Bsc Thesis - On the reasons for founding OSS projects

Wolfgang Sprenger
Institut für Informatik
FU Berlin
05.02.2009
Outline

I. Motivation for this study
II. What is founding?
III. Methodology Part I (Literature review)
IV. Methodology Part II (E-mail Study)
V. Result
VI. Analysis
VII. Conclusion
VIII. Discussion
I. Motivation for this study

- FLOSS, WIDI and other studies show only why people participate in Open Source Software projects*

- No other study looks at the reasons for founding

- Is 'Scratching an Itch' the only reason (Raymond)?

* OSS project = OSI license, open ended, ownership
II. What is founding?

- Differentiate between 'idea', 'founding', 'announcement' and '1\textsuperscript{st} public release'

- Special case: out-sourcing fork

- Founding in this study: 1\textsuperscript{st} public release

\textbf{Caution}: This will exclude all projects which never have a release.
III. Methodology Part I

- Literature review:
  - Academic literature
  - 10 well known OSS projects and their founding stories
  - Reasons for founding derived from community literature
III.a) 10 well known OSS projects

- Apache Webserver
- Mozilla (Firefox)
- Gcc
- Samba
- Cups
- Linux Kernel
- Eclipse
- Open Office
- LaTeX (TeX)
- BSD
III.b) Example Mozilla

- Netscape lost more and more market share to IE
- Netscape Communications made two important announcements on January 23rd, 1998:
  - First, that the Netscape Communicator product would be available free of charge
  - Second, that the source code for Communicator would also be free
- Mozilla was the original code name

(Source: http://www.mozilla.org/mission.html)
III.c) Reasons for founding:

- Scratching an itch
- Plan to close existing software gap
- Strategic reasons
- Exercise
- Academic research project
- Fundamental reason
- Existing project growing too big
- To reach a goal fast
III.d) 10 well known OSS projects

- Apache Webserver -> Scratching an itch
- Mozilla (Firefox) -> Strategic reasons
- Gcc -> Scratching an itch
- Samba -> Scratching an itch
- Cups -> Scratching an itch
- Linux Kernel -> Exercise
- Eclipse -> Strategic reasons
- Open Office -> Strategic reasons
- LaTeX (TeX) -> Scratching an itch
- BSD -> Fundamental reason
IV.a) Methodology Part II

- E-mail Study (100 each):
  - Debian packet list
  - Sourceforge.net list (through sql queries)
  - Intersection of Debian packet list and Sourceforge.net list

- What data is interesting:
  - Reasons for founding
  - Date of 'having the idea', founding and 1\textsuperscript{st} release
  - Number of developers
  - What kind of communication tools are used
IV.b) Debian packet list

- approx. 20,000 Pakete
- list is downloadable from website
IV.c) Sourceforge.net list

- hosted by University Notre Dame
- monthly data ~25 GB
- SQL-queries via webinterface
IV.d) Gathering information – Part I

1\textsuperscript{st}: look at the project website

Samba
Opening Windows to a Wider World

10 years of Samba!
Who would have believed that this announcement would eventually lead to the project that is now Samba? Thanks to the involvement of countless individuals over the years Samba has come a long way in the last decade, and is set to continue improving for the foreseeable future.

The first version
The first tentative version was written in the first few days of 1992. You can read the gory details at http://ftp.samba.org/ftp/unpacked/samba/docs/history or you can grab that very first version from http://ftp.samba.org/ftp/samba/old-versions/server-0.5.tar.Z.

That first version still brings back happy memories, and some cringing at the awful code. It was my first sockets program, and it really shows. Last year while Jeremy and I were at a conference we downloaded that first version and managed to coax NT4 into getting a directory listing out of it, but apart from that it really is quite useless with current clients. The protocol has changed too much.

\texttt{nbserver 1.5}

After a long break, \texttt{nbserver 1.5} was released in December 1993. This version saw the addition of "nmbserver" (now known as \texttt{nmbd}) and \texttt{smcclient} and the renaming of the primary daemon to "smbserver". This is the first version that was influenced by a SMB specification, the earlier versions were based purely on sniffer dumps. This version was also the first to sport the ability to act as a print server.

The \texttt{nbserver 1.5} was released under the GNU General Public License, unlike earlier version which were released under a "do what you like" style of license. This largely reflects the fact that I had become a huge fan of Linux and was impressed by the progress being made by the Linux community. I've never regretted this decision, and I think the GPL has been very good for Samba.
IV.e) Gathering information – Part II

1st: look at the project website
Hello <Name of Developer>,

I'm studying the reasons why Open Source Software projects get founded. In the last couple of days I've been studying your project. I'm almost done with gathering data, I just want to ask if you can help me with some concluding questions.

<Listing of all manually gathered data>

Is this information correct?
IV.g) Sending e-mail (cont.)

Can you briefly describe why you started the project?

Approx. when did you have the idea for your project? (E.g. August 2006)

Approx. when did you start your project?

Approx. when did you publish your project?

Approx. when did you publish your first release?

(here only which data is still not known from the website)

How many developers are actually involved in your project?

What kind of tools do you use in your project to communicate respectively share source code, documentation, etc.?

Information will be made anonymous for this study.

Please let me know if you have questions and thanks for helping me out,
V. Result – Reasons for founding

For all projects (84):

<table>
<thead>
<tr>
<th>Reasons for founding</th>
<th># of projects</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>scratching an itch</td>
<td>69</td>
<td>82.14%</td>
</tr>
<tr>
<td>exercise</td>
<td>7</td>
<td>8.33%</td>
</tr>
<tr>
<td>curiosity</td>
<td>6</td>
<td>7.14%</td>
</tr>
<tr>
<td>research</td>
<td>5</td>
<td>5.95%</td>
</tr>
<tr>
<td>spinoff</td>
<td>3</td>
<td>3.57%</td>
</tr>
<tr>
<td>hobby</td>
<td>2</td>
<td>2.38%</td>
</tr>
<tr>
<td>fundamental reason</td>
<td>1</td>
<td>1.19%</td>
</tr>
<tr>
<td>other</td>
<td>2</td>
<td>2.38%</td>
</tr>
<tr>
<td>no data</td>
<td>2</td>
<td>2.38%</td>
</tr>
</tbody>
</table>
VI.a) Analysis – Number of developers

- 69 projects founded by 'Scratching an itch' projects seems to be founded by a need of a person
- 39 of these have one developer
- over all projects 47 projects have only one developer
- only 8 projects have no 'Scratching an itch' as reason
- Nearly the half of the projects 'Scratching an itch' have only one developer

→ Majority of projects are founded and maintained by one person to satisfy their needs
→ Raymond is (almost) right with his thesis
VI.b) Analysis - Starting Points

Correlation Starting Points and Reasons for Founding

- 'Scratching an itch'
  - from idea to start the project: average 1.4 months
  - start of the project to 1st release: average 3.6 months

- another reason for founding
  - from idea to start the project: average 7.2 months
  - start of the project to 1st release: average 2.8 months

- data not very detailed

   → outliers have significant impact on the results

   larger database to absorb outliers
VI.c) Analysis - Collaboration Means

- no direct cover between more collaboration means when more than one developer is involved in a project

- 45% of all projects use a revision control system
- Mailing list are used by about 25% of all projects
- No tools are used by about 19% of all projects
VI.d) Analysis – Each Databasis

- **Sourceforge:**
  - 14 of 16 projects have 'Scratching an itch' as reason
  - 10 of 16 projects have one developer
    - whereby 9 projects have 'Scratching an itch' as reason

- **Debian:**
  - 22 of 26 projects have 'Scratching an itch' as reason
  - 12 of 26 projects have one developer
    - whereby 12 projects have 'Scratching an itch' as reason

- **Both:**
  - 32 of 42 projects have 'Scratching an itch' as reason
  - 18 of 42 projects have one developer
    - whereby 18 projects have 'Scratching an itch' as reason
VII. Conclusion

- Main reason for founding OSS projects in this study is ‘scratching an itch’
- But this is not the only reason
- Other reasons are not as distinct as ‘scratching an itch’, but there are still common reasons

- Only scratched on the surface
- Deeper, more detailed and wider spread survey might confirm the results and give deeper insight into the topic
Thank you!

Any further questions?