"RADIOGRAFÍA" DEL ESTADO DE LA REUTILIZACIÓN DE SOFTWARE EN COLOMBIA

"Snapshot" of the State of Software Reuse in Colombia "Radiografia" do estado da reutilização de software na Colômbia

> Luisa Fernanda Restrepo Gutierrez ¹ Elizabeth Suescún Monsalve ² Raúl Mazo ^{3 4} Paola Vallejo ⁵ Daniel Correa ⁶

RESUMEN

Debido a los mercados competitivos, el negocio del software quiere soluciones más rápidas, mejores y más baratas en un período corto de tiempo. La reutilización de software surge como una solución viable para estas demandas, ya que ofrece importantes beneficios, como mayor calidad y eficiencia, menores costos y esfuerzos de desarrollo y menor tiempo de comercialización. En este artículo se pretende analizar el estado de la práctica de la reutilización de software en Colombia, realizar comparaciones con trabajos relacionados y ofrecer con este resultado un instrumento para la toma de decisiones en empresas que adoptan estas prácticas. Para llevar a cabo los objetivos anteriores se plantearon tres etapas. En la primera etapa se definieron las preguntas de investigación.

¹ Universidad EAFIT, Medellín, Colombia, Contacto: lrestr61@eafit.edu.co

^{2.} Universidad EAFIT, Medellín, Colombia, Contacto: esuescu1@eafit.edu.co

^{3.} Universidad EAFIT, Medellín, Colombia, Contacto: rimazop@eafit.edu.co

^{4.} École nationale supérieure de techniques avancées (ENSTA) Bretagne, France, Contacto: raul.mazo@ensta-bretagne.fr

^{5.} Universidad EAFIT, Medellín, Colombia, Contacto: pvallej3@eafit.edu.co

^{6.} Universidad EAFIT, Medellín, Colombia, Contacto: dcorreab@eafit.edu.co

exitosas y barreras de adopción en el contexto de la industria de software colombiana. Finalmente, los resultados fueron analizados y reportados. El estudio mostró y evidenció las expectativas, barreras de adopción y factores que influyen en el éxito de la reutilización de software en entornos industriales en Colombia. El presente estudio muestra el estado actual de las prácticas de reutilización en la industria de software en Colombia. Asimismo, la experiencia en el desarrollo de este trabajo sirve como hoja de ruta para otras regiones que quieran analizar el estado actual de la reutilización. Sin embargo, cada organización necesita determinar sus capacidades y encontrar los factores adecuados para adaptarlos a su contexto.

Palabras clave: reutilización de software, encuesta, prácticas de éxito, barreras de adopción

ABSTRACT

Due to competitive markets, the software business wants faster, better, and cheaper solutions in a short amount of time. Software reuse emerges as a viable solution to these demands. It offers significant benefits such as increased quality and efficiency, lower development costs and effort, and shorter time to market. This research aims to study and understand the state of the practice of software reuse in Colombia. In addition, it will offer an instrument for decision-making in companies that adopt these practices. The method to reach this objective was organized in three stages. In the first stage, we defined the research questions. In the second stage, we developed, evaluated, and carried out a survey to validate successful practices and adoption barriers in the Colombian software industry context. Finally, the results were analyzed and reported. The paper compiles the survey participants' answers, showing the expectations, adoption barriers, and influence factors on the success of software reuse in Colombia. Also, this work serves as a roadmap for regions that want to analyze the current state of reuse. Nevertheless, each organization needs to determine its capabilities and find the factors appropriate to adopt to its context.

Keywords: Software reuse, survey, success practices, adoption barriers

RESUMO

Devido aos mercados competitivos, o negócio de software deseja soluções mais rápidas, melhores e mais baratas em um curto espaço de tempo. A reutilização de software surge como uma solução viável para essas demandas, pois oferece grandes benefícios, como maior qualidade e eficiência, menores custos e esforços de desenvolvimento e menor tempo de lançamento no mercado. Esta pesquisa tem como objetivo estudar e compreender o estado da prática de reuso de software na Colômbia e fazer uma comparação com trabalhos relacionados. Além disso, oferecerá um instrumento para a tomada de decisão nas empresas que adotarem essas práticas. O método para atingir este objetivo foi organizado em três etapas gerais. Na primeira etapa, definimos as questões de pesquisa. Na segunda etapa, desenvolvemos, avaliamos e realizamos uma pesquisa para validar práticas bem-sucedidas e barreiras de adoção no contexto da indústria de software colombiana. Por fim, os resultados foram analisados e relatados. O artigo compila as respostas dos participantes da pesquisa, mostrando as expectativas, barreiras de adoção e fatores de influência no sucesso da reutilização de software no contexto da Colômbia. Além disso, a experiência no desenvolvimento deste trabalho serve como um roteiro para regiões que desejam analisar o estado atual de reutilização. No entanto, cada organização precisa determinar suas capacidades e encontrar os fatores adequados para adotar em seu contexto.

Palavras-chave:

Reutilização de software, pesquisa, práticas de sucesso, barreiras de adoção

INTRODUCTION

The competitiveness in software development has increased over the last five years (Ramirez, 2019); however, besides the advancements, software industries require novel techniques to produce faster and better systems in a short time. As a result, product line engineering and other reuse-based methodologies such as duplication-based reuse, component-based reuse, delta-based reuse, and configuration management have gained increasing attention over recent years (Chikh, 2017; Mazo, 2018; Nogueira Teixeira, et al., 2018; Renault, 2014). Colombia is not an exception to this global trend. The conditions of the software industry, such as its transverse and globalized nature that supports any organization type, and the low cost of raw material investment, make it attractive to both domestic and foreign investors (MINTIC, 2019). In addition, the software industry is highly employment-generating, in design, development, maintenance, updating, and support phases. According to Fedesoft (2019), in 2018, the software industry in Colombia employed 109,000 people in 6,096 companies that billed 3 million euros per year with an average annual growth of 16.7% during the last six years in which 90% are micro and small enterprises, and 40% have five years old or fewer. Due to this growth, we consider it necessary to determine the state of the practice and identify

the technical and non-technical factors that influence software reuse practices in the Colombian industrial context.

This paper presents a survey that attempts to relate software development organizations' characteristics in Colombia with software reuse practices. The main objective of this research is to determine which factors have more influence on software reuse success. Twenty-six factors were identified in the literature and then considered in the survey. These factors were divided into four perspectives: organizational factors, business factors, technological factors, and process factors. In this survey, we also recorded recent experiences reported by industry professionals that already obtained success (or not) in reuse adoption in their organizations. Therefore, we collected information about adoption barriers in software reuse. This information helps identify trends, problems, and possible improvements to the reuse strategies currently used in the software industry. The survey involved Colombian software organizations' participation; however, we consider that other countries could take the results as a reference or carry out a similar study with the conditions of another country's context. It is because we found similar results in other surveys carried out in countries such as the US, Brazil, and others in Europe (Baharom, 2020; Barros-Justo et al., 2019; Bass et al., 2000; Frakes & Fox, 1995; Garcia et al., 2007; Karma et al., 2006; Kwon et al., 2015; Lucrédio et al., 2008; Morisio et al., 2002; Palomares et al., 2017; Rine & Nada, 2000; Rine & Sonnemann, 1998; Rothenberger et al., 2003; van der Linden et al., 2007).

The present document is structured as follows. Section 2 presents the research questions and the methodology used to develop the survey. Section 3 presents the survey results and describes the significance of the finding. Finally, Section 4 presents the conclusions of this research, and discusses possible future work.

METHODOLOGY

To make decisions (public policies and private initiatives) to sustain the software industry sector in Colombia, it is necessary to know its current state. Currently, this state is unknown, and to the best of our knowledge, there are no similar studies about the software industry sector in Colombia. A detailed snapshot of software reuse would enable us to identify the industry's opportunities, trends, and weaknesses. Therefore, it is necessary to know (i) the state of practice in software reuse in Colombia and (ii) identify the software reuse strategies (i.e., how it is applied, how it is used, which are its implications for practice, which are the research trends, which are the open problems and areas for

improvement) in the software industry sector in Colombian. Taking advantage of the key factors described in (Baharom, 2020; Barros-Justo *et al.*, 2019; Bass *et al.*, 2000; Frakes & Fox, 1995; Garcia *et al.*, 2007; Karma et al., 2006; Kwon *et al.*, 2015; Lucrédio *et al.*, 2008; Morisio *et al.*, 2002; Palomares *et al.*, 2017; Rine & Nada, 2000; Rine & Sonnemann, 1998; Rothenberger *et al.*, 2003; van der Linden *et al.*, 2007). We defined five global research questions for the Colombian scope:

- RQG1. What are the characteristics of software production (business factors)?
- RQG2. How is software reuse promoted in companies (organizational factors)?
- RQG3. How is software reuse applied and controlled (processes factors)?
- RQG4. How is software reuse supported (technological factors)?
- RQG5. What are the expectations and adoption barriers in the context characteristics (organization size)?

Each global research question (RQG1, RQG2, RQG3, and RQG4) was composed of a set of research questions. These research questions are listed in Table 1.

Tabla 1

	Research questions
Id	Research question
	RQG1-Business
RQ01	Which are the software application domains that have the most influence on the software reuse success?
RQ02	Which are the types of software developed in Colombia that have the most influence on software reuse success?
RQ03	Is using of a particular reuse approach as product family approach a factor that influences software reuse success?
RQ04	Is the implementation of a domain analysis a factor that influences software reuse success? RQG2-Organizational
RQ05	What are the organization sizes and the software team sizes that have the most influence on software reuse success?
RQ06	What team experience levels have the most influence on software reuse success?
RQ07	Which types of legal issues have the most influence on software reuse success?
RQ08	Is economic feasibility a factor that influences software reuse success?
RQ09	Is having organizational incentives strategies a factor that influences software reuse success?
RQ10	Is the commitment of top management a factor that influences software reuse success?
RQ11	Are training and motivation a factor that influences software reuse success?
	Is having an independent team that develops reusable software assets a factor that influences software
RQ12	reuse success?
	RQG3-Process
RQ13	Which are the quality models that have the most influence on software reuse success?
RQ14	Is the adoption of a software reuse process a factor that influences software reuse success?
RQ15	Is the measure of software reuse a factor that influences software reuse?
RQ16	Is having a certification process a factor that influences software reuse success?
RQ17	What are the types of developed assets for reuse that have the most influence on software reuse success?
RQ18	What are the configuration management and change control assets that have the most influence on software reuse success?
RQ19	Which are the types of reused assets that have the most influence on software reuse success?
RQ20	Is having a pre-development of reusable assets a factor that influences software reuse success?
RQ21	Which are the types of origins of reusable assets that have the most influence on software reuse success?
RQ22	Is having a specific role in the software reuse process a factor that influences software reuse success? RQG4-Technological
RQ23	Which are the software development approaches that have the most influence on software reuse success?

RQ24 Which are the programming languages that have the most influence on software reuse success?

RQ25 Is the use of repositories a factor that influences software reuse success?

RQ26 Is the use of CASE tools (language, frameworks, etc.) a factor that influences software reuse success?

A survey was used to answer research questions, because it allows us to reach a larger audience and collect and analyze data quickly from respondents in different regions of Colombia. The survey was carried out to define the state of the practice of software reuse in Colombia and to answer the research questions defined in Table 1. We carried out a survey following the guidelines of Pfleeger and Kitchenham, 2001. The survey follows a methodological process divided into seven steps: Step 1: Survey questions definition. Step 2: Survey preliminary evaluation. Step 3: Survey target population identification. Step 4: Survey data collection. Step 5: Survey data analysis. Step 6: Survey threats to validity. Step 7: Survey results. Steps 1 to 6 are described in the remainder of this section, and Step 7 is discussed in Section 3.

1. Survey questions definition

In the first step, we defined the survey questions. These questions were based on the 26 factors found in the literature (L. F. Restrepo Gutierrez *et al.*, 2021). The survey has 54 questions. It was built through an iterative process. Its final version was obtained after carrying out the preliminary evaluation, which is shown in subsection 2.2. The survey has closed and opened questions, distributed in three groups. Distribution is presented in the scheme sheet (L. F. Restrepo Gutierrez, 2021):

- Survey main question. One survey question allowed us to investigate and measure the software reuse success level and respond to the information collected through the other questions.
- **Survey research questions.** We defined 34 survey questions; each was related to a research question in table 1, which at the same time was related to a factor. These responses and their relationship with the main question allow us to answer the 26 research questions.
- Survey complementary questions. We defined 19 questions to collect general information about the business and the person who takes the survey, such as company location, software application domain, number of employees, staff experience, and person's role. Therefore, some of these questions were used to collect information about company tools, frameworks usage, programming languages usage, among others.

The survey was designed to be completed in less than 20 minutes; the whole questionnaire is available online (L. F. Restrepo Gutierrez, 2021).

2. Survey preliminary evaluation

In the second step, the survey questions were preliminary evaluated by six research experts in software engineering (internal review) and five software industry experts (external review).

Internal Review. For the internal review, we asked the six research experts to evaluate each survey question according to "a list of aspects to avoid in a survey" defined by Barbara & Shari (2002) and Leung (2001). List available in replication package on (L. F. Restrepo Gutierrez, 2021). Therefore, we asked the research experts to send feedback through the mail; in the end, it was analyzed and applied some changes in the survey questions based on that feedback. This internal review resulted in minor changes such as (i) clarifying terms and (ii) completing questions. Regarding the confusing terms, experts identified elements of the survey with unclear definitions resulting in misunderstanding. These elements were redefined to clarify them. For example, one of the experts suggested defining the systematic approach term since it was not understood to refer to software quality models. Regarding the incomplete questions, experts suggested adding more items to the answer lists for some questions. For example, in the list of services offered by companies, add hosting and software integration services.

External Review. After the internal review with research experts, we conduct a review with software industry experts. This evaluation included people with a high degree of experience in the field (head of the technology department, product manager, software innovation leader, and business analyst roles). These people were associated with recognized software companies in Colombia, located over several states of Colombia. To evaluate the survey, we held a one-hour meeting with each industry expert. We presented the survey content in each meeting, and we explained the research goal to each industry expert. We also requested each industry expert complete the survey aloud "say everything you think" (GAO, 2009). Then, we asked them to find issues regarding the survey questions. We recorded these issues, and we reviewed them. We also discussed the survey structure (topic coverage, survey length, sensitivity, and missing questions) with the industry experts. The final survey version was finished and shared with the external reviewers.

3. Survey target population identification

The survey population target was companies in Colombia whose main economic activity is software development. To identify this population, we got information from the Colombian chamber of commerce. This entity has a complete database that includes information that allows us to identify this population. We got information from the Colombian chamber of commerce of companies established in Colombia grouped by economic activity type. Using the database, we got 5818 companies whose principal economic activity was "Software development". Filters used are available in a replication package on (L. F. Restrepo Gutierrez, 2021). From these 5818 software companies, a sample of 361 companies was determined to be surveyed, considering a confidence level of 95% and a margin error of 5% (according to the general rule to acceptable margins of error mentioned by Kotrlik & Higgins, 2001). Since the response rates are typically well below 100% (Kotrlik & Higgins, 2001; Salkind, 2017) recommended oversampling when he stated that "If you are mailing out surveys or questionnaires, count on increasing your sample size by 40%-50% to account for lost mail and non-responders". For this reason, the oversampling was used to anticipate the response rate. For calculating the response rate, the first week's results after sending the survey were used to estimate how many additional responses were required. As the response rate was 9% of the initial sample of 361 companies, then it was established that the survey would be randomly sent to 4011 companies.

4. Survey data collection

We defined a standardized email to send the survey to the target population. We applied the following strategy to send the standardized email: (i) if the company provided an email address (for example, on its website), we sent the standardized email to that email address, or (ii) if the company did not provide an email address, we tried to contact it through the "contact us" website section, phone calls, and social networks, such as Facebook, Messenger, Linkedin, and WhatsApp Business. Two thousand five hundred thirty-four software companies were contacted through email and 1477 by other means. To motivate the respondents to share their experiences freely, it was mentioned that their answers would remain anonymous and that the results will be shared when they were published. It is important to note that we internally recorded each company's name and the company country state. This information was used to guarantee that the survey was answered by not more than one representative contact person per company. To manage

the survey responses, we used an online survey platform. In the end, 367 companies fully completed the online survey complying with the sample needed, with an average time responding of 15 minutes and a withdrawing rate of 37%.

5. Survey data análisis

This project was based on quantitative survey data, allowing us to do statistical analyses to test our research questions. Descriptive statistics and multinomial and ordinal logistic regressions were used to analyze the research questions. This analysis was used to determine which of the responses of the "survey research questions" (predictor variables) were the most influential over the "survey main question" responses (response variable), and, in the end, to be able to answer the research questions. Logistic regression models describe the relationship between a response variable and one or more explanatory variables. The strength of the logistic regression model is its ability to handle many variables, some of which may be on different measurement scales (Harrell, 2015). There are different logistic regression models as multiple, multinomial, and ordinal logistic regression models. The multiple logistic regression model considers having more than one independent variable. The ordinal logistic regression model considers the ordinal nature of the outcome. The multinomial logistic regression handles the case where the outcome variable is nominal with more than two levels (Harrell, 2015). As there is only one independent variable in the study, the multiple logistic regression model was ruled out. Instead, multinomial and ordinal logistics regression models were chosen to describe the relationship between variables.

The response variable ("survey main question" responses) was an ordinal numerical scale from 1 to 10 where 1 was "unsuccessful", and 10 was "very successful". This scale was grouped by two values, obtaining a scale of 5 levels from "no success" to "strong success" (no success, limited success, moderate success, substantial success, and strong success), where the values 1 and 2 of the initial scale correspond to the "no success" level and so on. Some of the predictor variables ("survey research question" responses) are also ordinal that goes from "strongly disagree" to "strongly agree" (questions with Likert scale answer). Other predictor variables ("survey research question" responses) were defined as categorical. The main question (related to the response variable) and an example of a survey research question (related to a predictor variable) can be found in the replication package on (L. F. Restrepo Gutierrez, 2021). The type of variable for each survey question can be viewed in the scheme sheet (L. F. Restrepo Gutierrez, 2021). For analyzing opened questions, each answer was hand-coded into one of several categories defined as presented on the sheet questions (L. F. Restrepo Gutierrez, 2021); these categories were created with a general review of the answers.

It should be noticed that multiple-choice questions cannot be taken into account as a single variable, so each item of those questions counted as an independent variable. Therefore, there were 135 independent variables, and one dependent variable in each model (L. F. Restrepo Gutierrez, 2021). Although some variables may be correlated, this analysis is not contemplated within the scope of this article. Each model required the response variable and the predictor variables. For the multinomial logistic regression, the response variable was entered as nominal; the response variable was ordinal for the ordinal logistic regression. The predictor variables associated with Likert scale answers could be entered as nominal or ordinal; for this reason, four models were created with the variation of taking response variable and predictor variables as ordinal and nominal. In order to create the models, the polr and multinom commands from the MASS package of the R tool (Ripley, 2020) were used. For estimating the best logistic regression model, an iterative model evaluation was made with the Step command from the R tool's stats package to fit the model. Step is an automated method based on the AIC (Akaike Information Criterion) model selection that aims to find out the best fit model from all possible subset models and returns back the optimal set of features (Hosmer et al., 2013). The four models were evaluated with the AIC. The replication package on (L. F. Restrepo Gutierrez, 2021) has the models created and evaluated. As a result of this process, model 3 and model 4 had the lowest AIC and the same value. The lowest AIC between models 1 and 2 is taken into account for deciding which model to choose, resulting in model 1, a model where the predictor variable is nominal. Model 4 is chosen where the response variable is ordinal, and the predictor variables are nominal. Results of the regression are available on (L. F. Restrepo Gutierrez, 2021).

6. Survey threats to validity

Threats to the survey's validity can be analyzed from the points of view of construct, internal, external, and statistical conclusion validity.

Construct validity. Construct validity addresses how well whatever is purported to be measured actually has been measured (Lavrakas, 2008). Verifying survey questions for construct validity becomes essential, especially when these questions are self-developed. We followed three steps to verify the construct validity. First, we performed a series of

validation tasks for identifying and calibrating ambiguously worded surveys. Second, four experienced researchers not involved in the questionnaire design were consulted to discuss each question. Then the survey was revised based on the feedback collected to ensure that the closed questions are clearly interpretable and sufficiently complete. Finally, a pretest was conducted to finetune the survey; two researchers administered the survey. They were asked to comment on their understanding of each item after filling out the questionnaire. Based on their feedback, minor revisions were performed.

Internal Validity. Internal validity considers whether the experimental design can support conclusions on causality or correlations (Morisio *et al.*, 2002). Questionnaires should have items logically interrelated to elicit information about the constructs under analysis without defects that would distort the information. The questionnaire presented in (L. F. Restrepo Gutierrez, 2021) was evaluated for internal consistency by computing the Cronbach's alpha test. It is the most widely used measure of reliability. It describes the internal validity (Hinkin, 1998), which is how much each measured element correlates to each other, or the overall consistency of the questionnaire (Cronbach, 1951). According to Hinkin, 1998, an acceptable value for Cronbach alpha is 0.7, but a minimum of 0.8 is the most desirable value. The evaluated questionnaire has a Cronbach's alpha of 0.9. Therefore, the questionnaire's internal coherence is acceptable, and the results presented in Section 3 are valid.

The accuracy of the answers is difficult to calculate due to the anonymity of the respondents (Singh, 2011). To face this, no incentives were given for responding to the survey then self-interest encouraged respondents to complete it. Also, answers were reviewed. As a result of this process, 222 incomplete surveys and six completed surveys were rejected due to inconsistency in the data, like company name not within our list of software companies or meaningless data in opened questions. Besides, the impact on the data depends on the size of the group surveyed. In the research, we have 367 reviewed and completed surveys. It would take several fifty people entering the wrong information to make a noticeable difference. In this case, the use of a survey protects us against bad actors. The dates of the studies found in the literature are different, so the comparisons are approximate since the state may have changed over time. The only way to avoid this threat to the validity of the comparison is to do the study in parallel. Such a study is costly, and we do not have the resources to do it.

External Validity. "External validity refers to the extent to which the research findings based on a sample of individuals or objects can be generalized to the same population that the sample is taken from or to other similar populations in terms of contexts, individuals, times, and settings" (Lavrakas, 2008). The sample defined represents the population in our survey, and the participants were chosen using a random sampling method. Therefore, our research can be generalized to describe software organizations in Colombia. Also, to avoid survey non-response effects, the response rate was maximized by sending reminders emails and re-sending the uncompleted survey.

Statistical conclusion validity. Statistical conclusion validity refers to the accuracy of a conclusion regarding a relation between or among variables of interest. As the sample size is considerable, there is no bias in the assumptions made of the descriptive statistics, but due to a large number of variables in the logistic regression model, there may be a bias in the results given by model 4. For this reason, a chi-squared-type statistic to test the validity of the logistic regression model was applied, using the anova command from the stats package of the R tool (Harrell, 2015; Hastie & Pregibon, 2017), which compares the model 4 before executing the Step command and the resulting model after running the command. The residual deviance of the model 4 was 521.8, for the resulting model was 477, and the Pr(Chi) was 0.9998, nearly 1, which indicates that the resulting model is better than model 4 and concludes that the resulting predictor variables have significant interactions with the independent variable. Another threat to the statistical conclusion validity is that the survey results focused on the impact of individual factors on the dependent variable, but some influences may imply combinations of factors. Finally, and as it was not contemplated, a correlation analysis within this article's scope is considered a threat that will be mitigated in future work.

RESULTS

This section presents the survey results and the answer to the 26 research questions.

Fig. 1 shows the number of participants of the survey grouped by Colombia regions. The software organizations that answered the survey were mainly located in the Colombian Middle-East region because 64.4% of software companies are located there. The Coffeebelt region follows this with 18.2% of the survey population, the Pacific region with 9.5%, and the other regions with 7.9% of the survey population.

Llanos with 40%, Coffee-belt with 67%, Middle-east with 72%, Pacific with 73%, Caribbean with 76%, and South-center with 80% reported a reuse success level that

ranges from strong to substantial, which indicates that Colombian software organizations are at similar levels in terms of software reuse (68% on average) regardless of the region where they are located. It is stressed that 68% of the participants indicated that reuse success is associated with companies that measure reuse level; the remaining companies made subjective analysis.

The participants represented a broad range of software domains (see Table 2). The categories are not mutually exclusive, and one participant can be placed in multiple software domains. Most of the participants work in software companies dedicated to developing custom products. Table 3 presents the participant roles. Again, the categories are not mutually exclusive, and one participant can play multiple roles. Most of the participants had the CIO role as their main role while playing a complementary role. In Colombia, mainly in Startups, the CIO can hold multiple roles, as back-end, front-end software developer, and architect at the same time. In large, medium, and small companies, the organizational structure is very well-defined, so people generally do not occupy more than one role.



Figure 1: Software organizations participants by regions of Colombia

Table 2
Study population's software domains

Table 3 Participant roles

, , , ,		1	
Software domain	Percentage	Role	Percentage
		Chief Information Officer (CIO)	31.3%
Custom software development	68.1%		24.00/
Development of new products	48.2%	Software development analyst	24.8%
Software and application	48.2%	Back-end software developer	24.3%
integration's		Software architect	21.0%
integration s		Front-end software developer	19 3%
Computer consultancy activities 47.4%			10.070

Software maintenance	36.2%	Chief Technology Officer (CTO)	14.2%
In house development	35.1%	Database administrator	13.9%
Administration of computer	24.8%	Consultant	10.9%
facilities		Component's developer	9.5%
Software testing	24.5%	Test analyst	8.4%
Hosting	21.3%	Business Analyst (BA)	6.3%
IT (Information Technology)	20.2%	Requirements engineer	6.0%
training programs		Other	5.7%
SaaS (Software as a Service)	2.4%	Coach	1.9%

Table 4 summarizes the influence of the studied factors on software reuse success. For Factors associated with a survey question with multiple choice answers, the influence column shows the categories that had the most influence according to the regression results. To see more details in the results for each factor see replication package – file "Survey Detailed Results" (L. F. Restrepo Gutierrez, 2021).

Research questions and factors		Influence	
RQ01	Application domain	Communications; Construction and Engineering; Financial; Legal;	
		Telecommunications; Tourism and entertainment	
RQ02	Type of software developed	Applications for mobile devices; Web applications; Applications for data analytics; Digital Content; Information systems for business management; IoT systems; Solutions based on artificial intelligence; Virtual and/or augmented reality	
RQ03	Product family approach	Yes	
RQ04	Domain analysis	Yes	
RQ05	Software organization and team size Team size - No	Startups, micro-enterprise, small and medium organizations sizes	
RQ06	Project team experience	All senior; more senior than junior; All junior.	
RQ07	Legal problems	legislation; contractual issues.	
RQ08	Economic feasibility	Yes	
RQ09	Reward and incentives	Yes	
RQ10	Management commitment	Yes	
RQ11	Software reuse education	Yes	
RQ12	Independent reusable assets development team	No	
RQ13	Quality models usage	CMMI-1 to CMMI-4; ISO 15504	
RQ14	Systematic reuse process	No	
RQ15	Software reuse measurement	Yes	
RQ16	Software certification process	Yes	
RQ17	Development of assets for reuse	User stories; tests plans.	
RQ18	Configuration management	Deployment artifacts; user documentation; detailed design.	
RQ19	Reused assets type identification	Requirements; user histories; high-level design (architecture); software components	
RQ20	Previous development of reusable assets	No	

Table 4 Factors that influence software reuse

RQ21	Origin of the reused assets	Artifacts are enhanced copies of existing work; Artifacts are developed from the re-engineering of existing products, not necessarily their own products; Artifacts are unmodified existing project work
RQ22	Specific function in the software reuse process	No
RQ23	Software development approach (method, methodologies, practices, and paradigms)	own methodology; FDD; Scrum; Aspect-oriented programming; procedural programming, constraint-based programming; reactive programming
RQ24	Programming language	C#; C++; SQL
RQ25	Repository systems usage	Yes
RQ26	CASE Tools usage	Yes

DISCUSSION

The results for every RQG are reported in the following items.

- RQG1. What are the characteristics of software production (business factors)? As its main production characteristic, the software industry in Colombia has the development of various custom software products that share characteristics. Furthermore, most organizations do not focus on a specific application domain; only 14% focus on one application domain. Therefore, companies have an excellent opportunity to take advantage of the software product lines approach. However, it should be noted that the product family approach factor was not thoroughly investigated because it is a software reuse strategy that contains many concepts, processes and practices (Mazo, 2018) that are beyond the scope of this investigation.
- **RQG2.** How is software reuse promoted in companies (organizational factors)? The companies promote software reuse through incentives, training and motivation, and top management's support to introduce and maintain software reuse. Legal issues can occur when working for Colombia's banking, financial, government, and health domains. Most participants believe that software reuse is economically feasible in his company.
- **RQG3.** How is software reuse applied and controlled (processes factors)? Software reuse is applied through a systematic reuse process established by the company not associated with quality models; in this systematic reuse process is identified that:
 - The planning of reusable assets development is not done from the beginning of the project; it is performed during the system's development.
 - Companies do not usually have a role or independent team dedicated to developing reusable assets. It is the task of all members of the development teams.

- The most created reusable assets are source code, detailed designs, software components, and libraries. The assets are under a configuration management platform.
- The certification process of these reusable assets to be reliable is not strictly used in Startups and Small companies.
- The most reused assets are the software components, architectures, test cases, and detailed designs.
- The origin of the reused assets is mainly from existing projects, which are improved to use. – Almost half of the companies measure their progress in software reuse.
- **RQG4. How is software reuse supported (technological factors)?** Software reuse is supported through the use of repositories and CASE Tools. It should be noted that the most used programming paradigm is object-oriented, and the most used methodology is Agile.
- **RQG5.** What are the expectations and adoption barriers in the context characteristics (organization size)? It was identified that cost reduction, decreased labor needs, and productivity increase are significant software reuse expectations. And the principal adoption barriers are the diverse running projects, the lack of time and resources, and lack of communication between employees.

Concerning state of the art, a high level of similarity between the result of this study and the literature was found such as (Baharom, 2020; Barros-Justo *et al.*, 2019; Bass *et al.*, 2000; Frakes & Fox, 1995; Garcia *et al.*, 2007; Karma *et al.*, 2006; Kwon *et al.*, 2015; Lucrédio *et al.*, 2008; Morisio *et al.*, 2002; Palomares *et al.*, 2017; Rine & Nada, 2000; Rine & Sonnemann, 1998; Rothenberger *et al.*, 2003; van der Linden *et al.*, 2007). Some factors led to the same conclusion of influence on software reuse success: application domain, product family approach, domain analysis, management commitment, economic feasibility, team size, the origin of the reused assets, configuration management, and reused assets type, and development of assets for reuse. Furthermore, it was evidencing that they have been conserved over time given the new development practices.

Particularly for the Colombian case, the results concluded that software organization size impacts on software reuse success. In the literature, this factor was treated in conjunction with the team size factor, but we found that it has a different influence, which suggests that this factor should be separated for further investigations. Also, it was found that a systematic reuse process, an independent reusable assets development team, and the previous development of reusable assets factors do not influence software reuse success. This shows that a software reuse process has not been formally adopted in Colombian software companies; the reuse of assets is made opportunistically, that is, reactive and unplanned. Despite, it is noted that benefits have been obtained since great work has been done from the organizational factors (incentives, training, top management support, etc.), but all companies need to formalize these reuse practices as part of their process model activities, to get to a point where reuse is planned and adopted by the entire development group. We hope that what is related here helps show the importance of these reuse practices and encourages others to replicate this work in other countries where the software industry is emerging and see these reuse practices accelerate the growth, quality, and rapid delivery of software.

CONCLUSIONS

This paper presented the results of the survey for software reuse in Colombian software organizations, providing important factors to consider for success in the adoption of this approach, also, experiences, current state, adoption barriers and expectations in software reuse, information that can be used to introduce reuse in all organizations sizes to obtain the desired benefits. Besides, the work presented here results from collaborative work with the software industry in Colombia, people interested in this result, and the appropriate skill to provide us with evidence of what is happening in organizations related to this issue. To sum up, we hope this work has helped identify aspects where software reuse can incorporate or improve. Nevertheless, each organization needs to determine its capabilities and find the factors appropriate to adopt to its context. With this work, we also want to motivate other regions to prepare this type of study to have a more global understanding of reuse and its adoption factors so that more companies have support information for decision-making.

As future works, (i) a deeper analysis should be made among the factors to present them in the form of a typology because there may be relationships among them (some may be sub-types of others and some may complement others), (ii) an analysis of correlations between the variables will be carried out, (iii) a comparison between the regions of Colombia will also be made, (iv) the resulting factors should be further put into practice through the use of case studies with companies that want to go deeper into these factors to have a most refined understanding, and (v) a capability maturity model will be produced with these results to describe particular software development processes to recommend feasible successful practices to enable and foster the adoption of assets reuse, then, this proposed model will be validated with software industries.

ACKNOWLEDGEMENTS

The authors would like to thank Vicerectoria de Descubrimiento y Creación from Universidad EAFIT. Universidad EAFIT supported this research. The authors would also like to thank Luisa Rincón, Henry Laniado, and Henry Velasco for their early comments and suggestions on this research and the anonymous participants.

BIBLIOGRAPHIC REFERENCES

- Baharom, F. (2020). A survey on the current practices of software development process in Malaysia. *Journal of Information and Communication Technology*, *4*, 57–76.
- Barbara, K., & Shari, P. (2002). Principles of survey research part 4. ACM SIGSOFT Software Engineering Notes, 27(3), 20–23. https://doi.org/10.1145/638574.638580
- Barros-Justo, J. L., Olivieri, D. N., & Pinciroli, F. (2019). An exploratory study of the standard reuse practice in a medium sized software development firm. *Computer Standards & Interfaces*, 61, 137–146. https://doi.org/https://doi.org/10.1016/j.csi.2018.06.005
- Bass, L., Buhman, C., Comella-Dorda, S., Long, F., & Robert, J. (2000). Volume 1: Market Assessment of Component-Based Software Engineering. Software Engineering Institute. http://www.dtic.mil/docs/citations/ADA395250
- Chikh, A. (2017). Component-based approach for requirements reuse. *The Knowledge Engineering Review*, 32, e11. https://doi.org/10.1017/S0269888917000030
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.
- Fedesoft (2019). ¿Cómo es la industria de Software y TI colombiana? https://fedesoft.org
- Frakes, W. B., & Fox, C. J. (1995). Sixteen questions about software reuse. Communications of the ACM, 38(6), 75-ff. https://doi.org/10.1145/203241.203260
- GAO. (2009). *Questionnaire Pretest Procedures*. https://www.ignet.gov/sites/default/files/files/14_Questionnaire_Pretest_Procedure s.pdf
- Garcia, V., Lucrédio, D., Alvaro, A., Santana De Almeida, E., Fortes, R., Fortes, M., Romero, S., & Meira, L. (2007). Towards a Maturity Model for a Reuse Incremental Adoption. *SBCARS*. https://doi.org/10.6084/M9.FIGSHARE.96661
- Harrell, F. E. (2015). Ordinal Logistic Regression. In *Regression Modeling Strategies* (pp. 311–325). Springer. https://doi.org/10.1007/978-3-319-19425-713
- Hastie, T. J., & Pregibon, D. (2017). Generalized Linear Models. In *Statistical Models in S* (pp. 195–247). Routledge. https://doi.org/10.1201/9780203738535-6

- Hinkin, T. R. (1998). A Brief Tutorial on the Development of Measures for Use in Survey A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires Questionnaires. https://doi.org/10.1177/109442819800100106
- Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). Introduction to the Logistic Regression Model. In *Applied Logistic Regression* (pp. 1–30). John Wiley & Sons, Ltd. https://doi.org/10.1002/0471722146.ch1
- Karma, S., Radha, A., & Zhangxi, L. (2006). Resources and incentives for the adoption of systematic software reuse. *International Journal of Information Management*, 26(1), 70–80. https://doi.org/10.1016/J.IJINFOMGT.2005.08.007
- Kotrlik, J., & Higgins, C. (2001). Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, *19*(1), 43.
- Kwon, Y., Kim, E., & Lee, N. (2015). Key factors on software reuse of e-Government common framework. *International Conference on Advanced Communication Technology, ICACT*, 2015-Augus, 774–781. https://doi.org/10.1109/ICACT.2015.7224900
- Lavrakas, P. J. (2008). Encyclopedia of survey research methods. Sage Publications.
- Leung, W.-C. (2001). How to design a questionnaire. *BMJ*, 322, 106187. https://doi.org/10.1136/sbmj.0106187
- Lucrédio, D., dos Santos Brito, K., Alvaro, A., Garcia, V. C., de Almeida, E. S., de Mattos Fortes, R. P., & Meira, S. L. (2008). Software reuse: The Brazilian industry scenario. *Journal of Systems and Software*, 81(6), 996–1013. https://doi.org/10.1016/J.JSS.2007.08.036
- Mazo, R. (2018). Guía para la adopción industrial de líneas de productos de software -Editorial EAFIT / Colecciones - Universidad EAFIT. Editorial EAFIT. https://www.eafit.edu.co/cultura-eafit/fondo-editorial/colecciones/Paginas/guiapara-la-adopcion-industrial-de-lineas-de-productos-de-software.aspx
- (MINTIC), M. de T. de la I. y las C. (2019). *Fortalecimiento de la industria TI FITI*. https://www.mintic.gov.co/portal/inicio/14404:Fortalecimiento-de-la-industria-TI-FITI
- Morisio, M., Ezran, M., & Tully, C. (2002). Success and failure factors in software reuse. *IEEE Transactions on Software Engineering*, 28(4), 340–357. https://doi.org/10.1109/TSE.2002.995420
- Nogueira Teixeira., E., Vasconcelos., A., & Werner., C. (2018). OdysseyProcessReuse -A Component-based Software Process Line Approach. *Proceedings of the 20th International Conference on Enterprise Information Systems - Volume 2: ICEIS*, 231–238. https://doi.org/10.5220/0006672902310238
- Palomares, C., Quer, C., & Franch, X. (2017). Requirements reuse and requirement patterns: a state of the practice survey. *Empirical Software Engineering*, 22(6), 2719–2762. https://doi.org/10.1007/s10664-016-9485-x
- Pfleeger, S. L., & Kitchenham, B. A. (2001). Principles of Survey Research: Part 1: Turning Lemons into Lemonade. SIGSOFT Softw. Eng. Notes, 26(6), 16–18. https://doi.org/10.1145/505532.505535

- Ramirez, J. M. (2019). *Industria nacional de software | Opinión | Portafolio*. https://www.portafolio.co/opinion/juan-manuel-ramirez-m/industria-nacional-desoftware-533306
- Renault, O. (2014). Reuse / variability management and system engineering. *CEUR* Workshop Proceedings, 1234, 173–194.
- Restrepo Gutierrez, L. F. (2021). *Replication package for: "Snapshot" of the State of Software Reuse in Colombia. 1.* https://doi.org/10.17632/2HDX42X6WC.1
- Restrepo Gutierrez, L. R., Suescún Monsalve, E., Mazo, R., Correa, D., & Vallejo, P. (2021). Success factors and adoption barriers in software reuse: A literature review. *Investigación e Innovación En Ingenierías* [To Be Published].
- Rine, D. C., & Nada, N. (2000). Empirical study of a software reuse reference model. *Information and Software Technology*, 42(1), 47–65. https://doi.org/10.1016/S0950-5849(99)00055-5
- Rine, D. C., & Sonnemann, R. M. (1998). Investments in reusable software. A study of software reuse investment success factors. *Journal of Systems and Software*, 41(1), 17–32. https://doi.org/10.1016/S0164-1212(97)10003-6
- Ripley, B. (2020). Ordered Logistic Or Probit Regression. https://www.rdocumentation.org/packages/MASS/versions/7.3-51.5/topics/polr
- Rothenberger, M. A. A., Dooley, K. J. J., Kulkarni, U. R. R., & Nada, N. (2003). Strategies for software reuse: a principal component analysis of reuse practices. *IEEE Transactions on Software Engineering*, 29(9), 825–837. https://doi.org/10.1109/TSE.2003.1232287
- Salkind, N. J. (2017). Exploring research (9th editio). Pearson Education Limited.
- Singh, L. (2011). Accuracy of Web Survey Data: The State Of Research on Factual Questions in Surveys. *Information Management and Business Review*, 3, 48–56. https://doi.org/10.22610/imbr.v3i2.916
- van der Linden, F., Schmid, K., & Rommes, E. (2007). *Software Product Lines in Action.* Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-71437-8