Information and Software Technology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

Content and structure of laboratory packages for software engineering experiments

Martín Solari[a,*], Sira Vegas[b], Natalia Juristo[b]

[a] Universidad ORT Uruguay, Cuareim 1451, Montevideo 11100, Uruguay
[b] Universidad Politécnica de Madrid, Campus de Montegancedo s/n, Madrid 28660, Spain

ARTICLE INFO

Keywords:
Software engineering
Experiment replication
Laboratory package
Knowledge transfer

ABSTRACT

Context: Experiment replications play a central role in the scientific method. Although software engineering experimentation has matured a great deal, the number of experiment replications is still relatively small. Software engineering experiments are composed of complex concepts, procedures and artefacts. Laboratory packages are a means of transferring knowledge among researchers to facilitate experiment replications.
Objective: This paper investigates the experiment replication process to find out what information is needed to successfully replicate an experiment. Our objective is to propose the content and structure of laboratory packages for software engineering experiments.
Method: We evaluated seven replications of three different families of experiments. Each replication had a different experimenter who was, at the time, unfamiliar with the experiment. During the first iterations of the study, we identified experimental incidents and then proposed a laboratory package structure that addressed these incidents, including document usability improvements. We used the later iterations to validate and generalize the laboratory package structure for use in all software engineering experiments. We aimed to solve a specific problem, while at the same time looking at how to contribute to the body of knowledge on laboratory packages.
Results: We generated a laboratory package for three different experiments. These packages eased the replication of the respective experiments. The evaluation that we conducted shows that the laboratory package proposal is acceptable and reduces the effort currently required to replicate experiments in software engineering.
Conclusion: We think that the content and structure that we propose for laboratory packages can be useful for other software engineering experiments.

1. Introduction

Experimentation is a way of maturing knowledge. It is at the heart of the scientific method. Experiments manipulate and observe reality in a controlled, rigorous and systematic manner in order to discover variables that somehow influence a specified phenomenon [1]. For this approach to be successful, experimentation must be iterative, taking earlier results to refine procedures for studying the phenomenon in more depth. As a research method in software engineering (SE), experimentation dates back only a few decades having first been formalized around 1986 [2].

Replication is a way of exploring experiments in more breadth and depth [3]. Replications serve two purposes: (i) they may study whether the results are valid for other contexts or whether other researchers are able to get the same results, and (ii) they may, if changes are made to an experiment, allow studying influencing variables and provide information about the bounds of the acquired knowledge [4]. Since replication is troublesome in SE, some researchers [5] have proposed alternative methods. Although there are many methodological uncertainties surrounding replication, its importance is clear [6].

Experiments are composed of concepts and of particular artefacts for experiment operation. The conceptual elements include the constructs, hypotheses and experimental design. The operational elements include the instruments, materials and procedures. On this ground, an experiment is a highly complex entity both conceptually and operationally. This is an obstacle to its replication, particularly when a researcher is replicating an experiment that he or she did not design [7].

There are some ways of facilitating experiment replication. They include collaboration among researchers and laboratory packages [8–10]. The use of such types of knowledge transfer mechanisms is widespread among researchers from other disciplines, such as physics, biology and the social sciences [11–13].

* Corresponding author.
E-mail addresses: martin.solari@ort.edu.uy (M. Solari), svegas@fi.upm.es (S. Vegas), natalia@fi.upm.es (N. Juristo).

https://doi.org/10.1016/j.infsof.2017.12.016
Received 9 December 2016; Received in revised form 20 December 2017; Accepted 27 December 2017

Laboratory packages (LP) contain the information and materials required to replicate an experiment. The term protocol is often used in natural science [14] to refer to the packaging of key information about an experiment for its transmission and use in a replication by others. Whereas the term protocol underscores strict adherence to a procedure, the term LP focuses on the transfer of information in self-contained units. Both components are necessary for conducting a replication.

The content of a LP is not static; it needs to be adapted to the needs of the researcher running the replication. A LP is usually task driven (that is, written like a *recipe*) and may use alternatives other than text like video or software. A scientific paper is unlikely to include this type of content as there are usually constraints on space and format [15]. Scientific publications usually stress the high-level conceptual aspects, omitting more specific operational issues that are helpful for replicating experiments. In some disciplines, this gap has been filled by journals dealing specifically with experimental protocols like *Nature Protocols*.[1] Open digital repositories of experimental protocols have also been set up, providing for discussion among researchers based on a forum system.[2] These repositories enable a faster information exchange than journals, as well as the use of alternative formats for answering questions and helping other researchers to replicate the experiment.

Although SE has made significant progress with regard to the conceptualization and running of replications, information packaging and transfer are still an open question. The use of LPs has been defended ever since the earliest systematic experiences of SE experiment replication [7]. The need for LPs as a key component for conducting families of experiments was again stated later on [8]. Collaboration between groups of SE experimenters was the focus of research project that proposed the use of a knowledge transfer model also based on LPs [16]. Interest in the topic appears to have cooled within the SE community, but this does not mean that how to package and transfer the information needed to run a replication has been solved. Although the number of replications has grown over the last few years, it is still not a common practice, even less so if researchers are trying to repeat an experiment at other sites [17]. Neither are there many cases of SE experiments whose materials are published for use by other researchers, and experiments whose procedures are shared with a view to conducting exact replications are even rarer still [18].

We aim to identify which information a LP for SE experiments should contain. Proper LPs would encourage more and better replications. Replications not only contribute to the body of SE knowledge, but also help researchers to learn about how to experiment with and observe the phenomenon. This helps to mature SE experimentation.

Our immediate goal is to improve the LPs used to replicate experiments within a SE experiment network. Although the research was undertaken in a small experimental research network, we believe this context is representative of the current replication and information exchange scenario between experimental researchers in SE.

We studied three experiments and seven replications iteratively. First, we observed the incidents in replications of one experiment run by researchers belonging to several research groups. From these observations we derived a preliminary LP structure. In later iterations, we refined the structure by observing more replications of the same experiment run by researchers belonging to the same and other research groups. Finally, we generalized the structure to build LPs for other two experiments.

The structure of the paper is as follows. Section 2 discusses related work, analysing the use of LPs and experimental protocols first in other disciplines and then in SE. Section 3 describes the research method. The other sections each address one of the stages of the research. Section 4 reports the diagnosis based on the observed incidents in experiment replications. Section 5 outlines the proposed LP structure. Section 6 describes the LPs generation for three experiments. Section 7 describes how LPs were evaluated by observing replications using them. Section 8 reflects on how the proposed LP might be generalized. Section 9 present further evolution plans for our proposal. Finally, Section 10 outlines the conclusions of the research.

2. Related work

2.1. LPs in other disciplines

It is common practice in many scientific disciplines to use LPs in different types of experimental replications. In the social sciences, for example, it is usual to share sampling, observation or intervention procedures for complex phenomena [13]. Disciplines like biology and medicine use repositories containing the protocols to be followed by experimenters [19]. These protocols facilitate the aggregation of studies replicated by different researchers at different sites.

Since the 1990s several scientific communities have helped to create new methods for speeding up research and promoting information exchange [20]. Digital media are not subject to the typical constraints on number of pages and format imposed by the print media. They also facilitate discussion and contacts among researchers.

Giles [15] argued in an opinion column that researchers visit each other to learn the protocol because much of the knowledge about the experiment is tacit. The column cites several scientific publication editors who stress the need to publish more detailed information and provide for direct exchange among the researchers that are going to run replications. Examples of journals that provide these exchange mechanisms are: *PLoS ONE*,[3] *Cold Spring Harbor Protocols*[4] and *Nature Protocols*.[5]

Protocol publication and study preregistration have become the norm in medicine. In recognition of the importance of detail in order to ensure the rigour of clinical trials, several initiatives try to standardize the content of such reports and protocols. The CONSORT (CONsolidated Standards of Reporting Trials) initiative promotes study reporting quality by means of a content checklist that was originally proposed in 1997 and has been updated on several occasions [21].

There is agreement among the editors of several medical journals about the importance of having access to the full protocol of published trials [14,22]. Journals like the *Lancet*,[6] *British Medical Journal*[7] and *PLoS Medicine*[8] demand that the protocols be submitted together with controlled experiment reports for inclusion in the review process. However, current guidelines on protocol content vary substantially with respect to recommendations [23]. Recently, the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) initiative proposed the standardization of the protocol content of experiments in medicine [24].

An exceptional example of LP or protocol repository is *Nature Protocol Exchange*.[9] This repository provides for the open exchange of experimental protocols and an open discussion. Replicating experimenters can submit queries and baseline experiment authors can provide the details that they had omitted to specify. Unlike *Nature Protocols*, the repository does not follow a process of arbitration before the protocols are uploaded. Research groups can sign up to share their own protocols or query the protocols of other groups. The repository is open to experiments from any scientific discipline. These procedures and data can be taken by other researchers to run replications or to perform secondary studies (systematic literature review (SLR), meta-analysis, etc.).

[1] http://www.nature.com/nprot/index.html.
[2] http://www.nature.com/protocolexchange/.
[3] http://www.plosone.org/.
[4] http://cshprotocols.cshlp.org/.
[5] http://www.nature.com/nprot/index.html.
[6] http://www.thelancet.com/protocol-reviews.
[7] http://www.bmj.com/about-bmj/resources-authors.
[8] http://www.plosmedicine.org/static/guidelines.
[9] http://www.nature.com/protocolexchange/.