The Role of Mailing Lists for Policy Discussions in Open Source Development

Masayuki HATTAa)

Abstract: This document analyzes the evolution of policy-related discussions in open source software by using several projects’ policy mailing list archives and focusing on the Debian Project. More specifically, it utilizes approximately 70,000 pieces of mail exchanged since the end of the 1990s, investigating the rise and fall in activity and what sort of topics was discussed. The results of this paper’s inquiry suggest that mail volumes peaked in 2005, that policy discussions were led and mainly contributed to by a relatively small subset of persons who only posted related to policy, and that overall mailing list traffic (not only related to policy) declined after 2006, possibly due to a transfer of discussion to Wikis, chats, and other such platforms.

Keywords: open source, organizational culture, mailing list, community governance

a) Faculty of Economics and Management, Surugadai University, 698 Azu, Hanno, Saitama, Japan, hatta.masayuki@surugadai.ac.jp
A version of this paper was presented at the ABAS Conference 2017 Summer (Hatta, 2017).
© 2018 Masayuki Hatta. This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Introduction

The word “open” has found itself attached to many terms in recent years—open source, open data, and open innovation, to name a few. However, due to the variance in contexts and users of this word, there are, at times, discrepancies as to what “open” specifically refers to, leading to fierce debates. A classic example of this conflict is the ideological rift between open source and free software (see Stallman, 2007).

The context for open source software has been defined by the interrelationship of being traditionally “open”—i.e., the essence of openness is codified in the software license—and the normalization/conventionalization of development project governance methods (Hatta, 2009). The value system developed by the projects formed in this way contribute greatly to community participation, continued motivation, and even advances in development (Mair, Hofmann, Gruber, Hatzinger, Zeileis, & Hornik, 2015). There is also quite a range of discussions about the similarities and differences in values between developers and users in open source and free software projects—we know that these do impact the progress of a project (Fujita & Ikuine, 2013, 2014). Ergo, this particular kind of discussion on “policy” in open source software is inseparable from daily development activities.

We should also note that, when a development project is undertaken by an existing organization operating over the long term, it is conceivable that values will change over time as they are affected by development advances, changes in the external environment, turnover in team members, and similar factors. For example, depending on the project, regardless of whether or not the software itself is open source, it is critical that funding does not rely too heavily on the government or other similar parties (S. Gardner, personal communication, September 27, 2015). A large body of
research exists on organizational culture and its impact, including Weick (1969, 1995), Hofstede (2001), Denison, Haaland, and Goelzer (2004), and Sato and Yamada (2004). Though these changes in traditional developer mindsets have come to fall under the scope of qualitative research, often employing interview methods, we felt that we could establish a clearer picture through a quantitative analysis of what kind of discussions are being had over time, using mailing archives of mailing lists—knowing their traditional role as primary forums for software development project discussions.

Method

In open source software development, mailing lists have become a primary channel for communication. There have been some prior studies (Guzzi, Bacchelli, Lanza, Pinzger, & von Deursen, 2013, etc.) that quantitatively analyze what development-related technical topics are being discussed in major mailing lists for specific projects. However, when it comes to the discussion of non-technical topics, it is difficult to fully grasp the overall picture due to the discussion’s dispersion across multiple media channels, including mailing lists, message boards, and chat. In addition, as pointed out by Guzzi et al., discussion of policy (social norms) on technical mailing lists for individual projects is rare.

In this study, we selected three relatively easily-accessible mailing lists, utilizing their archives for our analysis:

1. debian-legal (a mailing list for discussion of legal issues, including licenses, pertinent to the Debian Project, itself a very well-known and classic GNU/Linux distribution\(^1\) development project)

---

\(^1\) An aggregation of software with the Linux kernel to make it a functioning operating system.
2. fedora-legal (same as above, except the Fedora Project, also a well-known GNU/Linux distribution development project)
3. gnu-misc-discuss (mailing list where top-flight developers get together to discuss several issues related to “free software,” including but not limited to legal questions).

These lists are similar in that:

1. Since these lists are based not on individual software but on distributions, they attract a broad base of developers from different projects and with various backgrounds.
2. These developers discuss issues beyond coding and development.

In such ways, these mailing lists have similar characteristics. However, while the first two share the development organization of GNU/Linux distributions, Debian is a non-profit organization, whilst Fedora is supported by the RedHat company (despite being community-driven), and the gnu-misc-discuss list is populated by a number of developers who adhere to a belief in “free software,” establishing some distance from GNU/Linux distributions. We can therefore surmise that, since these three have slight differences in the wants and beliefs of their constituent developers, it would be appropriate to use these for comparison. Note that though the first two lists have “legal” in their titles, they are not limited to legal discussions and often discuss overall policy issues.

Next, we acquired all accessible mails for the aforementioned three mailing lists. From debian-legal, we obtained 36,589 mails from November 1998 to May 2016. From fedora-legal, we obtained 2,761 mails from May 2007 to May 2016. From gnu-misc-discuss, we obtained 20,146 mails from September 2000 to 2015. Since the archives for fedora-legal and gnu-misc-discuss are provided in mbox-format files, we were able to run an automated script using
Python’s mailbox library to extract actual mail text. Since debian-legal’s archive is not provided in an easily-processed format, we used html2text on the web archive (downloaded in HTML format) and discarded unnecessary portions using sed as needed.

Analysis

The collected data was analyzed using KH Coder (v2.00e), which is software primarily focused on quantitative text analysis (Higuchi 2015). Observations from this analysis are outlined below.

(1) Peak discussion periods

If we view the volume of mail for each mailing list, we can see that discussions peaked in 2005, excepting fedora-legal, which began operation only in 2007 (Figure 1).

We can see a divergent trend here versus other mailing lists in the same field that, for example, primarily discuss technical topics (e.g., debian-devel) (Figure 2). This suggests that, around 2005, policy

![Figure 1. The flow of the amount of mailing lists (unit: bytes)](image)

*Source*: Drafted by author.
issues saw greater discussion than technical ones.

There are several conceivable reasons for why policy-related discussions peaked in 2005. In early 2004, Microsoft began an anti-Linux campaign, while in late 2004, Novell began their “Unbending the Truth” campaign in response to the case of *SCO v. IBM* (began in 2003). These coincide with a period of increased interest in policy and legal issues.

However, we can see that “Microsoft” and “Novell” are not necessarily frequent topics of discussion if we look at commonly-used words and actual mail contents. Therefore, it is most likely that this is a period of increased interest in Linux-related legal and policy-related issues spurred by actions taken by Microsoft and Novell (Figure 3). Also, if we look specifically at the Debian Project’s situation and timeline, we notice that Ubuntu, a project derived from Debian work and primarily developed by Canonical Ltd., was announced in October of 2004. This spurred a great deal of discussion about the shift from the traditionally volunteer-based Debian development to a commercially-led project. Furthermore, Debian 3.1 was released in June 2005, leading to discussions

**Figure 2.** The comparison of debian-legal and debian-devel

*Source:* Debian Project.
regarding Mozilla-related software licensing issues.

(2) Classifying discussion members

Examining the number of posts per individual poster, we can see that each of the in-scope mailing lists has a relatively small cadre of “regulars” posting a large portion of mails, exhibiting a “long-tail”-like trend (Figure 4).

If we examine these regulars more closely, we can see that the lineup changes around 2005 (Table 1). Taking debian-legal as an example, we can see that pre-2005, the top posters are largely comprised of people who had been active in technical areas as Debian developers (top posters for 2003 included those later chosen as

---

**Figure 3.** The co-occurrence network graph of debian-legal in 2005

*Source: Drafted by author.*
**Figure 4.** Senders (horizontal axis) and the number of mail he/she sent

*Source:* Drafted by author.

**Table 1.** The top senders

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>452</td>
<td>B.R.</td>
<td>253</td>
<td>M.R.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>241</td>
<td>A.D</td>
<td>251</td>
<td>F.P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>224</td>
<td>H.M.</td>
<td>122</td>
<td>A.T.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>219</td>
<td>T.B.</td>
<td>110</td>
<td>G.M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>149</td>
<td>D.A</td>
<td>96</td>
<td>M.P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>148</td>
<td>S.L.</td>
<td>94</td>
<td>D.A.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>134</td>
<td>A.S.</td>
<td>80</td>
<td>A.D.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>G.M.</td>
<td>79</td>
<td>N.N.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>129</td>
<td>B.S.</td>
<td>72</td>
<td>S.L.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>M.R</td>
<td>70</td>
<td>J.S.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

debian-devel, 2003
debian-devel, 2005

*Source:* Drafted by author. Blue and red texts indicate persons who were not official developers at the time. The repeating initials represent the same person.
Debian Project leaders as well as leaders of other open source software development projects). However, around 2005, we see non-developers joining the ranks of most active posters. For example, in 2005, the second most-active poster (by number of posts) is not a public Debian developer. Over time, we see a trend where more and more posters are observed discussing solely policy-related topics.

(3) Opinions on the Free Software Foundation

The Free Software Foundation (FSF), an organization bent on “software freedom” and run by Richard Stallman, is both considered the originator of open source and yet continues to maintain a contentious existence. More specifically, the revision of the GNU GPL (General Public License) by FSF from 2006 into 2007 triggered much controversy in what can be called a redefinition of an entire value set and not simply a single license revision—in that it boldly addressed issues such as how to handle contentious DRM and software patents within the community. In a statement published by major Linux kernel developers in September 2006, the new version of GPL, GPLv3, “portends the Balkanization of the entire Open Source Universe upon which we rely.” (LWN.net, 2006)

However, in our analysis, we find that even by around 2000 and particularly in the debian-legal mailing list, opinions toward FSF had turned quite harsh. At the turn of the century, words that were found commonly alongside FSF included “ironic,” “suck,” and “mistake” among others. Reviewing mail content, we can see that this is because of dissatisfaction brought on by FSF’s GNU Free Documentation License (FDL), announced in 2000.

Thus, rather than the GPLv3 revision engendering “Balkanization,” it would likely be more accurate to say that the growingly active discussion up to 2005 pushed FSF to begin their GPLv3 revision process and that this revision led to the evident manifestation of the long-building divergence in values in the community.
Declining Use of Mailing Lists as a Communication Channel

Though open source continued to flourish after 2005 as a method for software development, the common acceptance of GNU/Linux led to a sort of obsolescence in open source as a discussion topic, reaching a state of commoditization. As depicted in Figure 3, though open source licensing (including GPL and copyleft) had been a mainstay of discussion topics within the policy realm, the emergence of several thinkpieces and broadened understanding amongst legal experts led to this being a much less perplexing topic. This broadened understanding also conceivably contributed to a general downturn in discussion activity pertaining to policy.

However, what is perhaps much more interesting is that not only did policy-related mailing lists see decreased posting volumes at the peak period around 2005 but also did development-related mailing lists (Figure 2). This suggests that discussions on open source moved from mailing lists to other forums.

It is understood that open source development discussions have, in recent years, moved away from mailing lists to a number of different platforms, including development support site ticket (fault reporting) management systems such as GitHub, Wikis, and new chat systems like Slack. Although version control systems (VCS) had been used since the 1990s to smoothly edit source code simultaneously amongst multiple developers, Git, developed in 2005, gained a burst in popularity because of its adoption for use in the Linux kernel. Although GitHub, a software development platform utilizing Git, did not emerge until 2008, it only took a couple of years for it to become the most popular “social coding” website (Git Project, 2009).

On these platforms, policy issues are often discussed on ticket

---

2 In an interesting twist, Git was a product of the extremely policy-heavy discussion about the right and wrong of using a proprietary (non-open source) VCS—BitKeeper—in Linux development (Bonilla, 2011).
The role of mailing lists for policy discussions in open source development

managements as if they were a kind of bug, and each project has a Wiki or other similar repository for information; this means that there are many instances where discussions are entirely supplanted by non-mailing list forums, and mailing lists are not even set up in the first place.

Conclusion

In this paper, we quantitatively analyzed mail archives to reveal the interests and values of developers at a given point in time between different open source projects. Through our analysis, we determine that while policy-related discussions peak around 2005, this coincided with a shift from open source as a hobbyist topic to a more accepted mainstream concept, and with open source brushing up against new, “non-technical” issues such as commercialization and licensing. Furthermore, the changes in active mailing list posters suggest that the development community diversified around 2005, with non-developers potentially coming to wield influence over policy issues where developers were once the sole power. This also coincided with more commercial implementations of open source—as was suggested by the emergence of Ubuntu at the end of the previous year—changes in development organizations, and even changes in terms of who actually handles development.

Though there is likely some room to improve in terms of our quality of data and analysis methods, we may conclude with some certainty—that we were able to present a hypothesis about the changes in values seen as open source became commonly accepted. This means that analysis of more recent trends will require establishment of new data acquisition methods. Furthermore, for Japan specifically, since nothing similar to the mailing lists we have placed in the scope of our research exists, it would be challenging at present to acquire the same type of
data. We would like to further expand our data collection methods in the future, including interviews, and conduct a similar analysis as such.

Acknowledgments

This work was supported by JSPS Grant-in-Aid for Publication of Scientific Research Results, Grant Number JP16HP2004.

References


The role of mailing lists for policy discussions in open source development


