industry	Practitioners			Interview					
A21	2006	Yes	Empirically-based	Industry	Practitioners		Case Study	Interview	More than 4	Offshore	Insourcing	Far	Large

APPENDIX III – THE PROPOSED EXTENDED TAXONOMY

