Free and Open Source Software Development and Research: Opportunities for Software Engineering

Fabio Kon Paulo Meirelles Antonio Terceiro
Christina Chavez Nelson Lago Manoel Mendonça
University of São Paulo and Federal University of Bahia, Brazil

Software Engineering research

- Advancing our understanding of the process of software development and its outcomes
- Should be based on the study of existing real-world software
- FLOSS offers SE researchers the opportunity of basing their research on abundant and publicly available data, freely accessible data analysis, and software
  - These are basic building blocks for reproducible and extensible science

Why FLOSS?

- Based on a collaborative, efficient, and open development process
- Allows users to use, study, modify, and redistribute the software
- Centered around some publicly-accessible source code
- Benefits from social, ethical, technical, and economic points of view
- Source code access and sharing eases learning and promotes participation in the development of software
- Development model that may result in high quality code that gets adapted quickly to different situations with lower direct costs

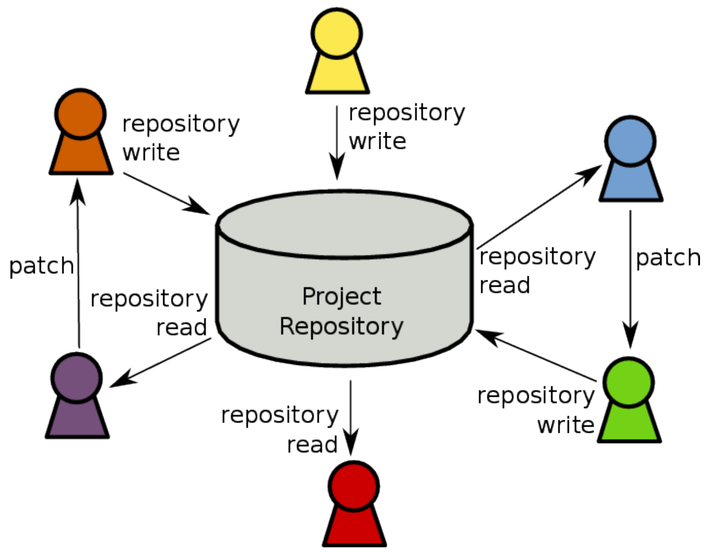

Figure : FLOSS development by means of a VCS repository

- Source code availability: projects have a publicly-accessible version control repository
- User/developer symbiosis: the developers are also users of the software, and they also provide requirements
- Non-contractual work: there is no central management with control over all of the developers' activities
- Work is self-assigned: volunteer developers tend to work on the parts of the project that most appeal to them, and employed developers will work on the parts that are of most interest to their employers
- Geographical Distribution: in most FLOSS projects the developers are spread among several different locations in the world

Opportunities

- Using FLOSS for SE Education
  - Involving students and faculty in large-scale FLOSS projects to provide them with real-world experience
  - FLOSS promotes project- and problem-based learning, developers/students work on projects that interest them
  - FLOSS is normally highly modular and APIs are well documented, for teaching principles and good practices
  - Provides quantitative data from real and freely available source code on which to perform analysis and base decisions
- Using data from FLOSS projects in SE Research
  - The social structure of development communities
  - Communication and workflow patterns in FLOSS projects
  - Developer evolution and participation
  - Attractiveness of FLOSS projects
  - Attributes of the source code and their impact on project success
- FLOSS software products in research activities
  - Releasing the source of software produced in research activities meets the basic scientific principle of reproducibility
  - Being open source allows research prototypes to be enhanced into production-class products without the researchers having to be involved forever

FLOSS in SBES (Brazil)

- 378 papers analyzed: 206 main track papers between 1999 and 2010, 142 tools session papers between 2001 and 2010, and 30 FEES papers from 2008 to 2010
- Matchers: software(s) livre(s), ferramenta(s) livre(s), ferramenta(s) aberta(s), software(s) aberto(s), código aberto, repositório(s) de software, free software, open source, open software, libre software, software repository, OSS, FLOSS, FOSS, and OSSD

Table : Main track papers analyzed, by year

Year	Analyzed	Mention FLOSS	Relative frequency
1999	26	0	0.00
2001	20	0	0.00
2002	18	2	0.11
2004	17	1	0.06
2005	21	2	0.10
2006	19	0	0.00
2007	23	2	0.09
2008	19	1	0.05
2009	24	6	0.25
2010	19	11	0.58

Table : Analysis of research papers mentioning FLOSS

Category	2002	2004	2005	2007	2008	2009	2010	Total
Reference	0	0	0	1	0	1	1	3
Example or Comparison	1	0	0	0	0	1	1	3
Running FLOSS tools	1	1	2	0	1	1	0	6
FLOSS tool as target	0	0	0	1	0	0	3	4
Data from FLOSS	0	0	0	0	0	2	3	5
Research about FLOSS	0	0	0	0	0	1	3	4
Total	2	1	2	2	1	6	11	25

Table : Tool papers analyzed, by year

Year	Analyzed	Mention FLOSS	Relative frequency
2001	18	0	0.00
2002	19	3	0.16
2004	15	2	0.13
2005	12	2	0.17
2006	25	7	0.28
2007	14	2	0.14
2008	11	3	0.27
2009	12	4	0.33
2010	16	7	0.44

Table : Analysis of tools papers mentioning FLOSS

Context	2002	2004	2005	2006	2007	2008	2009	2010	Total
Reference	1	0	0	0	1	0	0	0	2
Promise to be FLOSS	0	0	1	1	0	1	0	1	4
use FLOSS	1	1	0	0	0	0	1	1	4
FLOSS (fake*)	0	1	0	5	1	2	2	3	14
FLOSS	1	0	1	1	0	0	1	2	6
Total	3	2	2	7	2	3	4	7	30

- The Brazilian Software Engineering community could be taking more advantage of the FLOSS research opportunities
  - We are late in comparison with the international scenario

Agenda

- Public Data: Encouraging that software research uses public data and FLOSS tools to process and analyze data
- Venues: Increase the number of venues welcoming results from FLOSS research
- Tools: Creating a repository for publishing the FLOSS tools presented in the tools sessions of Brazilian software engineering conferences and workshops
- Educational Material: creating a repository of educational material and a forum where SE educators can discuss and share experiences in the use of FLOSS (see: **softwarelivre.org/sbc**)
- FLOSSCC: Fostering the creation of FLOSS Competence Centers throughout the country

ccsl.ime.usp.br | www.dcc.ufba.br {fabio.kon,paulormm,lago}@ime.usp.br | {terceiro,flach,manoel.mendonca}@dcc.ufba.br