

Software Integration Challenges in Global Software Development Environment: A Systematic Literature Review Protocol

Muhammad Ilyas^{1, 2, 4}, Siffat Ullah Khan^{1, 2, 3}

¹Software Engineering Research Group (SERG),

²Department of Computer Science and IT,

³Department of Software Engineering,

University of Malakand, Khyber Pakhtunkhwa, Pakistan.

⁴Federal Government College Batkhela,

Ministry of Defence, Government of Pakistan

Abstract : CONTEXT – Global Software Development (GSD) is an emerging trend to perform software development activities across national boundaries. Despite of the benefits gained from GSD, there are certain challenges associated with it. Amongst these challenges one of the critical challenge, which has comparatively little attention in software engineering research community, is the integration of products and components developed by different teams at different locations.

OBJECTIVE – The objective of this paper is to design SLR protocol to identify the challenges associated with software/product integration based on project size and product type in GSD environment from vendor's perspective and to identify the factors to be addressed before, during and after the integration process by the vendor organisations.

METHODOLOGY – We will conduct a Systematic Literature Review (SLR) to identify the challenges and factors mentioned above. SLR is different from ordinary review because it is comparatively unbiased and is based on a predefined protocol.

EXPECTED OUTCOME – The expected outcome of the review will be identification of challenges and factors for software integration in GSD. The identified factors through the review will lead us to develop a Software Integration Model (SIM) that will assist vendors to appropriately address the challenges and factors associated with software/product integration in GSD for different sized and types of projects and products.

Keywords - Global software Development, Integration Challenges, Software Integration, Systematic Literature Review, Vendors

I. Introduction

Global Software Development (GSD) is an emerging trend to perform software development activities across national boundaries. In GSD the clients and vendors may be separated at different geographical locations. The development team may itself be spread at different locations of the globe. The popularity and continuous growth of GSD is due to the following advantages [1, 2]:

- Exploitation of a 24-hour working day
- Decrease the cost by offshoring the development work to low wages countries
- Solve local I.T skills shortage
- Show the presence of vendors at different local markets
- Create space for vendors in new markets
- Advantage of competition
- Capitalizes on global resource pools
- Geographically closer to the end consumer
- Acquisition and mergers

Despite of these benefits, GSD is still immature and in a developing stage [3]. There are certain challenges and critical barriers like diverse cultures, time zone differences, language differences, inadequate communication, knowledge management, poor contract management, poor relationship management etc [1, 4, 5].

One of the major challenge, which have got very little attention and needs further research, is the software/products integration in GSD [6, 7]. In the process of software development, integration is one of the most important phase because most of the hidden problems of other phases start appearing in this phase [7].

These problems not only increases the work load but also decrease the overall quality of the final product and may also lead to cost and time over run of the entire software project. Most of the side effects of GSD environment appear in the form of integration problems [8]. Larsson [7] have identified three main problems of product integration:

- Components needed for integration may be not ready
- There may be no well defined interfaces between components
- There may be no well prepared environment for integration

These problems become more critical in GSD environment due to little communication and coordination of the different development teams working at different geographical locations [9]. Each team in GSD will be responsible to work on its own component without having a complete picture of the overall system [6]. Similarly each team may be using a different process model and standards which will further complicate the integration process at the end [8]. In GSD, considering the integration problem late will make the project more problematic and will also increase the overall development cost [10]. Zafar, Ali et al [6] have identified 14 success and 11 failure factors regarding product integration in GSD environment.

The objective of this research is to identify the challenges/risks associated with the integration of products based on project size (*small, medium and large*) and product type (*bespoke, of-the-shelf, customize-of-the-shelf*) in GSD environment from vendors perspective and to identify the factors to be addressed before, during and after the integration process by vendors based on project size (small, medium and large) and product type (bespoke, of-the-shelf, customize-of-the-shelf etc) . Thus we have formulated the following research questions.

RQ1: What are the factors that play a positive role, as identified in the literature, to be addressed by vendors at various stages of the product integration, i.e. before during and after the integration process in GSD?

RQ2: What are the challenges, as identified in the literature, faced by vendors in product integration based on the project size and product type in GSD?

RQ3: What are the real world practices, as identified in the literature, for GSD vendors to be successful at all stages of the product integration?

II. Background

Software integration is a key phase as most of the software projects, during testing, get delayed due to the complexities and incompatibilities found between components during the integration process [11].

Integration means to assemble or combine or merge two or more components or products into one system or subsystem. For example integration of Mouse, Keyboard, Monitor and CPU form a desktop computer system.

In the glossary of EIA (Electronic Industries Alliance) 731.1 Standard (SYSTEMS ENGINEERING CAPABILITY MODEL) integration has been defined as:

“Integration: The merger or combining two or more elements (e.g., components, parts, or configuration items) into a functioning and higher level element with the functional and physical interfaces satisfied”.

Herbsleb and Grinter [12] have defined integration as “all the work necessary to assemble the product from its components”.

McConnel [13] describes integration as “the software development activity in which you combine separate software components into a single system”.

While Larsson [7] have defined product integration as follow:

“The product integration process is a set of procedures used to combine components into larger components, subsystems or final products and systems. Product integration enables the organization to observe all important attributes that a product will have; functionality, quality and performance. This is especially true for software systems as the integration is the first occurrence where the full result of the product development effort can be observed. Consequently, the integration activities represent a highly critical part of the product development process”.

A number of researchers have worked in the area of software integration. e.g.,

Herbsleb et al., [12] in a case study on “Splitting the organization and integrating the Code”, have considered the integration phase as one of the most difficult phase of the GSD project. They have pointed out that in a multisite development, due to the loss of communication and coordination, integration process is badly suffered. The components needed for integration may not be available according to the planed schedule or may have not well defined interfaces.

Similarly Van Moll et al., [8] have conducted an exploratory study on “Identifying Pitfalls of System Integration”. They have identified that in more than 50% of the GSD projects integration phase is one of the difficult and problematic phase. It is suggested that better planning, monitoring and control, and

explicitly assigning the responsibilities to all members may help to minimize the integration problems. They have further suggested considering integration and integration testing as a separate process during software development because it is often considered as a formality only.

Bosch et al., [11] in a case study “From Integration to Composition: On the impact of Software Product Lines, Global Development and Ecosystem” have concluded that many of the challenges in large-scale software development exists due to the integration-centric approach. These challenges and complication can be simplified through a transition from integration-centric approach to composition-oriented approach to the development of software. It is suggested that composition-oriented approach is not always preferable and there are many cases where integration-centric approach is preferable.

Zafar et al., [6] have conducted a systematic review and Delphi survey to investigate integration challenges and solutions in GSD. They have pointed out that integration challenges usually remain hidden at the initial stages of the projects but create problems as the projects move towards completion. If these challenges are not properly addressed at the initial stage, it will not only make the project time and cost overrun but will also decrease the overall quality of the project leading to its failure. They consolidated the identified 79 failure factors and 26 success factors into 11 failures factors and 14 success factors. They identified from the various studies that integration problems vary from project to project. They concluded that further research is necessary because the integration problem can occur at any phase of GSD. One limitation of this study is that most of the factors were collected from the Delphi study and a limited number of experts.

Cataldo and Herbsleb [14] have performed an empirical analysis on a development organization (responsible for producing navigation system for automobiles) to identify “Factors Leading to Integration Failures in Global Features Oriented Development”. They have analyzed that cross-feature interaction, which is a measure of architecture dependencies between two product features, is one of the major failure of product integration. One of the major limitations of this study is that they have examined only a single system developed by a single organization.

Most of the studies have presented a problem oriented report rather than suggesting solutions and practices. Although [6] have identified success and failure factors but have not specified the practices needed to address these factors. Similarly the need for considering Integration as a separate process is also stressed. None of the studies has considered integration problems in depth based on the project size and product type. They also have not categorized what factors vendors should consider before, during and after the integration process. The novelty of our research work is that we will not only investigate the integration problems and success factors based on the project size and product type but will also investigate what factors vendor should consider before, during and after the integration process. We will also identify the real world practices, from the literature, to be successful at all stages of software integration.

III. Systematic Literature Review Protocol For Software Integration Challenges In Global Software Development Environment

The purpose of this paper is to design the systematic literature review (SLR) protocol for software integration challenges and success factors for vendors working in GSD environment by following the SLR guidelines [15]. We have also studied two other SLR protocols [5, 16] for guidance during the development of the SLR protocol.

There are three main phases of SLR as defined by Kitchenham: planning the review, conducting the review and reporting the review. This paper presents the protocol, a plan for the review. SLR is protected from researcher’s bias through a predefined protocol although it cannot control the publication bias [15]. The various steps in the protocol have been discussed in the remaining sections of the paper. Figure 1 presents the SLR process.

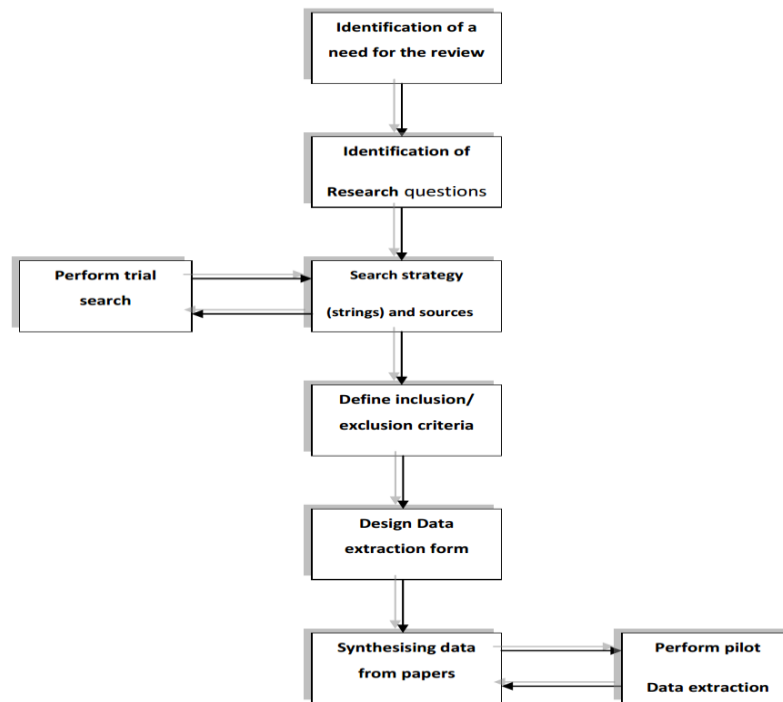


Figure.1 Systematic Literature Review process

IV. Research Questions

RQ1: What are the factors that play a positive role, as identified in the literature, to be addressed by vendors at various stages of the product integration, i.e. before during and after the integration process in GSD?

RQ2: What are the challenges, as identified in the literature, faced by vendors in product integration based on the project size and product type in GSD?

RQ3: What are the real world practices, as identified in the literature, for GSD vendors to be successful at all stages of the product integration?

4.1 Construction of Search Terms

Here we define some terms which will help us in designing the search string for our research questions.

Population: Global software development vendors.

Intervention: Integration challenges/risks/factors/problems.

Outcome of relevance: Successful and effective product integration.

Experimental design: SLRs, empirical studies, case studies, theoretical studies and expert opinions.

Thus using the above terms we can write our research questions as follow:

RQ1: [What are the factors,] **INTERVENTION**
 that play a positive role, to be addressed by
 [GSD vendors] **POPULATION**
 at various stages of the
 [software integration] **OUTCOMES OF RELEVANCE**

RQ2: [What are the challenges,] **INTERVENTION**
 faced by
 [GSD vendors] **POPULATION**
 in
 [software integration] **OUTCOMES OF RELEVANCE**
 based on the project size in GSD environment?

V. Search Strategy

5.1 Trial Search

We have conducted a trial search using the following search string in Google scholar and IEEExplore digital library:

((“Software Integration” OR “Product Integration”) AND (“Global software development” OR “Global software engineering” OR “Distributed software development” OR GSD) AND (challenge OR factor OR Issue) AND (Vendor OR Supplier OR Developer))

The papers retrieved through this search string will be used as guidelines for the development and validation of the major search terms.

5.2 Identification of Search Terms

We will use the following plan for the identification of search terms:

- i. First we will derive the major search terms from the research question by identifying population, intervention and outcome.
- ii. Then we will find the alternative spellings and synonyms for the major terms derived in the above step
- iii. Next we will verify the key words in some relevant paper
- iv. Finally we will use
 - a) the “OR” operator for concatenation of synonyms and alternative spellings
 - b) and the “AND” operator for concatenation of major terms
- v. Integrate the search string into a summarized form, if required

It should be noted that some databases (e.g. Google scholar, CiteSeer) do not allow lengthy search strings; therefore we will break the string formed in the above steps into two or more smaller strings if required.

We got the following results after execution of the steps in the above plan.

Result for (a)

RQ1: Global software development, factors, product integration, vendors.

RQ2: Global software development, integration challenges, product integration, vendors.

RQ3: Global software development, integration practices, vendors.

Result for (b)

Product Integration: “Product integration” OR “Software integration” OR “System integration” OR “Component integration” OR “Product assembling” OR “Software assembling” OR “System assembling” OR “Component assembling” OR “Product merger” OR “Software merger” OR “System merger” OR “Component merger”

Global Software Development: “Global software development” OR “Global software engineering” OR “Distributed software development” OR “Distributed software engineering” OR “Multisite software development” OR “Multi-site software development”

Challenge(s): Challenge OR risk OR problem OR issue OR barrier OR trouble

Factor(s): factor OR success factors OR key factors OR critical success factors

Vendor(s): Vendor OR Supplier OR Developer OR service-provider

Practice(s): Practice OR solution OR exercise OR best practices OR advice OR implementation initiatives

Result for c)

Software Integration, product Integration, global software development, global software engineering, distributed software development, GSD,GSE, DSD, challenges, factor, barrier, issue, vendor, supplier, developer

Result for d)

RQ1 and RQ2

((“Software integration factors” OR “Product integration” OR “System integration” OR “Component integration” OR “Product assembling” OR “Software assembling” OR “System assembling” OR “Component assembling”) AND (“Global software development” OR “Global software engineering” OR “Distributed software development” OR “Distributed software engineering” OR “Multisite software development” OR “Multi-site software development”) AND (Challenge OR risk OR problem OR issue OR barrier OR trouble OR factor OR “success factors” OR “key factors” OR “critical success factors”) AND (Vendor OR Supplier OR Developer OR service-provider))

RQ3: ("Software integration factors" OR "Product integration" OR "System integration" OR "Component integration" OR "Product assembling" OR "Software assembling" OR "System assembling" OR "Component assembling") AND ("Global software development" OR "Global software engineering" OR "Distributed software development" OR "Distributed software engineering" OR "Multisite software development" OR "Multi-site software engineering") AND (Practice OR solution OR exercise OR best practices OR advice OR "implementation initiatives") AND (Vendor OR Supplier OR Developer OR service-provider)

Note: We have constructed one full search string for both RQ1 and RQ2, because three major terms are the same and only one major term (i.e. challenges and factors) in each research question is different. We have combined these two terms and their synonyms into one group using Boolean OR operator.

5.3 Search String Breakup

As already mentioned that some digital libraries does not support long search string, therefore, we have divided the full search string into three substring and will try each of them separately.

Substring 1

("Software integration factors" OR "Product integration" OR "System integration") AND ("Global software development" OR "Global software engineering") AND (Challenge OR risk OR problem OR issue OR barrier OR trouble) AND (Vendor OR Supplier OR Provider)

Substring 2

("Component integration" OR "Product assembling" OR "Software assembling") AND ("Distributed software development" OR "Distributed software engineering") AND (factor OR "success factors") AND (Seller OR Developer OR service-provider)

Substring 3

("Software integration factors" OR "System assembling" OR "Component assembling") AND ("Multisite software development" OR "Multi-site software engineering") AND ("key factors" OR "critical success factors") AND (Vendor OR dealer OR trader OR marketer)

5.4 Resources to be searched

We will try to search the following digital libraries/databases.

- IEEE Xplore (<http://ieeexplore.ieee.org>)
- ACM Portal (<http://dl.acm.org>)
- Science Direct (www.sciencedirect.com)
- Springer Link (<http://www.springerlink.com/>)
- Google Scholar (<http://scholar.google.com>)
- CiteSeer Digital Library (<http://citeseerx.ist.psu.edu>)
- Emerald (<http://www.emeraldinsight.com/>)

5.5 Search constraints and validation

We will search all the available published work relevant to our research questions and will put no date boundaries. A prior search was conducted using the following set of major search terms and we found 169 papers on Google Scholar and 43 papers on IEEEExplore digital libraries. Prior to undertaking the review process, some of relevant papers will be used for the validation of our search strings.

("Software Integration" OR "Product Integration") AND ("Global software development" OR "Global software engineering" OR "Distributed software development" OR GSD) AND (challenge OR factor OR Barrier OR Issue) AND (Vendor OR Supplier OR Developer)

5.6 Search Documentation

Search results will be documented properly according to the following format:

Name of the database	Search Strategy	Search phase	Date of Search	Year covered by search	No. of publication found	Initial selection decision	Final selection decision
Google Scholar	((“Software Integration” OR “Product Integration”) AND (“Global software development” OR “Global software engineering” OR “Distributed software development” OR GSD OR GSE) AND (challenge OR factor OR Barrier OR Issue) AND (Vendor OR Supplier OR Developer))	Trial Search	11/06/2012	All	169		
IEEEExplore					43		

5.7 Search Result Management

The primary source references will be stored electronically in a separate directory.

VI. Publication selection criteria

The following figure explains the publication selection criteria:

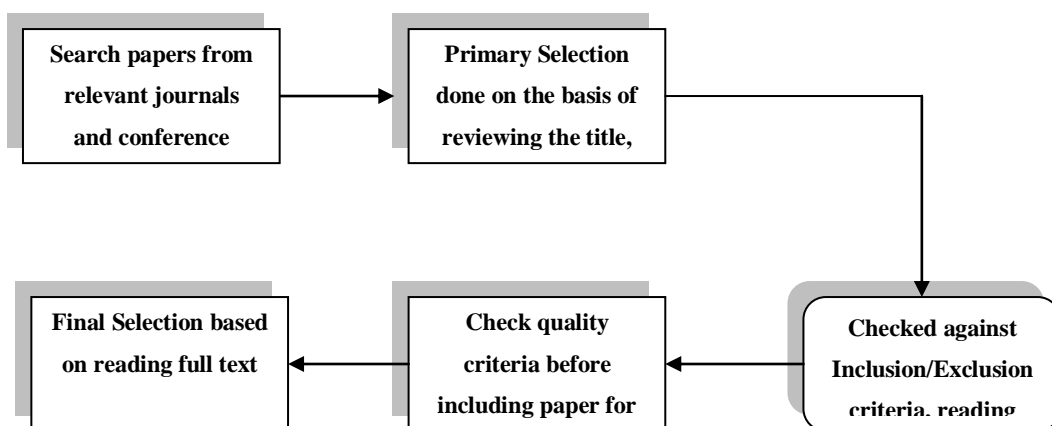


Figure.2 Publication selection

The publication selection criterion is used to select only those publications which are relevant to the research questions. We will consider only those papers which are relevant to software integration in GSD. Papers relating to hardware integration will be ignored.

6.1 Primary selection

Initial selection of the primary sources will be performed by reviewing the title, keywords and abstract. The purpose is to ignore only those results which have no relevance to the problem or research questions. The primary sources chosen in the initial selection process will be checked against the below mentioned inclusion/exclusion criteria by reviewing thoroughly full text of the studies. The source will be sent to the secondary reviewer, for review in case of any uncertainty regarding the inclusion or exclusion decision.

6.2 Inclusion criteria

The inclusion criteria we used to determine which piece of literature (papers, technical reports, or grey literature etc.) found by the search term will be used for the data extraction. We will consider only those papers which are related to global/distributed software development/engineering written in English language. The criteria are listed below:

- i. Studies that describe factors that have a positive impact on any stage of the integration process.
- ii. Studies that describe software/product integration challenges/issues in GSD environment based on project size or product type.
- iii. Studies that describe vendor role for effectiveness of software integration.
- iv. Studies that describe criteria for successful product integration in GSD.
- v. Studies that describe software attributes that assist in product integration.

6.3 Exclusion criteria

This section describes the exclusion criteria in order to decide which piece of literature found by the search term will be excluded/ignored. The criteria are listed below:

- i. Studies that are not relevant to the research questions.
- ii. Studies other than software integration i.e. hardware integration like telecommunication or electronic component integration will not be considered.
- iii. Studies that don't describe integration success factors or integration challenges/problems in global software development.
- iv. Studies conducted in environment other than GSD.

6.4 Publication quality assessment

The quality assessment will be performed after final selection of publications. The publications quality will be evaluated in parallel with data extraction. The quality will be gauged on the basis of the following questions:

- i. Is it clear how the integration challenges were identified for GSD vendors?
- ii. Is it clear how the success factors for software integration were identified for GSD vendors?
Each of the above factors will be marked as "YES" or "NO" or "Partial" or "N.A".
My supervisor, as a secondary reviewer, will randomly score a small subset for validation.

VII. Data Extraction Strategy

1.1 Primary Study Data

The focus of the study is to collect the data, from publications, which satisfy the research questions. We will extract the following data from each finally selected publication:

- Data about the publication i.e. Title, Authors, Journal/Conference title, etc
- Data that will address the research questions

The following data will be extracted to address the research questions:

RQ1: Background information and success factors i.e. the factors which will have a positive impact on GSD vendors at any stage of the integration process.

RQ2: Background information and the challenges/barriers that GSD vendors need to address based on the size of project and the type of product.

The following data will be captured during data extraction:

Data to be extracted
Date of review
Title
Authors
Reference
Database
Methodology (interview, case study, report, survey, etc)
Sample Population
Publication Quality Description
Organization Type (Software house, University, Research institute etc)
Company size (small, medium, large)
Project size (i.e. small, medium, large)
Product type (i.e. bespoke, off the shelf, customize of the shelf)
SPI Certification
Integration stage (i.e. before, during or after the integration process)
Success factors i.e. the factors which will have a positive impact on GSD vendors at any stage of the integration process.
Challenges/barriers that GSD vendors need to address based on the size of project and the type of product.
Practices for GSD vendors to be successful at all stages of integration.

7.2 Data Extraction Process

The review will be carried out by a single researcher, who will be responsible for the data extraction. In case of an issue the secondary reviewer will be approached for guidance concerning the data extraction. After completing data extraction process by primary reviewer, the inter-rater reliability test will be performed by secondary reviewer. The secondary reviewer will select few publications randomly from the list of publication already chosen by the primary reviewer. The secondary reviewer will independently extract data from the randomly selected publications. The results will then be compared with the results produced by the primary reviewer.

Primary Reviewer: Muhammad Ilyas
 Secondary Reviewer: Dr Siffat Ullah Khan

7.3 Data Storage

The summarized data for each publication will be kept as SPSS/Microsoft Word document and will be stored electronically.

VIII. Data Synthesis

Due to two research questions, the synthesis will also be categorized into two parts. For the Research Question1, the data will be synthesized by creating one summary table having the columns (S.No, Success factors, Frequency, Percentages) showing the list of all the success factors classified into various stages of integration process i.e. before, during and after the integration process along with their frequencies and percentages. The complete detail of every success factor mentioned in the Summary table will be recorded in a separate table which will hold the following columns (Success factors group name, S.No of reference, Success factor subgroups, Paper reference/Paper title).

For the Research Question2 and 3 the same process will be performed as for the RQ1 mentioned above.

IX. Validation of the Review Protocol

The protocol was initially submitted to my research fellows Abdul Wahid and Muhammad Ilyas Azeem. After that it was reviewed by my supervisor Dr. Siffat Ullah Khan. Finally it was presented to Software Engineering Research Group at University of Malakand (SERG_UOM).

X. Software Integration Challenges in Global Software Development Environment: A Systematic Literature Review Protocol

XI. Divergences

In case of any divergence from the protocol, which may occur during the study, we will record any change in a new Appendix to this document.

XII. Acknowledgments

We are thankful to software engineering research group (SERG_UOM), the reviewers and my fellows Abdul Wahid and Muhammad Ilyas Azeem for their assistance in the review process.

References

- [1] M. Romero, A. Vizcaíno, and M. Piattini, "Toward a Definition of the Competences for Global Requirements Elicitation," *ACM 978-1-60558-115-6/08/06.*, 2008.
- [2] B. H. C. Cheng and J. M. Atlee, "Research Directions in Requirements Engineering.," in *FOSE*, 2007.
- [3] D. Šmite, C. Wohlin, T. Gorschek, and R. Feldt, "Empirical evidence in global software engineering: a systematic review," *Journal of Empirical Software Engineering*, vol. 15, no. 1, pp. 91-118, 2010.
- [4] S. U. Khan, M. Niazi, and R. Ahmad, "Barriers in the selection of offshore software development outsourcing vendors: An exploratory study using a systematic literature review," *Information and Software Technology 53*, pp. 693–706, 2011.
- [5] M. I. Azeem and S. U. Khan, "Intercultural Challenges in Offshore Software Development Outsourcing Relationships: A Systematic Literature Review Protocol.," presented at IEEE The 5th MALAYSIAN SOFTWARE ENGINEERING CONFERENCE, "MYSEC2011", THE PUTERI PACIFIC, JOHOR, Malaysia., 2011.
- [6] A. Zafar, S. Ali, and R. K. Shahzad, ""Investigating integration challenges and solutions in global software development", presented at Frontiers of Information Technology (FIT), 2011, 2011.
- [7] S. Larsson, "Key Elements of the Product Integration Process," Malardalen University Sweden, 2007, pp. 78.
- [8] J. H. Van Moll and R. W. M. Ammerlaan, ""Identifying Pitfalls of System Integration -- An Exploratory Study", presented at IEEE International Conference on Software Testing Verification and Validation Workshop, 2008.
- [9] M. Cataldo and S. Nambiar, "On the relationship between process maturity and geographic distribution: an empirical analysis of their impact on software quality," in *Proceedings of the the 7th joint meeting of the European software engineering conference and the ACM SIGSOFT symposium on "The foundations of software engineering"*. Amsterdam, The Netherlands: ACM, 2009, pp. 101-110.
- [10] C. Farcas, E. Farcas, I. H. Krueger, and M. Menarini, "Addressing the Integration Challenge for Avionics and Automotive Systems; From Components to Rich Services," *Proceedings of the IEEE*, vol. 98, pp. 562-583, 2010.
- [11] J. Bosch and Petra Bosch-Sijtsema, "From integration to composition: On the impact of software product lines, global development and ecosystems," *Journal of Systems and Software*, vol. 83, pp. 67-76, 2010.
- [12] J. D. Herbsleb and R. E. Grinter, ""Splitting the Organization and Integrating the Code: Conway's Law Revisited", presented at Proc. of the 21st International Conference on Software Engineering, 1999.
- [13] S. McConnel, *Code Complete 2*, 2nd ed: USA: Microsoft Press Redmond, Wa, 2004.
- [14] M. Cataldo and J. D. Herbsleb, ""Factors leading to integration failures in global feature-oriented development: An empirical analysis", presented at Software Engineering (ICSE), 2011 33rd International Conference on DO, 2011.
- [15] B. Kitchenham and S. Charters, ""Guidelines for performing Systematic Literature Reviews in Software Engineering", Keele University, UK 2007.
- [16] S. U. Khan and M. Niazi, "Systematic Literature Review Protocol for Software Outsourcing Vendors Readiness Model (SOVRM), Technical Report: TR/08-01, ISSN: 1353-7776, School of Computing and Maths, Keele University, UK.," 2008.