From the Inside Out: Organizational Impact on Open-Source Communities and Women's Representation

Hana Frluckaj*
hanafrla@utexas.edu
University of Texas
USA

Huilian Sophie Qiu*
sophie.qiu@kellogg.northwestern.edu
Northwestern University
USA

Bogdan Vasilescu
vasilescu@cmu.edu
Carnegie Mellon University
USA

Laura Dabbish
dabbish@cmu.edu
Carnegie Mellon University
USA

ABSTRACT

The involvement of companies and public institutions in open-source software (OSS) has become widespread. While studies have explored the business models of for-profit organizations and their impact on software quality, little is known about their influence on OSS communities, especially in terms of diversity and inclusion. This knowledge gap is significant, considering that many organizations have the resources to enhance diversity and inclusion internally, but whether these efforts extend to OSS remains uncertain. To address this gap, we conducted interviews with maintainers of community-owned and organization-owned OSS projects, revealing tensions between organizations and their projects and identifying the impact of internal policies on OSS communities. Our findings reveal that, on the one hand, organization-owned projects often restrict external contributions due to stringent operating procedures and segmented communication, leading to limited external engagement. On the other hand, these organizations positively influence diversity and inclusion, notably in the representation and roles of women and the implementation of mentorship programs.

1 INTRODUCTION

According to GitHub's 2021 report [23], over 50% of GitHub contributors are affiliated with private companies, governments, or other organizations. This reflects the increasing involvement of organizations in the open-source software (OSS) movement. While OSS initially relied on volunteer-driven projects, many organizations have now established their own OSS initiatives or made parts of their source code OSS [6, 82], increasing the professionalization of the whole OSS ecosystem. Typically, such company-owned OSS projects are primarily developed and maintained by paid employees.

Conflicts can arise in OSS communities due to differing resource requirements and priorities between organization-owned OSS projects
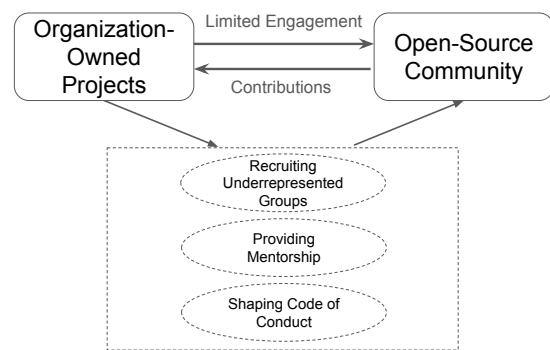
*These authors contributed equally to this work and were listed alphabetically.

Figure 1: Illustration of organizations' impacts on their OSS community. Themes in dotted lines are the supports organizations can provide for their OSS community.

and their proprietary, profit-driven counterparts. Organizations often hold decision-making power, influencing what is accepted into the code base and the project's future direction [55]. Consequently, these projects are sometimes seen as attempting to benefit from the labor of volunteers while deviating from the collaborative spirit of OSS participation.

While organizational involvement in OSS projects may raise concerns, it can also bring benefits to the communities. Organi-zation-owned OSS products reportedly have a faster time-to-market [72], faster development [80], and free and speedy user feedback [55]. Organizations also improve code quality [9], provide funding [7, 21], and coordinate logistics and marketing [9, 20]. More importantly, organizations can influence OSS projects' governance practices [9, 16, 25] and achieve sustained productivity [8].

Little research has examined the relationship between organizational involvement and OSS diversity- and inclusion-related practices. A notable exception is the recent interview study by Guizani et al. [27], which finds some high-level evidence that "companies invest time and effort in addressing OSS diversity and inclusion issues" as part of broader work on what motivates companies to contribute to OSS and how they tend to do so. Meanwhile, the overall OSS ecosystem suffers from a lack of demographic diversity, with clear evidence at least regarding gender [53, 57]. This not only

perpetuates a social inequity but also severely limits the pool of potential OSS contributors, negatively affecting the success and sustainability of OSS projects [15] and the overall health of OSS communities [10, 56]. Thus, in this study, we aim to answer the following central research question: *How do organizations' involvement influence the representation of women in their OSS projects?*

We conducted interviews with OSS project maintainers from 21 projects with different ownership structures, including small and large companies, public institutions, and community-led projects. These interviews explored contribution management and inclusion strategies. Based on these interviews we examined the influence organizations exerted on their open source project's inclusion in terms of interaction with external contributors and representation of women contributors.

The highlights of our results are shown in Figure 1. In our study focusing on women's representation within OSS projects, we found that organization-owned projects tend to limit external contributor engagement and enforce stringent quality requirements, resulting in segmented communication channels. However, these same organizations positively influence demographic diversity. They not only show a greater inclusion of women in significant roles but also actively recruit members from underrepresented groups. Additionally, these projects are more inclined to adopt Codes of Conduct, a practice often driven by organizational policies. This dual effect underlines a nuanced dynamic: organizational involvement, while restricting some aspects of open participation, plays a crucial role in enhancing demographic diversity within OSS projects.

2 RELATED WORK

This section focuses on the complex interplay between organizational strategies and community-driven OSS development, exploring themes like power dynamics, contribution patterns, and project governance. Additionally, the literature reflects a growing concern about demographic diversity, particularly gender diversity, and inclusive practices within OSS communities.

2.1 Organizational Involvement and Influence on OSS

For-profit organizations take part in OSS in various ways. Although OSS is provided to users unrestrictedly, companies provide services such as provision and maintenance for monetary gain [62]. Firms contribute to the OSS community in testing [25], reporting bugs [25], suggesting functional requirements [25], animating forums [25], getting in touch with potential users [25], employing core project developers [7], funding projects [6, 7, 20], joining project steering committees [7], providing logistic and marketing support [20], and organizing promotional activities [5]. The most important and most potent involvement, nevertheless, is to have employees directly commit to OSS projects [5]. Nguyen Duc et al. [44] reported that paid employees could resolve more issues than volunteers. Capiluppi et al. [8] argued that an important factor of an OSS project's success depends on whether it is owned by a company because company-owned OSS projects are found to have sustained productivity and increased amounts of output and number of new developers.

In addition to participating in community-founded OSS projects, many organizations, such as companies or public institutes, also founded their own OSS projects [82]. West and O'Mahony [82] classified OSS projects into *individually-founded* and *organizationally-founded* and designated the former as "organic" and the latter as "synthetic". West and O'Mahony [83] considered OSS projects founded by organizations as sponsored projects where one or more organizations control the projects' short- or long-term activities. The authors identified two types of openness: *transparency* and *accessibility*. Transparency allows external contributors to understand what is happening and to use the source code, whereas accessibility allows external contributors to participate in the decision-making process. West and O'Mahony further identified three dimensions that can influence an organization-owned project's openness: the organization of production, community governance, and intellectual property.

There are multiple studies on how companies manage their OSS projects. Harutyunyan and Riehle [30] proposed a set of best practices for companies' OSS governance and compliance in the areas of product analysis, transition policy, transition organization, IP-at-Risk analysis, and communication and capabilities. Alami and Wąsowski [2] classified different models when companies interact with OSS, including active participation for mutual support and benefits; latent participation, where companies delay the release of internal projects until an economic gain is guaranteed; and passive participation, where companies leverage OSS benefits without contributing back. Henkel [31] reported that companies might partially open a selective part of their intellectual property; some of the companies perceive active participation as overly open and thus unsuitable for a company. Forrest et al. [21] revealed that many organizations focused their contributions mainly on contributing codes and neglected other aspects, such as bug reporting or triaging, and thus their code might not address the need of the rest of the OSS community.

Relatively little is known about how organizations deal with external contributors. Although Alami and Wąsowski [2] studied how companies exploit software developed by external OSS contributions in their internal products, these OSS projects are not necessarily founded and managed by the companies. Pinto *et al.* [47] found that only a few corporate projects experienced growth in long-term external contributors. After analyzing five well-known company-owned OSS projects, Dias *et al.* [14] found that external contributors are making more contributions (56.7%), both code and non-code, than internal developers (43.3%), but contributions from external contributors take longer to be accepted.

Collaboration in organization-owned OSS projects typically follows a hybrid model, combining both online and offline interactions among project members [60]. However, in this distributed development setting, there can be a lack of awareness about the status of other project components due to factors such as physical distance, time-zone differences, language barriers, and limited face-to-face communication. This low awareness can result in misunderstandings, errors, and overall reduced performance within the project [29, 66]. Additionally, the division between online and offline collaborations can inadvertently exclude remote contributors from important decision-making processes and limit their overall engagement in the project.

Our study expands on this trend of prior works and explores tensions between organization-owned OSS projects and their external contributors.

2.2 Diversity and Inclusion in OSS

2.2.1 The Status of Diversity in OSS. Most studies on the diversity overview have focused on gender diversity [53, 58]. Overall, the percentage of women and non-men contributors has been very low in OSS, less than 10%. Many prior studies computed the percentage of different genders (see Qiu *et al.* [53] for an overview of the prior studies), but they were using a sample of the OSS population. Two recent studies that analyzed the gender distribution across the entire OSS population confirmed the low gender diversity [53, 58]. However, they also reported an optimistic observation that the percentage of non-men contributors has been increasing in the past decades. This is true in all ecosystems, such as PyPI, npm, and CRAN [53], and in most continents [48, 57].

2.2.2 Methods for Improving Diversity and Inclusion. Researchers and practitioners tried various ways to improve diversity and inclusion (DEI) in OSS projects.

Mentorship. Providing mentorship for newcomers is effective in making the OSS teams more inclusive to newcomers. Fagerholm et al. [17] found that deliberate mentorship, e.g. a mentor at a Hackathon, can make newcomers more active in the early stage of their participation than the ones with a non-deliberate process, which is more common in many OSS projects. Mentorship for newcomers is more common in corporate organizations. Mentorship and onboarding programs for new employees helped them achieve the best onboarding outcome [4, 54], made them "feel at home," and helped avoid the barriers that newcomers may face Dagenais et al. [13]. Moreover, in startups that suffer from high turnover rates, Pratiwi et al. [49] found that it was important to provide new employees with the context of the company and sufficient assistance and mentoring to guide them on the initial project.

While most OSS projects do not have the resources for an explicit mentorship program, many of them perform implicit mentoring during the contribution process, *i.e.*, code reviews on pull requests, or during events such as hackathons [45]. Feng *et al.* [18] showed that 27.41% of pull requests in their dataset contains implicit mentoring comments, such as providing instructions, suggestions, or mechanisms to fix errors. They further discovered that a higher proportion of women contributors performed implicit mentoring.

One of the barriers for newcomers in OSS projects is finding a mentor, as many experienced contributors are primarily motivated by their own contributions rather than actively seeking mentoring roles [67]. Existing studies on knowledge sharing and expertise sharing have developed tools on expert recommendation and identification [1, 42, 43]. However, there is currently no implementation of a mentor recommendation system specifically tailored to the context of OSS.

Code of Conduct. The inclusion of a Code of Conduct (CoC) in OSS projects has been advocated as part of corporate social responsibility [40]. Research suggests that implementing a CoC can help reduce the gender gap and promote inclusivity in society [48]. Furthermore, having a CoC is seen as a signal of inclusion, particularly for women and minority groups [64]. However, a study examining 28 codes of conduct found that only 13 of them included specific rules regarding gender and minority-related issues [64]. It is worth noting that many OSS projects adopt standardized templates for their CoC, which may limit the depth of discussion and input from project members when creating the code [74].

Transparency signals. Researchers have also worked on developing tools to facilitate DEI practices. Social coding platforms like GitHub offer transparency through their user interfaces, allowing users and projects to signal various properties such as code quality or test coverage. These signals, often displayed through visible cues, contribute to making the management process transparent and public [12, 50, 77]. By leveraging these signals, contributors in a distributed software development environment can gain team awareness and effectively collaborate on OSS projects [28].

Researchers have developed metrics and tools aimed at assisting project maintainers in monitoring the health and community dynamics of their projects. These metrics are designed to reflect and promote DEI within projects. For example, some tools recognize and highlight a project's efforts in promoting diversity, while others acknowledge and encourage contributions from new contributors [24, 27, 51]. By providing these metrics, maintainers can gain insights into the DEI levels of their projects and take proactive steps to foster a more inclusive environment. These tools contribute to the overall goal of promoting DEI within OSS communities.

Organizations, such as for-profit companies, often have more resources to implement a broader range of effective management strategies to improve their community's diversity and inclusion. Our study explores organizations' strategies for improving DEI and how they expand their OSS projects.

3 METHODS

To investigate the impact of organizational involvement on women's representation in OSS projects, we conducted interviews with project maintainers, using a semi-structured approach to gain detailed insights into maintenance practices, newcomer engagement, and diversity-supporting initiatives. This method, recommended for exploratory studies in qualitative research [11], allowed us to compare practices across organization-owned and community-led projects effectively. Through qualitative coding, we discerned the specific influence of organizational ownership on these practices.

3.1 Recruitment

In this study, we aimed to assemble a diverse sample of project maintainers, focusing on three key criteria: representation of women, project size, and ownership model. Our selection process was guided by the objective of understanding the impact of these variables on project dynamics. We specifically focused on the representation of women, rather than a broader concept of gender diversity, to address the well-documented underrepresentation of women in OSS projects, as highlighted in previous studies [57, 78]. This focus aligns with our research question which investigates the role of gender dynamics within project teams.

To identify projects for our sample, we utilized the GitHub API to access a list of npm project maintainers. We then categorized projects based on the presence of women among the top 20% most active contributors from Aug. 2020 - Aug. 2021. This approach was

Hana Frluckaj*, Huilian Sophie Qiu*, Bogdan Vasilescu, and Laura Dabbish

Table 1: Project and Maintainer Characteristics

ID	Total No. of Contributors	Project Owner Type	Maintainer Type	No. of Core Women
P1	16-50	Community	Owner	1
P2	16-50	Community	Internal Contr.	1
P3	51-100	Small Co.	Internal Contr.	0
P4	16-50	Start-up	Internal Contr.	0
P5*	1001+	Big Co.	External Contr.	0
P6	16-50	Big Co.	Internal Contr.	0
P7	16-50	Small Co.	Internal Contr.	0
P8*	51-100	Small Co.	Internal Contr.	0
P9	1-15	Start-up	Internal Contr.	2
P10	1-15	Government	Internal Contr.	1
P11	101-200	Big Co.	Internal Contr.	2
P12*	16-50	Government	Internal Contr.	2
P13*	201-1000	Big Co.	Internal Contr.	2
P14	51-100	Big Co.	Internal Contr.	0
P15	101-200	Community	Internal Contr.	1
P16*	101-200	Small Co.	Internal Contr.	1
P17*	51-100	Small Co.	Internal Contr.	0
P18*	201-1000	Big Co.	Internal Contr.	1
P19*	1-15	Small Co.	Internal Contr.	0
P20	101-200	Community	Owner	0
P21	101-200	Big Co.	Internal Contr.	1

PX: Interview participant. Star (*) symbol denotes follow-up interviews. **Small Co.**: Companies with less than 500 employees. **Big Co.**: Companies with more than 500 employees.

chosen as a pragmatic method to quantify women's representation, given the challenges in measuring other forms of diversity, such as racial or experience diversity, in a quantitative manner.For the determination of contributors' gender, we employed the NamSor web API,* which infers binary gender based on names. This method has been reported to have high accuracy [59, 61]. We only kept gender inference with confidence higher than 0.7 (out of 1) from Namsor [53]. We acknowledge the limitations of binary gender classification but selected this approach due to its feasibility and established precedent in similar research.

In summary, the selection criteria of women's representation, project size, and ownership model were chosen for their relevance to our research question and their feasibility of measurement. These criteria are intended to provide valuable insights into the dynamics of project management and team composition in OSS development.

In addition, we recruited across repositories with different ownership types. Specifically, we recruited maintainers from both community-led and organization-owned repositories. Organization-owned repositories included both company-owned and public institution-owned repositories. To make this distinction, we manually inspected the GitHub description of the organizational account and its linked external website when present. For company-owned repositories, we also recorded the company type and number of employees from inspecting the top few Google search results with the company name. We distinguished between small companies (less than 5,000 employees) and big companies (over 5,000 employees).

We reached out to maintainers on projects that fit our selection criteria via the contact method listed on their GitHub profile or

_____
*https://namsor.app

personal website (email, Twitter, or LinkedIn), asking if they would participate in a 45-60 minute discussion regarding their experience as a maintainer. Participants in our study were not financially compensated for their involvement.

In total, we reached out to 152 maintainers and received responses from 28 of them (18.4% response rate). However, we only interviewed 21 respondents due to scheduling conflicts, language barriers, or time constraints. To optimize our research process, we adopted a concurrent approach of coding interviews while conducting new ones, allowing us to efficiently process and analyze the data. We ceased sending out further invitations once we identified that we had reached theoretical saturation with no new themes arising, ensuring that our data collection was both comprehensive and focused. Despite our intent to include a more balanced gender representation in our study, particularly aiming to engage more women, we encountered a significant challenge due to the limited presence of women in maintenance roles. This resulted in the inclusion of only one woman in our participant pool. This outcome, while disappointing, underscores the ongoing issue of gender disparity in this field.

3.2 Overview of Participants

Among our 21 interviewees, 10 were drawn from projects without women, and 11 were drawn from projects with a woman as a core contributor (10 projects had cisgender women, 1 project had a transgender woman). Among the projects, 17 were managed by an umbrella organization while four were community led. and usually included organization employees among the maintainers: 7 were owned by big companies (P5, P6, P11, P13, P14, P18, and P21), 8 by small companies or start-ups (P3, P4, P7, P8, P9, P16, P17, and P19), and 2 by public institutions (P10 and P12). The other 4 projects (P1, P2, P15, and P20) were community-driven. Among these community-driven projects varied in terms of size (P1 and P2 were small projects; P15 and P20 were big) and women's representation (only P1, P2, and P15 had women). Although we aimed for a balance between projects with and without women core contributors, only one of the maintainers we interviewed identified as a (cis) woman (P10). See Table 1 for an overview of our participants.

3.3 Interview Protocol

There are five main sections in our interview protocol:

1. Background and Involvement: We started interviews by asking participants about their background, initial involvement with and general participation in OSS. Starting with questions about the participant's background and initial involvement in OSS helps establish a context for their perspectives on diversity and participation, particularly of women.

2. Experience with a Focal Project: We asked about their experiences with a specific focal project on which they were a maintainer, focusing on their contribution process management and diversity and inclusion, allowing for detailed insights into the project's operational dynamics.

3. Contribution Process Management: We asked maintainers how they managed incoming contributions and then asked them to walk us through a recently accepted and rejected code contribution. For each contribution, we asked about its purpose, what they knew

about the external contributor and how and why they decided to interact with the contributor. This is crucial for understanding how different types of ownership might influence project management and participation, including gender diversity. By examining how maintainers manage incoming contributions and make decisions about code contributions, we can gather data on if there are biases or systemic practices that impact women's participation.

4. DEI Practices: In order to understand DEI practices, we asked how the project attempted to attract new contributors, how new contributors joined the project, and how maintainers communicated and collaborated with new contributors. To ensure response accuracy, we asked maintainers to walk us through a recent interaction with a new/external contributor during a code review process or issue discussion.

We avoided direct questions about gender diversity to prevent socially desirable responses. We rather focused on behavior and practice-oriented questions, providing more genuine insights into the actual implementation of DEI principles within projects. This indirect approach is crucial for uncovering the true stance on diversity, beyond mere performative statements. Questions about attracting new contributors and collaborating with them are directly relevant to understanding the inclusivity of the project environment. This can highlight whether organizational ownership fosters more inclusive practices, which may encourage women's participation. We then asked about management strategies on the project, specifically inquiring about the provision of documentations [50] and a Code of Conduct (CoC) [37, 40], as well as the rationale behind these practices. Moreover, we asked whether the project had a mentorship program since these were found to be useful for newcomers by prior studies [4, 54].

5. Demographic Composition: Finally, we asked about their team's demographic composition. As we only used automatically inferred gender for sampling, we obtained the project's diversity by directly asking maintainers about the demographic dimensions of their contributors. This is a direct measure of women's participation and how it may be influenced by organizational ownership. Appendix A includes our interview protocol.

After our first round of interviews, our initial data analysis revealed the important influence of organizational involvement on project diversity. Therefore, we performed another round of follow-up interviews to dive deeper into organizational effects. We interviewed 8 maintainers of organization-owned projects from the previous round, including company-owned projects (P5, P8, P13, P16, P17, P18, and P19) and one public institution-owned project (P12) again. This round of interviews focused on how the involvement of the parent organization affected the way maintainers manage their projects. We asked maintainers how rules, policies, or guidelines from their organizations impacted their OSS project and about the DEI initiatives in their organizations. See Appendix B for the full follow-up interview protocol.

3.4 Data Analysis

Our analysis approach was rooted in grounded theory [71], which guided our data coding and categorization processes. We transcribed the interviews using Temi, a transcription API, and then hand-corrected them.

First, two authors performed *open coding* of the transcripts. For this initial coding process, we opted for "key-point coding" [3] rather than coding word-by-word to avoid making the coding task overwhelming. Two of the authors open-coded transcripts of the first 5 interviews and met to compare codes that emerged to those from the interviews. Then we refined the coding scheme, working through the transcripts, applying existing codes, and consolidating similar codes until theoretical saturation was reached and no new codes emerged from new interviews. Through this process, the authors obtained a set of 45 open codes in total.

Next, we performed *axial coding* to structure these open codes into broader conceptual categories, discerning relationships among them. This process involved revisiting the data, recoding with the emergent concepts, and conducting five rounds of coding. Two-author teams collaborated, discussing and resolving interpretation disagreements, refining our codebook until achieving a high level of inter-rater reliability (0.75) across three transcripts. Many of these codes were general OSS management strategies that were not particular differences caused by organizational involvement. Therefore, we kept only codes from 2 umbrella categories encompassing 7 subcategories that are relevant to our research question (Table 2).

Focusing on the codes relevant to organizational involvement, we used *selective coding* on the axial codes to discern contrasts between community-led and organization-led projects. This examination brought forth themes that highlighted the influence of organizational involvement on project communities.

To refine these themes, we employed a comparative analysis, juxtaposing responses from maintainers of both types of projects. This comparison enabled us to discern nuanced themes, revealing the distinct impact of organizational involvement on project dynamics. Themes were continuously reviewed and refined through discussions among researchers, ensuring that they were representative of the data and aligned with our research question.

4 RESULTS

We identified multiple ways a parent organization's policies and practices *influence the representation of women in their OSS projects?*, both positively and negatively. In this section, we present our results based on the two themes that emerged from our interviews: the tension between organizations and their external contributors; and their influence on the projects' DEI strategies.

4.1 Inhibitors of Community and Engagement

We found that organizations' operating procedures restricted participation by external contributors and the potential for a cohesive community around their OSS projects. For many organizations in our sample engagement with external contributors was limited, stringent quality requirements prevented accepting external contributions, and communication between organization contributors and the community was segmented.

4.1.1 Limited External Engagement. We observed limited engagement with external contributors in company-owned OSS projects. While these projects made their source code publicly available, they often lacked significant involvement from external contributors. In our sample, several organization-owned projects were primarily developed and maintained by the organization's employees where

Hana Frluckaj*, Huilian Sophie Qiu*, Bogdan Vasilescu, and Laura Dabbish

Table 2: Observed Themes Highlighting Differences Between Organization-Owned and Community-Owned OSS Projects

Category	Theme	Definition	Example
Inhibitors of Community and Engagement	Limited External Engagement	Organizational developers focused on and prioritized internal concerns and development priorities over external community concerns.	*"[the code owned by the organization] happens to be open source" (P6)*
	Strict Quality Requirements	Organizationally-defined contribution policies making the code contribution process onerous for external contributors.	*"[Company] requires external contributors to assign copyright...A bot makes sure they've signed a license agreement to assign copyright back to [Company]...There was some issue where this person's repository had commits with different email addresses, and it confused the bot..and rather than get it all fixed with a forced push, they closed the PR and started again." (P21)*
	Segmented Communication	Communication access is unequal between organization members and external contributors.	*"We have Slack, but that is for the internal team, it is not open for the [community]" (P3)*
Organization Ownership Enhancing Inclusion	Women Employees (as Hidden Figures)	Women employees in the organization became involved in the open source project and played crucial roles in project and community management.	*"And I put that into code of conduct, looks like [Woman Maintainer] updated it...Basically [Woman Maintainer] took up the pen and wrote [the CoC] up based on whatever research she did and then we passed it by legal." (P12)*
	Recruiting Members of Underrepresented Groups	Actively recruiting members of underrepresented groups and making concerted efforts to enhance diversity and inclusion within the broader OSS community.	*"As a team we've generally supported ...a volunteer sort of workshop scheme, that women can go and do free workshops in [programming language]". (P5)*
	Mentorship Availability	Organizations had internal formal mentorship programs, and offered mentorship for external contributors to their OSS projects through special programs and events.	*"This [Name] program is effectively a mentorship program...once we identify someone as being a keen collaborator, we effectively invite them to be more engaged with the team. They would be given...a point of contact. And that person will kind of act as a mentor...check in with them and see if they're happy with what they're working on and make sure that they understand how to go about doing that" (P5)*
	Code of Conduct	Internal organizational policies and practices regarding diversity and inclusion had a notable influence on the adoption of Codes of Conduct.	*[Code of Conduct is] "standard practice in the open-source community" (P11)*

"*[the code] happens to be open source*" (P6). The presence of the organization on platforms like GitHub was often viewed as "part of their branding and marketing" (P18). In certain projects that allow external contributions, non-employees were unable to join the core teams or access the internal repository due to security policies (P5).

Despite the limited external contributions, many projects in our sample made efforts to promote their projects to a wider audience through online platforms like Twitter and offline events such as Hackathons (P12, P16, and P18) and conferences (P4, P5, P7, P12, P16, P18, and P21). They also engaged in sponsorships (P5, P6, and P10) to raise awareness about their projects.

While some projects aimed to attract external contributions, maintainers explicitly prioritized internal resources over external contributions. Certain companies in our sample offered OSS bundled with support services, and provided dedicated channels for

paid users to report and address issues promptly (P6, P8). Paid users often expected that bug fixes and patches were the responsibility of company employees, as they were paying for the service and "*not a community thing*" (P8). Consequently, maintainers did not prioritize addressing issues raised by external users or contributors on platforms like GitHub (P8).

In some cases, access to OSS projects was restricted exclusively to internally-registered machines, hindering external contributors' ability to participate. As one interviewee expressed, they "*[had] more access browsing GitHub anonymously than [they] do log into [their] GitHub account on a non-enrolled device*" (P18). Additionally, the authority to merge code into the company's code base often lay solely with internal employees, excluding external contributors from essential processes (P5, P6). Thus OSS's historical ability to circumvent organizational pre-approval processes (e.g. passing

through an organizational hiring processes), which has enabled alternative involvement and involvement from under-represented contributors, is increasingly hindered.

4.1.2 Strict Quality Requirements. Organizationally-defined contribution policies made the code contribution process onerous for external contributors. Both companies and public institutions in our sample implemented complex code quality assurance processes triggered by requirements such as backward compatibility, security, integration with internal products, or adherence to specific standards (P5, P8, P12, P13, P16, and P17). These processes resulted in lengthy code contribution procedures, causing delays before a change could be merged. For instance, one interviewee acknowledged that, if not for working on a company-owned project, they "*would probably not have tested as thoroughly.*" While going over a PR they had accepted, P17 remarked that most PRs received are of a low quality: *"This [PR] is quite an ideal example, but maybe not a common one. Quite often we get lower quality issues... no, this is not very typical. This is better than the norm."*

The strict quality requirements imposed by organizational contribution policies and code quality assurance processes restricted the contribution process and in turn, limited external contributors' active involvement. This overhead to the contribution process may contribute additional barriers for underrepresented contributors, especially for women developers who experience higher confidence-competence gaps [81].

4.1.3 Segmented Communication. organization-owned OSS projects experienced segmented communication between internal and external channels. Communication access was unequal between organization members and external contributors; internal contributors have access to more information whereas external contributors are left out from most of the project discussions. While many of these projects utilized public communication platforms like Gitter, Discord, or Slack for external contributors, the core project team relied on internal company channels for key discussions (P5, P6, P7, P8, P9, P13, and P14). For instance, in one organization in our sample, a significant amount of project communication occurred "*just in passing or just in meetings, which not everyone was necessarily at*" due to employees being located in the same physical space, excluding external contributors (P5). Some projects used platforms like GitHub for external communication with the community around their project but resorted to private internal chats for employee discussions about substantive changes to the code and important design decisions (P6). Consequently, challenges arising from internal codebase issues or decisions were difficult to effectively communicate back to external contributors, who might have perceived gaps in communication as a lack of responsiveness (P5).

Although some organization-owned projects made efforts to keep conversations exclusively on GitHub, private discussions still occurred when additional in-depth discussions were needed (P13). Maintainers recognized that conflicts arising from segmented communication were nearly "*impossible to avoid this kind of conflict in projects like this*" (P5). This segmented communication created tension between internal maintainers and external contributors by contributing to information asymmetry and reduced transparency [83]. However, there were exceptions, such as a government-owned

project that strove for transparency by openly discussing and making changes visible (P10).

Conversely, community-led projects prioritized public and asynchronous communication channels like Gitter, ensuring transparency and inclusion for external contributors (P2, P15, P20). These projects aimed to avoid private back channels as they were seen as "*a problem in open-source projects*", believing that public conversations benefit everyone involved (P13, P20). Maintainers of community-led projects often moved specific implementation discussions to private chats to avoid cluttering public discussions without adding substantial value (P20).

In summary, communication was segmented in organization-owned OSS projects due to the use of both internal and external channels. This segmentation led to information asymmetry and lower transparency. In contrast, community-led projects prioritized public communication channels, striving for transparency and inclusion by avoiding private discussions whenever possible.

4.2 Organization Ownership Enhancing Inclusion

We observed several practices that transferred from organizations to their OSS projects and, as a result, improved the projects' DEI: women employees as hidden figures contributing to the project in ways that weren't traceable in public repositories, recruiting members of underrepresented groups, mentorship availability, and widespread use of codes of conduct.

4.2.1 Women Employees as Hidden Figures. In organization-owned projects, we observed a greater presence of women contributing compared to community-led projects in our sample. While these women employees may not appear prominently in trace data as top committers, they played crucial roles in project and community management (P4, P5, P7, P10, P12, and P17). For instance, participants highlighted the contributions of women employees in project management positions, even in cases where the majority of developers were male (P10). Additionally, specific roles such as a dedicated "community engineer" (P4) or "head of marketing and communication" (P7) were mentioned, highlighting the significance of these hidden figures in maintaining OSS projects. Women employees in our sample undertook important tasks such as unit testing and documentation, which were vital for newcomers interested in contributing to the project (P17). These roles were essential for project sustainability and success but tended to remain unrecognized and undervalued, aligning with previous research on hidden contributions in OSS projects [22, 75].

4.2.2 Recruiting Members of Underrepresented Groups. Several organizations in our sample actively recruited members of underrepresented groups and made concerted efforts to enhance DEI within the broader OSS community. For instance, P5 described sponsorship programs implemented by their organization, providing opportunities for individuals from underrepresented groups to attend conferences, receive mentoring, and more. Similarly, P6 discussed career development programs that focused on supporting "*women who are in their second career phase and decided to enter the technology field*", offering financial assistance and scholarships for their development programs.

In contrast, attracting a diverse team was not prioritized in the community-led projects in our sample. P15 expressed that prioritizing diversity would be more feasible if the project were larger, as the current size and attractiveness of the project did not warrant active outreach efforts (P15). P1 also mentioned the lack of time to prioritize DEI considerations. P20 openly acknowledged being unsure about improving diversity without imposing restrictions on potential contributors.

4.2.3 Mentorship Availability. A notable distinction emerged between the mentorship opportunities available in organi-zation-owned projects compared to community-led projects. Organizations in our sample had formal mentorship programs and resources, while community-led projects in our sample relied more on informal and *ad hoc* mentorship, if any was available at all.

In organization-owned projects, developers who were also employees of the organization had access to internal onboarding and mentorship programs. One participant described specific resources in their organization for onboarding and mentorship to help developers *"navigate the company"* (P6). These programs ranged from generic resources to personalized initiatives offering one-on-one mentorship at different tenure levels. In their OSS projects, there was a focus on making newcomers comfortable and confident, with one participant explaining that they review code in a private repository initially to *"make folks the most comfortable and to get them in and get them productive, get them confident"* (P6).

While many organizations in our sample had internal formal mentorship programs, they also occasionally offered mentorship for external contributors to their OSS projects through special programs and events. Some companies in our sample extended mentorship through events like hackathons (P18) and programs such as internships and Google's Summer of Code (P12, P15). Employed maintainers acted as guides, describing their role as a *"sort of mentor"* who *"checked in with them and see if they're happy with what they're working on and make sure that they understand how to go about doing that"* (P5). P12 considered their summer internship program as an informal mentorship opportunity, with company employees assisting the intern in coding and progress tracking.

Conversely, our community-led projects reported limited or no formal mentorship programs. Both P15 and P20 reported not having time for mentoring or office hours (P15), while P1 commented that their conversations were focused on problem-solving, without *"personal interactions."*. The lack of dedicated mentorship resources was attributed to the voluntary nature of maintaining community-led projects and the overwhelming workload for maintainers. However, some community-led projects in our sample mentioned participating in external programs as an approximation of mentorship. For example, one participant stated that they participate in Google Summer of Code as a way to provide mentorship to students" (P15).

4.2.4 Code of Conduct. In organization-owned OSS projects in our sample, internal organizational DEI policies and practices had a notable influence on the adoption of CoCs. Many projects explicitly adopted CoCs due to their affiliation with larger organizations, either by mandate or with the assistance of the organization's legal team (P3, P4, P6, P8, P11, P13, P18, and P21). Some maintainers mentioned that the CoC was introduced as part of a broader push by the parent organization (P18). In other cases, maintainers took

the initiative, but the organization's legal team provided guidance and approval for the CoC language (P6, P8, P12, P18, and P20).

However, we also observed varying levels of attention and familiarity with CoCs among organization-owned project maintainers. While many recognized their importance and considered them *"standard practice in the open-source community"* (P11), some maintainers admitted to limited knowledge of their CoC's content (P8) and did not edit the content after receiving it from their organization (P18). P8 referred to their CoC as *"kind of hidden"* and asserted that *"[only] some [of the project members] will have read it"*. P21 was an exception, and spoke passionately about proactively creating an inclusive and welcoming environment for any contributors through the presence and enforcement of a CoC: *"By having a CoC and having it be visible, we're making a commitment to anybody who might want to be involved that you're welcome and that we're not going to be horrible to you, and we're not going to let anyone else be horrible to you and continue to participate...I meant it sincerely when I added it. I hope it means something to someone who's reading it."*

In contrast, among the community-led projects we interviewed, only two had a CoC (P15 and P20) but were very engaged in its adoption and contents. P1 commented that he *"didn't think it was really important"* because *"it was a small project."* P15 added a CoC, well aware of the intention and the initiatives behind it and wanted to *"apply the same best practices to this open source projects"* and *"create an inclusive space for contributors."* P20 saw CoC in conversations in other repositories and created one for his project by copying from The Contributors Covenant.

5 DISCUSSION

We explored how organizational ownership influences the dynamics of external contributor involvement and women's representaion in OSS. In organizationally owned projects, maintainers had limited engagement with the external community, strict quality requirements for contributions that deterred external participation, and segmented communication on internal project decisions and discussions hidden from community view. Parallelly, organizational involvement enhanced DEI as women employees became engaged with OSS, organizations recruited underrepresented employees, provided mentorship opportunities, and included CoCs in the project.

Our results suggest that organizations need to think explicitly about how their internal practices affect the community of external contributors around a project. Organizationaoperating procedures can unintentionally block or restrict external contributors from becoming involved with a project community. As maintainer attention is focused internally, communications are hidden and quality requirements deter contribution. Organizations' DEI efforts in recruiting underrepresented groups, providing mentorship, and standardizing policies like codes of conduct can enhance DEI inside and outside the organization.

5.1 Organizational Ownership and External Contribution

While increasing corporate involvement in OSS has provided mutual benefits between companies and the OSS community at large [6], tensions may still arise on projects owned by companies where priority is given to organization requirements over the wants of

the external community. We observed that although the companies make their source code open to the public, their OSS projects often have few external contributors.

5.1.1 Organizational Barriers to Contribution. Since the commercialization of OSS projects began to gain traction, there have been concerns about company influence. OSS advocates voiced concerns that companies would dominate key technical and strategic decision-making processes to increase their control of projects and volunteers would increasingly feel limited and that they no longer have any stake in the project [79]. Moreover, since companies' goals are making profits, feature requests and directional choices made by the company are prioritized over requests made by the external community. As contributor involvement wanes, this has implications for the broader community. Oh and Jeon found that a developer leaving would influence their neighbors to follow suit; once enough key volunteers have left the community, a snowball effect is expected to occur and the project is abandoned [46]. Our findings connect with observations by Tiwani stating that if the architecture of a platform is at odds with decision rights, despite reducing coordination costs for developers, then developers are more likely to leave the platform [73].

External contributors have limited opportunities to communicate with organizational employees as they are not on the companies' internal communication channels. Lack of involvement in decision-making processes and relative neglect of external contributor interests is demotivating; Li *et al.* [38] found that lack of response from maintainers, not being treated seriously by them, and other tasks taking higher priority were significant factors in pull request desertion. Low levels of communication in virtual teams wear down trust over time [32, 34]. These effects become more salient in less structured environments [34], such as the decentralized environment of OSS, potentially further alienating the external community.

Instead of bridging the gap between community-led and company-led projects, the increasing presence of companies in OSS appears to sustain the separation of these two spheres. This can make company presence appear performative for some — more focused on branding rather than technological collaboration and progress.

5.1.2 Organizations as Open Source Stewards. Guizani *et al.* recognize mutual benefits of collaboration between companies and OSS projects but also place emphasis on company reciprocity [26]. As direct benefits to their OSS communities, companies have invested in DEI initiatives, focused on the geolocational diversity of their projects, and by participating in mentorship experiences. Mentorship is particularly crucial for encouraging underrepresented contributors [22, 52]. A recent survey in the Apache ecosystem found that respondents across all demographics sought out a mentor/parental figure [19], therefore companies are helping fill an important gap for OSS communities. Nevertheless, as Guizani *et al.* highlight, companies and organizations must also recognize their responsibility in supporting the sustainability and longevity of a project [26]. In paying employees for mentorship, companies can help to attract more volunteers and therefore provide smoother onboarding of newcomers to OSS projects [65]. Companies and organizations largely outrank OSS projects in terms of access to resources; as these organizations benefit from the labor and volunteerism of external contributors, they have a social obligation to

contribute back. Additionally, they should encourage efforts to host OSS projects within their companies. P21 noted how much more difficult it had become for a suggested OSS project to be approved by the company since he was an intern, e.g. needing approval from multiple managers at different levels now.

5.1.3 Organizational Policies and Inclusion. Our results suggest that project owners and organizations may unintentionally restrict their external community and the diversity of contributors. We found that projects impose barriers to contribution as a function of carrying out their organizational routines and complying with regulations and policies required in their domain. This could be disproportionately harmful to under-represented contributors; women developers tend to have higher confidence-competence gaps [81] and the addition of more rigorous organizational quality checks may deter their sustained involvement. Future work should examine whether and how to mitigate some of the tensions around organizational operation and inclusion in OSS.

Our results suggest organizations may wield substantial influence over OSS projects through their human resources (HR) policies and practices. In the projects examined, hiring and recruiting, DEI initiatives, and mentorship programs directly increased diversity. Communication practices and contribution policies set at the organizational level can significantly impact project direction and the dynamics within the contributing community. Future research should examine the extent of organizational influence on OSS project culture, particularly how HR policies shape project interactions and whether and how organizations can directly and intentionally influence or control community nature and evolution.

5.2 Hybrid Ownership Models for Open Source

Hybrid ownership in OSS projects can manifest in various forms, ranging from solely community-driven initiatives to involvement from multiple organizations. Understanding the effectiveness of these models and their impact on project sustainability is crucial, as they can significantly impact governance, decision-making processes, and community involvement. Future research in this area should explore the spectrum of hybrid ownership structures to comprehend their implications and impact on project dynamics.

5.2.1 Representing Community Interests. Finding effective ways to represent the interests of external communities within company-owned OSS projects is an important area for future work. Governance choices by platform owners and the dynamics of their exogenous environment influences their evolutionary dynamics [73] and effective governance choices help improve the data sustainability of projects [33]. Establishing governance models that balance the needs of both the owning organization and external contributors is essential for sustained project growth and community engagement. Research should focus on identifying governance models that foster inclusivity and participation while maintaining project coherence. Defining whose interests should take priority (company vs community) and what constitutes equitable representation within these hybrid models is an important challenge. We need research that examines the trade-offs between different representation models and their impact on community involvement.

5.2.2 Transparency and Distributed Collaboration. Distributed virtual teams have struggled with collaboration challenges for decades. Virtual teams often have to contend with time zone differences, lean online communication channels, and different native languages and proficiency levels [35]. Language differences amongst team members can lead to social categorization and biases, which can severely damage team dynamics and accentuate communication difficulties [36]. As mentioned before, effective communication is critical for building relationships and trust in virtual teams [34].

Challenges associated with virtual teams are aggravated for external OSS contributors. Organizations exert uniformity in the tools, applications, processes, etc. to which external contributors aren't privy. Furthermore, as seen amongst our projects, communication channels are usually segmented between those within the organization and those outside of it. External contributors are excluded from "secure" internal conversations and as a result, have diminished voice in decision-making processes and are several steps behind.

We have yet to fully solve the issues associated with globally distributed software development across multiple sites. As hybrid models of work proliferate, it is imperative to reduce the effect of these issues on external contributors. More open, transparent, and readily accessible communication between organizations and their external communities could help mitigate the debilitating effects on OSS contributors. As a result, building up communication and trust will also pave the way for stronger connections and more sustainable project communities [32]. Future research can explore initiatives to reduce the inequities of participating in virtual software development, with particular emphasis on supporting involvement from underrepresented contributors.

5.2.3 Accountability to Community. The question of accountability within emerging hybrid governance models is an important area of future research. Future work should consider whom these projects should be accountable to, be it specific organizations, community members, or other projects, and what accountability looks like. In addition, work is needed to examine different mechanisms of accountability ranging from transparency to financial accountability for project impacts on external contributors and users.

5.2.4 Alternative Incentives for External Participation. Given that company-owned projects depend on external contributors, exploring alternative models of support and recognition for community members is vital. Research is needed to develop innovative forms of incentives and compensation that acknowledge and enhance the careers of contributors in unique ways. These kinds of incentives could boost community engagement and sustainability of OSS projects. Understanding how to reward contributions effectively could support community health around a project which is pivotal for long-term project success. Future work could also explore how organizations can leverage their resources to make volunteering through OSS more sustainable for contributors. Special attention is needed on supporting underrepresented contributors who may come from lower socioeconomic backgrounds.

5.3 DEI Initiatives in OSS

Despite increased academic research attention to DEI in OSS, participation by women and underrepresented groups remains low (e.g.

[70, 75, 76, 78]. This gap is attributed to various barriers, including socio-cultural factors, implicit biases, and systemic challenges within the OSS environment [68, 69]. Previous work highlights strategies to promote inclusivity including mentorship programs [65], outreach initiatives, and the establishment of codes of conduct to ensure a welcoming atmosphere for contributors from all backgrounds [52, 63, 70]. Many of the DEI strategies cited by previous work were employed by organizationally owned projects in our sample. There is little to no empirical evidence, however, on the effectiveness of these different DEI initiatives particularly under different project ownership models. Much of the guidance on DEI in OSS assumes community ownership or is directed at a single source while this is less often the case as corporate involvement increases. Future research should examine how DEI initiatives and programs can be designed to leverage organizational resources and connections and bridge the public/private divide.

Recent work on organizational involvement in OSS validates our findings that company-led DEI initiatives can have spillover effects into their OSS project communities [26]. Guizani *et al.* observed that companies-sponsored contributor events, summits, and geolocation-specific contributions can enhance project-level diversity.

Organizations have a vested interest to encourage external participation to support the sustainability and diversity of their projects [39]; by not engaging deeply with external contributors, companies may miss out on hiring new talent and may discourage users from getting more involved in projects. Company-owned projects prioritizing contributors from employees over external developers could spell trouble for the many developers who rely on GitHub to sharpen their skill set, build up their network and connections, and find and receive job opportunities[41]. OSS provides an alternative career path with particular significance to those coming from underrepresented backgrounds.

6 CONCLUSION

Our study explored how organizational policies and practices impact women's participation in open-source projects. Through interviews with maintainers in organizationally owned and community-owned projects, we identified ways that organizational ownership influences engagement with external contributors. We found that internal organizational practices and policies can shape the level and nature of community engagement and influence project contributor diversity. We consider how organizations can leverage internal resources to cultivate a more diverse and inclusive open-source community. By addressing barriers to external contributors, building external communities, promoting mentorship programs, and fostering open discussions on community-oriented efforts, organizations can contribute to a more inclusive open-source ecosystem. Further research on open-source governance models and the implementation of these strategies will be valuable in promoting a more inclusive and accessible open-source community for all.

7 ACKNOWLEDGEMENTS

REFERENCES

[1] Mark S Ackerman, Juri Dachtera, Volkmar Pipek, and Volker Wulf. 2013. Sharing knowledge and expertise: The CSCW view of knowledge management. *Computer Supported Cooperative Work (CSCW)* 22 (2013), 531–573.

[2] Adam Alami and Andrzej Wąsowski. 2019. Affiliated participation in open source communities. In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. IEEE, 1–11.

[3] George Allan. 2003. A critique of using grounded theory as a research method. *Electronic journal of business research methods* 2, 1 (2003), pp1–10.

[4] Andrew Begel and Beth Simon. 2008. Novice software developers, all over again. In *Proceedings of the fourth international workshop on computing education research*. 3–14.

[5] Evangelia Berdou. 2006. Insiders and outsiders: paid contributors and the dynamics of cooperation in community led F/OS projects. In *IFIP International Conference on Open Source Systems*. Springer, 201–208.

[6] Benjamin Birkinbine. 2020. *Incorporating the digital commons: corporate involvement in free and open source software.* University of Westminster Press.

[7] Simon Butler, Jonas Gamalielsson, Björn Lundell, Per Jonsson, Johan Sjöberg, Anders Mattsson, Niklas Rickö, Tomas Gustavsson, Jonas Feist, Stefan Landemoo, et al. 2018. An investigation of work practices used by companies making contributions to established OSS projects. In *International Conference on Software Engineering: Software Engineering in Practice*. 201–210.

[8] Andrea Capiluppi, Klaas-Jan Stol, and Cornelia Boldyreff. 2012. Exploring the role of commercial stakeholders in open source software evolution. In *IFIP International Conference on Open Source Systems*. Springer, 178–200.

[9] Eugenio Capra, Chiara Francalanci, Francesco Merlo, and Cristina Rossi-Lamastra. 2011. Firms' involvement in Open Source projects: A trade-off between software structural quality and popularity. *Journal of Systems and Software* 84, 1 (2011), 144–161.

[10] Gemma Catolino, Fabio Palomba, Damian A Tamburri, Alexander Serebrenik, and Filomena Ferrucci. 2019. Gender diversity and women in software teams: How do they affect community smells?. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. IEEE, 11–20.

[11] John W Creswell and J David Creswell. 2017. *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage publications.

[12] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. 2012. Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on computer supported cooperative work (CSCW)*. 1277–1286.

[13] Barthélémy Dagenais, Harold Ossher, Rachel KE Bellamy, Martin P Robillard, and Jacqueline P De Vries. 2010. Moving into a new software project landscape. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering-Volume 1*. 275–284.

[14] Luis Felipe Dias, Igor Steinmacher, and Gustavo Pinto. 2018. Who drives company-owned OSS projects: internal or external members? *Journal of the Brazilian Computer Society* 24, 1 (2018), 1–17.

[15] Nadia Eghbal. 2016. *Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure.* Ford Foundation.

[16] Robert English and Charles M Schweik. 2007. Identifying success and tragedy of FLOSS commons: A preliminary classification of Sourceforge. net projects. In *First International Workshop on Emerging Trends in FLOSS Research and Development (FLOSS'07: ICSE Workshops 2007)*. IEEE, 11–11.

[17] Fabian Fagerholm, Alejandro Sanchez Guinea, Jürgen Münch, and Jay Borenstein. 2014. The Role of Mentoring and Project Characteristics for Onboarding in Open Source Software Projects. In *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. ACM, International. https://doi.org/10.1145/2652524.2652540 Volume: Proceeding volume: ; Unknown host publication ; Conference date: 01-01-1800.

[18] Zixuan Feng, Amreeta Chatterjee, Anita Sarma, and Iftekhar Ahmed. 2022. Implicit Mentoring: The Unacknowledged Developer Efforts in Open Source. *arXiv preprint arXiv:2202.11300* (2022).

[19] Zixuan Feng, Mariam Guizani, Marco A Gerosa, and Anita Sarma. 2023. The State of Diversity and Inclusion in Apache: A Pulse Check. *arXiv preprint arXiv:2303.16344* (2023).

[20] Karl Fogel. 2005. *Producing open source software: How to run a successful free software project.* " O'Reilly Media, Inc.".

[21] Darren Forrest, Carlos Jensen, Nitin Mohan, and Jennifer Davidson. 2012. Exploring the role of outside organizations in Free/Open Source Software projects. In *IFIP International Conference on Open Source Systems*. Springer, 201–215.

[22] Hana Frluckaj, Laura Dabbish, David Gray Widder, Huilian Sophie Qiu, and James D Herbsleb. 2022. Gender and participation in open source software development. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–31.

[23] GitHub. 2021. The State of the Octoverse: The State of the Octoverse explores a year of change with new deep dives into writing code faster, creating documentation and how we build sustainable communities on GitHub. https://octoverse.github.com/.

[24] Sean P Goggins, Matt Germonprez, and Kevin Lumbard. 2021. Making open source project health transparent. *Computer* 54, 8 (2021), 104–111.

[25] Ron Goldman and Richard P Gabriel. 2005. *Innovation happens elsewhere: Open source as business strategy.* Morgan Kaufmann.

[26] Mariam Guizani, Aileen Abril Castro-Guzman, Anita Sarma, and Igor Steinmacher. 2023. Rules of Engagement: Why and How Companies Participate in OSS. *arXiv preprint arXiv:2303.08266* (2023).

[27] Mariam Guizani, Thomas Zimmermann, Anita Sarma, and Denae Ford. 2022. Attracting and retaining oss contributors with a maintainer dashboard. In *Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society*. 36–40.

[28] Carl Gutwin, Reagan Penner, and Kevin Schneider. 2004. Group awareness in distributed software development. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW)*. 72–81.

[29] Carl Gutwin, Kevin Schneider, David Paquette, and Reagan Penner. 2005. Supporting group awareness in distributed software development. In *Engineering Human Computer Interaction and Interactive Systems: Joint Working Conferences EHCI-DSVIS 2004, Hamburg, Germany, July 11-13, 2004, Revised Selected Papers*. Springer, 383–397.

[30] Nikolay Harutyunyan and Dirk Riehle. 2019. Getting started with open source governance and compliance in companies. In *Proceedings of the 15th International Symposium on Open Collaboration*. 1–10.

[31] Joachim Henkel. 2006. Selective revealing in open innovation processes: The case of embedded Linux. *Research policy* 35, 7 (2006), 953–969.

[32] Sirkka L Jarvenpaa, Celeste Cantu, and Shi Ying Lim. 2017. Trust in virtual online environments. *The Wiley Blackwell handbook of the psychology of the internet at work* (2017), 103–130.

[33] Sirkka L Jarvenpaa and Anna Essén. 2023. Data sustainability: data governance in data infrastructures across technological and human generations. *Information and Organization* 33, 1 (2023), 100449.

[34] Sirkka L Jarvenpaa, Thomas R Shaw, and D Sandy Staples. 2004. Toward contextualized theories of trust: The role of trust in global virtual teams. *Information systems research* 15, 3 (2004), 250–267.

[35] Alfredo Jimenez, Dirk M Boehe, Vasyl Taras, and Dan V Caprar. 2017. Working across boundaries: Current and future perspectives on global virtual teams. *Journal of International Management* 23, 4 (2017), 341–349.

[36] Anders Klitmøller, Susan Carol Schneider, and Karsten Jonsen. 2015. Speaking of global virtual teams: Language differences, social categorization and media choice. *Personnel Review* 44, 2 (2015), 270–285.

[37] Renee Li, Pavitthra Pandurangan, Hana Frluckaj, and Laura Dabbish. 2021. Code of Conduct Conversations in Open Source Software Projects on Github. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–31.

[38] Zhixing Li, Yue Yu, Tao Wang, Gang Yin, Shanshan Li, and Huaimin Wang. 2021. Are you still working on this? An empirical study on pull request abandonment. *IEEE Transactions on Software Engineering* 48, 6 (2021), 2173–2188.

[39] Zhifang Liao, Libing Deng, Xiaoping Fan, Yan Zhang, Hui Liu, Xiaofei Qi, and Yun Zhou. 2018. Empirical research on the evaluation model and method of sustainability of the open source ecosystem. *Symmetry* 10, 12 (2018), 747.

[40] David López Jiménez, Eduardo Carlos Dittmar, and Jenny Patricia Vargas Portillo. 2021. New directions in corporate social responsibility and ethics: codes of conduct in the digital environment. *Journal of Business Ethics* (2021), 1–11.

[41] Jennifer Marlow and Laura Dabbish. 2013. Activity traces and signals in software developer recruitment and hiring. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW)*. 145–156.

[42] David W McDonald. 2003. Recommending collaboration with social networks: a comparative evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 593–600.

[43] Audris Mockus and James D Herbsleb. 2002. Expertise browser: a quantitative approach to identifying expertise. In *Proceedings of the 24th international conference on software engineering*. 503–512.

[44] Anh Nguyen Duc, Daniela S Cruzes, Claudia Ayala, and Reidar Conradi. 2011. Impact of stakeholder type and collaboration on issue resolution time in oss projects. In *IFIP International Conference on Open Source Systems*. Springer, 1–16.

[45] Alexander Nolte, Linda Bailey Hayden, and James D Herbsleb. 2020. How to support newcomers in scientific hackathons-an action research study on expert mentoring. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–23.

[46] Wonseok Oh and Sangyong Jeon. 2007. Membership herding and network stability in the open source community: The Ising perspective. *Management science* 53, 7 (2007), 1086–1101.

[47] Gustavo Pinto, Igor Steinmacher, Luiz Felipe Dias, and Marco Gerosa. 2018. On the challenges of open-sourcing proprietary software projects. *Empirical Software Engineering* 23, 6 (2018), 3221–3247.

[48] Gede Artha Azriadi Prana, Denae Ford, Ayushi Rastogi, David Lo, Rahul Purandare, and Nachiappan Nagappan. 2021. Including everyone, everywhere: Understanding opportunities and challenges of geographic gender-inclusion in oss. *IEEE Transactions on Software Engineering* (2021).

[49] Putu Yudia Pratiwi, Ridi Ferdiana, and Rudy Hartanto. 2018. An analysis of the new employee onboarding process in startup. In *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE, 603–608.

[50] Huilian Sophie Qiu, Yucen Lily Li, Susmita Padala, Anita Sarma, and Bogdan Vasilescu. 2019. The signals that potential contributors look for when choosing open-source projects. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–29.

[51] Huilian Sophie Qiu, Anna Lieb, Jennifer Chou, Megan Carneal, Jasmine Mok, Emily Amspoker, Bogdan Vasilescu, and Laura Dabbish. 2023. Climate Coach: A Dashboard for Open-Source Maintainers to Overview Community Dynamics. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[52] Huilian Sophie Qiu, Alexander Nolte, Anita Brown, Alexander Serebrenik, and Bogdan Vasilescu. 2019. Going farther together: The impact of social capital on sustained participation in open source. In *International Conference on Software Engineering (ICSE)*. IEEE, 688–699.

[53] Huilian Sophie Qiu, Zihe H Zhao, Justin Wang, Tielin Katy Yu, Alexander Ma, Hongbo Fang, Laura Dabbish, and Bogdan Vasilescu. 2023. Gender Representation Among Contributors to Open-Source Infrastructure - An Analysis of 20 Package Manager Ecosystems. In *International Conference on Software Engineering, Software Engineering in Society (ICSE SEIS)*. ACM.

[54] Ayushi Rastogi, Suresh Thummalapenta, Thomas Zimmermann, Nachiappan Nagappan, and Jacek Czerwonka. 2015. Ramp-up journey of new hires: Tug of war of aids and impediments. In *2015 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. IEEE, 1–10.

[55] Dirk Riehle. 2007. The economic motivation of open source software: Stakeholder perspectives. *Computer* 40, 04 (2007), 25–32.

[56] Gregorio Robles, Laura Arjona Reina, Jesús M González-Barahona, and Santiago Dueñas Domínguez. 2016. Women in free/libre/open source software: The situation in the 2010s. In *IFIP International Conference on Open Source Systems*. Springer, 163–173.

[57] Davide Rossi and Stefano Zacchiroli. 2022. Worldwide Gender Differences in Public Code Contributions. In *International Conference on Software Engineering, Software Engineering in Society (ICSE-SEIS)*. ACM.

[58] Davide Rossi and Stefano Zacchiroli. 2022. Worldwide Gender Differences in Public Code Contributions: and how they have been affected by the COVID-19 pandemic. *Proceedings of the 44th International Conference on Software Engineering (ICSE 2022) - Software Engineering in Society (SEIS) Track* (2022).

[59] Lucía Santamaría and Helena Mihaljević. 2018. Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science* 4 (2018), e156.

[60] Meera Sarma and Thomas Matheus. 2015. 'Hybrid'open source software virtual communities of practice–a conceptual framework. *Technology Analysis & Strategic Management* 27, 5 (2015), 569–585.

[61] Paul Sebo. 2021. Performance of gender detection tools: a comparative study of name-to-gender inference services. *Journal of the Medical Library Association: JMLA* 109, 3 (2021), 414.

[62] Shahrokh Shahrivar, Shaban Elahi, Alireza Hassanzadeh, and Gholamali Montazer. 2018. A business model for commercial open source software: A systematic literature review. *Information and Software Technology* 103 (2018), 202–214.

[63] Dan Sholler, Igor Steinmacher, Denae Ford, Mara Averick, Mike Hoye, and Greg Wilson. 2019. Ten simple rules for helping newcomers become contributors to open projects. *PLOS Computational Biology* 15, 9 (2019), e1007296.

[64] Vandana Singh and William Brandon. 2019. Open source software community inclusion initiatives to support women participation. In *IFIP International Conference on Open Source Systems*. Springer, 68–79.

[65] Igor Steinmacher, Sogol Balali, Bianca Trinkenreich, Mariam Guizani, Daniel Izquierdo-Cortazar, Griselda G Cuevas Zambrano, Marco Aurelio Gerosa, and Anita Sarma. 2021. Being a Mentor in open source projects. *Journal of Internet Services and Applications* 12, 1 (2021), 1–33.

[66] Igor Steinmacher, Ana Paula Chaves, and Marco Aurélio Gerosa. 2013. Awareness support in distributed software development: A systematic review and mapping of the literature. *Computer Supported Cooperative Work (CSCW)* 22 (2013), 113–158.

[67] Igor Steinmacher, Tayana Conte, Marco Aurélio Gerosa, and David Redmiles. 2015. Social barriers faced by newcomers placing their first contribution in open source software projects. In *Proceedings of the 18th ACM conference on Computer supported cooperative work & social computing (CSCW)*. 1379–1392.

[68] Igor Steinmacher, Marco Gerosa, Tayana U Conte, and David F Redmiles. 2019. Overcoming social barriers when contributing to open source software projects. *Computer Supported Cooperative Work (CSCW)* 28, 1 (2019), 247–290.

[69] Igor Steinmacher, Marco Aurélio Gerosa, and David Redmiles. 2014. Attracting, onboarding, and retaining newcomer developers in open source software projects. In *Workshop on Global Software Development in a CSCW Perspective*.

[70] Igor Steinmacher, Gustavo Pinto, Igor Scaliante Wiese, and Marco Aurélio Gerosa. 2018. Almost there: A study on quasi-contributors in open-source software projects. In *2018 IEEE/ACM 40th International Conference on Software Engineering (ICSE)*. IEEE, 256–266.

[71] Anselm Strauss and Juliet M Corbin. 1997. *Grounded theory in practice.* Sage.

[72] Matthias Stuermer, Sebastian Spaeth, and Georg Von Krogh. 2009. Extending private-collective innovation: a case study. *R&d Management* 39, 2 (2009), 170–191.

[73] Amrit Tiwana. 2015. Platform desertion by app developers. *Journal of Management Information Systems* 32, 4 (2015), 40–77.

[74] Parastou Tourani, Bram Adams, and Alexander Serebrenik. 2017. Code of conduct in open source projects. In *2017 IEEE 24th international conference on software analysis, evolution and reengineering (SANER)*. IEEE, 24–33.

[75] Bianca Trinkenreich, Mariam Guizani, Igor Wiese, Anita Sarma, and Igor Steinmacher. 2020. Hidden figures: Roles and pathways of successful oss contributors. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–22.

[76] Bianca Trinkenreich, Igor Wiese, Anita Sarma, Marco Gerosa, and Igor Steinmacher. 2021. Women's Participation in Open Source Software: A Survey of the Literature. *arXiv preprint arXiv:2105.08777* (2021).

[77] Asher Trockman, Shurui Zhou, Christian Kästner, and Bogdan Vasilescu. 2018. Adding Sparkle to Social Coding: An Empirical Study of Repository Badges in the npm Ecosystem. In *International Conference on Software Engineering (ICSE)*. ACM, 511–522. https://doi.org/10.1145/3180155.3180209

[78] Bogdan Vasilescu, Daryl Posnett, Baishakhi Ray, Mark GJ van den Brand, Alexander Serebrenik, Premkumar Devanbu, and Vladimir Filkov. 2015. Gender and tenure diversity in GitHub teams. In *ACM Conference on Human Factors in Computing (CHI)*. 3789–3798.

[79] SJ Vaughan-Nichols. 2005. Commercializing open-source stirs debate. eWeek (August 4).

[80] Claudio Vitari and Aurelio Ravarini. 2009. A longitudinal analysis of trajectory changes in the software industry: the case of the content management application segment. *European journal of information systems* 18, 3 (2009), 249–263.

[81] Zhendong Wang, Yi Wang, and David Redmiles. 2018. Competence-confidence gap: A threat to female developers' contribution on github. In *Proceedings of the 40th International Conference on Software Engineering: Software Engineering in Society*. 81–90.

[82] Joel West and Siobhan O'Mahony. 2005. Contrasting community building in sponsored and community founded open source projects. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. IEEE, 196c–196c.

[83] Joel West and Siobhán O'Mahony. 2008. The role of participation architecture in growing sponsored open source communities. *Industry and innovation* 15, 2 (2008), 145–168.

A FULL INTERVIEW PROTOCOL

Hello, thank you for taking the time to talk with us today.

- We are doing a research study on <how people use GitHub to manage and maintain open source projects>.
- The researchers working on this project, including myself, work for ><University Name>.
- The interview takes around 45 minutes to an hour.
- Participation is not compensated.
- With your permission, we will also collect public data from contributions and data that are brought up during the interview.
- Everything will be anonymous and confidential. No one will be identified by name or any other specific characteristic.
- There are no "right" or "wrong" answers
- But please refrain from discussing sensitive information about yourself or third parties that would put them at risk for civil or criminal liability or damage to their financial standing, employability, or reputation.
- We really appreciate your participation
- We'd like to record the audio of this interview and your screen, primarily for internal note taking purposes, meaning only the members of our research team will have access to these recordings. Is that OK with you?
- You can let us know to stop the tape during the interview if you say anything you would like removed from the record.

A.1 Maintainers' Background

(1) Tell me a bit about who you are and what you do.

A.2 Project Background

(2) What project have you been most involved with in the past year?
(3) What is your role on Project X?

A.3 Contribution Management Process

A.3.1 Making Decisions.

(4) Are you the sole maintainer or are there others involved?
(5) How is the project managed?
(6) How do decisions get made?
  (a) Can you tell me about a recent decision you were involved in?

A.3.2 Community.

(7) How do people on Project X interact with newcomers?
  (a) Can you think of an example of a successful interaction?
  (b) Can you think of an example of an unsuccessful interaction?
(8) What are the communication channels within the project?
  (a) Is there a slack channel? Discord? Mailing list?
(9) What kinds of community events does your project have?
  (a) Who goes to them?
(10) Can you tell me about community interactions within your project, specifically with newcomers?
(11) How has COVID-19 affected your project?

A.4 Diversity and Inclusion strategies

A.4.1 Attracting New Members.

(12) Does your project engage in any specific activities to attract new members?
  (a) How important is it for your project to attract new members?

A.4.2 Joining Process.

(13) How do people join Project X?
  (a) Can you tell me about a recent example of someone joining and how they got to where they are now?

A.4.3 New Contributions.

(14) How does your project handle new contributions?
(15) Can you walk me through a recent pull request (PR) from a new contributor and your thought process while interacting with it?
  (a) What was the PR? What were they trying to accomplish?
  (b) What do you know about this person?
  (c) What happened after they submitted the PR?

(d) Did you communicate with this person about their commit? Why/why not?
(e) Did you interact with any of the comments? If not, can you show me an example that had some discussion?
(f) Is this a typical example of how PRs get handled?
(16) Accepted/Rejected PR
(a) (If previous example was an accepted PR) Can you show me an example from a new contributor where the PR was rejected?
(b) (If previous example was a rejected PR) Can you show me an example from a new contributor where the PR was accepted?
(17) Have you had any surprising or interesting PRs from new contributors?

A.4.4   Mentorship.
(18) Does your project have a mentorship program?
(19) Have you mentored any new contributors? Can you give an example?
(20) Does your project engage in any specific activities to retain new members?
(21) Are there any examples of someone joining and then leaving the project?
(a) (If so) Why did this happen? Has this happened before?

A.4.5   Documentation.
(22) Given documentation status on Project X
(a) (If visible documentation exists) I see Project X has a README, do you have any other documentation elsewhere?
    (i) How was Project X's documentation created?
(b) (If no visible documentation) What kind of documentation does Project X have?
(23) How do decisions get made regarding documentation?

A.4.6   Code of Conduct.
(24) Given Code of Conduct status on Project X
(a) (If visible Code of Conduct exists) I see Project X has a Code of Conduct, why did your project decide to include one?
    (i) Can you walk me through how it got created?
    (ii) How did you decide on its contents?
    (iii) How is it enforced?
(b) (If no visible Code of Conduct) Does Project X have a Code of Conduct? Why/why not?

A.5   Team Diversity

(25) How diverse is your project?
(a) In terms of gender?
(b) In terms of ethnicity of member backgrounds?
(26) Does your project engage in any specific activities to enhance diversity or inclusion?

A.6   Final Words

(27) Is there anything else you think we should know about your involvement in Project X?

B   FOLLOW-UP INTERVIEW PROTOCOL

Hello again, we spoke last year about your role as a maintainer on Project X. I have a couple of follow up questions from our discussion last year that I hope you can answer.

(1) Can you think about a time when you did something in a certain way as the maintainer of Project X because of rules, policy, or guidelines from <company/organization>?

(2) Can you think about a time when you had to undo or rework the way you were doing something as a maintainer because of rules, policy, or guidelines from <company> policies and if so, can you briefly explain what you changed and why?

(3) Can you think of any other company policies or guidelines that affected your experience as a maintainer of Project X? And if so, how?

(4) Have there been any events or incidents within your company that have impacted your project?

(5) Does your company or organization have any internal diversity and inclusion initiatives or activities? If so, what are they and how have they impacted your project?

(6) What are some initiatives that you're aware of in other company or organization-owned open source projects that have been effective in improving their involvement in open source?

Those are all my questions. Thank you so much for answering. Is there anything else you'd like to add?