

An Empirically Based Terminology and Taxonomy for Global Software Engineering

DARJA ŠMITE, CLAES WOHLIN, ZANE GALVIŃA, RAFAEL PRIKLADNICKI

Abstract: Many organizations nowadays strive for utilization of benefits offered by global software engineering (GSE) and sourcing strategies are thus discussed more often. Since there are so many variations of the attributes associated with global software projects a large amount of new terms has been introduced. The diversity in sourcing jargon however has caused difficulties in determining which term to use in which situation, and thus causing further obstacles to searching and finding relevant research during, for example, systematic literature reviews. The inability of judging the applicability of the research in an industrial context is another important implication on the transferability of research into practice. Thus the need for accurate terminology and definitions for different global sourcing situations emerges as a way for the community to build upon each other's work and hence making progress more quickly. In this paper we first investigate the state of the use of the GSE jargon concluding that terminology is very diverse (many synonyms used to describe the same phenomena), often confusing (same terms used to describe different phenomena) and occasionally ambiguous (few terms used to describe several phenomena). In order to address the identified problems, we conducted a Delphi-inspired study with ten well-established researchers in GSE and developed an empirically based glossary for the key concepts in global software engineering. We then propose a taxonomy for GSE by categorizing the selected terms based on generalization-specialization relationships and illustrate how the taxonomy can be used to categorize and map existing knowledge. The proposed contribution targets future researchers, who will publish or synthesize further empirical work and practitioners, who are interested in published empirical cases. Therefore this work is expected to make a contribution to the future development of research in the GSE field, and alleviate understandability and transferability of existing and future knowledge into practice.

Keywords: Global Software Engineering, Global sourcing, Offshoring, Outsourcing, Taxonomy, Terminology

1 Introduction

Global Software Engineering (GSE) becomes a part of the everyday business and so does the use of different terms related to sourcing strategies such as *outsourcing*, *offshoring*, *nearshoring*, *farshoring*, *rightshoring*, *bestshoring*, etc. While the research literature addressing the challenges and benefits of different sourcing strategies matures, the readers are often forced to deduce the information from descriptions that are sometimes scarce, ambiguous or unclear (Smite et al. 2010; Prikladnicki and Audy, 2010). It is in many cases hard to understand the context of a study, which makes it hard for both researchers and practitioners to identify cases that may be of interest for them. The introduction of a clear and concise terminology and taxonomy would help mitigate these challenges. Being experienced readers in the field we have come across different cases of confusion in relation to the use of the GSE jargon, such as cases of unclear or insufficient combination of terms, and cases of using the same terms differently. More importantly, we have faced difficulties in searching and finding relevant research since the keywords used in this field are very diverse. Carmel and Tjia (2005) note that even the popular term *offshore information technology outsourcing* is replete with misleading usage. The inconsistency in terminology can have certain implications, including inability to judge the applicability and thus transferability of the research into practice, when learning from studies with a poorly described or unclear context.

This paper contributes by basing a terminology and a taxonomy of global software engineering on a thorough literature study to identify potential terms and then a set of experts are asked to help in determining a best set of terms. These terms are presented in a glossary for GSE. The experiences from the

expert survey form the basis to formulate a taxonomy, which provides a set of sourcing strategies that also can be used to identify commonalities between specific studies or cases.

1.1 Related work

Taxonomy of distribution developed by Gumm (2006) is one of the most related attempts in the same direction as ours. In this work the multi-dimensional taxonomy is used to describe the ways in which people or artifacts are distributed. The dimensions are: physical, organizational, temporal and distribution among stakeholder groups measured on the scale of High-Medium-Low. However, the work is more explanatory in nature and builds on the casual dependencies among distribution risks and necessary coping mechanisms rather than providing concrete rules that can be used to instantiate the taxonomy or classify the cases.

Several classifications of sourcing strategies in particular exist. These are often based on the combination of geographic location and relationship structure between companies as in Ågerfalk and Fitzgerald (2008), Robinson and Kalakota (2004) and Höfner et al. (2011), see Figure 1. Interestingly, the terminology used in these classifications only partly overlaps, the terms often being derived from the words *offshore*, *onshore* and *outsource*. Although all three models use the same two dimensions (location and relationship structure) we chose not to align them in order to highlight the diverse use of the same starting point for classification. Also note that the models use two different concepts associated with the in-house/captive and onshore quadrant. In Model A and C this characterizes traditional development and internal supply, which we interpret as “under one roof”, while in Model B this is associated with a domestic supply (see Figure 1).

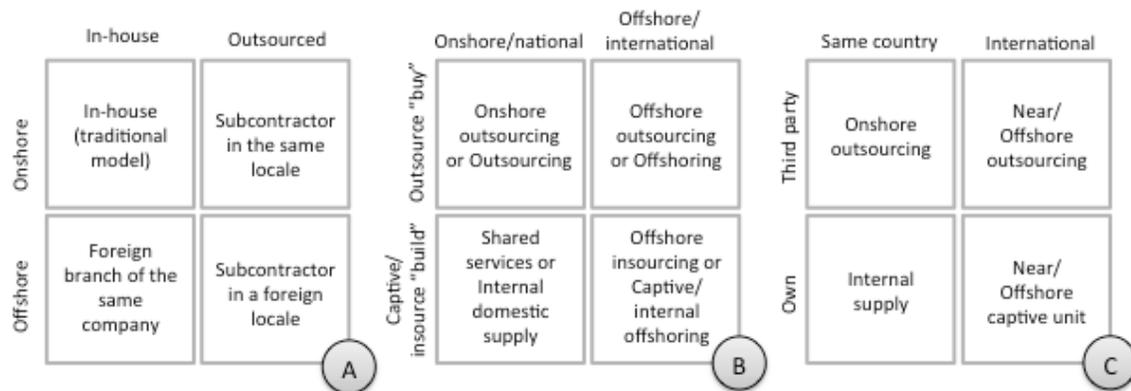


Figure 1: Existing classifications of sourcing strategies (A – adapted from Ågerfalk and Fitzgerald (2008), B – adapted from Robinson and Kalakota (2004), C – adapted from Höfner et al. (2011))

Carmel and Abbott (2007) suggest that as the outsourcing and offshoring phenomena mature, an increased differentiation on the basis of location emerges. Accordingly, the term *nearshore* was introduced to describe nearby locations as opposed to the main sourcing destination (India), which was seen as *farshore*.

Classification of knowledge in the GSE field is also performed in recent systematic reviews (Prikladnicki and Audy, 2010; Smite et al., 2010; Hossain et al., 2009). Prikladnicki and Audy (2010) distinguish between *offshore outsourcing* (when development is moved to an external third party in another country), *internal offshoring* (when development is moved to a division of a specific company established in another country) and *offshoring* (used as a generic term when the relationship is unknown). Smite et al. (2010) and Hossain et al. (2009) classified collaboration modes into two categories: *inter-organizational* (between two companies) and *intra-organizational* (within the same company).

To the best of our knowledge, there is no unified empirically based glossary or reference model established in the field, which contains a clear description of the various terms that are used to describe sourcing strategies, or a taxonomy that would explicitly define the relationships between the different terms. On the contrary, our investigation repeatedly came across, for example, oceanography descriptions when searching for sourcing terms in the dictionaries. Thus, we emphasize that the importance of clarity in terminology in

the GSE field is currently underestimated. This effort is thus intended to be an important step to make progress more quickly in this rather recent field of research.

1.2 Objectives and research questions

The work reported in this paper targets researchers, who publish or synthesize empirical work in GSE and practitioners, who are interested in published empirical cases. The aim of our research is to investigate the state of the use of terminology that characterizes sourcing strategies in GSE in relation to consistency in spelling and meaning, and in result provide a systematically accumulated set of terms and categorize them in the form of taxonomy. In particular, research questions that drove our investigation were as follow:

RQ1: Which terms and combinations of terms are used to characterize different sourcing strategies?

RQ2: Do the authors define the mentioned terms or not? Are these terms defined consistently?

RQ3: Can field experts agree on the terms to be used in GSE?

RQ4: What are the interrelationships between the identified terms?

A map compound of existing classification of GSE strategies (see Figure 1) is the starting point for our investigation. The map (see Figure 2) contains dimensions, which we aimed to include in the proposed terminology and taxonomy.

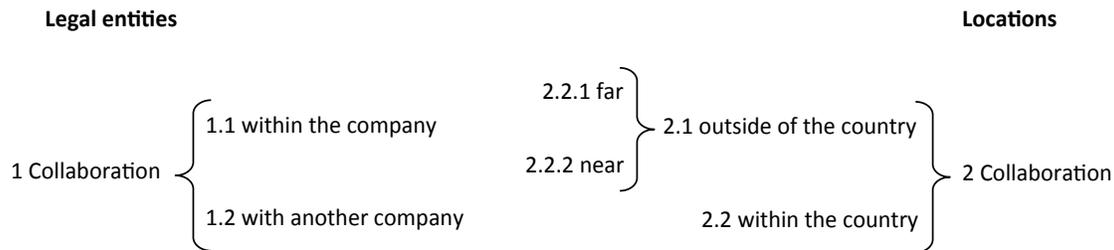


Figure 2: Necessity for terms

The remainder of the paper is structured as follows. In Section 2 we present our research process. Section 3 is dedicated to our findings: the state of the use of the sourcing terminology (Section 3.1), results of the expert survey (Section 3.2), GSE glossary (Section 3.3) and taxonomy and its validation (Section 3.4). Finally, Section 4 provides the conclusions, and outlines the target audience, future research and implications for practice.

2 Research overview

2.1. Literature Review

In order to investigate the state of the use of the GSE terminology, we chose to systematically review research papers from the major conferences and thematic special issues from different journals. This step was conducted in Q4 2009 – Q1 2010 with refinements in Q4 2011. We have extracted the papers from the following venues:

- IEEE International Conference on GSE (ICGSE) 2006-2011,
- International Conference on Software Engineering (ICSE) 2000-2011,
- ACM SIGSOFT Symposium on the Foundations of Software Engineering (FSE) 2000-2011,
- Communications of ACM special issue – 49(10)/2006,
- SPIP special issues – 8(4)/2003, 13(3)/2008, 13(6)/2008, 14(5)/2009,
- IST special issue – 49(9)/2006,
- IEEE Software special issues – 18(2)/2001, 23(5)/2006.

The selection resulted in 296 papers after excluding editorials summaries of workshops and panels, non-GSE articles, and presentation summaries. Notably, the majority of included papers (about 2/3) come from

ICGSE conference series. This is understandable, since the conference is dedicated to the topic and major exclusions were done for other conferences and journals. Nonetheless, we believe that these are venues that provide a representative view of the research field related to global software engineering in particular since the objective was not to capture every single term but rather to find the most commonly used terms. The full papers were downloaded and registered in an MS Excel tool for further analysis.

In our analysis we have chosen to focus on sourcing strategies as a combination of geographic location and relationship structure between companies as suggested by Ågerfalk and Fitzgerald (2008), Robinson and Kalakota (2004) and Höfner et al. (2011). Each paper was read by one of the researchers (Galviņa) and analyzed using selective coding techniques. All relevant definitions or context descriptions that indicate the use of the terms were coded and extracted. Then two researchers (Smite and Prikladnicki) performed parallel analysis of the terms and definitions and linked the terms used in the reviewed articles to the combination of geographic location and relationship structure. Involvement of two researchers in the analysis is expected to increase the credibility of the results. This way, single researcher's bias was mitigated by these two independent analyses and during the expert review, in which new terms could be added. The following information was captured:

- All mentioned terms, which could be linked to sourcing strategies as well as characteristics of these strategies (described in Figure 2);
- Definitions of the terms or indication that the definition is missing;
- If the definitions were not available, the context of the term in use;
- General information about the literature source, including authors, year, venue of publication and country of origin of the authors.

It is worth noting that initially we also extracted terms used to characterize different sourcing strategies (adjectives) and activities related to exercising these strategies (verbs). For example, we aimed at finding a term that describes team members working in the same location (candidates were *co-located*, *same-site*, *co-site*, *single site*, *onsite*, *intra-site*, etc.), or team members from another location (candidates were *remote*, *distant*, *multi site*, *inter site*, *cross site*, *offshore*, *off-site*, etc.). Examples of the verbs are: distribution of work across locations (candidate: *distribute*), relocation of work (candidate: *transfer*), keeping the work (candidates: *keep*, *insource*, *retain*). These terms can be characterized by a wide variety of possible synonyms and are often used interchangeably in the text. This is perhaps the reason for the lack of convergence among the experts and inability to reach a consensus during the survey. Finally, we decided to exclude the adjectives and verbs from the final report and not restrain the researchers in their choices.

From the extracted material we sought to find answers to research questions RQ1 and RQ2. The extracted definitions of terms and the context descriptions were analyzed using axial coding techniques. In other words, to determine the consistency of the terms all definitions and contexts of use of the same terms were compared. The findings from the parallel analysis performed by the two researchers were compared and discussed. Additional observations were produced and supported by concrete examples from the literature. This also resulted in cross-term relationships that produced a list of synonyms. The results of this step are described in Section 3.

2.2. Expert survey

The motivation for organizing an expert survey was twofold. First of all it was driven by the findings from the literature review that demonstrated a divergence rather than convergence in the use of the terminology. Second of all, we have decided to adopt the process used in Project 610 (Radatz, 1988) that was used to establish a standard dictionary for computer terminology, in order to eliminate subjectivity and avoid suggesting terminology on behalf of the GSE society. In other words, the objective was to create an empirically based terminology and hence taxonomy. Furthermore, by basing the terminology and taxonomy on experts, we wanted to increase the chances of adoption both by the researchers involved and by others. This step was conducted in Q3 2010 – Q1 2011.

2.2.1. Expert selection

To increase the credibility of the results we aimed to involve key experts in the field. An expert is anyone especially knowledgeable in the field at the level of detail (Meyer and Booker, 2001). To find the experts we searched the GSE authors of prime research articles in Google Scholar. To diversify the list we added several knowledgeable and active organizers in the IEEE series of the International Conferences on Global Software Engineering (ICGSE), as well as experts from Information Systems research related to global sourcing in software organizations. Twenty experts were identified and invited through email. The experts were informed about the necessary effort and potentially lengthy rounds of survey. Ten experts accepted the invitation and committed to participate. This satisfied our goals as the number of experts for a Delphi study can be in the range of 10-18 (Okoli and Pawlowski, 2004). The profiles of the experts are given in Table 1. The ten experts have a mix of different countries, background, culture and experiences, from both Software Engineering and Information Systems communities. The expertise of the invited researchers can also be illustrated by their involvement in the Organizing Committee (OC) and Program Committee (PC) of the series of ICGSE conferences.

Table 1: Expert profiles

| Expert | Affiliation | Country | Involved in GSE research | Involved in ICGSE OC | Involved in ICGSE PC |
|-------------------|---|------------------|--------------------------|----------------------|----------------------|
| Sarah Beecham | LERO | Ireland | Since 2009 | 2010 | |
| Aurora Vizcaino | University of Castilla-La Mancha | Spain | Since 2004 | 2010 | Since 2007 |
| Rini van Solingen | Delft University of Technology | The Netherlands | Since 1994 | | Since 2007 |
| Erran Carmel | American University | USA | Since 1995 | | Since 2007 |
| Jürgen Münch | University of Helsinki, Fraunhofer IESE | Finland, Germany | Since 2003 | 2010 | Since 2007 |
| Daniela Damian | University of Victoria | Canada | Since 1998 | 2006, 2009 | Since 2006 |
| Yael Dubinsky | IBM Research, Haifa Lab | Israel | Since 2005 | 2010, 2011 | Since 2008 |
| Aybuke Aurum | University of New South Wales | Australia | Since 2005 | | |
| Casper Lassenius | Aalto University | Finland | Since 2000 | 2011 | |
| Maria Paasivaara | Aalto University | Finland | Since 2000 | 2011 | Since 2007 |

2.2.2. Process

The process of developing the GSE glossary consisted of six steps (see also Figure 3):

1. Select terms
2. Prepare draft definitions
3. Refine draft definitions
4. Discuss among experts. Revise
5. Perform consistency check. Revise
6. Check on response bias

In the first step, the terms identified through the literature review were selected. Draft definitions were prepared by one of the researchers on the basis of the gathered definitions. The other researchers helped refining the terms to achieve consistency in the formulations. A Delphi-inspired survey approach was then implemented for the fourth step, which was iterated several times to achieve trustworthy results. The rounds were repeated until a substantial agreement or a poor agreement with no convergence was detected.

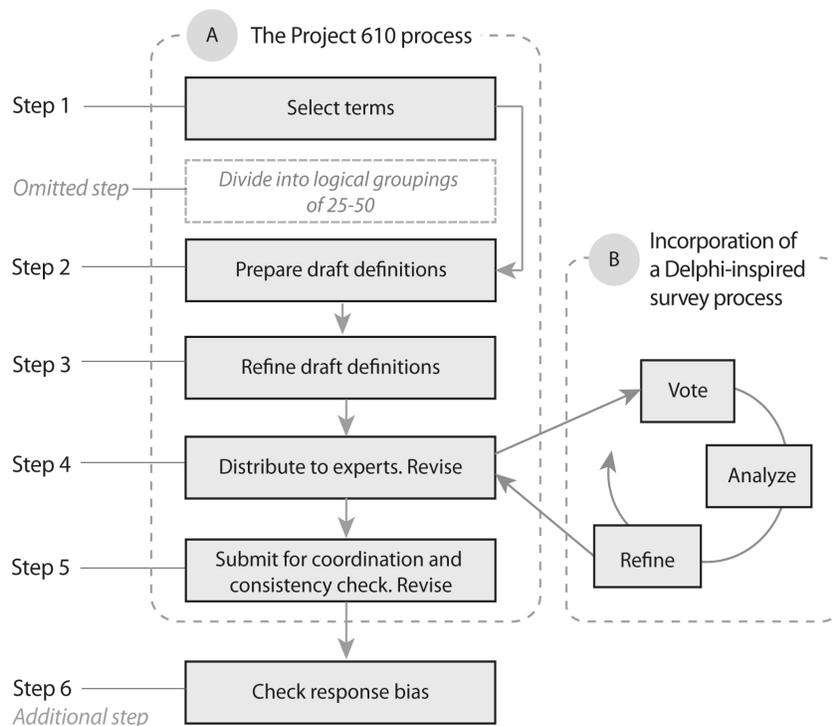


Figure 3: Expert survey process

All votes remained anonymous throughout the process and could not be traced back to the participants even after the completion of the survey. This was done to enable the free expression of opinions. In the fifth step, the accepted terms that achieved the necessary agreement or no final convergence were examined. We aimed at selecting one term for each of the previously identified sourcing strategies (see Figure 2). Thus, in some cases terms with poor agreement were nonetheless proposed wherever no alternatives were accepted. In the sixth and the last step the glossary was sent to two additional experts that were initially invited but did not accept to participate in the full expert survey. This was done to check the response bias. Response bias is the effect of non-respondents on survey estimates. If, the responses of non-respondents substantially influence the overall results of the survey this means that the results are biased or unstable. Our goal was to evaluate if the additional experts' responses differ substantially from the respondents.

2.2.3. Expert review

In the first round the experts received the list of terms and definitions for voting. These definitions were based on the state of the use of terminology in the research articles identified as described above. The terms were divided into three sections: sourcing strategies, adjectives and verbs. Because of disagreement among the experts we have chosen not to report the results for the latter two as explained and exemplified above (see Section 2.1). Due to a variety of terms and synonyms, key terms were selected by the researchers and proposed as “preferred”. A category “To be avoided” was also put forward. All experts were instructed to:

- Review the terms (Approve/Y, Disapprove/N, Suggest to avoid/A)
- Add new terms
- Review definitions (Approve/Y, Disapprove/N)
- Suggest changes to the definitions.

The received responses were reviewed and synthesized. Expert agreement was calculated and communicated in the next round together with the additions and changes that were open for voting.

In the second round the experts received a list of approved terms, new terms and a list of proposed changes together with justifications for these changes and other comments and opinions, which were submitted anonymously. The experts were offered an opportunity to revise their votes or express opinions. Here the responses were obtained for the following steps:

- Approve the terms (Approve/Y, Disapprove/N)
- Suggest to avoid the terms (Approve/Y, Disapprove/N)
- Add new terms
- Review definitions (Approve/Y, Disapprove/N)
- Suggest changes to the definitions.

The votes and comments were again reviewed and synthesized, followed by the assessment of the degree of agreement.

2.2.4. Data analysis

There are several methods for estimating inter-rater reliability in a study (Jordan and Miller, 2003). According to Gwet (2010), the reliability estimate quantifies the distance of scores assigned by a group of raters to the same subjects; the closer the scores assigned by each rater, the higher the reliability (Jordan and Miller 2003). For an estimate of the consistency between two or more raters, which describes our case, Fleiss' generalized Kappa is suggested as one of the possible methods (Gwet, 2010). All responses in the expert survey were characterized in terms of a Degree of agreement (DA) index, calculated by the frequency of each vote. Blank responses, in cases when experts sustained, were analyzed as a separate category and therefore the calculation method for unbalanced data was adopted. Fleiss kappa was calculated for each term and definition. In summary, we analyzed the responses with three votes rated by ten experts and with one or more treatments. The following interpretations of reliability coefficients based on a scale suggested in (Landis and Koch, 1977) are used in the analysis of the expert agreement:

- $\kappa = 1.00$ Perfect agreement
- $0.81 \leq \kappa < 1.00$ Almost perfect agreement
- $0.61 \leq \kappa < 0.80$ Substantial agreement
- $0.41 \leq \kappa < 0.60$ Moderate agreement
- $\kappa \leq 0.40$ Poor agreement

In our expert survey we have aimed at agreeing on one term for each GSE concept to keep the terminology clear and concise, and achieving an agreement about coherent definitions for these terms. During each round the experts were offered a variety of terms to choose from, firstly expressing their preferences by choosing terms from a set of synonyms, secondly voting for disapproving or suggesting avoiding the non-preferred term, along with reviewing the proposed definitions for approval, disapproval or change. The decisions were made using the following arguments:

- Terms were approved on the basis of
 - expert agreement;
 - as a term compound of the approved terms (during consistency check);
- Terms were suggested to be avoided on the basis of expert agreement;
- Terms were rejected on the basis of
 - disagreement among the experts;
 - inconsistency with the other terms (during consistency check);
 - on the basis of expert agreement.

The process for seeking expert agreement for definitions of the terms was similar. Experts were offered initial definitions that were formulated by the researchers on the basis of reviewed GSE literature. Each expert could vote for the initial definition or propose his/her own changes. Improvements that achieved substantial consent were implemented in the final version of the definitions.

2.2.5. Validity

The next and the last step in our research before proposing the glossary and taxonomy aimed at validating the findings from the expert survey through a check on response bias. We organized the survey results in two tables (terms and definitions) and invited two experts from the list of non-participants to express their opinion in terms of agreement or disagreement with the results. We invited one researcher from Europe and

another from the United States. Both of them have several years of experience with GSE and are very active in the field, including participation in the PC and OC in the series of ICGSE conferences.

Due to time constraints we did not perform a complete cycle of response bias and limited our validity check to the final results from the last round of the survey. This means that we did not perform statistical assessment of the effect of non-participants' responses per se. Instead, we asked them to agree or disagree with the final list of terms and definitions, and evaluated the potential threats to validity using the feedback received. The results of validation are discussed in Section 3.2.1.

The experts agreed with most of the definitions, but questioned some of the decisions. For example, one of the experts did not agree with the definition provided for *Global Software Engineering*, arguing that the word *global* usually refers to a different country and we should be using a more generic word to define the area, such as *distributed*. However, *global* is a reference to the globe or around the world, but around the world does not exclude within the same country, and hence we have chosen to use *Global Software Engineering* throughout the paper.

Another potential validity threat is related to the conflict of interest related to the actions of the experts or the authors of this paper in favor of particular terms that they have previously used in their own publications. Inviting independent researchers from various research groups to jointly develop an empirically based terminology mitigated the influence of the authors of this paper on the results. While the expert influence is not and cannot be completely avoided, we believe that the high level of agreement on the terms demonstrates a certain level of validity. In fact, during the expert rounds we have evidenced cases when experts were able to agree on the meaning of particular terms, which was different to what they previously used in their publications.

2.3. Taxonomy creation

A taxonomy adds to a terminology by describing the relationships between terms. By depicting the relationships from generalization to specialization a taxonomy may be created, which helps connecting terms in an area or for a topic. Thus, a taxonomy is a hierarchical classification of a topic or area. Originally, taxonomies come from describing animal species and their relationships as described in (Carl Linnaeus in Wikipedia, 2011). In the 18th century the Swedish scientist Carl von Linné (or Carl Linnaeus) created taxonomies for animals, plants and minerals respectively. He grouped them when they had things in common, which he then refined into smaller groups on a lower level. His taxonomies still form the basis for the classification of animals, plants and minerals today.

If allowing ourselves to simplify the view of animals and, in particular, birds for illustration purposes, and only use four levels, a simplistic branch through such a tree may look as follows (with four levels): Animal – Bird – Finch family – Chaffinch. On each of the three latter levels it is possible to imagine other branches, for example, birds are not the only type of animal, birds contain many different families of birds (the Finch family is one such family of birds) and the family contains many related birds including the Chaffinch. It may be noted that all terms may be used separately. For example, if talking to an ornithologist that person knows what a Chaffinch is and hence there is no need to explain it from the beginning of the tree. However, if explaining it to a child, you may choose to say that it is an animal and it is a bird, and the name of the bird is Chaffinch. In this case, you probably skip the name of the family of birds. In summary, the whole idea with the classification is to capture similarities and differences through the taxonomy.

The above reasoning can be applied to global software engineering too. In the case of the animals, we now have a well-established taxonomy, but in GSE we do not. The objective in GSE would be to build a taxonomy to identify similarities and differences between sourcing strategies (not species). On the one hand, a taxonomy in GSE provides a set of sourcing strategies (cf. species) and on the other hand it can be used to classify individual cases (cf. observations of individual birds). Based on a taxonomy, new observations, studies or cases could be compared to the existing classification. It would potentially result in improvement of our understanding of individual strategies and we could possibly identify causes for observed effects. For example, by observing the commonalities between cases that could be classified in the same way in a GSE taxonomy. Thus, the intention is to develop a taxonomy for GSE, which captures

relationships between terms identified through the expert survey.

This step was conducted in Q2 2011. The GSE strategies and relationships emerged from the literature analysis and the specific terms were identified through the expert survey. During taxonomy creation we organized brainstorming meetings and used email communication among all four of us to discuss different graphic representations and inclusion/exclusion of different levels as well as positioning of each level in the taxonomy.

2.4. Taxonomy validation

A taxonomy can be validated by demonstrating orthogonality of its dimensions, benchmarking against existing classifications and as described by Zelkowitz and Wallace (1998) through demonstrating its utility to classify existing knowledge. Orthogonality of taxonomy's dimensions is implied by design of the taxonomy, as discussed in Section 2.3. Existing classifications drove our work and the resulting taxonomy is discussed in relation to Robinson and Kalakota (2004), Ågerfalk and Fitzgerald (2008), and Höfner et al. (2011). Finally, to illustrate the use of the taxonomy and explain how it can help synthesizing existing knowledge, we classified 296 articles from the literature study according to the dimensions defined in the taxonomy. In the latter we used the following process. Furthermore, we added a scenario for usage of the taxonomy when searching for papers related to specific setting.

This step was conducted in Q1 2012. To classify existing knowledge from the 296 articles reviewed in the literature study, we started with identifying the papers containing empirical cases, i.e. description of projects or collaborations between two or more partners. We first evaluated ten randomly selected papers to define and discuss inclusion/exclusion criteria, and data extraction form. Then all remaining papers were equally split among the four researchers. The process started with selection of empirical papers with actual cases. We excluded papers that were non-empirical or empirical papers based on expert opinions, or consolidated findings from diverse cases, interviewees or survey respondents that do not provide any traces of evidence to the actual cases. In result we included 85 papers after the first screening. During the next step we divided the remaining papers so that each included paper is reviewed by a different researcher. By doing so we addressed the single researcher bias implied in the first step. During this step the second reviewer additionally evaluated each paper against the inclusion/exclusion criteria. 21 papers were excluded at this point. Data necessary to classify the papers was then extracted from the included papers. In particular, we extracted the information necessary to map the cases to the taxonomy, and to evaluate understandability of the case context, i.e. whether the authors of the papers clearly describe collaborations, provide details that can be used to deduce the necessary information, or leave the context unclear. Because we have found several cases of misused terminology, we did rely on the authors' classifications, but carefully judged the context of each study. In total, 64 papers were included in the final analysis and classified according to the taxonomy (see Section 3.4).

3. Findings

3.1. Research literature: state of the use

In this section we present the results of the first three steps of the process defined in Section 2.2.2. While there is a wide variety of terms in the GSE jargon our investigation aimed at understanding which are more popular terms, whether the terms are defined by the authors or not, and whether there is a common agreement in the usage of the terms. In total we have read 296 articles (two from 2000, 11 from 2001, one from 2002, seven from 2003, one from 2004, two from 2005, 62 from 2006, 32 from 2007, 39 from 2008, 58 from 2009, 48 from 2010, and 33 from 2011). From these articles we have extracted 23 different terms used to characterize sourcing strategies and 46 definitions. The following are the terms that are used to describe the sourcing strategies with the frequency of their occurrence in the analyzed data set (one vote per paper) and the number of collected definitions:

| | References / Definitions | | References / Definitions |
|----------------------|--------------------------|------------------------|--------------------------|
| Outsourcing | 87 / 13 | Backsourcing | 1 / 0 |
| Offshoring | 49 / 6 | Multisourcing | 1 / 0 |
| Offshore outsourcing | 29 / 8 | Onshore insourcing | 1 / 1 |
| Nearshoring | 11 / 5 | Onshore outsourcing | 3 / 2 |
| Global sourcing | 4 / 1 | International sourcing | 1 / 0 |
| Internal offshoring | 3 / 2 | National sourcing | 1 / 0 |
| Offshore insourcing | 6 / 2 | Domestic sourcing | 1 / 0 |
| Offshore sourcing | 3 / 2 | Farshoring | 2 / 0 |
| Onshoring | 3 / 1 | Rightshoring | 1 / 0 |
| Insourcing | 4 / 2 | Right-sourcing | 1 / 0 |
| Captive offshoring | 1 / 1 | Internal outsourcing | 1 / 0 |
| | | External outsourcing | 1 / 0 |

The list of reviewed articles is available in the Appendix.

In total we collected 23 different terms. The analysis suggests that sourcing strategies usually describe the movement or relocation of work or work contracts. The following factors can help to classify different strategies:

- legal entities between the “source” site and the receiving site (same/different company)
- location of the receiving company (same/different country)
- distance till the receiving site (nearby/far country)
- number of receiving sites (one/many).

Detailed analysis of the terms and their definitions led to several interesting conclusions. Comparative analysis of the extracted definitions and contexts led to deriving the following confusing and/or ambiguous uses of the terms:

- C1: The same terms are at occasions used to describe different sourcing strategies;
- C2: The same terms are at occasions used with orthogonal or mutually exclusive meanings;
- C3: Unconventional terms are used without definitions or information that helps deduce the context.

We illustrate our findings by classifying the terms in a quadrant that links them to the company and country (same/different) dimensions (see Figure 4) and later discuss examples of the confusing use referring to the three problems identified (C1, C2 and C3).

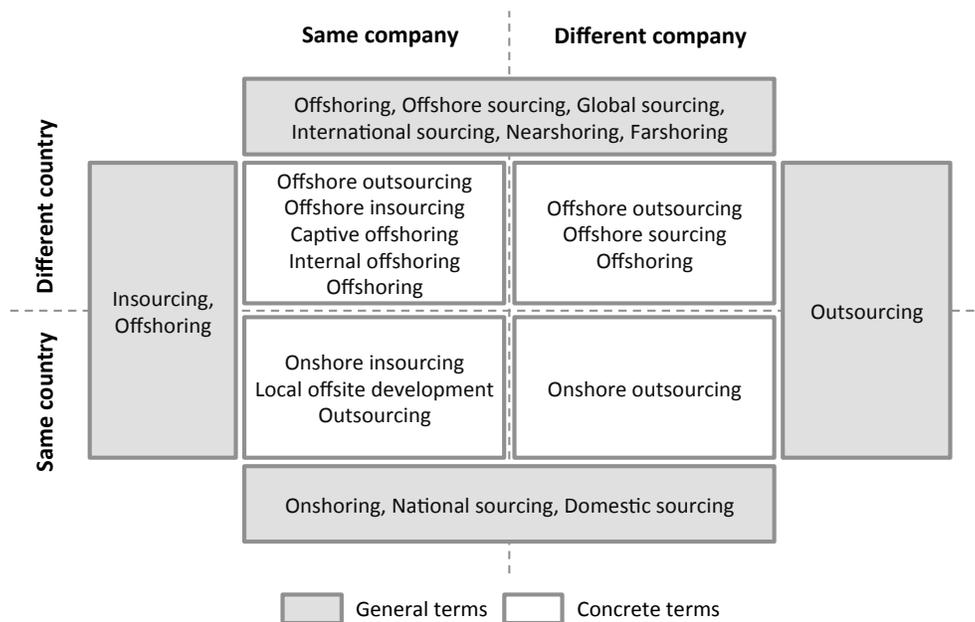


Figure 4: Use of different terms

We allocated these terms using the original definitions of the authors wherever possible. The terms that were not defined in the studied papers (*farshoring*, *international sourcing*, *national sourcing* and *domestic sourcing*) were allocated using the definitions given for the opposite terms (for example, *nearshoring* versus *farshoring*) or using patterns from other definitions and literal interpretations (for example, *domestic sourcing* – in the same country, unknown legal relationship). We were unable to locate several terms – *backsourcing*, *rightshoring*, *right-sourcing* and *multisourcing* because these terms cannot be characterized along the same dimensions, and *internal outsourcing* and *external outsourcing*, because we could not deduce their meaning.

Among the articles read, our attention was caught by the confusing use of the terms *offshoring* and *outsourcing*. The majority of the sources agree that *offshoring* is attributed to any legal entity, but different country, while *outsourcing* is commonly used to mean collaboration with a different organization in the same or different country. Nonetheless, we have identified several other uses of these terms. In particular, Casey and Richardson (2008) specify the legal entity when referring to *offshoring* (C1). They define *offshoring* as collaboration with remote divisions of the same company and oppose it to *outsourcing* (collaboration with a different organization). “Offshoring and outsourcing” are used as opposed or complementary terms explicitly or implicitly in a number of articles (Babar and Niazi, 2008; Berenbach and Gall, 2006; Burger, 2007; Casey and Richardson, 2006; Casey and Richardson, 2008; Richardson et al., 2008). This use is confusing because the authors do not specify the relationship between the terms (C1). Another confusion with the term *offshoring* is related to its use as a synonym of *offshore outsourcing* (Braun, 2007; Tervonen and Mustonen, 2009), which would directly clash with the previously described use of the term (within the organization) (C2). To make things more confusing, we found the latter combination to be used with a yet another meaning — Braun (2007) investigate *offshore outsourcing* “(in short: offshoring)” as delivery to a subsidiary and define *offshore* to mean location in a remote foreign country and *outsourcing* to refer to transferring whole work packages in contrast to pure staff augmentation or assignment of tasks (C2). Another unconventional use of the terms can be found in (Bavani, 2011), who distinguishes the work with captive centers in another location into *outsourcing* (referring to the same country), and *offshoring* (referring to “off the shore”) (C2). In summary, we have found the word *offshoring* to mean four different setups — 1) same company and same or different country, 2) same or different company and different country, 3) same company and different country, and 4) different company and different country. Similarly we have found *outsourcing* to mean 1) different company and same or

different country, and 2) same company and same country.

Even the term *onshore* is used in different ways. Szymanski and Prikladnicki (2007) use the term *onshore insourcing* to refer to “a department in the company premises or a subsidiary in the same country (onshore), which provides software development services throughout internal projects (insourcing)”. In this example confusion comes from grouping onsite projects and onshore projects under the same category (C2). We noticed that the use of the term *nearshoring* in different articles varied — in contrast to the usual meaning of geographical proximity, this term was also used to describe sourcing strategies between countries close in terms of culture and language, but geographically far, or closer to clients and end users as compared to other strategies such as offshoring (C1). Several other confusing terms were identified without any definitions or information that could be used to deduce the context (C3). For example, Hawthorne and Perry (2005) mention *internal and external outsourcing*, which might mean international collaboration within and outside the company (which would mean that the word outsourcing is used in an unconventional way). Finally, Prikladnicki et al. (2007) similarly to us describe the confusing use of the terminology in their paper by emphasizing common misconceptions related to the interchangeable use of the terms *outsourcing*, *offshore outsourcing* and *offshoring*.

It is worth noting that the general definitions tend to be ambiguous and thus the context of the GSE strategies is often unclear (this is further exemplified in Section 3.4). In practice, some terms are alleviating classification of the referred to strategies by explicitly indicating the particular case, while others are more general. Further analysis of the terms and their definitions suggests that many authors in their articles refer to different sourcing strategies, at the same time using only a few of the available terms to describe these strategies. Abstraction in the use of terminology has however an important implication — it allows the researchers to hide the diversity of the population used in empirical studies and thus endangers validity. For example, if the researchers perform analysis of diverse cases together and do not specify that the conclusions are derived from mixing, for example, experiences from collaborations with subcontractors and those with own subsidiaries, the validity of these conclusions might be questionable. However, if the researchers hide the diversity of the cases under a more general concept, the validity threat might not even be evident for the reader.

We also noticed an emerging pattern in the terminology concerning the two main classification criteria (legal relationship and location), which we refer to as “In-/Out-, On-/Off-“. In particular, we found the choice of opposing *INSourcing* versus *OUTsourcing* and *OFFshoring* versus *ONshoring* to be convenient and easy to comprehend. This pattern was further explored through the expert survey. This is in line with the general layout of quadrants discussed in Section 1.1 despite the differences in content.

Finally we note that although a certain level of convergence in the terminology exists, it goes slowly and without formalization efforts the necessary clarity and coherence might never be reached. New researchers entering the field might use terms differently if terms are not clearly defined, and while divergence in terminology as such might not be a problem, ambiguous and confusing use of terms introduces obstacles for both researchers who attempt to synthesize related research findings and practitioners attempting to learn from existing work. We further address these issues by proposing an empirically based terminology and a taxonomy for classifying different collaborative contexts.

3.2. Expert survey: defining the key concepts

The next step in our research aimed at validating the findings from the literature review through an expert survey. We saw a threat in basing our terminology proposal on the state of the use, which contains a wide variety of terms and is also criticized for occasional misconceptions and different use of terminology. Therefore, it was important to explore whether field experts can agree on a common glossary. The survey was organized in two rounds, as described in Section 2.2. After the second round of the survey we achieved a substantial agreement for the terms and hence no third round was conducted, which hence helps in responding RQ3. The survey served two main purposes. First of all, the survey aimed at achieving an expert agreement about the GSE terminology. Second of all, expert judgment and comments were subject to qualitative analysis. This led to observations about the acceptance level and reasons of disagreements in

the use of the collected GSE terms.

3.2.1. Acceptance of the terms

The results of the expert survey are summarized in Table 2. Each term is supported by a decision (Approved, Rejected, Suggested to be avoided) and an explanation of the grounds for these decisions. Furthermore, the accepted terms are supported by the level of agreement. As a response to RQ3, it is worth noting that the experts were in most cases able to agree quite well.

The table content shall be read as follows. From the table we can see that the term *sourcing*, collected from the literature, was approved in the first round with a substantial agreement from the experts as being a generic term that does not specify legal relationship or location of the collaborating parties; while the term *multisourcing*, collected from the literature and associated with the same generic meaning, was suggested to be avoided by the experts who achieved substantial agreement on this decision in the second round; and so forth.

Notably, the key concepts that we referred to as “In-/Out, On-/Off-“ pattern, were well received by the experts. The results demonstrate a consensus about acceptance of the terms *insourcing*, *outsourcing* and *offshoring*, while the term *onshoring* was accepted with slightly lower but nonetheless substantial agreement. Interestingly, one of the experts liked the idea of having similar structure for these terms, but considered it somewhat uneasy. An expert admitted — *Instead of “onshore” I would use “domestic” and instead of “insourcing” I would use “in-house development”*. Similar opinions were expressed about several other terms, indicating that certain terms are more intuitive and easy to understand and that convergence prevails in the GSE jargon today. While researchers are likely to use different terms, the role of consistency in the GSE terminology, from our point of view, is often neglected. This however has one important implication on the inability to search and find related research literature and in some cases interpret the suitability of the results to the reader’s context. For example, the mentioned term *in-house development* as alternative to *insourcing* has been noted in the comments with two different meanings — as work performed within a company, and as work performed in a single site. Thus we emphasize the importance of a unified terminology and consistent use of the terms.

Confusions were also identified regarding the term *onshoring*, which might be the reason for the lack of consensus, and other compound terms such as *onshore insourcing* and *offshore insourcing*. In contrast to the widely used definitions both terms *onshoring* and *onshore insourcing* were confused with the concept of co-located development. The use of the adjective *onshore* instead of more common *on-site* was also noticed in our literature analysis.

Another expert commented on the use of the term *offshore insourcing* as being complex. The expert confessed that in previous research referred to the work performed in a different country within the same company as *offshoring*, and suggested to reserve this term with that meaning. The difficulty in using the term *offshoring* has been acknowledged earlier. In contrast to associating *offshoring* with collaboration within the boundaries of the company, Prikladnicki et al. (2007) noticed a common misconception when *offshoring* is attributed to the global application of *outsourcing*. Ågerfalk and Fitzgerald (2008) also observe that the two orthogonal terms *offshoring* and *outsourcing* are often used as synonyms. Despite the previous varying uses of the GSE terms we suggest focusing on the future use that aims at avoiding (or decreasing) further misunderstandings and confusion.

Table 2: List of the key sourcing strategies

| Activity | Legal entities | Locations | Terms (Approved terms highlighted) | Sources | Decisions* | Level of agreement** |
|---------------------------|----------------------|----------------------------------|---------------------------------------|---|--|----------------------|
| Collaboration | - Any - | - Any - | Sourcing | Literature review | Approved in Round 1 | Substantial |
| Collaboration | - Any - | - Global - | Multisourcing | Literature review | Suggested to be avoided in Round 2 | Substantial |
| Collaboration | Within the company | - Any - | Global Software Engineering | Experts, Round 1 | Approved in Round 2 | Substantial |
| Collaboration | Within the company | Within the country | Insourcing | Literature review | Approved in Round 1 | Perfect |
| Collaboration | Within the company | Within the country | Offshoring | Experts, Round 2 | Rejected due to a conflict with an accepted term | Substantial |
| Collaboration | Within the company | Outside of the country | Onshore insourcing | Literature review | Approved in Round 2 | Substantial |
| Collaboration | Within the company | Outside of the country | Domestic insourcing | Experts, Round 1 | Rejected in Round 2 due to a poor agreement | Perfect |
| Collaboration | Within the company | Outside of the country | In-house development | Experts, Round 1 | Rejected in Round 2 due to a poor agreement | Perfect |
| Collaboration | Within the company | Outside of the country | Offshore insourcing | Literature review | Approved in Round 2 | Perfect |
| Collaboration | Within the company | Outside of the country | Captive offshoring | Literature review | Rejected in Round 2 due to a poor agreement | Perfect |
| Collaboration | Within the company | Outside of the country | Internal offshoring | Literature review | Rejected in Round 2 due to a poor agreement | Perfect |
| Collaboration | Within the company | Far outside of the country | - | Not included in the survey since these can be compound from other terms | | |
| Collaboration | Within the company | Outside of the country but near | - | | | |
| Collaboration | With another company | - Any - | Outsourcing | Literature review | Approved in Round 1 | Perfect |
| Collaboration | With another company | Within the country | Onshore outsourcing | Literature review | Approved in Round 2 as compound term | Perfect |
| Collaboration | With another company | Outside of the country | Domestic outsourcing | Experts, Round 1 | Rejected in Round 2 due to a poor agreement | Substantial |
| Collaboration | With another company | Outside of the country | Offshore outsourcing | Literature review | Approved in Round 2 | Substantial |
| Collaboration | With another company | Outside of the country | Global outsourcing | Experts, Round 1 | Rejected in Round 2 due to a poor agreement | Substantial |
| Collaboration | With another company | Outside of the country | International outsourcing | Experts, Round 1 | Rejected in Round 2 due to a poor agreement | Substantial |
| Collaboration | With another company | Far outside of the country | - | Not included in the survey since these can be compound from other terms | | |
| Collaboration | With another company | Outside of the country, but near | - | | | |
| Collaboration | - Any - | Within the country | Onshoring | Literature review | Approved in Round 2 | Substantial |
| Collaboration | - Any - | Within the country | National sourcing | Literature review | Suggested to be avoided in Round 2 | Perfect |
| Collaboration | - Any - | Within the country | Domestic sourcing | Literature review | Rejected in Round 2 | Substantial |
| Collaboration | - Any - | Outside of the country | Collocated development | Experts, Round 1 | Rejected in Round 2 due to poor agreement | Perfect |
| Collaboration | - Any - | Outside of the country | Offshore outsourcing | Literature review | Rejected due to a conflict with related term | Perfect |
| Collaboration | - Any - | Outside of the country | Offshoring | Literature review | Approved in Round 2 | Perfect |
| Collaboration | - Any - | Outside of the country | Global sourcing | Literature review | Suggested to be avoided in Round 2 | Substantial |
| Collaboration | - Any - | Outside of the country | International sourcing | Literature review | Suggested to be avoided in Round 2 | Substantial |
| Collaboration | - Any - | Outside of the country | Offshore sourcing | Experts, Round 1 | Rejected in Round 2 due to poor agreement | Substantial |
| Collaboration | - Any - | Far outside of the country | Farshoring | Literature review | Suggested to be avoided in Round 2 | Substantial |
| Collaboration | - Any - | Outside of the country, but near | Nearshoring | Literature review | Approved in Round 1 | Substantial |
| Return from collaboration | - Any - | - Any - | Backsourcing | Experts, Round 1 | Rejected in Round 2 due to a poor agreement | Substantial |

* The terms are *approved* on the basis of 1) expert agreement (indicated in column Level of agreement) or 2) as a term compound of the approved terms; *suggested to be avoided* on the basis of expert agreement (indicated in column Level of agreement); and *rejected* on the basis of 1) disagreement among the experts or 2) inconsistency with the other terms, or 3) on the basis of expert agreement.

** The level of agreement is calculated using Fleiss kappa: 0.61-0.80 — Substantial agreement, 0.81-1.00 — Almost perfect agreement, while kappa ≤0.6 is regarded as disagreement.

3.3. Glossary

For complete support we supplement the key terms that describe sourcing strategies with definitions. While experts achieved substantial agreement on the majority of the GSE terms, slightly lower agreement was achieved for their definitions. The experts seemed to have a low agreement in their preferences for formulating definitions. One of the possible explanations could be related to the variety of suggestions expressed in the first round. These ultimately generated more options to choose from during the second round. Since the changes were not significant, we decided to propose the definitions based on what we have achieved through the related literature analysis, and incorporate the changes proposed and agreed by the experts that targeted the meaning rather than the choice of words.

With this glossary we would like to emphasize the importance of differentiating the combinations of terms used in the GSE research papers. Our findings indicate that there are stand-alone terms and terms used in a combination with other terms, and they can have a different meaning. This puts demands on the authors to use GSE jargon carefully. Our analysis suggests that these words and combination of words provide the most accurate characteristics of the sourcing strategies (in alphabetical order):

Global software engineering

Development of a software artifact across more than one location

Insourcing

Leveraging company-internal human resources

Nearshoring

Leveraging resources from a neighboring country

Offshore insourcing

Leveraging company-internal resources situated in a different country

Offshore outsourcing

Leveraging external third-party resources situated in a different country

Offshoring

Leveraging resources from a different country

Onshore insourcing

Leveraging company-internal resources situated in the same country

Onshore outsourcing

Leveraging external third-party resources situated in the same country

Onshoring

Leveraging resources from the same country

Outsourcing

Leveraging external third-party resources

Sourcing

Leveraging resources

3.4. GSE Taxonomy

The GSE taxonomy (see Figure 5) uses the terminology identified through the expert survey as a starting point. The taxonomy extends on the terminology by making relationships explicit and defining in particular geographical distance and time differences into different classes. The objective has been to formulate the taxonomy so that a researcher or practitioner could use the taxonomy, or the hierarchical structure to identify the branch or the branches that describe their specific situation. The taxonomy is derived in response to RQ4. At the end, it turns out that the first levels of the taxonomy use many of the terms defined by, for example, Robinson and Kalakota (2004), Ågerfalk and Fitzgerald (2008) and Höfner et al. (2011), and also described above. However, we have taken it further by formulating a taxonomy based on the terminology from a literature study (to identify potential terms), an expert survey (to identify a set of terms to recommend) and then complement the terms to extend them into a taxonomy. The intention is that it should provide direct support for users to identify their GSE strategy or to put their specific case into a broader context.

The taxonomy is divided into five levels and illustrated in Figure 5:

1. GSE – The starting point is sourcing, i.e. some form of external development from the viewpoint of those being responsible whether being for a project, product, system or service.
2. Location – The sourcing can be onshore, i.e. within the same country, or offshore that is in another country. Here it is definitively preferable to provide the names of the countries involved as well as describing the actual responsibilities of the different geographical locations involved. It should be noted that this is by no means delimited to two locations. We may have many participants in different geographical locations involved in the development.
3. Legal entity – The next level is concerned with whether the development is conducted within the same company or with another company, for example by subcontracting. Insourcing means at another site of the company doing the sourcing and outsourcing is with a different legal entity.
4. Geographical distance

The geographical distance is in many cases larger when considering offshoring even if some countries are quite large. Thus, we have decided to have different definitions for when it comes to geographical distance in onshoring and offshoring situations respectively.

Onshore: In addition, we would like to consider the geographical distance even if the development is done onshore. The actual development situation may be very different if it is possible to meet face-to-face on a frequent basis or if it requires air travelling. It is a big difference between being in the same city or being, for example, a collaborative venture between the west coast and the east coast in the US (with three hours time difference and five hours by air travelling). Thus, we would like to distinguish between the following situations and define:

- **Close** – No flights are needed and hence it is possible to have quite frequent face-to-face meetings.
- **Distant** – A flight is needed, which means that the costs and the time to meet increase.

Offshore: When it comes to offshoring situations it could be countries located close or further away, and hence we want to distinguish between nearshoring and farshoring. We would like to introduce the following definitions:

- **Near** (or nearshoring) – The flying time is less than two hours. This is motivated by the fact that it is possible to travel back and forth in a day and still have time for a 3-4 hour meeting.
- **Far** (or farshoring) – Flying time is two hours or more, and hence it normally includes staying overnight.

5. Time difference

Both in onshoring and offshoring situations, we may have time difference. However, they have some different characteristics. Thus, we would like to differentiate between offshore and onshore.

Onshore: In many countries there are no time differences, but in other larger countries we may have several hours time difference. Thus, in general small differences are expected within a country, but longer distances may happen too. We would like to define the following for time differences within a country:

- **Similar:** The time difference is one hour or less. This means that there is almost a full overlap in work hours at the different sites. One hour has been included in similar since two sites could be close, but in two times zones, for example in two neighbouring states in the US in

different time zones.

- **Different:** The time difference is more than one hour.

Offshore: In an offshore situation, a time difference is expected, and hence the limits for what is perceived as small and large have to be defined differently than in the onshore situation. For offshoring, non-overlapping work hours are expected and hence it is more a matter of how many overlapping hours there are between two sites. Thus, we would like to make the following definitions:

- **Small:** A small time difference is defined as being four hours or less. This limit is set since it means that at least half of a normal workday is overlapping.
- **Large:** Accordingly, a large time difference is defined as being more than four hours, which then imply a rather small overlap in work hours.

Given the above definitions, we obtain the taxonomy presented in Figure 5. It should be observed that some combinations become infeasible. In total, four combinations are impossible. These are marked separately in Figure 5. They are combinations where the geographic distance is either close (onshoring case) or near (offshoring case) and in combination with the time difference being different and large respectively. For example, a flight is needed if the time difference is more than one hour and hence the combination close and different is infeasible. A similar reasoning can be done for combination of near and large.

In addition to the taxonomy in Figure 5, it is essential to capture and describe cultural differences. It is well known that, for example, UK and Australia have quite a number of commonalities culturally even if they are far apart geographically, while the cultural differences between other countries that are closer may be considerably larger. Having said this, the taxonomy provides a starting point for others to also understand the cultural differences given that the geographical locations are included in the taxonomy.

The usage of the taxonomy for a researcher documenting a study may be illustrated with one of our own studies, Smite and Wohlin (2010), of a product being development jointly between two Ericsson sites. The development is done between a site in Sweden and a site in India, and hence it is an offshoring situation. Given that both sites are fully owned by Ericsson, it is an insourcing situation. With the countries involved it becomes farshoring (or far) based on the travel time by air being 9-10 hours. Finally, it becomes small given that the time difference is 3-4 hours. Thus, the branch for this example in Figure 5 is: Sourcing – Offshore – Insourcing – Far – Small. This information provides important information for anyone interested in understanding the study and to judge whether it is relevant from his or her perspective, whether conducting GSE research or looking for experiences that may be relevant for someone else moving into a sourcing situation.

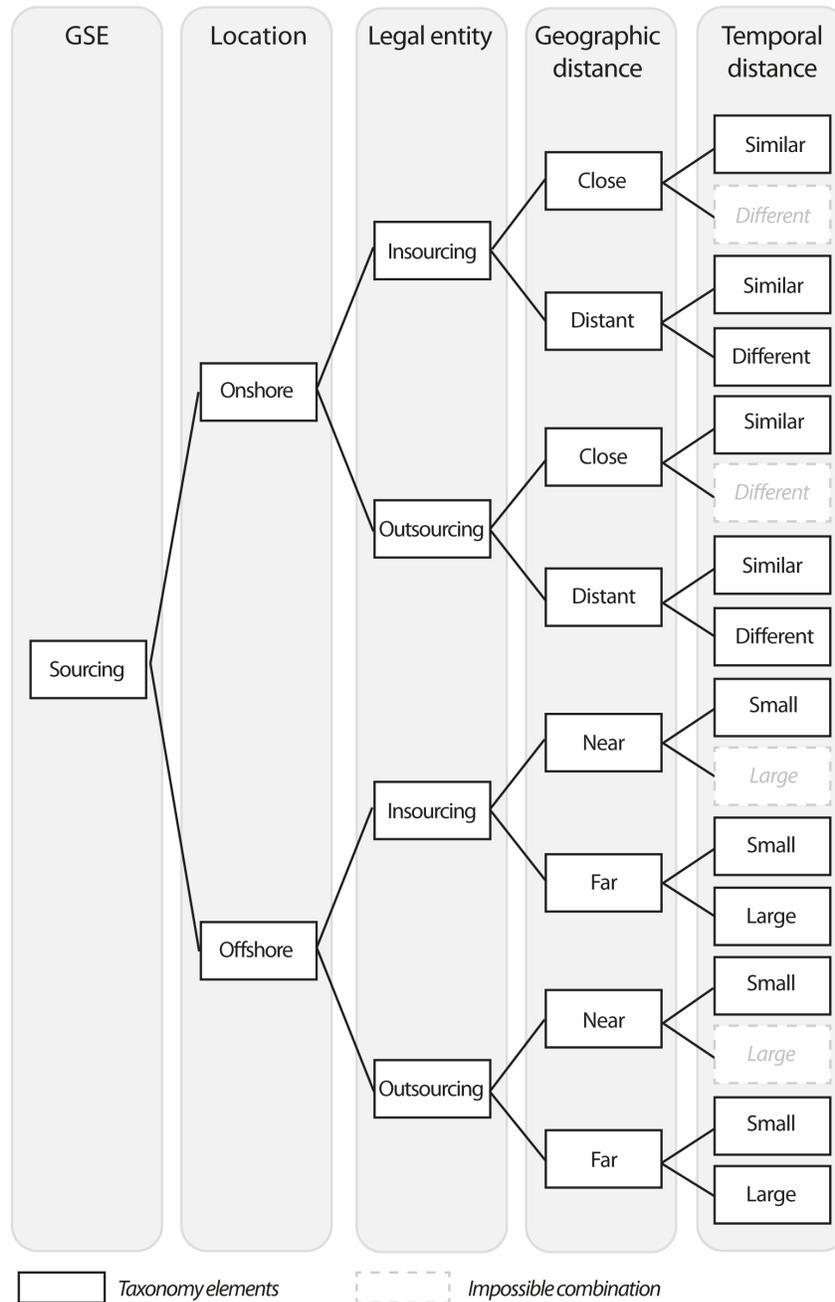


Figure 5: GSE Taxonomy

The taxonomy presented in Figure 5 captures eight different sourcing strategies and it provides support in classifying different GSE studies. The objective is that usage of the taxonomy should help researchers and practitioners alike, and hopefully ensure that it becomes easier to understand different papers documenting findings and experiences from GSE. The taxonomy has two primary usages. First of all, 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. Researcher can use it to synthesize existing knowledge, identify gaps in the literature and to find related work, and practitioners can use it to find evidence of a specific question in a context that is similar to their own. The use of a common terminology and taxonomy will simplify comparison between studies in different settings.

To validate the usage of the taxonomy we first demonstrate in detail how studies in four papers can be mapped to the taxonomy and then provide a summary of classifying research papers included in the literature study. The four research papers were selected to provide an in-depth understanding of the utility of the taxonomy for classification purposes (Höfner and Mani, 2007; Tervonen and Mustonen, 2009, Caprihan, 2006; Cristal et al., 2008). All four papers contain specific empirical cases, which are mapped to the taxonomy. The results are summarized in Figure 6, and each case is elaborated in some more detail below. For each paper, both the classification based on the taxonomy and the related description in the paper is provided. In several cases, the classification can be deduced from the contextual description, but it does require a detailed reading of each paper to find the relevant information. If the authors had provided a classification as done in Figure 6, it would have been much easier to capture the context of each study. In addition to the classification, it is important that the actual work conducted in each site is reported and that the responsibility of each site is provided, in particular in relation to the other site(s). Aspects to cover in a contextual description are further elaborated in Petersen and Wohlin (2009).

| Description and references | | Location | Legal entity | Geographic distance | Temporal distance | Setups |
|--|-----------|-----------|--------------|---------------------|-------------------|----------------------------|
| Paper 1: Höfner and Mani (2007) Single case | | | | | | |
| Germany - India | Taxonomy: | Offshore | Outsourcing* | Far* | Small* | GER-MANY — INDIA |
| | In paper: | Offshore | Deducible | Deducible | Deducible | |
| Paper 2: Tervonen and Mustonen (2009) Multiple case (here three cases) | | | | | | |
| (a) Finland – India | Taxonomy: | Offshore | Outsourcing | Far* | Small* | FIN-LAND — INDIA |
| | In paper: | Offshore | Outsourcing | Deducible | Deducible | |
| (b) Finland – India | Taxonomy: | Offshore | Outsourcing | Far* | Small* | FIN-LAND — INDIA |
| | In paper: | Offshore | Outsourcing | Deducible | Deducible | |
| (c) Germany – India | Taxonomy: | Offshore | Outsourcing | Far* | Large* | GER-MANY — INDIA |
| | In paper: | Offshore | Outsourcing | Deducible | Deducible | |
| Paper 3: Caprihan (2006) Staged case (here three sites) | | | | | | |
| USA 1 – USA 2 | Taxonomy: | Onshore* | Outsourcing* | Unclear | Unclear | USA 1 — USA 2 |
| | In paper: | Deducible | Deducible | Not described | Not described | |
| USA 2 - India | Taxonomy: | Offshore | Insourcing* | Far* | Large* | INDIA |
| | In paper: | Offshore | Deducible | Deducible | Deducible | |
| Paper 4: Cristal et al. (2008) Circular case (here three sites) | | | | | | |
| USA – Brazil | Taxonomy: | Offshore | Insourcing* | Far* | Unclear | USA — BRAZIL — CHINA |
| | In paper: | Offshore | Deducible | Deducible | Not described | |
| Brazil – China | Taxonomy: | Offshore | Insourcing* | Far* | Large* | BRAZIL — CHINA |
| | In paper: | Offshore | Deducible | Deducible | Deducible | |
| USA – China | Taxonomy: | Offshore | Insourcing* | Far* | Large* | BRAZIL — CHINA |
| | In paper: | Offshore | Deducible | Deducible | Deducible | |

* The values for taxonomy dimensions are deduced from the context

Figure 6: Illustration of classifications using the taxonomy

The first paper (Höfner and Mani, 2007) contains a single case and describes an offshore relationship, where most of the information is deducible but not expressed explicitly in the paper. A mapping of the single case and collaborations between two sites requires extracting information from the paper and to map the case to our four dimensions. This is the simplest example of taxonomy application.

In Tervonen and Mustonen (2009) three different cases are presented, and from the extracted information it is notable that the authors use the terms (offshore and outsourcing) as suggested in the terminology. However, the other information has to be deduced from the contextual description. Once again, it illustrates the challenges related to finding literature describing a similar situation as any company's specific

situation.

As described above, the taxonomy can be used even when having more than two sites. This is done by considering each combination of sites using the taxonomy as illustrated through the paper by Caprihan (2006). In Caprihan (2006), there is a two-stage relationship. One company outsources work to a company in the same country (USA). The company being outsourced to conducts most of the work at their site in India. This paper illustrates how the taxonomy can be used when more than two sites are involved. In reality, there are three potential relationships, but it seems from the paper, as there is no direct contact with the customer in the US and the people working in India. Thus, only two relations are described in Figure 6. The paper contains very little information about the relationship between the customer and the company being outsourced to in the US. The main focus is on the insourcing relationship in the company being outsourced to. It is noteworthy that by using the taxonomy, it becomes clear that quite a lot of information is missing to fully understand the case described.

An example of a more complex collaboration with three sites is illustrated by mapping the paper by Cristal et al. (2008). In this paper one of the two empirical cases describes a circular collaboration among three geographically dispersed sites of the same company, which we demonstrate through three pairs of sites mapped to the taxonomy.

To further demonstrate the utility of the taxonomy we have reviewed the same 296 research papers that were screened for the development of the terminology. We then selected empirical papers that contain actual cases (64 papers) and classified the sourcing relationships according to the taxonomy. We followed the data extraction process described earlier in this section (illustrated in Figure 6). Out of the 64 papers, the majority of the papers (36 papers) contained descriptions of one or more empirical cases on offshore insourcing, 21 papers described offshore outsourcing, six papers – onshore insourcing and one paper – onshore outsourcing relationship (see summary on Figure 7). We also identified studies with either unclear legal relationship or location (colored in grey in Figure 7). Note that one paper might appear in different categories if it contains several empirical cases. We have also found papers that contained clearly described and unclearly described cases. For example, Herbsleb et al. (2005) present nine global projects in Siemens; it was only possible to clearly classify two cases. Several other papers contained descriptions, which were so complicated that even after carefully reading the paper it was impossible to deduce the context and thus use the experiences to learn from.

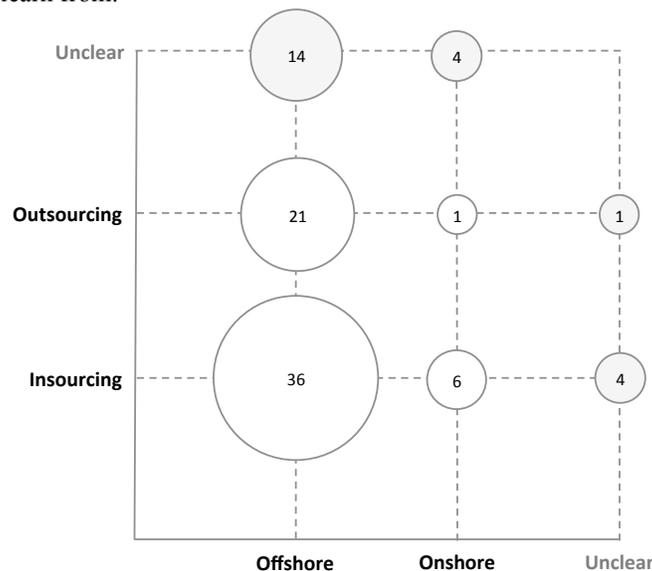


Figure 7: Summary of the papers

In summary, having a clear classification of each paper simplifies finding papers describing a specific sourcing situation. This is particularly important for researchers looking for related work or practitioners attempting to learn from existing research related to their contexts or topics of interest.

Finally, we use one particular scenario to demonstrate the utility of our taxonomy in relating existing knowledge to a particular situation. Consider that a company or a researcher looks for evidence from empirical experiences with establishing outsourcing collaboration between USA and India. In need of empirical research related to outsourcing from USA to India, the first step would be to conduct a search for papers containing keywords, for example, “offshore outsourcing”. However, challenges in interpreting the terminology and different cases of misuse discussed in Section 3.1 suggest that such search might lead to limited results. In the following we describe how a classification of GSE papers, which we performed for this study, helps to navigate through the classified knowledge and identify articles relevant in the described situation.

Figure 8 summarizes empirical cases (the 64 papers) that we were able to map to the taxonomy. We provide references to pair collaborations that we deduced from literature and provide respective references to the reviewed articles listed in the Appendix. According to the taxonomy outsourcing collaboration between USA and India would be classified as Offshore Outsourcing on a Far geographic distance and a Large temporal distance. To identify potentially interesting cases, we may choose to use different steps, here illustrated with three steps:

1. Look for cases with exact match offshore-outsourcing-far-large and USA-India.
On the respective leaf of the taxonomy we find references to four research papers (23, 73, 118 and 277) that are most relevant to the selected topic of interest and specifically address experiences from collaborations between USA and India.
2. Look for similar cases, for example North America – Asia.
If the findings from step 1 are judged insufficient, we may look at cases of offshore outsourcing collaboration between North America and India (from the same taxonomy leaf). We find one paper with collaboration North America and India (7), one paper with collaboration USA and Japan (23) and one paper with collaboration Canada and India (115). This step illustrates cases with one of the main countries of interest being involved (USA or India) and collaboration between the same continents.
3. Look for more remote cases, for example western countries to Asia.
In this step, the following cases can be identified from the classification in Figure 8: four papers describing collaboration between Germany and India (10, 104, 105, and 180) and one paper with collaboration between Belgium and Asia (278).
Other options exist when it comes to more remote cases and the actual choice of the most relevant cases depends on the actual setting, i.e. it must be judged on a case-to-case basis which cases are relevant to study. Some other examples include: outsourcing from USA, outsourcing to India, insourcing from USA to India and other possible experiences that could be indirectly related, such as collaborations on a far geographic distance and a large temporal distance.

This example results in finding 11 potentially interesting papers. We can of course not expect all papers to be classified in one place as in Figure 8. However, if papers being published were classified using the taxonomy then searching for relevant cases would be considerably simplified for researchers and practitioners when looking for similar cases to the context of their interest.

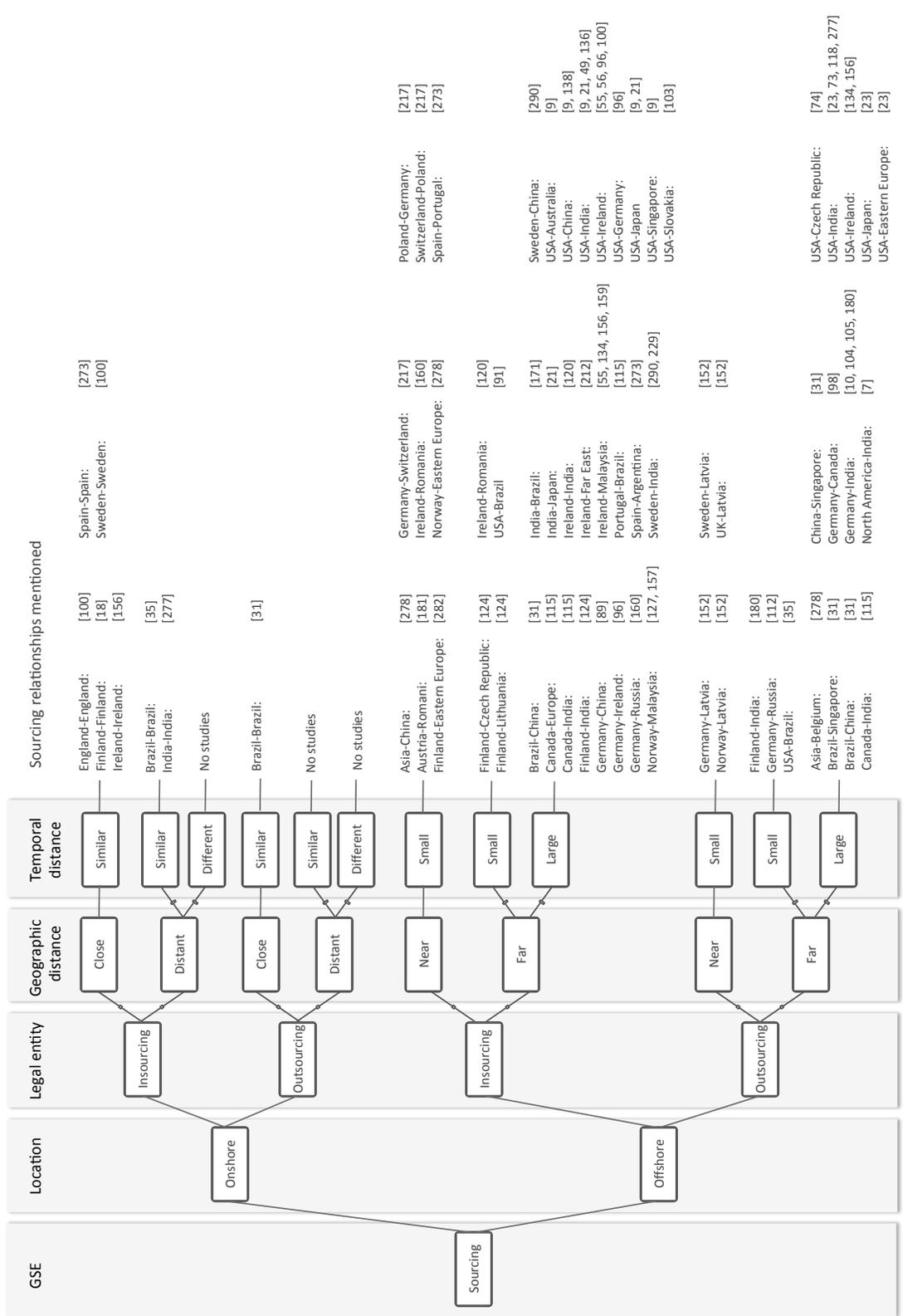


Figure 8: Summary of classified papers using the taxonomy

4. Conclusions

4.1. Principal findings

In response to **RQ1** we have found that there is a variety of terms used to describe different collaboration modes in GSE projects, and different terms are found to describe the similar concepts. The main difference in the state of the use of different GSE terms is related to the amount of detail provided by the authors, which indicates a possibility to define generalization-specialization relationships among the terms.

When it comes to **RQ2**, we observed that the terms were rarely defined in the studied papers, and the definitions for many terms cannot be found in dictionaries. Furthermore, terms are used inconsistently and authors have a tendency to use their own preferred terms. Analysis of the state of the use of terminology provided examples of misleading usage emphasized by Carmel and Tjia (2005). We have found that the same terms at occasions are used with orthogonal or mutually exclusive meanings. This makes it very difficult when trying to get an overview of the state of the art and potentially synthesize results from different research studies. Although a certain level of convergence exists, it goes slowly and without formalization efforts might not succeed to reach the necessary clarity and coherence. New researchers entering the field might continue to use terms differently if the terms are not clearly defined. These findings emphasize the need and the importance of our contribution. The need for consistent reporting in GSE was also stressed in Smite et al. (2008).

On the way towards building the GSE terminology, we organized a Delphi-inspired survey involving field experts and in relation to **RQ3** strived for an agreement on a common terminology. Our observations about the acceptance of different terms among ten GSE experts suggest that the key concepts were accepted with at least substantial agreement, while the vast majority of the synonyms were considered with varying opinions. We conclude that without a defined terminology, researchers are likely to choose different terms and thus complicate the search of related research for a reader, and ability to judge the applicability of research findings for the reader's context. Furthermore, it makes research synthesis as aimed for in systematic literature reviews overly complicated.

We have found that the proposed and accepted terms do not entirely illustrate the current use of the terms. Several experts have regarded some terms as complex, uneasy to understand, and reported confusion of these terms in their previous use. Nonetheless, we suggest that despite the past experiences, the GSE field yields for a common terminology and our contribution — the taxonomy, which is developed in response to **RQ4**, and glossary — shall support the consistent usage of the GSE jargon.

4.2. Target audience

While the main purpose of the terminology and taxonomy proposed is to facilitate the understandability and ability to transfer the knowledge from research into practice, the target audience for our contribution is twofold. First of all, we address researchers who will publish or synthesize further work (see Section 4.3 for implications for future research). Secondly, we address practitioners, who are interested in published empirical cases (see Section 4.4. for implications for practice). We emphasize the importance of contextual information in empirical research, which ought to be understandable and we see an inconsistent or diverse use of terminology as an obstacle. The terminology and taxonomy described in this paper are intended to serve as a roadmap for practitioners in distinguishing different contexts and recognizing related cases.

4.3. Implications for future research

The contribution of this paper may be viewed as twofold. First and foremost, the paper provides a terminology and taxonomy of GSE. Second, it illustrates how a terminology and taxonomy can be based on an empirical study and used to map existing empirical studies.

The implication of the first contribution is that terminology and taxonomy can be used in future research on GSE. Conducting literature reviews in a specific domain could be harder if terms are not well defined or

standardized. This is specifically the case for new areas of research such as GSE. We believe that common terminology and taxonomy in new research fields would make it easier for a community to build upon each other's work and hence making progress more quickly.

The taxonomy can be also used to summarize existing knowledge and identify the gaps by linking empirical cases to the branches of the taxonomy (Figure 5, Section 3.4) as illustrated in our example (Figure 6).

The method used to identify a terminology and hence a basis for proposing a taxonomy may be used in other areas of software engineering. Thus, through an investigation of the use of terminology in the relevant research literature it is possible to identify commonalities and differences in the use of terms in a field. Based on the identified terms, it is possible to conduct an expert survey to build an agreement on a common set of terms and definitions. Empirically based terminology and taxonomies can help in the maturation of an area and hence it should be possible to use or adapt the process described in Figure 3 in other domains or subareas of research in software engineering.

4.4. Implications for practice

While global arrangements are enabled through a number of diverse scenarios, it is fair to assume that what works in one context does not necessarily apply in another (Šmite et al. 2008). Understanding the applicability of research findings in a particular context is thus important. Our experiences from analyzing the state of the use of GSE terminology and attempts to map empirical cases to the taxonomy suggest that it is often difficult to clearly understand the context of the global arrangements described in the GSE literature to date. This means that learning from existing knowledge can be challenging.

Another implication for practice is that by having a formal taxonomy about GSE, it is possible to systematically accumulate experiences and navigate easily through existing knowledge, as illustrated in this paper with the offshore outsourcing situation from USA to India. It is also possible to develop lexical resources to automatically extract the information about contextual descriptions and integrate it into planning or decision support tools. For example, organizations can use the terminology defined and add the contextual information in the project management tool. Information about similar cases can be extracted and analyzed together with performance and quality measures, and then can be used to support decisions or perform simulations in industrial projects based on a common classification.

We believe that our observations about the diverse use of terminology will make practitioners aware of potential misleading use of terms and help becoming more attentive to contextual descriptions. Additionally, mapping projects to the taxonomy shall support a better understanding of the types of global arrangements that are relevant for a particular context, and those that are not. Distinguishing the different scenarios based on the dimensions of our taxonomy shall also warn practitioners of the potential threats to validity and trustworthiness of the findings from mixed studies, i.e. studies based on several diverse cases.

Acknowledgements

We are thankful to all experts who have kindly accepted our invitation and contributed to this initiative, and especially for their dedication and interest in the results. This kept us motivated. We would like to express our thanks to the reviewers, whose constructive comments helped improve the paper.

The research by Dr. Darja Šmite and Prof. Claes Wohlin is supported by the following research grants: R2D2 (2009/0249) and BESQ+ (20100311) from the Knowledge Foundation in Sweden. Dr. Darja Šmite and Zane Galviņa are also supported by the European Social Fund through a research project "Application of computer science and its links to quantum physics" at the University of Latvia. Prof. Rafael Prikladnicki holds research grants from CNPq (550130/2011-0, 483125/2010-5, and 560037/2010-4), from Ci&T, and from the PDTI program, financed by Dell Computers of Brazil Ltd. (Law 8.248/91).

References

- Babar, M, A, Niazi, M (2008) Implementing Software Process Improvement Initiatives: An Analysis of Vietnamese Practitioners' Views. In: proceedings of the IEEE International Conference on Global Software Engineering ICGSE, pp. 67-76.
- Barney, S, Wohlin, C, Chatzipetrou, P, Angelis, L (2011) Offshore Insourcing: A Case Study on Software Quality Alignment. In: proceedings of the 6th IEEE International Conference on Global Software Engineering ICGSE, pp. 146-155.
- Battin R D, Crocker R, Kreidler J, Subramanian K (2001) Leveraging resources in global software development. In: IEEE Software 18(2):70-77
- Bavani, R (2011) Governance Patterns in Global Software Engineering: Best Practices and Lessons Learned. In: proceedings of the 6th IEEE International Conference on Global Software Engineering ICGSE, pp. 50-54.
- Berenbach, B, Gall, M (2006) Toward a Unified Model for Requirements Engineering. In: proceedings of the 1st IEEE International Conference on Global Software Engineering ICGSE, pp. 237-238.
- Braun, A (2007) A Framework to Enable Offshore Outsourcing. In: proceedings of the 2nd IEEE International Conference on Global Software Engineering ICGSE, pp. 125-129.
- Burger, W (2007) Offshoring and Outsourcing to INDIA. In: proceedings of the 2nd IEEE International Conference on Global Software Engineering ICGSE, pp. 173-176.
- Caprihan, G (2006) Managing Software Performance in the Globally Distributed Software Development Paradigm. In: Proceedings of the 1st International Conference on Global Software Engineering, 83-91.
- Carl Linnaeus in Wikipedia (2011) http://en.wikipedia.org/wiki/Carl_Linnaeus
- Carmel E, Tjia P (2005) Offshoring Information Technology: Sourcing and Outsourcing to a Global Workforce. Cambridge University Press, NY
- Carmel E Abbott P (2007) Why 'Nearshore' Means That Distance Matters. In: Communications of the ACM, 50(10):40-46
- Carmel E, Agarwal R (2001) Tactical Approaches for Alleviating Distance in Global Software Development. In: IEEE Software, 18(2): 22-29
- Casey, V, Richardson, I (2006) Project Management within Virtual Software Teams. In: proceedings of the 1st IEEE International Conference on Global Software Engineering ICGSE, pp. 33-42.
- Casey, V, Richardson, I (2008) The Impact of Fear on the Operation of Virtual Teams. In: proceedings of the 3rd IEEE International Conference on Global Software Engineering ICGSE, pp. 163-172.
- Cristal M, D. Wildt and R. Prikladnicki, (2008) Usage of SCRUM Practices within a Global Company. In: Proceedings of the IEEE International Conference on Global Software Engineering, pp. 222-226.
- Ebert C, De Neve P (2001) Surviving global software development. In: IEEE Software 18(2): 62-69.
- Gumm, D C (2006) Distribution Dimensions in Software Development Projects: A Taxonomy. In: IEEE Software, 23(5): 45-51.
- Gwet, K L (2010) Handbook of Inter-Rater Reliability, Second edition. Advanced Analytics, LLC.
- Hawthorne, M, J, Perry, D, E (2005) Software Engineering Education in the Era of Outsourcing, Distributed Development, and Open Source Software: Challenges and Opportunities. In: Proceedings of the 27th International Conference on Software Engineering, pp. 643-644.
- Herbsleb J D, Paulish D J, Bass M (2005) Global Software Development at Siemens: Experience from Nine Projects. In: Proceedings of the 27th International Conference on Software Engineering, pp. 524-533.
- Hossain E, Babar M A, Pail H-Y (2009) Using Scrum in Global Software Development: A Systematic Literature Review. In: Proceedings of the 4th International Conference on Global Software Engineering, 175-184.

- Höfner, G, Mani, V S (2007) TAPER: A generic framework for establishing an offshore development center. In: Proceedings of the 2nd International Conference on Global Software Engineering, 162-172.
- Höfner, G, Mani, V S, Nambiar, R, Manoy A (2011) Fostering a High-Performance Culture in Offshore Software Engineering Teams Using Balanced Scorecards and Project Scorecards. In: Proceedings of the 6th International Conference on Global Software Engineering, 35-39.
- Jalali, S, Wohlin, C (2010) Agile Practices in Global Software Engineering - A Systematic Map. In: proceedings of the 5th IEEE International Conference on Global Software Engineering ICGSE, pp. 45-55.
- Jordan W J, Miller S R (2003) Inter-Rater Agreement in Analysis of Open-Ended Responses: Lessons from a Mixed Methods Study of Principals. *Knowledge Creation Diffusion Utilization*, 1-5.
- Landis J R, Koch G G (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 33, pp.159-174.
- Meyer M A, Booker J M (2001) Eliciting and Analyzing Expert Judgment: A practical guide. ASA-SIAM Series on Statistics and Applied Probability
- Okoli C, Pawlowski S D (2004) The Delphi Method as a Research Tool: An Example, Design Considerations and Applications. In: *Information & Management*, 42, pp.15–29
- Poikolainen T Paananen J (2007) Performance Criteria in Inter-Organizational Global Software Development Projects. In: proceedings of International Conference on Global Software Engineering ICGSE, 60-70
- Prikladnicki, R, Audy, J, L, N, Evaristo, R (2003) Global Software Development in Practice Lessons Learned. *Software Process: Improvement and Practice*, Volume 8, Issue 4, pp. 267–281.
- Petersen, K., Wohlin C. (2009) Context in Industrial Software Engineering Research, In: proceedings of the 3rd International Symposium on Empirical Software Engineering and Measurement, pp. 401-404.
- Prikladnicki, R, Audy, J, L, N, Evaristo, R (2006) A Reference Model for Global Software Development: Findings from a Case Study. In: proceedings of the IEEE International Conference on Global Software Engineering ICGSE, pp. 18-28.
- Prikladnicki R, Audy J L N, Damian D, de Oliveira TC (2007) Distributed Software Development: Practices and Challenges in Different Business Strategies of Offshoring and Onshoring. In: proceedings of the IEEE International Conference on Global Software Engineering ICGSE, pp. 262-274
- Prikladnicki, R, Damian, D, Audy, J, L, N (2008) Patterns of Evolution in the Practice of Distributed Software Development in Wholly Owned Subsidiaries: A Preliminary Capability Model. In: proceedings of the IEEE International Conference on Global Software Engineering ICGSE, pp. 99-108.
- Prikladnicki R, Audy, J L N (2010) Process Models in the Practice of Distributed Software Development: A Systematic Review of the Literature. *Information and Software Technology*, 52 (8), 779-791.
- Radatz, J (1988) A standard dictionary for computer terminology: Project 610, In: *Computer*, 21(2), 1988, pp. 72-77
- Richardson, I, Avram, G, Deshpande, S, Casey, V (2008) Having a Foot on Each Shore – Bridging Global Software Development in the Case of SMEs. In: proceedings of the 3rd IEEE International Conference on Global Software Engineering ICGSE, pp. 13-22.
- Robinson M, Kalakota R (2004) *Offshore Outsourcing: Business Models, ROI and Best Practices*. USA: Mivar Press
- Smite, D, Wohlin, C, Feldt R Gorschek T (2008) Reporting Empirical Research in Global Software Engineering: a Classification Scheme. In: proceedings of the IEEE International Conference on Global Software Engineering, pp. 173-181.
- Smite, D, Wohlin C (2010) Software Product Transfers: Lessons Learned from a Case Study. In: proceedings of the IEEE International Conference on Global Software Engineering, pp. 97-105.
- Smite, D, Wohlin, C, Feldt, R, Gorschek, T (2010) Empirical Evidence in Global Software Engineering: A

Systematic Review, *Empirical Software Engineering Journal*, 15 (1), pp. 91–118.

Szymanski, C, H, Prikladnicki, R (2007) The Evolution of the Internal Offshore Software Development Model at Dell Inc. In: proceedings of the 2nd IEEE International Conference on Global Software Engineering ICGSE, pp. 40-50.

Tervonen, I, Mustonen, T, (2009) Offshoring Test Automation: Observations and Lessons Learned. In: Proceedings of the 4th International Conference on Global Software Engineering, 226-235.

Zelkowitz M V, Wallace D R (1998) Experimental Models for Validating Technology. In: *IEEE Computer*, 31(5), p. 23-31.

Ågerfalk, P J, Fitzgerald B (2008) Outsourcing to an Unknown Workforce: Exploring Opensourcing as a Global Sourcing Strategy, *MIS Quarterly*, 32(2), pp. 385-409.