Realising Evidence-Based Software Engineering (REBSE-2) A Report from the Workshop held at ICSE 2007

David Budgen
Department of Computer Science
Durham University, U.K.
e-mail: david.budgen@durham.ac.uk

Pearl Brereton
School of Computing & Mathematics
Keele University, U.K.
e-mail: o.p.brereton@cs.keele.ac.uk

Abstract

Context: *The REBSE international workshops are concerned with exploring the adaptation and use of the evidence-based paradigm in software engineering research and practice, through a mix of presentations and discussion.*
Objectives: *These were to explore both experience with, and potential for, evidence-based software engineering (EBSE); to consider how this might affect empirical practices in software engineering; and to work towards creating a community of researchers to practice and promote EBSE.*
Method: *Three sessions were dedicated to a mix of presentations and interactive discussion, while the fourth was dedicated to summarising progress and identifying both issues of concern and actions to pursue.*
Conclusions: *While we identified a number of issues, a key need is clearly to have a central repository to both provide information and to maintain a record of activity in this area.*

Introduction

"Science is built up of facts as a house is built of stones, but an accumulation of facts is no more a science than a heap of stones is a house" [Poincaré]

If software engineering is to advance as an engineering discipline it needs to move away from its current dependence upon advocacy and analysis, towards employing more objective, systematic, and empirically-based approaches to developing an understanding of what works, why, and under what conditions. The *evidence-based paradigm* is now well established in clinical medicine as an objective and structured means of assembling and analysing the available data in order to answer research questions. It has also been adopted in other domains such as education, criminology and social policy, for which the empirical practices are closer to those deployed for software engineering.

The primary goal for this workshop was to provide another step towards the vision of creating widely-accepted evidence-based foundations for software engineering, following on from the workshop held at ICSE in 2005 [BK05]. Indeed, our concern is not just with the *evidence* itself, important as this is, but also with creating an associated community of researchers who will review, analyse and promulgate empirical results in different areas of software engineering. (For convenience, we will refer to the concept of *Evidence-Based Software Engineering* as EBSE for the rest of this report.)

The sub-goals for the workshop were therefore concerned with:

1. Developing and extending our understanding of the implications and potential of the evidence-based paradigm for empirical software engineering research and practice, drawing upon the experiences accumulated so far.

2. Identifying how the adoption of EBSE will affect empirical practices in SE (and the infrastructure needed for this).

3. Identifying practical steps to help with creating a community that can develop these ideas as well as encourage the adoption of evidence-based approaches in Software Engineering, both for academics and practitioners.

Organisation of the Workshop

Our call for papers elicited fewer papers than in 2005. This might have arisen from both the timing of the call (just before Christmas) and also as a consequence of the better-formed understanding that we now have of the evidence-based paradigm—or at least, of the processes involved in conducting Systematic Literature Reviews—meaning that both the call and the papers we did receive were generally more focused than in 2005.

All papers were reviewed by three members of the Programme Committee and the outcome was that three papers were accepted, two of which were used to introduce sessions, and the organisers also provided an introduction. As previously, we were concerned to create a workshop format that would allow for interaction. To achieve this, we allocated one session to each of three major topics: state of the art; forms of evidence and aggregating evidence; and infrastructure. Each of these was introduced via one of the workshop papers, followed by discussion, and then the fourth session was an open discussion on *future directions*. Records were kept via note-pads, flip-chart and voice recorder, with the workshop team acting as scribes as well as facilitators.

Reports from the Sessions

Of necessity we have kept these brief and have focused mainly upon the discussions following the presentations, since the papers themselves are available in the ICSE Workshop Proceedings.

Session 1: State of the Art

For this session, David Budgen presented a summary of some of the outcomes of an ongoing *tertiary study*, being led by Barbara Kitchenham, examining the extent to which the concepts of EBSE were being reflected in the literature.

The working definition of *evidence* that was used to underpin the presentation was:

> *That which demonstrates or establishes the truth of a point in question.*

(This is an amalgam of various dictionary definitions!) However, for the workshop, the main concern was more one of how the concept is to be interpreted within software engineering.

The baseline used in selecting the inputs for this tertiary study was 2004, since this is the year when the ideas of EBSE were introduced in a widely-cited ICSE paper by Kitchenham, Dybå and Jørgensen [KDJ04]. As for a secondary study, the tertiary study has adopted the form of a *Systematic Literature Review*, but identifying secondary studies and their form as the inputs, rather than primary studies.

Since a fuller report is available in the workshop proceedings [KBBT07], only a brief summary of some key findings is presented here.

The *research questions* for the tertiary study were:

1. How much EBSE activity has there been since 2004?
2. What research topics are being addressed?
3. Who is leading EBSE research?
4. What are the limitations of current research?

The study process involved a hand search of major journals and conferences, as well as contact with researchers known to be active in the field.

In answer to the first question, at time of reporting, the review had identified 23 secondary studies. Twenty of these were systematic literature reviews, one was a meta-analysis, and two were classified as being evidence-based guidelines. And in answer to the second, from these 23 reviews:

- 9 addressed research trends
- 9 were on cost estimation
- 4 were on software experiments
- 3 were on testing

This was felt to be disappointing, since research trends are not of direct interest to practitioners.

To answer the third question, we determined that:

- 17 reviews had European authors
- 4 reviews had North American authors
- 11 had authors from the Simula Laboratory (Norway)

Most were published in (a wide range of) journals, with three in conference proceedings.

Answers to the fourth question were still emerging, but included issues of quality as well as choices of topic. This to some extent formed a focus of the following discussion which drew out a number of issues and concerns. Some of these concerned researchers, while others were focused upon practitioners and their needs. Issues of concern to researchers included:

- The problem of getting longer papers (such as those reporting Systematic Literature Reviews) published due to journal and conference page limits.
- The question of which audience to address in reporting a review:
  - those interested in the *results*
  - those interested in the *process*
- The limited availability of primary studies on topics of interest.

In terms of what might interest a practitioner, the discussion identified the following concerns:

- The need for a better framework both for reporting to industry and also for obtaining reports of industrial experience as inputs.
- Many of the current corpus of reviews are not conclusive.
- The need to connect with decision makers and practitioners by publishing in the journals and magazines that they are likely to read.
- While industrial data does exist on questions of interest, this may well be subject to access restrictions due to its commercial confidentiality (a problem that may be specific to software engineering).
- The need for researchers to identify the topics that are of interest to industry. Related to this is the issue of timescale: conducting reviews can be quite time-consuming, but industry tends to want answers quickly.

In summing up this session, David Budgen also reported that the UK's Engineering & Physical Sciences Research Council (EPSRC) had approved funding for a further two-year study in this area (Evidence-based Practices Informing Computing–EPIC), and some of these issues were ones that the project was expected to address. In particular, the project would be aiming to develop the web site at

`www.ebse.org.uk`

to act as a repository and central resource for EBSE.

Session 2: Forms of Evidence & Aggregating Evidence

This topic was introduced by Tracy Hall, who examined the question as to whether the use of *simulation* could enhance our approach to doing experiments in software engineering, and in particular, as a means of testing out empirical evidence. Again, as a fuller discussion of this is available in the Workshop Proceedings [WH07], the description here has been confined to the main points.

A key role for the use of simulation in this context was as a motivator to trigger empirical research—that is, where there were gaps in the data needed for the simulation, this represented a 'desert' that needed to be addressed through further research. To illustrate this idea, she described the development of a simulation model of software development processes, and how this could then form the basis for deciding what empirical studies were needed to help understand the impact of such factors as the use of pair programming on meeting requirements, and determining which factors were important enough to be measured in such studies.

The strengths of simulation models were seen as being their usefulness for making predictions as well as for building hypotheses and theories, together with the scope to calibrate them with empirical data. An added benefit being the relatively low cost of simulation when compared with conducting experiments. However, there are pitfalls too, including the rapid escalation of complexity, the risk of not making any assumptions explicit enough, and over-stating the reliability of results from simulation.

Overall, the case made in the presentation was that simulation should be included as part of the portfolio of approaches used to motivate, collect, analyse and reason about software engineering data; to help with directing empirical studies; and to help justify such studies (and the related expense).

Discussion ranged around a number of aspects of simulation. Some of the points made were as follows.

- Practitioners tend to like such modelling and are well aware of its limitations. However, academics seek proof that it is sound and valid.
- Simulation can provide a useful framework for linking different forms of evidence, possibly providing an alternative to systematic literature reviews.
- Models can help with communication of ideas.

This in turn led on into discussion about the different stakeholders associated with the development of EBSE. It was felt that the use of systematic reviews may be more appropriate for policy makers, whereas modelling might be more suited to the needs of practitioners, where it can be parameterised to fit their own context.

Session 3: Infrastructure

This theme was introduce by a paper presented by Liming Zhu that addressed the question of the infrastructure needed to index and organise best practices [ZSG07]. The main challenge for this work was seen as being to maintain a repository of 'best practice' and the approach being investigated was to organise this as an 'information overlay' on top of existing repositories, rather than creating new ones.

Liming discussed some of the technical challenges involved in this, including the use of the semantic web and collaborative tagging, as well as some of the existing packages and systems that could contribute. He described experiences with building a prototype system for the topic of software design, and the possibilities inherent in this, including collecting expert knowledge. However, it was noted that while a number of useful repositories are available, it can be hard to know what exactly is available and where to find it—a possible role for a central directory.

Discussion identified some needs for support from an infrastructure, including the provision of templates for protocols. It was also observed that we might analyse the papers found in the tertiary review to identify useful digital libraries and those that were used most widely. The question of the grey literature was also a challenge, as well as applying quality criteria that were appropriate to the source of evidence ('PhD thesis', 'industry report',...). Other issues included:

- How do ACM/IEEE maintain their keyword lists (which seem to be too outdated to be of use) and how do we handle synonyms?
- Are there other ontologies that we could employ?
- The need for a place to register systematic literature reviews to avoid researchers duplicating scarce effort.
- Provision of teaching materials for reviews and EBSE in general. There is a need for a one-page summary of the review process that could perhaps also be sent to inform reviewers when papers are submitted?
- The need for a journal dedicated to systematic reviews.

Looking ahead, it was suggested that if we hold a REBSE-3, it might usefully focus on the various elements in the systematic literature review process.

Session 4: Discussion & Planning

For this session, we considered the following two questions:

- What do we want to do as a community?
- How might we influence others to support/employ EBSE (especially industry)?

We also identified some specific actions that we wished to pursue as outcomes of the workshop.

What do we want to do as a community?

We essentially answered this as a community of evidence-based researchers, and identified the following needs.

- The urgent need for a stable web site, both as the means of providing resources and also as a repository (echoing earlier concerns).
- The provision of guidelines that can be used to educate academia, as well as providing support for postgraduate students and their supervisors when wanting to perform a review as part of the preparation of a thesis topic.
- A set of reviews that provide positive support for existing practices (not just ones that are negative or inconclusive) that might help to convince others about the process.
- Determining where primary studies are particularly needed—and related to this, identifying the topics of interest to industry and practitioners. This might involve us in targeting industry-led events and leading edge issues likely to be of interest to industry.
- Encouraging government bodies and industry to fund systematic reviews.

One possible way of addressing some of these that was debated was organising a special issue of a journal, with the aim of containing tutorial material, process analysis papers and some actual reviews, perhaps centred upon a theme that might appeal to practitioners.

How might we influence others to support/employ EBSE (especially industry)?

Some of the ideas here inevitably overlapped with the previous list.

- The idea of a special issue (as above).
- Looking for 'bandwagon' topics, preferably ones that had an international aspect.
- Teaching our students (the next generation of practitioners)—and sharing material and experiences among ourselves to help support this.
- Providing an international repository (as above) and then creating national vehicles for dissemination.

Two (partially related) questions that we identified here were firstly whether we should really be 'selling' the empirical paradigm, not just EBSE; and also how we might get others to fund this type of work?

Summary

At the end of the workshop, we identified four key actions that we felt could be pursued immediately.

1. Development of `www.ebse.org.uk` as a resource with the immediate support of the EPIC project. It was also felt that we needed to have a formal launch of this, and to promote it more widely, perhaps with a distinctive logo too. David Budgen and Pearl Brereton, as principal investigators on EPIC would undertake to instigate this over the coming summer.
2. Provision of documentation to provide guidelines for new postgraduate students and researchers along the lines of 'EBSE for Dummies'. Tracy Hall, Guilherme Travassos and John Bailey agreed to form a group that would look into this.
3. Creating an EBSE mail-list, starting from the attendees at the first two REBSE workshops (and programme committee members).
4. Investigate the possibility of having a regular column in ACM's *Software Engineering Notes* to promote EBSE and empirical studies. David Budgen agreed to contact Will Tracz on this.

Acknowledgments: Part of this work has been supported by a grant from the U.K.'s EPSRC for research into the procedures appropriate for using evidence-based practices in software engineering. We also acknowledge the contributions made by our fellow Workshop Organisers (Barbara Kitchenham and Mark Turner), John Bailey (who helped with the event itself) together with the members of the Programme Committee.

References

[BK05] David Budgen and Barbara Kitchenham. Realising Evidence-Based Software Engineering: A Report from the Workshop held at ICSE 2005. *Software Engineering Notes*, 30(5):44, 2005. Full report published electronically by ACM Press.

[KBBT07] Barbara Kitchenham, David Budgen, Pearl Brereton, and Mark Turner. 2nd international workshop on realising evidence-based software engineering (REBSE-2): Overview and Introduction. In *Proceedings of REBSE-2 Workshop, ICSE 2007*, pages 1–5. IEEE Computer Society Press, May 2007.

[KDJ04] B.A. Kitchenham, T. Dybå, and M Jørgensen. Evidence-based software engineering. In *Proceedings of ICSE 2004*, pages 273–281. IEEE Computer Society Press, 2004.

[WH07] P. Wernick and T. Hall. Getting the best out of software process simulation and empirical research in software engineering. In *Proceedings of REBSE-2 Workshop, ICSE 2007*, May 2007.

[ZSG07] L. Zhu, M. Staples, and I. Gorton. An infrastructure for indexing and organizing best practices. In *Proceedings of REBSE-2 Workshop, ICSE 2007*, May 2007.