Tool diffusion in large open-source projects

Martin F. Krafft

Supervised by: Prof. Brian Fitzgerald
Research Area: GED
Project Title: Tool diffusion in large open-source projects

About me
- Self-employed, based in Switzerland
- Strongly involved in FLOSS culture
- Chose Lero for Ph.D. research

About the Debian Project
- Oldest GNU/Linux distribution: 14 years of age
- Among the largest FLOSS projects:
  - 1,100 developers + 2,500 contributors
  - 22,000 packages, 11 architectures
- Conservative & quality-oriented: prefer solid solutions to quick ‘n’ dirty ones
- Robust, secure, stable, Free
- 100% volunteers, flat, meritocratic

About the research
- Started: summer 2006; expected completion: summer 2009
- http://phd.martin-krafft.net

Motivation
I am a developer of the Debian GNU/Linux operating system, and an official member of the Debian project, one of the largest FLOSS projects. I use the Debian System on a daily basis, and most of my income is based on it.

The development tools and processes have not kept up with the immense growth of the project: development nowadays happens around the clock across all timezones, and most work is done in teams.

Unfortunately, most of the tools are not ready for team usage, or come with too steep learning curves and thus make it hard to contribute to the projects. In addition, few aspects of the development workflows are integrated with each other, or the integration is very brittle.

Research objectives
- To determine the salient influences to package maintainers’ adoption or rejection decisions, develop an unambiguous and orthogonal terminology of factors to capture these influences, and identify clusters of factors to produce a framework.
- To apply this framework to a number of previous or ongoing diffusions in the Debian Project, and compare the real-world adoption rates to the adoption rates emerging from this application.

The Delphi approach
- A method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem
- Aim for consensus or dissensus among a group of participants
- Optimal in a situation where participants cannot meet in the same physical location at the same time
- Characteristics:
  - Anonymity: reduces the effect of socially dominant individuals
  - Controlled feedback: prevents heated or personal debates
  - Statistical group response: reduce pressure to conformity and still yield meaningful results even with unanimous responses

Factors influencing adoption decisions
Whether a team or individual adopts a given tool depends on a number of factors, such as documentation, ease of use, how well the new tool fits with the existing (or desired) workflow, bug turnaround time, ...

I have processed mailing list archives, conducted a few informal interviews, and produced an initial set of factors from those. However, to prevent problems with bias, I chose the Delphi approach (see right side) instead. I may well use my list later, though.

Diffusion frameworks
A diffusion framework is a structuring of relevant aspects of a diffusion in an orthogonal and non-redundant way. Given a successful and an unsuccessful diffusion, a good framework should make the differences and reasons for success/failure clear.

A large number of diffusion assessment frameworks exist (mainly in sociology), e.g.:
- Rogers’ Diffusion of Innovations
- Davis’ Technology Acceptance Model and TAM2
- ...

These are almost exclusively about individual, voluntary decisions without any network externalities (dependencies on others), or authoritarian decisions.

No framework exists that can appropriately frame voluntary decisions with large network externalities.