

Empirical Case Studies in Industry: Some Thoughts

Nachiappan Nagappan

This position paper deals with empirical case studies from an industry perspective. Empirical Software Engineering at Microsoft spans several areas including traditional software engineering, program analysis, and HCI. From the perspective of transferring ideas, tools and techniques into the product teams Microsoft Research more or less behaves like an academic department, being closer to the problem. Some of the factors to conduct empirical studies from an industry view point to get a conversation started are described below. In general empirical studies are costly in industry due to the working environment constraints and the risks associated with failure.

Cost: Cost is an important factor in most industrial empirical studies. Commercial software developers like to get an initial estimate of the cost of deploying a new software system or process. This helps them make cost benefit decisions and perform risk analysis if they are willing to adopt the new techniques.

Scale: Scale is a very important factor than needs to be understood and acknowledged in most industrial case studies. Techniques that work well with smaller systems, groups of people might not work with large systems. Understanding the reasons for such failures early on can help avoid costly mistakes.

Logistics: Conducting empirical studies in an industry environment involve a considerable amount of organizational logistics. Very often it requires becoming part of the development organization to understand trends, data, tools and techniques in order to collect in-process, important contextual information.

Tools: Most data collection, not requiring any human input has to be automated to minimize intrusion into developer's normal work. For example, tools to collect code coverage information, measure code churn etc. Such tools need to scale for large systems and should work in reasonable time with a high degree of accuracy to be adopted by product teams.

Privacy: Most product teams expect a high degree of privacy while doing empirical studies. They do not wish to have any personally identifiable information mapped backed to them with respect to their development practices.

The above factors are by no means complete or comprehensive and are just to serve as a starting point for discussion in this area.