1. Introduction

There is a growing interest in empirical studies on software engineering (SE). It can be evidenced by the increasing number of papers describing results from empirical studies published elsewhere. The Experimental Software Engineering (ESE) team is a research group of the Systems Engineering and Computer Science Program (PESC) at COPPE/UFRJ. Its goals include the improvement of software engineering by the application of the scientific method (experimentation) for the construction of new methods and techniques to support software development, and researching of new models and approaches for planning, execution and packaging of SE experimental studies.

The ESE team has used the experimentation model defined by [1] to perform different experimental studies. The diversity of the packaged experimental studies made us able to observe that the conventional two-staged SE experiments taxonomy (in vivo/in vitro) was not enough to classify some of them. Therefore, a four-staged experiment taxonomy was defined [2] (in vivo, in vitro, in virtuo and in silico).

All the works performed by the ESE team have reinforced, despite their importance, how hard is to deal with knowledge regarding experimental studies’ planning, execution and packaging. Usually a lot of clerical activities take part in the whole experimentation process, making it tedious and repetitive mainly for the novices experimenters. It led us to look for ways to support some of these SE experimentation process activities by providing automated (or semi-automate) computer based tools integrated into a software engineering environment, highlighting the aim of identifying and defining an infrastructure that can allow the instantiation of SE environments for experimentation.

This paper describes a Doctorate degree research that aims at identifying and defining an infrastructure that allows instantiation of SE environments for experimentation, so called eSEE (experimental Software Engineering Environment). An eSEE must manage the software engineering experimentation process, including knowledge acquired when defining, planning, executing and packaging experiments.

In order to support the building of such infrastructure, the experience and knowledge acquired to build the TABA Project (www.cos.ufrj.br/~taba) are going to be reused. TABA is a meta-environment that allows the instantiation of Software Engineering Environments (SEEs) according to different domains and technologies application. TABA also includes a knowledge dimension that makes possible external CASE tools’ integration into the SEE [3].

2. eSEE’s Infrastructure

The proposed computerized infrastructure for eSEE is based on the experimentation process and package models described in [1]. To instantiate an eSEE for a specific study, the performed experiment type must be defined. Three levels of knowledge organization about the experimentation process have been identified: knowledge for any kind of experiments (meta level), knowledge for each type of experiment (configuration level), and knowledge for a specific experimental study (instance level). These levels reflect themselves at eSEE’s architecture, which was inspired by the TABA’s metaphor. All the information generated throughout the experimentation process must be packaged for future use. The layers composing eSEE’s architecture are:

- Meta-eSEE: the meta-level contains common knowledge regarding SE experimental studies, including software engineering knowledge. This knowledge may be captured by software engineering ontologies [3] integrated with an experimentation ontology. At this level, a standard experimentation process (SEP) is defined. The SEP represents the basis to instantiate standard processes for each type of experiment;

- Configured Meta-eSEE: the configuration level contains knowledge regarding specific experiments types. At this level, specific eSEE’s instantiation can be accomplished by choosing experimental study type and adding specific experimental study’s study characteristics. The instantiated environment,
then, is generated with the specific knowledge to that study type;

- eSEE: environment supporting for the definition, planning and execution of the specific experimental study type. For each new experimental study type, one eSEE should be instantiated to manage the experimentation process and knowledge.

Once the instantiated environments are created, some facilities must be available to support the accomplishing of experimental studies. For instance, to allow process tracking in each eSEE, a XML based process execution machine will be integrated to the environment. Nevertheless, experimental studies should be packed into experimental packages repositories throughout the experimentation process.

Some of the necessary data to perform experimental studies in SE may come from CASE tools. Therefore, these tools must be extended to collect experimental data [4]. Once CASE tools collect the data, data tool integration facilities based on XML and ontologies will allow data importing to eSEE’s repository.

Besides supporting experimental studies accomplishment, other services and tools are necessary to support eSEE activities, such as:

- Knowledge Management: the eSEE infrastructure must make available knowledge about the experimentation process, methods, techniques and tools to assist the software engineering researcher managing of the experimental study. Besides, it allows lessons learned to be stored and reused for future experimental studies’ planning activities (to prevent errors and to identify opportunities of experimentation process improvement);

- eCASE Tools: supporting experimentation life cycle activities, such as definition of questionnaires, assignment of participants, data collecting or data analysis;

- Data Visualization Tools: must be enclosed to the experimental environment to easy data analysis;

- Environment, Objects and Subjects Computerized Models: to perform in virtuo or in silico experiments, it is necessary to build objects, participants and environments models (simulators) used in the experimental study.

3. On Going and Future Works

Based on the observation that several experimental SE practices are similar to experimental SE ones, such as training, configuration management, process management, documentation and so on, we believe that the experimentation process could be improved by incorporating some of the ISO/IEC 12207 processes. An initial mapping of ISO’s processes into the experimentation life cycle was already made in order to organize the experimental process. For instance, as defined in [1], an experiment Planning stage’s sub-activity is Results Validation Adequacy. This sub-activity is concerned with the results’ validity for the population of interest. In this phase, threads to experimental results must be identified, managed and mitigated. Therefore, this sub-activity was mapped as ISO/IEC 12207’s Risk Management Process.

We defined a set of eSEE’s initial set of features and categorized them according to the integration dimensions of the TABA [3], adding a specific dimension related to experimentation. The reason to do so is that we believe that an environment for experimental software engineering may have similar characteristics to SEEs. Therefore, six features categories were defined: data, control, presentation, platform, knowledge and experimentation dimension.

An initial prototype to support experimental studies’ planning is being built. This prototype is based on the experimentation packaging process [1], exploring the Zope+Plone development platform.

The next steps of this work include the following:

- Experimentation process adaptation, regarding the four staged taxonomy [2] and the definition of a SEP to support each one of them;

- Experimentation ontology definition to organize knowledge about experimental software engineering in eSEE, used to identify general characteristics of each experimental study type.

4. Acknowledgments

The authors acknowledge CAPES and CNPq for the financial support to this work.

5. References


