Knowledge Management in Distributed Software Development: A Systematic Review

Fernanda Tamy Ishii¹, Gislaine Camila L. Leal¹, Edwin V. Cardoza Galdamez¹, Elisa Hatsue M. Huzita¹, Renato Balancieri¹ and Tânia Fátima C. Tait¹,

¹State University of Maringá, Maringá, Paraná, Brasil
fernanda.tamyi@gmail.com, {gclleal, evcgaldamez, ehnhuzita, rbalancieri2}@uem.br, tait@din.uem.br

Abstract. Software development is characterized as a knowledge intensive activity. Particularly, Distributed Software Development (DSD) is an approach that demands more attention for coordination and communication among members of distributed team, due to regional, cultural and infrastructure differences. Knowledge has being, increasingly, seen as the most important strategic resource in organizations. So, the management of this knowledge is critical to organizational success. Knowledge Management (KM) is a set of processes directed at creating, capturing, storing, sharing, apply, and reuse of knowledge, which are useful to decision making. The purpose of this paper is to present a systematic review carried out to identify the processes, techniques, methods, practices and/or tools adopted for Knowledge Management in Distributed Software Development. With this systematic review some interesting points for research were identified.

Keywords: Knowledge Management; Distributed Software Development; DSD; Cooperative Development; Systematic Review.

1 Introduction

The Distributed Software Development (DSD) is an approach for software development that comes to meet of globalization need, such as: increase of productivity; quality improvement and costs reduction. It added to software development challenges concerned to cultural differences, geographic dispersion, coordination and control, communication and team spirit, which intensified some problems found during the project lifecycle [1], [2].

Software development is, by itself, characterized as a knowledge intensive activity, including for example the knowledge about processes, products, skills of different professionals involved. The interaction and information sharing among teams members are important parts of the process of knowledge construction in DSD context. Furthermore, knowledge has being seen as the most important strategic resource in organizations.

Knowledge Management postulates that organizations must deal knowledge as a factor of richness. It is a discipline that promotes in an integrated way, management and sharing of all information assets that a company owns. This information can be found in a database, source code, documents, procedures, as well as on people,
through their experiences and skills. So, when software development is considered, it is essential to maintain the knowledge available for fast and easy access for whose that needs it in the organization or in a supply chain.

In this sense, knowledge management has been increasingly used in companies to organize, strategically, the knowledge of employees and external knowledge, which are essential for business success. This paper describes a systematic review that was carried out in order to identify the processes, techniques, methods, practices and / or the tools adopted to promote knowledge management in distributed software development environment.

The text is structured into more three sections beyond this introduction. The second section describes the methodology adopted. The third section presents the results and discussion. Finally, the fourth section, presents the final considerations.

2 Methodology

A. A Systematic Review

Systematic review is a planned and structured process that aims at finding work and research related to a specific research question or concern. One of its applications is best known in the evaluation, interpretation and synthesis of available data on given technology, treatment or procedure, allowing the identification of problems and the formulation of a new scenario or solutions for this study [4].

B. Method adopted

The procedure applied to this review was adapted from the model presented in Kitchenham [4]. Firstly the research purpose was defined. Based on that, the research questions, which contain criteria for inclusion and exclusion of papers were defined. After that, the database to be used to carry out searches, the language, the publication year were also defined.

For data collection, were considered keywords related to issues of research and based on them, query sequence were formulated. All papers found were archived and cataloged with the help of a program called JabRef. After registration of the papers, began the pre-selection stage of them, which consisted of rapid reading of titles, keywords and abstracts. With this, was evaluated the adequacy of them regarding of search criteria established. After the pre-selection stage, was performed a complete reading of the papers approved in the first stage. Then the synthesis and interpretation of data were performed.

C. Research Goals

This review aimed to:

- Identify the processes, techniques, methods, practices and / or tools adopted for knowledge management in distributed software development.
- Observe the results of the practices of Knowledge Management in Distributed Software Development.

D. Research Questions

Considering the goals above mentioned, questions containing appropriated criteria for inclusion and/or exclusion of papers were elaborated. These are:
Question 1: What processes, techniques, methods, practices and / or tools are adopted to promote knowledge management in distributed software development?

Question 2: What are the benefits of the knowledge management in distributed software development?

Question 3: What kinds of difficulties, limitations and / or problems occur in the use of knowledge management for distributed software development?

E. Search Strategy

- Sources: Electronic indexed database (IEEE, ACM, Compendex and ScienceDirect).
- Language: English, once it is the internationally accepted language for scientific papers.
- Year of publication: published papers in the period from 2000 to 2013.

G. Query strings

The keywords were determined considering the terms “distributed software development” and “knowledge management”. The query string was create using logical operators AND and OR.

(“Distributed Software Development” OR “Global Software Development” OR “Geographically Distributed Development” OR “Collaborative Development” OR “Distributed Development” OR “Distributed Software Project” OR “Global Software Engineering” OR “Globally Distributed Work” OR “Distributed Teams” OR “Global Software Teams” OR “Virtual Teams”) AND (“Knowledge Management” OR “Knowledge Acquisition” OR “Knowledge Achievement” OR “Knowledge Retention” OR “Knowledge Application” OR “Knowledge Sharing” OR “Knowledge Use” OR “Knowledge Integration” OR “Knowledge Discovery” OR “Knowledge Organizational” OR “Knowledge Transference”).

H. Inclusion criteria (ICi)

To address the research questions, the following criteria were defined:

- [IC1] Processes, techniques, methods, practices and / or tools adopted to promote knowledge management in distributed software development;
- [IC2] Scenarios and cases in which knowledge management has been applied to the distributed software development;
- [IC3] Difficulties or problems found to implement knowledge management in distributed software development.

I. Exclusion criteria (ECi)

To address the research questions, the following criteria were defined:

- [EC1] Processes, techniques, methods, practices and / or tools adopted to promote knowledge management those are not adequate for distributed software development context;
- [EC2] Scenarios and cases in which knowledge management has not been applied for distributed software development;
- [EC3] Difficulties or problems found in the implementation of knowledge management not related to distributed software development;
- [EC4] Papers written in language different from English;
- [EC5] Papers that were not available to perform the complete reading.
J. Preliminary Process
In the preliminary process of selection, the query string, composed by keywords and synonymous terms (Knowledge Management and Distributed Software Development), was submitted to indexed databases and search engines. Later, the title, abstract and keywords of found papers were read. Each paper selected was analyzed and was approved for complete reading if presented results that were relevant to research carried out. If there was not relevant data for research in question, the paper was not approved for complete reading.

K. Final Selection Process
At the end of selection process, the papers that had been approved in preliminary phase were undergone to a complete reading. So, when the subject (approach, tools, experiment) dealt was not related to research goal, the paper was discarded. At the end of the complete reading, if the paper was approved, the reader was responsible for summarize the data.

L. Quality assessment of the papers (Qi)
The assessment goal was not to classify the papers according to a total quality score. So, was not assigned any score. Thus to classify them the binary scale was used, considering the criteria defined by Dyba et al [5] as following:

- [Q1] Is there a clear statement of the research goals?
- [Q2] Is there an adequate description of the context in which the research was carried out?
- [Q3] Does the research design is appropriate to address the aims established for that?
- [Q4] Was there a control group with which could compare the results?
- [Q5] Does the data analysis is sufficiently rigorous?
- [Q6] Does the article is based on research or experiments?
- [Q7] Are there results which can be applied in practical situations?
- [Q8] Is there a clear statement of findings?
- [Q9] Are there artifacts generated as result from the search presented at the paper?
- [Q10] Does article presents some model for promoting the Knowledge Management?

3 Results and Discussion
After carry out the procedures defined in previous section, 38 primary studies were selected. Fig. 1 shows the method used and the results obtained at each stage.

![Fig. 1. Results of systematic review](image-url)
In Table 1 are highlighted properties and characteristics concerned to knowledge management discussed in the selected papers (SP).

Table 1. Properties and Characteristics.

<table>
<thead>
<tr>
<th>Properties and Features</th>
<th>Papers</th>
<th>Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>SP 1, SP 3, SP 4, SP 5, SP 8, SP 9, SP 10, SP 12, SP 13, SP 14, SP 15, SP 16, SP 17, SP 25, SP 26, SP 27, SP 29, SP 30, SP 33, SP 34</td>
<td>20 papers</td>
</tr>
<tr>
<td>Knowledge representation</td>
<td>SP 2, SP 3, SP 4, SP 5, SP 6, SP 7, SP 9, SP 11, SP 12, SP 13, SP 16, SP 17, SP 20, SP 21, SP 22, SP 23, SP 24, SP 25, SP 26, SP 27, SP 28, SP 30, SP 32, SP 34</td>
<td>24 papers</td>
</tr>
<tr>
<td>Knowledge dissemination</td>
<td>SP 10, SP 12, SP 17, SP 18, SP 19, SP 20, SP 23, SP 28, SP 34</td>
<td>9 papers</td>
</tr>
<tr>
<td>Support to decision making</td>
<td>SP 3, SP 6, SP 7, SP 16, SP 21, SP 37</td>
<td>6 papers</td>
</tr>
<tr>
<td>Use of repository</td>
<td>SP 1, SP 2, SP 8, SP 14, SP 17, SP 19, SP 21, SP 24, SP 27, SP 28, SP 29, SP 30</td>
<td>12 papers</td>
</tr>
<tr>
<td>Use of contextual information</td>
<td>SP 4, SP 10, SP 13, SP 15, SP 25, SP 30, SP 31, SP 35</td>
<td>8 papers</td>
</tr>
<tr>
<td>Communication Standard</td>
<td>SP 6, SP 7, SP 8, SP 10, SP 11, SP 12, SP 28, SP 31, SP 37, SP 38</td>
<td>10 papers</td>
</tr>
<tr>
<td>Capturing knowledge/experience</td>
<td>SP 2, SP 5, SP 7, SP 11, SP 17, SP 19, SP 20, SP 22, SP 25, SP 28, SP 29, SP 32, SP 34, SP 37</td>
<td>14 papers</td>
</tr>
<tr>
<td>Strategy for knowledge reuse</td>
<td>SP 14, SP 25</td>
<td>2 papers</td>
</tr>
</tbody>
</table>

Among the papers that presented models, 70% of these present a software for automation of them. The automation of the proposed model demonstrates its applicability and also facilitates its adoption.

Regarding the representation of knowledge was observed that most of them use resources such as templates and tags. Only one of the selected papers used ontology to promote knowledge management in distributed development. The ontology was used as a tool for standardization and quantification of the information and helped on externalization of knowledge. It is noteworthy that the benefits of using ontologies are related with the facility to carry out inferences and allow the reuse of information/knowledge. OntoDiSEN, is an ontology for distributed software development domain, which can be extended to include matters related to knowledge management [6].

The concept of reuse was also used as a justification of knowledge management, aiming at the classification of information and subsequent use of them.

E-mail, wiki, video conferencing and telephone are resources used for knowledge dissemination. However, it is emphasized that the use of appropriate tools to support knowledge dissemination is essential to ensure the knowledge availability on site and on time for correct decision making. In case of distributed software development the problems related to communication are exacerbated by socio-cultural and temporal questions.
The flow of knowledge was considered in 6 papers and demonstrated the importance of communication for development and maintenance of knowledge as intellectual capital owned by the company.

Regarding the use of repositories, the papers analyzed highlighted how important is to store the information concerned to the project, process, distributed teams and artifacts generated during whole development of distributed software. The papers also showed the importance of using of appropriated techniques to capture and represent of knowledge so that it can be processed and thus generate new knowledge.

In respect to communication standards, the papers presented a set of best practices, such as: knowing the sender to establish trust, standardization of documents to be sent, among others. Thus, adopting these best practices together with the formalization and use of automated tools could be a strategy to support knowledge reuse.

Studies point that culture has a great impact on promoting of Knowledge Management in DSD, since the information provided informally has higher interactivity than those obtained from tools and repositories. Based on this, it is emphasized that when dealing with distributed team, knowledge management cannot be disassociated from practices that promote the trust among team members. Another analysis was also performed to evaluate the year of paper publication. Fig. 2 shows 2010 as the year in which were found the most relevant publications presented in this review and also that there was no relevant publication in 2000 and 2001.

Fig. 2. Quantity of publications per year.

As for assessing the quality of work, the questions set out in methodology of the systematic review were answered. The results on Table 2, show the quantity of papers that met the particular issue of quality in the occurrence column.

Table 2. Quality Evaluation of Selected Papers.

<table>
<thead>
<tr>
<th>Question</th>
<th>Papers</th>
<th>Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is there a clear statement of the research goals?</td>
<td>SP 1, SP 2, SP 3, SP 4, SP 5, SP 6, SP 7, SP 11, SP 12, SP 13, SP 15, SP 16, SP 17, SP 18, SP 19, SP 20, SP 21, SP 23, SP 24, SP 26, SP 28, SP 29, SP 30, SP 31, SP 32, SP 33, SP 34, SP 35, SP 36, SP 37, SP 38</td>
<td>33 papers</td>
</tr>
<tr>
<td>Is there an adequate description of the context</td>
<td>SP 2, SP 3, SP 4, SP 5, SP 7, SP 8, SP 9, SP 10, SP 11, SP 12, SP 15, SP 16,</td>
<td>27 papers</td>
</tr>
<tr>
<td>Question</td>
<td>Answer</td>
<td>Count</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>Q1</td>
<td>in which the research was carried out?</td>
<td>SP 17, SP 18, SP 19, SP 20, SP 22, SP 23, SP 24, SP 25, SP 26, SP 27, SP 28, SP 31, SP 34, SP 35, SP 36, SP 37, SP 38</td>
</tr>
<tr>
<td>Q2</td>
<td>Does the research design is appropriate to address the aims established for that?</td>
<td>SP 1, SP 2, SP 3, SP 4, SP 5, SP 7, SP 8, SP 10, SP 11, SP 12, SP 13, SP 15, SP 16, SP 17, SP 18, SP 19, SP 20, SP 23, SP 24, SP 26, SP 28, SP 29, SP 30, SP 31, SP 34, SP 35, SP 36, SP 37, SP 38</td>
</tr>
<tr>
<td>Q3</td>
<td>Does the research design is appropriate to address the aims established for that?</td>
<td>SP 1, SP 2, SP 3, SP 4, SP 5, SP 7, SP 8, SP 10, SP 11, SP 12, SP 13, SP 15, SP 16, SP 17, SP 18, SP 19, SP 20, SP 23, SP 24, SP 26, SP 28, SP 29, SP 30, SP 31, SP 34, SP 35, SP 36, SP 37, SP 38</td>
</tr>
<tr>
<td>Q4</td>
<td>Was there a control group with which could compare the results?</td>
<td>SP 7, SP 11, SP 12, SP 15, SP 17, SP 34</td>
</tr>
<tr>
<td>Q5</td>
<td>Does the data analysis is sufficiently rigorous?</td>
<td>SP 2, SP 3, SP 5, SP 7, SP 11, SP 12, SP 15, SP 17, SP 19, SP 23, SP 24, SP 25, SP 26, SP 29, SP 30, SP 32, SP 35, SP 36, SP 38</td>
</tr>
<tr>
<td>Q6</td>
<td>Does the article is based on research or experiments?</td>
<td>SP 1, SP 2, SP 3, SP 4, SP 5, SP 6, SP 7, SP 8, SP 9, SP 10, SP 11, SP 12, SP 13, SP 14, SP 15, SP 16, SP 17, SP 18, SP 19, SP 20, SP 21, SP 23, SP 25, SP 26, SP 27, SP 28, SP 29, SP 30, SP 32, SP 33, SP 34, SP 35, SP 36, SP 37, SP 38</td>
</tr>
<tr>
<td>Q7</td>
<td>Are there results which can be applied in practical situations?</td>
<td>SP 1, SP 2, SP 3, SP 4, SP 7, SP 8, SP 11, SP 16, SP 18, SP 20, SP 24, SP 26, SP 29, SP 31, SP 32, SP 33, SP 34, SP 35, SP 37</td>
</tr>
<tr>
<td>Q8</td>
<td>Is there a clear statement of findings?</td>
<td>SP 1, SP 2, SP 3, SP 4, SP 5, SP 6, SP 7, SP 8, SP 10, SP 11, SP 12, SP 13, SP 15, SP 16, SP 17, SP 18, SP 19, SP 20, SP 21, SP 23, SP 24, SP 25, SP 26, SP 28, SP 29, SP 30, SP 31, SP 32, SP 33, SP 35, SP 36, SP 37</td>
</tr>
<tr>
<td>Q9</td>
<td>Are there artifacts generated as result from the search presented at the paper?</td>
<td>SP 1, SP 2, SP 3, SP 4, SP 9, SP 13, SP 16, SP 20, SP 24, SP 26, SP 29, SP 31, SP 35</td>
</tr>
<tr>
<td>Q10</td>
<td>Does article presents some model for promoting the Knowledge Management?</td>
<td>SP 1, SP 2, SP 3, SP 4, SP 5, SP 6, SP 8, SP 11, SP 24, SP 26, SP 27, SP 29, SP 31, SP 34, SP 35, SP 37, SP 38</td>
</tr>
</tbody>
</table>

It can be observed that 50% of papers present results of knowledge management practices in distributed software development environment. Only 6 papers stand out practical experiments containing a control group for comparison and validation of hypotheses. Among these papers that present practices, most quote communication as a difficulty, thus emphasizing the need to establish the sense of trust in the sender of the message for the effective acquisition and internalization of knowledge reported.
4 Final Considerations

Regarding the studies analyzed, we observed that they address somehow, techniques, practices, tools and methods to promote knowledge management in distributed software development. Most studies showed knowledge representations and definitions for creating models of knowledge management, since there is specific knowledge in a DSD. Other studies showed the use of data repository to capture and maintain knowledge in the organization. To meet the challenge of communication in geographically dispersed teams, the use of e-mail, phone and web conferencing practices were more suitable.

This paper aimed to present the results achieved through a systematic review, listing the works that dealt with processes, techniques, methods, and practices to promote KM tools in DSD. Based on the analysis contained in Section III, we can highlight some points that deserve further study:

• appropriate mechanisms for disseminating knowledge;
• explore data mining techniques, observing socio cultural aspects in DSD, from data stored in the repository in order to extract interesting knowledge;
• mechanisms to capture and represent data, aiming to generate knowledge. One idea that has been explored by our research group is to proceed with the extension of OntoDiSEN [6];
• integrate knowledge management practices to approaches for distributed software development, such as proposed in [7]. It is important emphasize how artifacts can be used to disseminate knowledge in distributed teams [8];
• explore the context information as way for dissemination of knowledge.

Therefore, the results from this systematic review, evidence that there are challenges in knowledge management for DSD. So, the integration of several areas for standardization and reuse in software, storage and dissemination of information and communication are points that could be explored. Furthermore, the results show the need for tools that promote knowledge management in DSD to enable the socialization of information. For each one of these points we can foresee an interesting point for future works to be developed as research.

References


SELECTED PAPERS


