

# OpenStack Gender Diversity Report

Daniel Izquierdo, Nicole Huesman, Alexander Serebrenik, and Gregorio Robles

*// As do many other open source communities, the OpenStack community reflected a lack of representation of females and minorities. Based on the lack of data with this respect, we conducted the comprehensive analysis of gender diversity within OpenStack, opening the possibility of extending insights to other communities. //*



**WHEN IT COMES** to gender, the field of software engineering is heavily skewed toward men; multiple studies show that the gender situation in the open source arena is even more lopsided.<sup>1,2</sup> This is bad for the industry because diversity, in all of its forms, is essential in open source communities and in the larger technology industry. Besides the ethical arguments that favor diversity, recent studies show that gender diversity promotes balance between the traits of feminine and

masculine roles,<sup>3</sup> makes teams more productive,<sup>2</sup> and benefits both industrial and academic teams.<sup>4</sup> Moreover, a more diverse development team is more likely to understand and represent the users' needs, contributing to the better alignment between software and its intended customers.<sup>5</sup>

Not surprisingly, multiple open source projects have tried to increase the diversity of their communities, specifically gender diversity, as evident in the formation of many "women in" groups established in Linux, Gnome, Ubuntu, and Debian. Despite these efforts, the situation

has not changed much in the last 15 years, resulting in what has been called a *lost decade* in the integration of women in the open source field.<sup>6</sup> Understanding the success or failure of these and similar activities starts with measuring the participation of females—in terms of both population and level of activity—within a specific open source project. Below we report on our experience with measuring gender diversity.

The proposed analysis and methodology have been developed to increase the understanding of diversity and inclusion across the open source ecosystem as a whole. In this case, we focus on OpenStack, which is a well-known open source project involving thousands of contributors, many of whom are employed by companies.

## Why Focus on OpenStack?

The OpenStack community has long recognized the importance of gender diversity. This is reflected in ongoing discussions and activities hosted during OpenStack Summits and in the creation of the Women of OpenStack Working Group, which aims to "increase the diversity of the OpenStack community by overcoming OpenStack's barrier to entry through educational sessions, professional networking, mentorship, social inclusion, and enhanced resource access." Measuring the effectiveness of activities aimed to increase gender diversity requires repeated measurement and examination of the results of these activities before (baseline), during, and after their implementation. Moreover, understanding how people interact and perform is essential to ensure transparency and fairness.<sup>7</sup> However, the OpenStack Foundation did not have baseline information about the involvement

of women. Indeed, obtaining such baseline information is not easy for open source foundations; unlike traditional companies, open source foundations do not typically have a human resources division with detailed records of participants, their activities, and their gender.

Hence, Intel and Bitergia, the latter a software analytics company, conducted research to produce these initial data sets related to participation of different genders and then shared the results with the community in an effort to create broader awareness of the current state of diversity and, in turn, narrow the gender gap. The results of this research, communicated during OpenStack Summits in 2016, 2017, and 2018, have been welcomed by the community. Subsequent diversity reports have been prepared after incorporating feedback to address the needs and expectations of the community. These reports measure participation by gender across both code and non-code contributions. While noncode contributions do not directly affect the OpenStack project's code base, they are considered essential to the organization and overall well-being of the project. Noncode contributions included leadership, governance, event representation, and event participation, among other factors.

## Challenges of Measuring Gender Diversity

Measuring gender diversity based on digital traces, such as logs of developers' communication and collaboration or records of their attendance and participation in events, involves multiple challenges. First, gender is a complex social construct; since digital trace data rarely record gender explicitly, most studies reduce gender to binary and employ heuristics to determine whether the individual is a woman or

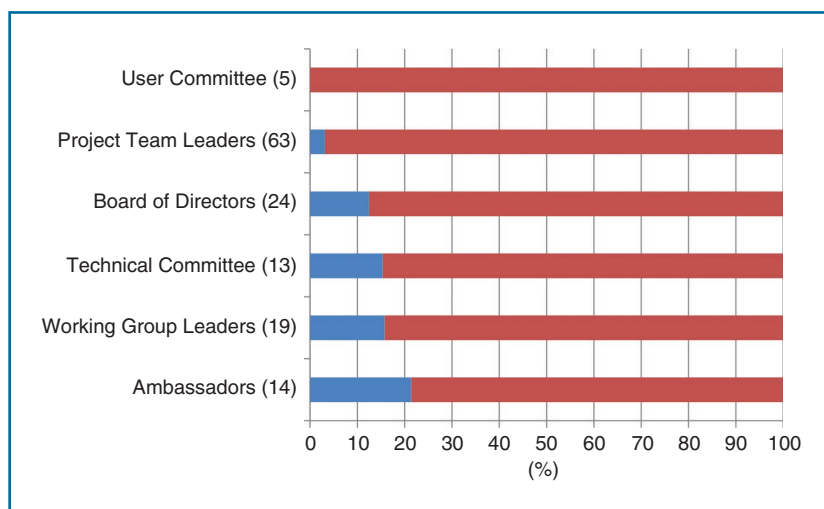
a man.<sup>8</sup> Risks implied by this strategy are related to the inherent imprecision of such heuristics<sup>8</sup> and to the marginalization of nonbinary individuals. In our study, the data and gender of the developers have been updated using a manual approach up to a certain threshold. The OpenStack project selection is defined by the Governance project. For these projects, the developers with the highest level of activity were manually inspected and their gender updated if needed. This was done in the same way, but working at the level of companies. However, given the size of the community and the total number of different identities (more than 7,000 across all Git repositories), the remaining number of "unknown" profiles is still high (around 25%), although their activity is lower (around 15%).

Another challenge was the difference between academia and industry. While academia has a long experience in measuring in software engineering, industry sees some of these methods as potentially innovative. Therefore, some industrial organizations are often reluctant to share the methodology or

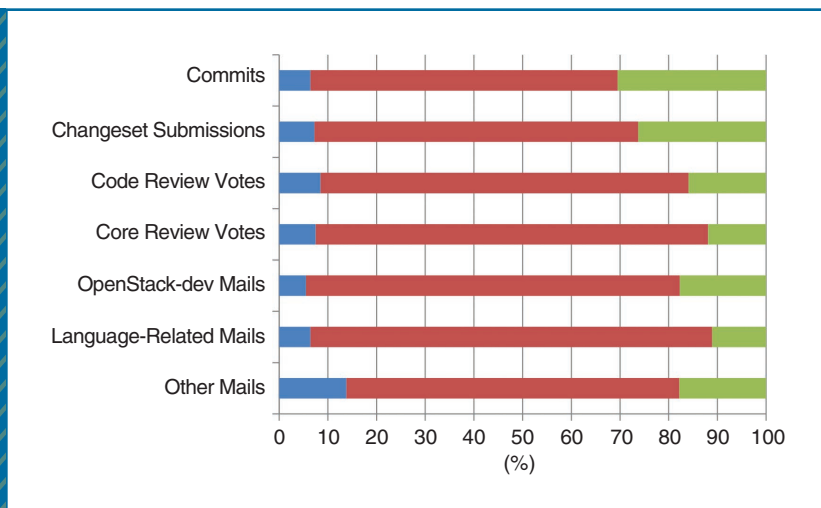
Measuring diversity in the OpenStack community is challenged by inherent imprecision of gender-identification approaches (whether manual or automatic), lack of data, and lack of methodology and process common to academia and industry.

results within the community and with other potential competitors. Thus, lack of common methodology and process may lead to discrepancies between internal company studies and the current work. This risk has been minimized by stating the definition of each metric, using open source software for reproducibility, and choosing the official projects and data sources (defined by the OpenStack Governance project).

Figure 1 shows the percentage of women in the governance structures



**FIGURE 1.** The percentage of women (blue) in governance and leadership positions. The numbers in parentheses are the total members of each group.



**FIGURE 2.** The percentages of code- and noncode-related artifacts contributed by women (blue), men (red), and individuals whose gender could not be identified (green). The percentage of women is 10–12%, depending on the data source and the analysis. Participation at the governance and leadership level increased remarkably, and participation in all of the code- and noncode-related contributions also increased.

of the OpenStack Foundation (October/November 2017). One of the lowest percentages is the technical position of the project team leaders. Being a member of a technical committee also counts as having a governance technical position. The latest data indicate that 15% of technical committee members are women. A previous analysis reported no women in that role. These roles are filled either by community elections or by direct representation of the companies.

When considering the percentage of women among attendees and speakers of the OpenStack Summits and Project Team Gatherings in 2016–2017, we observe a consistent pattern of participation of around 10–12%. We do not see differences in percentages of speakers and attendees.

The percentages of women among code and noncode contributors remains in the same range of 10–12%. The percentages of their contributions

(Figure 2) are usually lower. In particular, the lowest share of contributions is in the developer mailing list `openstack-dev`, while “other mailing lists” have the highest percentage of contributions by women. This observation concurs with the study of Drupal mailing lists.<sup>1</sup> Adding the time aspect to the picture reveals that participation of women increased in all of the code- and noncode-related contributions over time: Evidence shows that women tend to move from developers to maintainers and act as code reviewers.

## Discussion and Lessons Learned

We have presented the results of the OpenStack Gender Diversity Reports at several OpenStack and Linux community events—specifically, OpenStack Summits, Open Source Summits, CHAOSSCon (the Community Health Analytics Open Source Software Convention), and the Open Source Leadership Summit—in 2017

and 2018. The feedback received from the open source communities is overwhelmingly positive. The ability to measure and report on the level of female and minority representation in open source communities, as was done in the case of the OpenStack community, is key to having well grounded discussions about this issue. Data help increase awareness and understanding. During this time, discussions focused on this representation gap—and approaches to resolving it—have gathered momentum across the open source community as a whole.

Thanks to the dissemination of the results, we were able to collect feedback about the initial results, giving them guidance for altering the parameters of the research over time to address the needs and expectations of the community. This feedback can be categorized into two main areas: improving the current set of metrics and data sets and using those metrics and data sets for increasing gender diversity. Regarding the second area, the focus was on providing the community with actionable recommendations and extending the benefit of insights gained through this data, experience, and metrics beyond OpenStack.

Recommendations in the reports aimed to increase diversity and inclusion included an increased understanding of the importance of diverse, inclusive teams; continued tracking of female participation; enhanced collaboration with project team leads and other leaders across the community to identify, document, and communicate best practices; continued support of working groups; and improved onboarding guidelines and mentoring programs.

The research community has recently worked to identify barriers that

make it difficult for women to participate in software development,<sup>9–11</sup> and plans are underway to explore whether the recommendations derived from these studies, such as the use of GenderMag<sup>12</sup> to reveal gender-bias issues in software, can benefit open source communities.

The authors recognized the importance of being inclusive in these analyses. After all, diversity applies not only to gender but to underrepresented minorities as well, with multiple facets and dimensions. In addition, participation in open source communities is more than writing code; noncode contributions must also be considered and accounted for when assessing the overall health of a community. The authors view this analysis as a starting point. They intend to go further. The examination of such topics as project leadership, events inclusivity, and the creation and reinforcement of codes of conduct have helped shed light on the sensitivities within a community about diversity and inclusion; this is part of further work on improving inclusion of the analysis undertaken.


The privacy of the data was another critical consideration. The data sets are provided under restrictions meant to protect privacy, and presentations use aggregated information where there is not a direct match between developer and gender. However, it is hard to curate, improve, and, in general, manage a set of thousands of developers. Other ways to participate and help community members should be explored to keep improving the community data set as more and more people join the community each month.

This research has helped pinpoint areas where further research is needed. Further research, using qualitative methods, offers the

potential for deeper insight, interpretation, and understanding of these numbers. For example, the quantitative data can be used to identify projects that attract a more diverse group of contributors and, equally important, those that do not. Qualitative research is then needed to understand why this is the case.

There are still open questions in the results that are focused on providing more advanced data sets that may help provide greater insight into the idiosyncrasies of open source projects and communities. For instance, is there any difference between the development activities by gender? Is there unconscious bias toward contributions depending on the gender of the developers? At the communication level, is the collaboration taking place primarily between developers of the same gender? Are code reviews fair? Does the presence of a greater number of female participants within the project help attract more diverse participants to that specific project, when compared to other projects?

**T**he expectation of the authors is that the OpenStack community will enter the next phase now and use the numbers to drive deeper qualitative research,



Open source communities should continue monitoring participation of women, discover and learn best practices for attracting and retaining women, and broaden the focus to include other underrepresented groups.

learn from the most—and least—diverse projects and organizations, and deliver data-driven strategies to guide actions and activities to resolve this issue. The goal is to create a welcoming, inclusive environment that encourages diverse contributions across all constituents of a community.

With respect to extending the benefit of insights gained through this data, experience, and metrics beyond the OpenStack project, the authors have fostered collaboration with other projects, such as the Community Health Analytics for Open Source Software (CHAOSS) Project. This project, announced by the Linux Foundation in September 2017, was created to establish standard, implementation-agnostic metrics to gauge



## HIGHLIGHTS

- Of the total population of contributors to OpenStack, 10% are women
- Recommendations to improve diversity and inclusion within OpenStack
- Analysis of female representation in the OpenStack community
- Areas to consider when measuring diversity and inclusion: leadership, governance, event representation, code-and noncode-related contributions



**DANIEL IZQUIERDO** is the chief data officer at Bitergia, a company that provides software analytics for open source ecosystems. His research interests include the quality of the data, research of new metrics, analysis, and studies of interest for Bitergia customers. Izquierdo received a Ph.D. in free software engineering from the Universidad Rey Juan Carlos, Spain. He is an active member of the Community Health Analytics for Open Source Software Diversity and Inclusion Project, GrimoireLab, and the InnerSource Commons. Contact him at [dizquierdo@bitergia.com](mailto:dizquierdo@bitergia.com).




**ALEXANDER SEREBRENIK** is an associate professor of software evolution at the Eindhoven University of Technology. His research interests include a wide range of topics, from source-code analysis to collaborative and human aspects of software engineering. Serebrenik received a Ph.D. in computer science from Katholieke Universiteit Leuven, Belgium. He has coauthored *Evolving Software Systems* (Springer-Verlag, 2014) and more than 100 scientific papers and articles. He is a Senior Member of the IEEE. Contact him at [a.serebrenik@tue.nl](mailto:a.serebrenik@tue.nl).



**NICOLE HUESMAN** is a senior global marketing manager and open source advocate and an active member of the Community Health Analytics for Open Source Software Project's Diversity and Inclusion Work Group. She also helps lead speed-mentoring workshops across the OpenStack and Linux communities. Huesman has a master's in business administration, with an emphasis in marketing, from Marylhurst University. She is a frequent speaker and blogger in the areas of diversity and mentorship. Contact her at [nicole.k.huesman@gmail.com](mailto:nicole.k.huesman@gmail.com).



**GREGORIO ROBLES** is an associate professor at the Universidad Rey Juan Carlos, Spain. His research interests include free/open source software mainly by mining their repositories. Contact him at [grex@gsync.urjc.es](mailto:grex@gsync.urjc.es).

the overall health of open source communities. In December 2017, the CHAOSS Project's Diversity and Inclusion Working Group was formed with the goal of integrating the methods and processes of the OpenStack Gender Diversity Reports, among other initiatives, to create a proper body of knowledge about the topic of diversity and inclusion. 

## References

1. B. Vasilescu, A. Capiluppi, and A. Serebrenik, "Gender, representation and online participation: A quantitative study," *Interacting Computers*, vol. 26, no. 5, pp. 488–511, 2013.
2. B. Vasilescu, D. Posnet, B. Ray, M. G. J. van den Brand, A. Serebrenik, P. Devanbu, and V. Filkov, "Gender and tenure diversity in GitHub teams," in *Proc. 2015 ACM CHI Conf. Human Factors in Computing Systems*, 2015, pp. 3789–3798.
3. L. F. Capretz and F. Ahmed, "Why do we need personality diversity in software engineering?" *SIGSOFT Softw. Eng. Notes*, vol. 35, no. 2, pp. 1–11, Mar. 2010.
4. A. Rehman Gila, J. Jaffa, M. Omar, and M. Zahid Tunio, "Impact of personality and gender diversity on software development teams' performance," in *Proc. Int. Conf. Computer*,



- Communications, and Control Technology*, 2014, pp. 261–265.
5. M. J. Muller and S. Kuhn, “Participatory design,” *Commun. ACM*, vol. 36, no. 6, pp. 24–28, 1993.
  6. G. Robles, L. A. Reina, J. M. González-Barahona, and S. D. Domínguez, “Women in free/libre/open source software: The situation in the 2010s,” in *Proc. Int. Federation for Information Processing Int. Conf. Open Source Systems*, 2016, pp. 163–173.
  7. J. Gonzalez-Barahona, D. Izquierdo, S. Maffulli, and G. Robles, “Understanding how companies interact with free software communities,” *IEEE Softw.*, vol. 30, no. 5, pp. 38–45, Sept. 2013.
  8. B. Lin and A. Serebrenik, “Recognizing gender of Stack Overflow users,” in *Proc. 13th Int. Conf. Mining Software Repositories*, 2016, pp. 425–429.
  9. D. Ford, A. Harkins, and C. Parnin, “Someone like me: How does peer parity influence participation of women on Stack Overflow?” in *Proc. IEEE Symp. Visual Languages and Human-Centric Computing*, 2017, pp. 239–243.
  10. D. Ford, J. Smith, P. J. Guo, and C. Parnin, “Paradise unplugged: Identifying barriers for female participation on Stack Overflow,” in *Proc. ACM SigSoft Int. Symp. Foundations of Software Engineering*, 2016, pp. 846–857.
  11. C. J. Mendez, H. S. Padala, Z. Steine-Hanson, C. Hilderbrand, A. Horvath, C. Hill, L. Simpson, N. Patl, A. Sarma, and M. M. Burnet, “Open source barriers to entry, revisited: A sociotechnical perspective,” in *Proc. 38th Int. Conf. Software Engineering*, 2018, pp. 1004–1015.
  12. M. M. Burnet, S. Stumpf, J. Macbeth, S. Makri, L. Beckwith, I. Kwan, A. Peters, and W. Jernigan, “GenderMag: A method for evaluating software’s gender inclusiveness,” *Interacting Computers*, vol. 28, no. 6, pp. 760–787, 2016.



# ITProfessional

TECHNOLOGY SOLUTIONS FOR THE ENTERPRISE

## CALL FOR ARTICLES

*IT Professional* seeks original submissions on technology solutions for the enterprise. Topics include

- emerging technologies,
- cloud computing,
- Web 2.0 and services,
- cybersecurity,
- mobile computing,
- green IT,
- RFID,
- social software,
- data management and mining,
- systems integration,
- communication networks,
- datacenter operations,
- IT asset management, and
- health information technology.

We welcome articles accompanied by web-based demos. For more information, see our author guidelines at [www.computer.org/itpro/author.htm](http://www.computer.org/itpro/author.htm).

**WWW.COMPUTER.ORG/ITPRO**

