Definition of a Support Infrastructure for Replicating and Aggregating Families of Software Engineering Experiments

Efraín R. Fonseca C.
Universidad Politécnica de Madrid
Boadilla del Monte 28660
Madrid, Spain
erfonseca@espe.edu.ec

ABSTRACT

Experimental software engineering includes several processes, the most representative being run experiments, run replications and synthesize the results of multiple replications. Of these processes, only the first is relatively well established in software engineering. Problems of information management and communication among researchers are one of the obstacles to progress in the replication and synthesis processes. Software engineering experimentation has expanded considerably over the last few years. This has brought with it the invention of experimental process support proposals. However, few of these proposals provide integral support, including replication and synthesis processes. Most of the proposals focus on experiment execution. This paper proposes an infrastructure providing integral support for the experimental research process, specializing in the replication and synthesis of a family of experiments. The research has been divided into stages or phases, whose transition milestones are marked by the attainment of their goals. Each goal exactly matches an artifact or product. Within each stage, we will adopt cycles of successive approximations (generate-and-test cycles), where each approximation includes a different viewpoint or input. Each cycle will end with the product approval.

Keywords

Infrastructure, Families of Experiments, Support, Replication, Aggregation, Transfer, Knowledge

1. INTRODUCTION TO THE RESEARCH

1.1 Motivation

We have detected the need to improve the replication and synthesis processes. A tool that provides integral support for the software engineering (SE) experimental research process and provides experimenters with support for handling and managing information within a family of experiments [4] has yet to be developed. Such an instrument should make it easier to extract pieces of knowledge from replications that are members of a family of experiments.

1.2 Background

Experimental software engineering (ESE) includes several processes. The most representative are experimentation, replication and synthesis. Of these, only the first is relatively well established in SE. Hundreds of experiments have now been run covering virtually all areas of SE [28], and this number is growing all the time [35] [10]. The number of experiments run does not mean that experimentation is a fully institutionalized methodology, but signs of maturity are gradually emerging.

The replication process is, however, quite a different matter. SE experimentation is necessary, but complex. It is fed by mechanisms for motivating studies and integrating results, which, in turn, rely on researchers replicating studies [4]. To do this, the researcher that ran the baseline experiment has to pass on all the relevant information to the researcher that is to run the replication. This information is usually transmitted by means of experimental reports (for example, articles describing experiments published in journals or conferences) or, at best, using experimental packages [30].

Experimental reports are sometimes drafted according to guidelines, such as those proposed by Jedlitschka and Pfahl [15], although compliance is not always strict. Even reports that do follow guidelines omit some of the information required for replication, which is often confined to an outline of the experiment, design, some details on the statistical tests run, experimental objects and a list of replications of the experiment run and the experiences gathered from these replications [33].

Experimental packages are tools used to guide researchers through the process of replicating an experiment. They also provide the materials and information required at replication time and thereby somewhat minimize differences with the results of other similar replications [17]. There is no standard experimental package formulation [33], and individual researchers build packages at their discretion, including whatever information they consider to be relevant for replication [18].

Even if packages are used instead of publications, several problems still tend to occur:

IDoESE September 21, 2012, Lund University, Sweden

- Packages do not usually provide raw data or specify the details of the replications to improve synthesis (accurately discover secondary factors, for example) [29].
- Packages do not make a distinction between possible experimental package users (experimenters, replicators, meta-analysts, etc.), who have different views and needs and do not always use the same terms to refer to concepts.
- Experimental packages tend not to record many details that are relevant for replication [3], such as, for example, contextual variables. This information belongs to the category of implicit knowledge. There are different types of knowledge. In this case, the most important is the knowledge that the researcher knows of and fails, for any number of reasons, to pass on. Researchers refer to this type of knowledge differently, using terms such as non-explicit, uncoded, individual, personal , strategic, etc. [9] [25]. As experimental packages do not transfer this knowledge, it may not be possible to identify why the results of the replication differ from the original experiment.
- Experimental packages do not account for the adaptation of the original experiment to the new context in which the replication is to be run [33].

Solari interviewed replicating experimenters and found that it is very hard to replicate an experiment from the information supplied in experimental packages [30]. Generally, the original and replicating researchers have to meet and, in an open discussion, review the experimental protocol step by step to specify the knowledge, which experimental packages tend not to do, and adapt the original experiment to the replicator's environment [33].

The circumstances surrounding replication and synthesis are not substantially unlike. On the one hand, there are well-defined procedures, such as systematic review [19], that are useful for searching for primary studies, extracting data and aggregation. However, very few aggregations have actually been completed in the last few years [8]. The number of statistical aggregations (meta-analyses), considered to be the most reliable procedure [6], is even smaller. The synthesis of replications belonging to a family should help to extract new relevant variables [17], but this is not common practice in ESE [18].

1.3 Problem Statement

Replication and synthesis activities in ESE suffer from a problem of between-researcher information exchange. This problem has different manifestations depending on the activity (replication or synthesis), but the effect is in both cases to obstruct or prevent execution, as the relevant information (context, raw data, etc.) is not readily available.

The above weaknesses of the experimental process have evidenced the problem that our research will address. Generally, the aim is to facilitate the management and communication of explicit and implicit information about the experimental process to help to run replications and aggregate experimental data gathered from a family of experiments. More specifically, the problems to be solved are:

- There is no common reference framework (terminological and operational) to facilitate communication between researchers and the activities that researchers perform in the roles that they play in the experimental research process.
- Relevant information about an experiment for either replication or aggregation with other experiments is not always accessible.
- There are no experimental research process support tools for running replications and synthesizing the results of studies that are members of a family of experiments.

2. RELATED WORK

Experimentation in SE has expanded considerably over recent years. This has led to proposals to support the experimental process. However, there is no proposal that provides an infrastructure to support experimenters in the SE experimentation process, which is the simplest of the above problems.

We have completed a literature review concerning both the experimental research process and support tools. In the future, we intend to apply the systematic review process proposed by Kitchenham and Charters [19] to refine this initial review. However, the research conducted so far has revealed the following facts.

- Apart from standard texts on experimentation, like Wohlin et al. [34] or Juristo and Moreno [16], there are no proposals or papers that address how to conduct the experimental process in SE. The standard texts propose not a process in the strict sense but a set of general guidelines about how to plan, run and analyse an experiment. Other proposals, like Kitchenham and Charters [19], have dealt with methodological guidelines for the synthesis process. Synthesis is embedded within a systematic literature review, unconnected with the work of the SE experimenter. Additionally, synthesis is detached from the execution and replication of experiments.
- The best mechanisms existing today for transmitting relevant information among researchers are: (1) experimental reports, in compliance with standards such as the guidelines proposed by Jedlitschka et al. [15], and (2) experimental packages, such as those built by Lott [22], Shull [2], Solari [30], etc. However, as mentioned earlier, the information that these mechanisms provide is still unsatisfactory from the viewpoint of the replicating experimenter.
- There are no repositories from which to extract experimental data with a view to a potential aggregation process, except more general-purpose repositories like PROMISE [26] and reports published by experimenters in journals and conferences.
- There are several experimental process support tools, like SIR [11]. But these tools generally specialize in one part of the experimental process (for example, provide experimental objects) and do not provide integral support for the problem identified in this research. There are also praiseworthy attempts at posting information on the web about specific experiments for the purpose of replication by Genero [12] and Lott [22].

In the following, we describe the points mentioned above in more detail.

2.1 Proposals for Formalizing the Experimental Process

Apart from the above two standard texts [34, 16], there are not many proposals for supporting the development of a formal experimental process in SE. Kitchenham et al.[20] proposed a more formal approach to experimental research in SE based on guidelines developed by medical researchers and grounded on their own SE experience. However, as they themselves claim, "There needs to be a wider debate before the software engineering research community can develop and agree on definitive guidelines".

Finally, there are guidelines for performing a systematic literature review [19], which include the synthesis process. This report proposes guidelines for systematic literature review. The last stage is of this systematic review is to synthesize the identified experimental results. It is not designed, however, to be a methodological proposal on how to synthesize experiments, and there is a risk that not all the identified papers will contain the necessary information for results synthesis.

2.2 Reporting Standards

The situation of reporting standards and the circumstances of theoretical proposals of formalization of the experimental process are not far removed. We have found few reporting standard proposals. On the one hand, Singer [27] proposes the use of the American Psychological Association (APA) Style Guidelines to report experimental results in ESE. The APA guidelines were devised to solve reporting problems related to replication and meta-analysis in psychology. Adaptation to ESE again requires acceptance and validation by the ESE community.

On the other hand, Jedlitschka et al. [15, 14] propose guidelines for reporting experiments aimed primarily at minimizing the amount of relevant information that is not available for evaluating an experiment. Unless the tool is used in routine practice, however, it will not fully eradicate the problem that it sets out to solve. Additionally, it does not target results synthesis.

2.3 Replication Packages

Lott's experimental package [22] provides the elements necessary to replicate an experiment that compares testing and reading techniques. This package is now operational and accessible via web. Generally, this package offers the following information: technical reports and publications about the replications run, experimental materials (lecture notes, instructions and data collection forms) and experimental objects used in both the training and execution phase (specification, code, example, failures and faults).

Shull's package [2] describes an experiment that has been designed to compare software reading techniques. This package is now operational and accessible via web. This package offers the following information: a description of the experimental design, analysis techniques used, experimental materials and experiences gathered from the replications run using this package.

We have located other packages apart from the above, like Genero [12].

Generally, experimental packages do not include detailed information about the original experiment. This could generate a conceptual gap in the replicating researcher. The conceptual gap could be between the definition of the original experiment and the experimental elements used in the replication. Without knowledge of the factors, response variables, metrics (presented in the associated publication), for example, the importance and utility of the data collection instruments and the experiment measurement instruments are unclear. With access to the detailed information of the original experiment, it would be quite feasible to analyse the instruments implemented for the replication from a different angle and improve them.

Another weakness that we have found in experimental packages is the ambiguity of the information about the data analysis techniques. They tend to omit sections that detail the analysis techniques used in past replications or that provide advice on which technique to use based on the details of the original experiment.

None of the identified experimental packages provides support for collaborative work, storage for raw data from all the replications run in a family, help for adapting the experiment to the new context in which the replication is to be run, etc.

In sum, we have not found a replication package that provides interactive support for a group of researchers interested in enacting the experimental research process from start to finish: experiment, replications, synthesis (with identification of variables).

2.4 Experimental Information Repositories

eSEE [21] is a knowledge repository capable of instantiating SE environments to manage the knowledge of the definition, planning, execution and packaging of SE experimental studies. This proposal reportedly promises to provide support covering much of the experimental process. However, it does not refer to the replication and experimental synthesis processes. The eSEE repository is available via web, but we have not been able access and validate all the features that are described in the report.

CeBASE [5] is an apparently now obsolete repository of experiences. This repository is able to reconcile several phenomenological software models in a common framework in order to pursue empirical software engineering research and organize the results in a useful experience base.

ViSEK [13] is an on-line repository for promoting intra- and inter-organizational learning in small- and medium-sized German software enterprises. This repository is now operational, although no new information has been added since it was set up. It is capable of recording experiments, instantiating technologies and exchanging experiences through a network of SMEs.

SIR [11] is a repository that provides Java and C programs for use in experimentation with testing techniques and materials that facilitate their use. SIR is now operational and work is ongoing, with continuous improvements and extensions.

The analysed repositories generally serve the purposes for which they were built. Nevertheless, none of these repositories sets out to provide integral support for the experimental process in SE. From this viewpoint, the analysed repositories are conceptually incomplete, as is the information recorded about the experimental research process.

2.5 Experimental Process Support Tools

SESE [1] is a web-based SE experimentation support environment for professional developers that perform experimental tasks as part of their routine work. This tool is apparently still in production, but we have been unable to find out how to gain access. We have not found any indication specifying the mechanism for future experiment replications or signalling data analysis or synthesis.

Ginger2 [31] is a computer-assisted ESE environment composed of three elements. The first element is an ESE process model. The second element is an ESE decision-making support model. The third element is a framework-based architecture composed of a toolkit for each process model phase, process management mechanism and data integration and control mechanisms. However, the environment core does not appear to be the experimental process, and there is no model outlining the experimental process and its main elements. We have not found any information on how and where to access Ginger2 and are, therefore, unfamiliar with its current status.

Giants [32] is an on-line web portal proposal providing a data collection service, algorithms, experimental settings and millions of empirical evaluations submitted by data mining tools and individual users. It is a search resource designed to help researchers to set up and execute new experiments. This tool is just a proposal. We have not been able to locate the proposal for evaluation purposes and are, therefore, unfamiliar with its current status.

In sum, there have been some attempts to provide software support for the experimental process in SE. However, none of the tools appear to be accessible save in the paper in which they are reported. Neither do they appear to provide support for every stage of the experimental process in the shape of either guidance for experimenters, data analysis, experimental materials management or instrument and raw data storage. Our findings suggest that all the tool proposals are sound and worth analysis, but very few, if any, have been followed up, and their current status is unclear.

3. RESEARCH OBJECTIVES

The aim of this research is to provide the necessary technological capabilities to facilitate the exchange of information and knowledge among experimenters working on a family of SE experiments. The detailed research objectives are:

- Study the current situation in order to contextualize the problem to be addressed. This study is divided into two parts: a study of the state of the art (related research work) and a study of the state of practice (what researchers actually do when they perform experimental research) in order to take note of, identify and thoroughly understand the complications of information and knowledge exchange among experimenters working on a family of experiments.

- Define conceptual models to formalize the three aspects of the experimental process in SE: running experiments, running replications and synthesizing results.

- Structure an information repository (based on conceptual models) for handling and managing explicit and implicit information and knowledge on the experimental process in SE.

- Build a technology tool to manage the processes, activities, tasks and support services that are part of the experimental research process in SE.

4. PROBLEM-SOLVING APPROACH

Based on the analysis of the state of the practice and related work, we will propose a technological solution consisting of an infrastructure to support the experimental process in SE focused on a family of experiments [4].

The experimental research process is complex and tends to rely on the tacit and intrinsic knowledge of the researchers collaborating on the investigation of a family of experiments. The study of the state of practice has revealed that there are several roles in collaborative experimental research, and each role takes a different (albeit complementary) view of the experimental process. Each role performs a group of activities, and each researcher can play one or more roles in the experimental process. The fact that there is neither a common terminology nor a standardized means of managing the information and knowledge that experimenters handle is an obstacle to or prevents communication, often leading to an information-researcher dependency (where information is concealed from other experimenters and is the exclusive property of a single experimenter). The consequences of these problems are what complicate the replication and synthesis processes.

The experimental process needs to be formalized by explicitly defining conceptual models that specify the information and knowledge used by researchers from the viewpoint of each role. Albeit on a smaller scale, these conceptual models aim to play the same role as controlled vocabularies do in other disciplines like medicine (for example, MED [7]), namely, facilitate communication and understanding among the professionals involved, in this case, experimenters working together on a family of experiments.

The conceptual model will structure an ontology that clearly defines the elements of the experimental process (experiments, replications and synthesis) and their relationships. This model should facilitate information communication and exchange among experimenters working together on a family of experiments.

The conceptual model is useful for structuring and storing all the information handled in an experimental collaboration in a central repository (CR). The created ontology will help researchers to develop the activities proper to the role that they are playing and also specify all the knowledge that could support the experimental process, as well as communication with other experimenters with whom they are working on the family of experiments.

Remember that all researchers have their own research style and way of managing information about their activity as part of any of the different roles they play. The CR will integrate all the information required for the experimental process, including the information that each researcher handles locally. To do this, we will implement a mechanism for mapping the data from researchers'local repositories to the ontology. Viewed from the global perspective of collaboration, this will fracture the information-research dependency, and more importantly will not much alter each experimenter's modus operandi.

Apart from building a conceptual model, we also have to model the processes, activities and tasks involved in the experimental research process. The definition of this process model will standardize the experimental research process, making the activities more repeatable.

The CR models and functionality will be implemented in a tool that regulates the processes, activities and tasks involved in the experimental research process, as well as its inputs and outputs. As part of the tool, we propose several experimenter support processes:

- The tool will provide experimenters with decision-making support on the best options to take under certain circumstances in order to execute the most relevant processes, especially critical tasks. For example, the tool will analyse the experimental elements and, depending on their meaning, suggest the best design options.
- The tool will include data analysis services based on free software tools. This service will offer users support based on previously input statistical procedures. These procedures will operate according to the selected data and analysis options, and the link with the experimental elements.
- The tool will provide easy access to the experimental elements, experimental materials and their link and to a host of experimental details. The tool will offer the option of enacting a guided replication based on the details of earlier replications, especially details of events that marked any important milestone. This would help experimenters to make decisions conducive to improving replications and ultimately the validity of the results.
- The tool will enable aggregation and ultimately meta-analysis of the results of a family of experiments. It will not account for aggregation and meta-analysis outside a family of experiments, for example, as a result of a systematic review. The proposed tool will in particular support the synthesis activity by assuring raw data maintenance. This will promote the application of alternative experiment aggregation strategies, such as, for example, an experiment-blocked ANOVA.

5. RESEARCH METHOD

The research has been divided into stages. The stage transition milestones are marked by the achievement of their objectives. Each objective is equivalent to a specific artefact or product (for example, conceptual model, repository, etc.) to be generated. Consequently, the outcome of a stage is represented by an artefact built during that stage. The research stages are:

- Stage 1: State Problem
- Stage 2: Define Conceptual Model
- Stage 3: Structure Information Repository
- Stage 4: Build Management Tool

Within each stage, we will adopt cycles of successive approximations (generate-and-test cycles) where each approximation includes a different viewpoint or contribution. Each cycle will end with the approval of the product.

Figure 1 illustrates the research method, including the estimated completion dates. The planned stages are detailed below.

- **Stage 1 - State Problem** The objective of this stage is to study the current situation to gather in-depth knowledge of the problem in question. To do this, we will carry out the following activities:

  Define the Problem: In a first approximation, we define the problem based on the needs of a group of experimenters collaborating on a particular family of SE experiments [17], as well as an initial literature review. The experimenters are members of the Empirical Software Engineering Research Group (GrISE)[1]. GrISE has about ten members from different higher education institutions like the Universidad Politécnica de Madrid, Universidad ORT de Uruguay, Universidad Autónoma de Madrid, etc.

  Review the Literature: Based on this problem definition, we initially review the experimental process support systems in a second approximation to evaluate whether any existing infrastructure provides the type of support that the experimenters need.

  Elicit Knowledge: Taking the review of the relevant literature, we interview, in this approximation, experimenters collaborating on a family of experiments in order to learn about the state of practice. Specifically, we elicit the knowledge and experience of expert SE researchers that are members of GRISE. The preliminary result was a provisional structure which was then revised and validated by the same experts.

  Systematically Review the Literature: In the third approximation, we undertake a systematic review of existing systems that support the experimental process in SE.

  The product of this first stage will be an understanding of the current situation, expressed as both the state of the art and the state of practice.

- **Stage 2 - Define the Conceptual Model**

  The objective of this stage is to define conceptual models to formalize the experimental research process in SE. To do this, we will carry out the following activities:

  Review the Conceptual Framework: In a first approximation, we conduct a review of the literature on the experimental process in SE to discover theoretical knowledge, specifically, processes, elements and relationships, actors, etc.

  Outline the Conceptual Model: We map the experimental research elements (experiments, replications, synthesis) to the conceptual model. The model will be composed of the conceptual submodels representing the tasks of the different experimental research roles.

  Experts Evaluate the Structure: The model is built and evaluated iteratively by experimenters with different roles. Specifically, the most experienced GrISE members will be responsible for evaluating the structure by inspecting the model, case studies and prototypes.

  The product of this second stage is the conceptual model, which will be used to structure the information repository.

[1]The GrISE website is available at `http://www.grise.upm.es/`

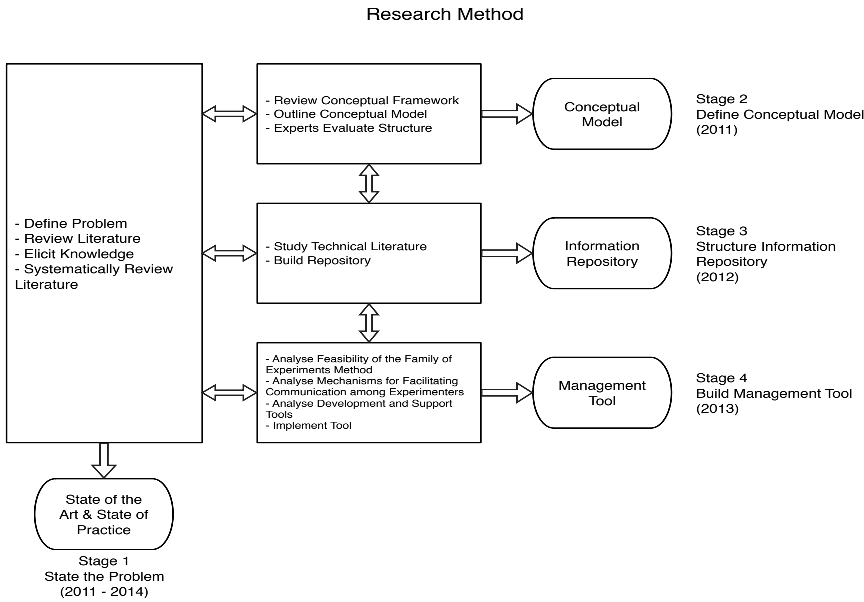

Figure 1: Artefacts Output by the Research

- **Stage 3 - Structure the Information Repository**

  The objective of this stage is to structure the information repository to facilitate the handling and management of explicit and implicit information about the experimental process in SE based on the conceptual models. To do this, we will carry out the following activities:

  Study the Technical Literature: In this first approximation, we study the technical literature to determine the best tool for building the information repository.

  Build the Repository: Having decided which tool to use to build the information repository, we go ahead and build the repository in this approximation.

  The artefact output by this stage is the information repository.

- **Stage 4 - Build the Management Tool**

  The goal behind this stage is to build a technology tool to manage the processes, activities, products and support services for the experimental research process in SE. To do this, we will carry out the following activities:

  Analyse the Feasibility of the Family of Experiments Method: The first approximation involves analysing whether the infrastructure in question is a feasible option for dealing the family of experiments method, as some special characteristics are required to maintain a family of experiments. These characteristics should be taken into account for management purposes, especially synthesis of experiments and knowledge exchange among experimenters. This analysis will be conducted by building mock-ups that will simulate the use of several families of experiments.

  Analyse Mechanisms for Facilitating Communication among Experimenters: The second approximation involves analysing the different mechanisms for making the exchange information and knowledge among experimenters practicable, taking into account the characteristics of the repository, the information that the experimenters handle and possible support tools that could coexist with the infrastructure.

  Analyse Development and Support Tools: In the third approximation, we analyse several commercial tools for both developing the tool and supporting the management of the infrastructure for the experimental process in SE.

  Implement the Tool: In this approximation, we go ahead and implement the tool that will manage the information in the central repository, the exchange of information and knowledge among experimenters, as well as the support services for the experimental process in SE.

  The final product of this stage will be the tool, which, combined with the above artefacts, will make up the target infrastructure. This will mark the end of the research for this thesis, although the infrastructure will continue to be evaluated and upgraded by researchers from both GrISE and other research groups.

6. SUMMARY OF THE CURRENT STATUS OF THE RESEARCH AND PLANNED STEPS

6.1 Current Status of the Research

Of the first stage, we have defined the problem, completed the preliminary literature review of tools existing to support the experimental SE process and elicited knowledge. We are now starting the systematic literature review.

Regarding the second phase, we have reviewed the conceptual framework with respect to the experimental process in SE and its elements. We have obtained a conceptual model based on study - elicit - test - modify cycles. We now have three complementary conceptual models that are equivalent to the research manager, experiment manager and experimenter roles. Together, they define most of the experimental research process as regards running experiments. Based on this, we have been able to model a first approximation of the CR. The replicator and synthesizer roles are outstanding.

In the third phase, we have run a literature and technical study of data analysis tools and database management systems preferentially released under BSD or GNU GPL licences, which are popular and supported by a recognized community, have help forums in most languages and are easy to integrate. We opted to run preliminary tests with R statistical software and the MySQL database management system. We used experimental data to test the operation and validate the data analysis results of the selected tools. The results of this exercise demonstrated that the cohesion between the two tools is robust. From the above preliminary studies, we learned what information an experiment manages, as well as specialized tools for information storage and statistical analysis.

As part of the fourth stage, we are now analysing the feasibility of the software ecosystems paradigm (SECOs) as a possible mechanism for integrating the information and knowledge of the different experimenters working together on a family of experiments. SECOs is a recent and promising approach for improving inter- and intra-organizational reuse, involving end users in the process [24]. These relationships are often backed by a common technology platform and operate through information, resource and artefact exchange [23].

6.2 Planned Steps

We plan in the immediate future to carry out several activities in parallel in compliance with the established schedule for 2012. To be precise, the tasks to be executed are:

- Continue with the feasibility study of the SECOs paradigm
- Analyse the applicability of the families of experiments method in SE, building prototypes to instantiate conceptual groups of experiments with common characteristics in the model in order to synthesize the results and information exchanged by experimenters.
- Complete the systematic literature review of SE experimental research process support systems
- Study the technical literature to determine the best tool for building the central information repository
- Build the central information repository.

Having built the repository and completed the feasibility study of the operational aspects of experimentation that will guarantee information management, we will focus in 2013 primarily on the developing the management tool using incremental prototyping.

Finally, we will put together the final research reports in 2014.

7. ACKNOWLEDGMENTS

This research has been funded by a grant from the Armed Forces Technical School (ESPE), Republic of Ecuador National Higher Education, Science, Technology and Innovation Secretary's Office (SENESCYT) and partially funded by the Spanish Ministry of Economics and Competitiveness project TIN2011-23216.

8. REFERENCES

[1] E. Arisholm, D. I. Sjøberg, G. J. Carelius, and Y. Lindsjørn. A web-based support environment for software engineering experiments. *Nordic Journal of Computing*, 9(3):231–247, September 2002.

[2] V. Basili, S. Green, O. Laitenberger, F. Lanubile, F. Shull, S. Sørumgård, and M. Zelkowitz. Lab package for the empirical investigation of perspective-based reading. Available at `http://www.cs.umd.edu/projects/SoftEng/ESEG/manual/pbr_package/manual.html`, Last visited: May 15th, 2012.

[3] V. R. Basili. The past, present, and future of experimental software engineering. *Journal of the Brazilian Computer Society*, 12(3):7–12, December 2006.

[4] V. R. Basili, F. Shull, and F. Lanubile. Building knowledge through families of experiments. *IEEE Transactions on Software Engineering*, 25(4):456–473, July/August 1999.

[5] B. Boehm and V. Basili. The cebase framework for strategic software development and evolution. *Third International Workshop on Economics-Driven Software Engineering Research (EDSER-3 2001)*, May 2001.

[6] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein. *Introduction to Meta-Analysis*. Wiley & Sons, Ltd, 2009.

[7] J. J. Cimino, P. D. Clayton, G. Hripcsak, and S. B. Johnson. Knowledge-based approaches to the maintenance of a large controlled medical terminology. *Journal of the American Medical Informatics Association*, 1(1):35–50, January/February 1994.

[8] D. S. Cruzes and T. Dybå. Research synthesis in software engineering: A tertiary study. *Information and Software Technology*, 53(5):440–455, May 2011.

[9] T. de Jong and M. G. Ferguson-Hessler. Types and qualities of knowledge. *Educational Psychologist*, 31(2):105–113, 1996.

[10] O. Dieste, E. Fernandez, R. G. Martinez, and N. Juristo. Comparative analysis of meta-analysis methods: When to use which? In *15th Annual Conference on Evaluation and Assessment in Software Engineering (EASE 2011)*, volume 15, pages 36–45, April 2011.

[11] H. Do, S. Elbaum, and G. Rothermel. Software-artifact infrastructure repository. Available at `http://sir.unl.edu/content/sir.php`, Last visited: January 12th, 2012.

[12] M. Genero, J. A. Cruz-Lemus, S. Abrahão, E. Insfrán, J. A. Carsí, and D. Caivano. A controlled experiment for assessing the influence of stereotypes on the comprehension of uml sequence diagrams. Available at `http://alarcos.esi.uclm.es/ExpStereotypes/`, Last visited: March 15th, 2012.

[13] B. Hofmann and V. Wulf. Building communities among software engineers: The visek approach to intra- and inter-organizational learning. In *Advances in Learning Software Organizations*, volume 2640 of *Lecture Notes in Computer Science*, pages 25–33. Springer Berlin / Heidelberg, April 2003.

[14] A. Jedlitschka, M. Ciolkowski, and D. Pfahl.

Reporting experiments in software engineering. In *Guide to Advanced Empirical Software Engineering*, pages 201–228. Springer London, 1st edition, 2008.

[15] A. Jedlitschka and D. Pfahl. Reporting guidelines for controlled experiments in software engineering. *4th International Symposium on Empirical Software Engineering (ISESE 2005)*, pages 95–104, November 2005.

[16] N. Juristo and A. M. Moreno. *Basics of Software Engineering Experimentation*. Kluwer Academic Publishers, 2001.

[17] N. Juristo and S. Vegas. Using differences among replications of software engineering experiments to gain knowledge. In *3rd International Symposium on Empirical Software Engineering and Measurement*, ESEM '09, pages 356–366, 2009.

[18] N. Juristo and S. Vegas. The role of non-exact replications in software engineering experiments. *Empirical Software Engineering*, 16:295–324, 2011.

[19] B. Kitchenham and S. Charters. Guidelines for performing systematic literature review in software engineering. Technical Report EBSE-2007-01, Keele University and University of Durham, July 2007.

[20] B. A. Kitchenham, S. L. Pfleeger, L. M. Pickard, P. W. Jones, D. C. Hoaglin, K. El Emam, and J. Rosenberg. Preliminary guidelines for empirical research in software engineering. *IEEE Transactions on Software Engineering*, 28(8):721–734, August 2002.

[21] V. P. Lopes and G. H. Travassos. Knowledge repository structure of an experimental software engineering environment. *23th Brazilian Symposium on Software Engineering*, pages 32–42, October 2009.

[22] C. M. Lott. Comparing reading and testing techniques. Available at `http://web.archive.org/web/20071026204353/http://www.chris-lott.org/work/exp/`, Last visited: December 12th, 2011.

[23] D. G. Messerschmitt and C. Szyperski. *Software Ecosystem: Understanding an Indispensable Technology and Industry*. MIT Press, 2003.

[24] O. Pettersson, M. Svensson, D. Gil, J. Andersson, and M. Milrad. On the role of software process modeling in software ecosystem design. In *Fourth European Conference on Software Architecture: Companion Volume*, ECSA '10, pages 103–110, 2010.

[25] M. Polanyi. *The Tacit Dimension*. Doubleday and Company INC., 1966.

[26] J. Sayyad Shirabad and T. Menzies. The promise repository of software engineering databases. Available at `http://promise.site.uottawa.ca/SERepository`, School of Information Technology and Engineering, University of Ottawa, Canada, Last visited: July 4th, 2005.

[27] J. Singer. Using the american psychological association (apa) style guidelines to report experimental results. In *Workshop on Empirical Studies in Software Maintenance (WSESE 1999)*, pages 71–75, Oxford, England, September 1999.

[28] D. I. Sjøberg, J. E. Han Hannay, O. Hansen, V. B. Kampenes, A. Karahasanovic, N.-K. Liborg, and A. C. Rekdal. A survey of controlled experiments in software engineering. *IEEE Transactions on Software Engineering*, 31(9):733 – 753, September 2005.

[29] M. Solari and S. Vegas. Classifying and analysing replication packages for software engineering experimentation. *7th International Conference on Product Focused Software Process Improvement (PROFES 2006)-Workshop Series in Empirical Software Engineering (WSESE)*, 2006.

[30] M. Solari B. *Propuesta de Paquete de Laboratorio para Experimentos de Ingeniería de Software*. PhD thesis, Universidad Politécnica de Madrid, march 2012.

[31] K. Torii, K. Matsumoto, K. Nakakoji, Y. Takada, S. Takada, and K. Shima. Ginger2: an environment for computer-aided empirical software engineering. *IEEE Transactions on Software Engineering*, 25(4):474–492, July/August 1999.

[32] J. Vanschoren and H. Blockeel. Stand on the shoulders of giants: Towards a portal for collaborative experimentation in data mining. *International Workshop on Third Generation Data Mining at ECML PKDD*, pages 88–89, September 2009.

[33] S. Vegas, N. Juristo, A. Moreno, M. Solari, and P. Letelier. Analysis of the influence of communication between researchers on experiment replication. In *International Symposium on Empirical Software Engineering*, ISESE '06, pages 28–37, 2006.

[34] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering An introduction*. Kluwer Academic Publishers, 2000.

[35] A. Zendler. A preliminary software engineering theory as investigated by published experiments. *Empirical Software Engineering*, 6(2):161–180, June 2001.