5-21-2019

“Something to talk about” Exploring open source design spaces

Kevin Lumbard
University of Nebraska at Omaha, klumbard@unomaha.edu

Vinod Kumar Ahuja
University of Nebraska at Omaha, vahuja@unomaha.edu

Matt Germonprez
University of Nebraska at Omaha, germonprez@gmail.com

Sean Goggins
University of Nebraska at Omaha, sgoggins@gmail.com

Follow this and additional works at: https://aisel.aisnet.org/mwais2019

Recommended Citation
Lumbard, Kevin; Ahuja, Vinod Kumar; Germonprez, Matt; and Goggins, Sean, ““Something to talk about” Exploring open source design spaces” (2019). MWAIS 2019 Proceedings. 18.
https://aisel.aisnet.org/mwais2019/18

This material is brought to you by the Midwest (MWAIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in MWAIS 2019 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
ABSTRACT

Open source projects are dynamic environments where individuals and organizations collaborate to accomplish mutually beneficial design tasks through the creation of shared systems, architectures, and platforms (Chesbrough, 2006; Germonprez et al., 2017). Open source projects are the foundation of the modern digital infrastructure (Eghbal, 2016). Thus, organizations

LITERATURE REVIEW

Open Source Design

Open source projects are dynamic environments where individuals and organizations collaborate to accomplish mutually beneficial design tasks through the creation of shared systems, architectures, and platforms (Chesbrough, 2006; Germonprez et al., 2017). Open source projects are the foundation of the modern digital infrastructure (Eghbal, 2016). Thus, organizations

RQ1: What is the relationship between discussion and implementation spaces as evident in the design of open source software?
are investing resources in open source projects as evidenced by research that suggests approximately half of open source designers are now acting as agents on behalf of organizations (Riehle, Riemer, Kolassa, & Schmidt, 2014). As organizations engaging in open source design find creative ways to balance internal and external design while working with open source projects (West & Gallagher, 2006), project dynamics are altered through new governance and coordinating mechanisms (Feller, Finnegan, Fitzgerald, & Hayes, 2008). In this context, open source designers coordinate their design in design spaces, including discussion and implementation spaces (Sack et al., 2006).

**Open Source Discussion Spaces**

Open source discussion spaces include email lists, chats, video conferencing, web platforms (e.g., GitHub and Stack Overflow), private messaging modes for sensitive communication, and in-person conferences and hackathons. Discussion spaces provide support for formal and informal design communication in open source projects (Boden, Rosswog, Stevens, & Wulf, 2014). Open source discussion spaces have evolved from their initial form where they were primarily used to track and archive identified design flaws into a platform that allows designers to discuss new design concepts and coordinate large interdependent design tasks (Bertram, Voida, Greenberg, & Walker, 2010). Discussion spaces are beneficial - enabling trust to be built between designers, clarifying problems with design, and providing feedback, which may lead to design improvements (Haraty, McGrenere, & Bunt, 2017). Timely and visible feedback allow open source designers to reap the advantages of temporal and geographic distance (Dabbish, Stuart, Tsay, & Herbsleb, 2012; Lundell, Lings, Ågerfalk, & Fitzgerald, 2006) rather than the associated problems due to locality and spatiality in distributed engagements (Olson & Olson, 2000). As discussion spaces have evolved to become more central to the design process, they have begun to service many of the conversational, archival, and organizational needs of projects (Bertram et al., 2010). From their central role in open source design, discussion spaces are expected to shape open source design tasks.

**Open Source Implementation Spaces**

Open source projects complete design tasks within design spaces through the implementation of proposed design changes into an expanding codebase (Howison & Crowston, 2014). Proposed changes may result from implicit coordination in the implementation space or explicit coordination in discussion spaces (Bolici, Howison, & Crowston, 2016). To manage interdependencies in design tasks, implementation spaces often incorporate or link to discussion spaces to share knowledge (Lindberg, Berente, Gaskin, & Lytyinen, 2016). Implementation spaces further utilize discussion spaces for identification of design flaws by capturing bug reports from users and designers (Fogel, 2005). This shows that implementation spaces are often dependent upon discussion spaces for guidance on open source design tasks.

**Relationship Between Discussion and Implementation Spaces**

Well defined design processes create constraints that promote the layering of design tasks into small manageable units leading to code that is more reusable, higher quality, and easier to understand (Howison & Crowston, 2014). As projects self-organize, implicit design processes--trust, mental maps, and shared frames--and explicit design processes--written rules, plans, and feedback --are defined (Bird, Pattison, D’Souza, Filkov, & Devanbu, 2008; Bolici et al., 2016) Open source projects often create documents for how to contribute, joining scripts, and project roadmaps to reduce the barriers of entry for new designers and increase success outcomes (Steinmacher, Conte, Gerosa, & Redmiles, 2015). Further, many open source projects define explicit connections between discussion spaces and implementation spaces. For example, the published design processes of many open source projects recommend starting design tasks within discussion spaces prior to moving the task to the implementation space to increase the likelihood of success (https://kubernetes.io/docs/contribute/). This indicates that linking discussion spaces to implementation spaces can play an important role in completing design tasks.

**METHOD**

To investigate the relationship between discussion spaces, implementation spaces, and the completion of design tasks in open source projects, this research explores design in organizational-communal open source projects under the umbrella of the Linux Foundation and focuses on design processes of three projects -- Kubernetes, GRPC, and Zephyr. Kubernetes is an open source system for managing of containerized applications. GRPC is an open source remote procedure call framework used to connect devices across different platforms. Zephyr is an open source operating system for “Internet of Things” devices. These projects were selected because they are active Linux Foundation projects and share similar levels of organizational interest. Additionally, project work for all three occurs on the GitHub platform making trace data collection repeatable for each project.

---

1. https://kubernetes.io/
2. https://grpc.io/
3. https://www.zephyrproject.org/
4. https://github.com/
GitHub is an established source of data for research (Cosentino, Luis, & Cabot, 2016). The GitHub discussion space includes social features and a transparent workflow that make information about users and their work accessible. Designers can make issue comments and code comments in online forums built around the workflow. The GitHub implementation space provides a mechanism for designers to share and merge code bases via pull requests. Pull requests allow designers to create a thread for proposing code changes and discussing it with others. Within a pull request, proposed design changes can be viewed, commented, edited, voted on, and finally accepted or rejected.

Throughout a seven-year research project, the research team has become active participants in numerous open source projects. This gives us a unique position to understand and report on open source project research broadly. From this broad understanding, time was spent specifically exploring the three aforementioned projects on GitHub to understand specific components of their design process. The three projects selected show different levels of activity providing some variation in the contexts within which the research question is addressed. (see Table 1).

### Table 1: Project Activity

<table>
<thead>
<tr>
<th>Project</th>
<th>Pull Requests</th>
<th>Pull Request Files changed</th>
<th>Code &amp; Comment Contributors</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kubernetes</td>
<td>42,485</td>
<td>727,218</td>
<td>13,423</td>
<td>464,834</td>
</tr>
<tr>
<td>GRPC</td>
<td>10,749</td>
<td>223,699</td>
<td>3,613</td>
<td>59,109</td>
</tr>
<tr>
<td>Zephyr</td>
<td>7,078</td>
<td>87,858</td>
<td>932</td>
<td>22,474</td>
</tr>
</tbody>
</table>

#### Research Approach

Digital ethnography was used to extract and examine trace data and project documentation from Kubernetes, GRPC, and Zephyr. Digital ethnography is a participant-observation research approach adjusted to an online world (Kozinets, 2015), in this case the lens aimed explicitly at both the online work space and the online talk space (Goggins, Mascaro, & Valetto, 2013). Online data sources are used to arrive at understandings and representations of online social experiences, through observation, exploration of archival data, and quantitative analysis. Human communications were interpreted – not merely words but also digital artifacts - “under realistic contexts, in situ in native conditions of interaction” (Kozinets, 2015, p. 5).

#### Data Extraction and Analysis

Data was extracted from GitHub using its GraphQL interface on January 15th, 2019. RStudio was used to clean the data by removing data that did not fit the described schema, outlier data that were the result of automatic processes malfunctioning (i.e., bots), and pull requests that did not include changes to the repository. To operationalize the link between discussion and implementation spaces, R and regular expressions were used to extract whether a pull request was explicitly linked to an issue from the body text of the pull request.

To understand the relationship between discussion spaces and implementation spaces, this research explored how posted issues (discussion spaces) are related to pull requests (implementation spaces) that have been merged (success) or closed (non-success). Within pull requests, data was extracted about, whether the pull request was explicitly linked to an issue and merge success (see Figure 1).

#### Figure 1. Comparing merge success rates of pull requests that originate in issue discussions to pull requests that do not.

Based on open source literature (McDonald, Blincoe, Petakovic, & Goggins, 2014), it is expected that discussion in issues will influence pull request merge success. To determine the correlation between issue discussion and merge success, descriptive statistics were used to explore the relationship between the variables. To determine if the relationship is statistically significant and to calculate odds ratios, a logistic regression was performed. Logistic regression was chosen because the two variables in the analysis, pull request merge and pull request issue link are both binary.
RESULTS

Discussion Spaces are viewed as essential parts of the design process

All of the projects explored referenced discussion spaces in their documented design processes and encourage planning and discussion first in the discussion space and then continuing into the implementation space when working on design tasks. The following are examples from each project documentation:

“Even for small changes, it is often a good idea to gather feedback on an issue you filed, or even simply ask in the appropriate SIG's Slack channel to invite discussion and feedback from code owners.” -- Kubernetes

“For speculative changes, consider opening an issue and discussing it first. If you are suggesting a behavioral or API change, consider starting with a gRFC proposal. Provide a good PR description as a record of what change is being made and why it was made. Link to a GitHub issue if it exists.” -- GRPC

“Before starting on a patch, first check in our issues Zephyr Project Issues system to see what’s been reported on the issue you’d like to address. Have a conversation on the Zephyr developer mailing list (or the #zephyrproject IRC channel on freenode.net) to see what others think of your issue (and proposed solution). Send a message to the Zephyr developer mailing list to introduce and discuss your idea with the development community.” -- Zephyr

Linking discussion spaces to implementation spaces results in design tasks being completed at a higher rate

Design tasks - that include discussion spaces linked to implementation spaces - are successfully completed at a higher rate, than those with no link. For example, in Kubernetes, 51.50% of successfully merged pull requests are linked to an issue whereas 43.15% of successfully merged pull requests are not linked to an issue (See table 2).

Table 2: Design Tasks Link to Discussion Spaces

<table>
<thead>
<tr>
<th>Project</th>
<th>Linked to Discussion Space (Issue)</th>
<th>Successfully Completed (Merged Pull Request)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kubernetes</td>
<td>Yes</td>
<td>51.50%</td>
</tr>
<tr>
<td>No</td>
<td>43.15%</td>
<td></td>
</tr>
<tr>
<td>GRPC</td>
<td>Yes</td>
<td>33.86%</td>
</tr>
<tr>
<td>No</td>
<td>28.21%</td>
<td></td>
</tr>
<tr>
<td>Zephyr</td>
<td>Yes</td>
<td>19.98%</td>
</tr>
<tr>
<td>No</td>
<td>18.69%</td>
<td></td>
</tr>
</tbody>
</table>

The calculated odds ratios from the logistic regression show that design tasks where discussion spaces are linked to implementation spaces are more likely to be successfully completed for all three projects. Pull requests linked to issues are 1.39 times more likely to be merged on Kubernetes, 1.29 more likely on GRPC and 1.08 more likely on Zephyr. For Kubernetes and GRPC the results are statistically significant. While this positive relationship is observable in Zephyr, the results are not significant, suggesting that there may be other factors that influence successful task completion, or, possibly a different discussion space such as an email list or routine meetings where pull requests are discussed (See table 3).

Table 3: Logistic Regression

| Project | Coefficients | Estimate | Std. Error | Z Value | Odds Ratio | Pr (>|z|) |
|---------|--------------|----------|------------|---------|------------|----------|
| Kubernetes | (Intercept) | 1.24433 | 0.01644 | 75.69 | 3.4706 | <2e-16 *** |
| Link to an Issue Success | 0.3357 | 0.02457 | 13.66 | 1.3989 | <2e-16 *** |
| Null deviance: 42201 on 42484 degrees of freedom |
| GRPC | (Intercept) | 1.56431 | 0.01362 | 50.134 | 4.7794 | <2e-16 *** |
| Link to an Issue Success | 0.25611 | 0.05771 | 4.438 | 1.2919 | 9.09e-06 *** |
| Null deviance: 9490.2 on 10714 degrees of freedom |
| Zephyr | (Intercept) | 1.95298 | 0.04024 | 48.528 | 7.04965 | <2e-16 *** |
| Link to an Issue Success | 0.0831 | 0.09273 | 0.896 | 1.08665 | 0.37 |
| Null deviance: 57237 on 60277 degrees of freedom |

DISCUSSION

This research shows that the discussion space - implementation space relationship plays a key role in open source design processes for three specific projects on GitHub, supporting and elaborating more specifically on prior research (Bertram et al., 2010). The role of discussion spaces supporting implementation spaces to share knowledge, manage interdependencies, and

5 https://kubernetes.io/docs/contribute/
6 https://github.com/grpc/grpc/blob/master/CONTRIBUTING.md
7 https://docs.zephyrproject.org/latest/contribute/index.html#pull-requests-and-issues
implement design changes is surfaced in prior work (Bertram et al., 2010; Howison & Crowston, 2014; Lindberg et al., 2016). This research contribution shows the specific influence of discussion spaces on design task completion success.

In organizational-communal projects (Germonprez et al., 2017), discussion spaces are used for planning and coordination of design tasks. Well defined project design processes let contributors know that changes to the code base should be proposed and discussed in discussion spaces prior to making a change. Activity within discussion spaces often leads to work within implementation spaces and when it does, it has a higher completion success rate than work that does not originate in discussion spaces. Analysis of Kubernetes, GRPC, and Zephyr indicate that it is more likely that the work of a designer will result in successful completion of a design task if the task stems from a combination of work in both discussion and implementation spaces.

Further, the occurrence rate and the completion success rates of linked discussion and implementation spaces may vary by project activity. For example, in projects with large amounts of activity, like Kubernetes, more pull requests are linked to issues and there is a strong positive relationship between issue linking and pull request merge success, whereas the two projects showing less activity also have fewer occurrences and a weaker positive relationship. In this context, project activity appears to be a moderating variable that affects the relationship between design spaces. Future research could explore the impact of project activity and project size on design spaces.

LIMITATIONS

This research focused on strategically important open source projects at the Linux Foundation. This research does not explore open source projects outside of the Linux Foundation or volunteer-driven projects, nor is it meant to negate prior research on volunteer-driven communities. Further, this research is not intended to generalize design for all open source projects but merely explore design through the study of organizational-communal open source projects. Additionally, bots were identified in Kubernetes (4 bots), GRPC (2 bots) and Zephyr (1 bot) but they were not removed from the research. Bots were included as participants and may add comments to pull requests and participate in discussion spaces. This research sees bots as agents acting on behalf of project leadership. In this, they communicate design guidelines, enforce rules, and review pull requests for compatibility with interdependent files. As such, they are participants in design. Removal of the bots from this research is unlikely to change the results drastically but their existence should be noted.

CONCLUSION

This research explored design in three open source software projects. This work contributes to design literature and open source literature by exploring the relationship between discussion spaces and implementation spaces as evident in the design of open source software. Discussion spaces are an increasingly important part of design in organizational-communal open source projects. Understanding how discussion spaces and implementation spaces affect task completion can help open source designers create well defined design processes that promote success outcomes and improve designer engagement for all.

ACKNOWLEDGEMENTS

This project received funding from the Alfred P. Sloan Foundation Digital Technology grant on Open Source Health and Sustainability, Num 8434 (https://sloan.org/grant-detail/8434).

REFERENCES


