Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

Community Tapestry: An actionable tool to track turnover and diversity in OSS

Mariam Guizani [a,1], Zixuan Feng [b,1,*], Emily Arteaga [c], Katie Kimura [b], Diane Mueller [d], Luis Cañas Díaz [d], Alexander Serebrenik [e], Anita Sarma [b]

[a] Queen's University, Kingston, Ontario, Canada
[b] Oregon State University, Corvallis, OR, United States
[c] Stanford University, Stanford, CA, United States
[d] Bitergia, Leganés, Spain
[e] Eindhoven University of Technology, Eindhoven, Netherlands

ARTICLE INFO

Keywords:
Dashboard
Open source
Turnover
Participatory design

ABSTRACT

Context: A healthy open-source software (OSS) community is one that has a diverse contributor base and is sustainable by retaining its contributors. Project leaders, therefore, must understand their community's turnover and diversity makeup.
Objectives: This study aims to investigate how to support project leaders in monitoring OSS community health. Specifically, we examine the role of an interactive dashboard in enhancing awareness of contributor turnover and diversity.
Methods: We designed and developed **Community Tapestry**, a dynamic, daily-updated dashboard, using **Participatory Design (PD)** sessions with stakeholders from the Apache Software Foundation (ASF), Community Health Analytics in Open Source Software (CHAOSS), and Bitergia Analytics. We initially evaluated Community Tapestry by engaging contributors from our PD partners' OSS projects. To further validate our findings, we conducted a confirmatory study with a prominent OSS project under the Cloud Native Computing Foundation (CNCF). Contributors from both projects explored a personalized version of the dashboard that uses their own up-to-date project data.
Results: Our results demonstrate that Community Tapestry enhanced participants' awareness of their community's turnover and diversity state. It enabled them to identify areas for improvement and provided actionable insights to foster a more inclusive and stable community.
Conclusion: Community Tapestry offers OSS project leaders an actionable approach to monitor turnover and diversity state, enabling data-driven governance and fostering more inclusive and sustainable communities. Our PD approach provides practical insights into how community-driven interventions can be developed and adopted.

1. Introduction

We know by now that "rain is wet"–extensive research has shown that Open Source Software (OSS) projects grapple with high turnover rates [1–3] and low diversity [4–7], factors that significantly impact their sustainability.

High turnover rates in OSS projects have been a key challenge for the sustainability of many OSS projects [8], which can become a crisis when the turnover is due to a company decision to withdraw [9]. Researchers have investigated the reasons for turnover in OSS, which include low organizational commitment [10] and dissatisfaction with the OSS community [11]. Turnover can also negatively impact the quality and productivity of software development [12].

The positive impacts of diverse contributors are well-researched [4, 13–17]. Gender diversity, for instance, has been investigated for its effect on productivity [5] and creating a welcoming environment [18],

* Corresponding author.
E-mail addresses: mariam.guizani@queensu.ca (M. Guizani), fengzi@oregonstate.edu (Z. Feng), arteagae@oregonstate.edu (E. Arteaga), kimuraka@oregonstate.edu (K. Kimura), dmueller@bitergia.com (D. Mueller), lcanas@bitergia.com (L.C. Díaz), a.serebrenik@tue.nl (A. Serebrenik), Anita.Sarma@oregonstate.edu (A. Sarma).
[1] Co-first author.

https://doi.org/10.1016/j.infsof.2025.107871
Received 21 March 2025; Received in revised form 10 August 2025; Accepted 11 August 2025

while organizational affiliation diversity has been shown to improve productivity, product quality, and project sustainability [19,20].

We are not alone in putting out a call to action to create interventions to help OSS projects be healthy and a welcoming environment. For instance, Murphy-Hill et al. [21] investigate anonymizing code review in practice. Ahmed et al. [22] have designed machine learning algorithms to identify toxicity in code reviews, and Murphy-Hill et al. [23] aimed at reducing code review toxicity by implementing respectful code review reminders. Multiple studies have used the GenderMag method [24] to build processes and tools to find and fix inclusivity bugs [25–28]. Researchers [29,30] and OSS communities (e.g., CHAOSS [31]) have deduced a host of metrics to monitor project health. Thus far, research [32–34] and industry-driven dashboards [35, 36] for monitoring OSS project metrics have focused on project activity metrics, such as commits, issues, pull-requests, none specifically monitor turnover or diversity aspects. Additionally, none have been comprehensively evaluated in the field to identify what information can help project leaders understand their project health and take future actions to improve retention.

We close this gap by designing, implementing, and deploying a daily-updated project monitoring tool, Community Tapestry. The goal of Community Tapestry is to (1) signal project health by showing the project's turnover rate and diversity makeup, (2) give project leaders information to take actions relevant to their project needs, and (3) help monitor the effects of their actions over time.

Because OSS contributors and project leaders are resource constrained, it is crucial to provide actionable tools that they *want and can readily use*. This necessitates working together with OSS stakeholders to guide the design and development of the intervention. Thus, we opted to use Participatory Design (PD) [37] to actively involve our stakeholders, comprehend their needs and constraints, and collaboratively design a solution that best matches the situation on the ground. An additional benefit of PD is that having the stakeholders as partners enhances interest in the tool, which in turn can increase community adoption.

We partnered with the ASF, CHAOSS, and Bitergia Analytics to *design, implement, deploy, and evaluate* Community Tapestry with a first OSS project. We used a large ASF project, Beam as the partner OSS project. Beam is a large Big Data project with more than 1,000 current contributors. The Project Management Committee (PMC) chair of Beam, an OSS practitioner, and the then Diversity and Inclusion vice-president of the ASF brought to the table the needs and concerns of the ASF. Two OSS community members from CHAOSS brought their expertise about project health metrics. A lead engineer with Bitergia Analytics [38], the infrastructure on which we implement Community Tapestry, helped us design within the capabilities of the underlying infrastructure (Grimoirelab [38]). Bitergia Analytics was given a (paid) contract by the ASF to implement the dashboard. We refer to this group of people from now on as our PD partners.

Through discussions with our PD partners, we prioritized project health aspects of interest (i.e., turnover, gender, and organization affiliation). We then used PD principles [39] to collaboratively design the dashboard, which we implemented and deployed on Bitergia infrastructure (Fig. 3 shows two of the dashboard pages from Community Tapestry). We then evaluated Community Tapestry for (1) inclusivity using the Why/Where/Fix approach [25], (2) usefulness with a user study of 15 participants from Beam, and (3) transferability with a user study of 8 participants from OpenShift, a large project in Cloud Native Computing Foundation (CNCF). The PD study spanned 20 months (See Fig. 1 for method overview).

Our key contributions span three core areas: process for developing OSS interventions, implementation of a dashboard, and analysis of how the dashboard facilitates data-driven decision-making, which we expand on below.

- Our process offers one of the first detailed accounts of how to effectively combine a systematic inclusivity debugging process and participatory design in the context of OSS to develop an inclusive and stakeholder-driven tool. This approach ensures the tool is inclusive, aligns with diverse community needs, and highlights trade-offs in decision-making among stakeholders.
- We implemented Community Tapestry, a daily-updated dashboard that centralizes project community turnover and diversity metrics, enabling project leaders to observe community trends and take data-driven action. Community Tapestry bridges the gap between practical interventions and empirical research on OSS diversity and turnover.
- Community Tapestry's ability to empower a data-driven governance was highlighted in an in-depth analysis of free explorations of the Community Tapestry by participants from two structurally distinct OSS communities (ASF and CNCF).

2. Background and related work

The role of diversity in OSS sustainability. Diversity is widely recognized as a critical factor for the sustainability of OSS communities. Rodríguez-Pérez et al. [40] defined perceived diversity as innate individual diversity factors and highlights the importance of diverse Software Engineering teams. Gender diversity has received considerable attention, with studies linking gender diversity enhances productivity and performance [4,6,13,40,41]. Diversity goes beyond gender, encapsulating other aspects such as race, age, neurodiversity and disability [40,42,43].

With the changing OSS landscape, OSS has evolved from a volunteer-based community to a hybrid environment where company-affiliated and volunteer contributors coexist [20,44,45]. Contributors' diverse affiliations are vital for OSS projects sustainability [46], as competition motivates companies to join, and some donate projects to OSS foundations to signal neutral governance and increase participation [19].

While these studies have highlighted the importance of diversity in domains such as gender and company affiliation, there is no mechanism for community managers to assess how contributor diversity affect their projects. We close this gap by translating empirical insights into actionable, quantitative metrics, thus enabling data-driven diversity management.

Turnover as a measurement of OSS community health. Another important dimension of measuring the health of OSS communities is turnover, as it reflects the stability and sustainability of contributor engagement. High turnover signals challenges in attracting and retaining contributors [30,47,48] as well as maintaining project momentum [48]. Steinmacher et al. [49,50] identified 58 barriers faced by newcomers and analyzed how the answers to newcomers' first emails influenced their likelihood to stay [51]. Pinto et al. [52] found that nearly half of a project's contributors submitted a single contribution and never returned [52], with many never having a single pull request (PR) accepted [53]. With the high rate of turnover in OSS [1–3] and its negative influence on team cognition and performance [54,55], Steinmacher et al. [51] have found that most newcomers (as high as 80% in some projects) do not become long-term contributors.

While these studies shed light on the importance of monitoring turnover as an indicator of OSS health, our work builds on these findings to actively help community managers do three things: manage the state of diversity, monitor turnover through the lens of diversity, and understand contributor turnover in their projects.

Existing interventions enhancing OSS Community Health. OSS community has seen the development of a few interventions aimed at enhancing community health, streamlining the onboarding process, and increasing the inclusiveness of the tool itself (e.g., OSS projects) [24–26, 56]. Steinmacher et al. [57] introduced a portal to facilitate newcomers' integration into OSS projects, streamline newcomer orientation, and simplify the contribution process.

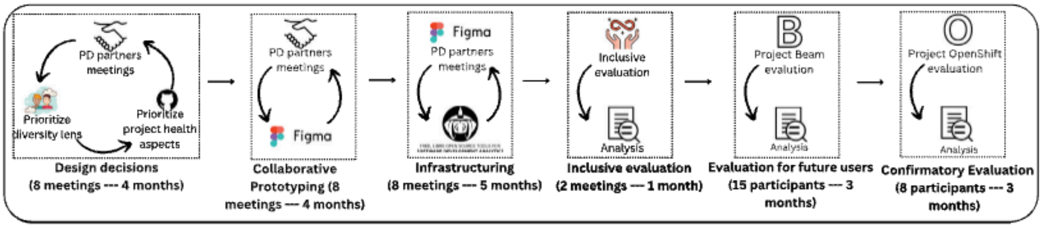

Fig. 1. Method Overview: Our PD methodology includes six stages: (1) Making design decisions with PD partners, (2) Prototyping with PD partners, (3) Implementing and deploying system infrastructure with PD partners, (4) Evaluating design for inclusivity (i.e. accommodating different methods of information-processing), (5) Conducting future user evaluation with 15 participants from ASF Beam project, (6) Conducting confirmatory evaluations with 8 participants from CNCF OpenShift project.

Table 1
Demographics of participatory design partners.

ID	Organization	Gender	Role
1	ASF	Woman	ASF DEI committee Vice-President (during the time of the study)
2	CHAOSS	Man	CHAOSS community health specialist
3	Google	Woman	OSS practitioner
4	Bitergia	Man	Software engineering professional
5	Bitergia	Man	Software engineering professional
6	ASF	Man	Chair of ASF Beam
7	CNCF	Woman	Director of community development of CNCF OpenShift

Qiu et al. [32] developed a dashboard that assists maintainers in understanding project activity metrics. Guizani et al. [33] designed a dashboard to assist maintainers in attracting and retaining contributors by highlighting project goals and recognizing active newcomers. Ramchandran et al. [34] built a time-stamped dashboard to track the sustainability trajectories of incubator projects within the Apache Software Foundation. They found that projects often experience downturns due to factors such as declining contributor activity and lack of documentation. Moreover, Emerging Decentralized Autonomous Organizations (DAOs) [58], a subset of OSS communities, have also been studied for sustainability, including metrics such as voter turnover [59] with tools like DeepDAO [60] providing real-time analytics to monitor these metrics.

While a few actionable tools exist to address specific community needs, such as onboarding newcomers, tracking the sustainability of incubator projects, none focus on monitoring diversity and contributor turnover. The closest effort, ClimateCoach [32], is a community health dashboard focused primarily on responsiveness (e.g., time to close issues and pull requests) and inclusiveness (e.g., tone of conversations). In contrast, our dashboard uniquely combines repository activity with turnover metrics through the lens of community diversity. This provides community managers with a comprehensive understanding of both disengagement and diversity dynamics, enabling them to make informed, data-driven decisions to improve contributor retention and foster more diverse participation.

3. Design and implementation

We followed a participatory design (PD) approach [61] to create a user-centered turnover and diversity monitoring tool designed to meet the needs of the OSS community managers and integrated with the technology being used by the project. Fig. 1 presents an overview of our approach. Our PD partners were recruited through the ASF DEI Working Group who recommended Beam as a case study as it has been an active participant within the ASF and recently graduated from incubation, indicating strong community engagement and project maturity. Therefore, we recruited one ASF Beam Chairperson (P6) and one Director of Community development of CNCF OpenShift Project (P7). Note, Community Tapestry was then evaluated with 15 other contributors to Beam. To further validate our findings, we conducted a confirmatory study with a different community from the CNCF (OpenShift), which operates under a different governance model and contributor base. This evaluation helped ensure that our design and insights were not limited to a single community or organizational context. Table 1 shows the demographics of PD partners.

3.1. Design decisions

The first step in PD is to understand the needs of the stakeholders and make design decisions. Different stakeholders may have differing expectations of the proposed tools and what can be implemented in the current infrastructure. Through design discussions, which serve as a mutual learning process, we created an initial set of requirements. We met bi-weekly with our PD partners over a course of 4-months to finalize (1) project health aspects of interest and (2) the project metrics (outcome variables) that would be shown in Community Tapestry.

A critical aspect of sustainability is managing *turnover*, which can be deciphered through trends of newcomers joining or contributors leaving. Retaining contributors is key to OSS project health [62,63]. Contributors including newcomers face many challenges when contributing to OSS [15,49,57,64]. Therefore, our CHAOSS PD partners recommended that we monitor turnover based on the threshold from GrimoireLab [65] : (1) A new contributor has made their first recorded contribution (e.g., commit, pull request, or issue comment) within the last three months; (2) An active contributor has made at least one contribution in the last three months, excluding new contributors (i.e., they joined more than three months ago); (3) An inactive contributor for who has made no contributions for six months or more; (4) A contributor is considered potentially leaving if they have been inactive for the last three to six months. Such turnover trends can signal the declining sustainability of an OSS project and trigger follow-up action by community managers.

Diversity plays an essential role in driving innovation which in turn can help cultivate a healthy community. Diversity is a multidimensional construct that includes attributes such as gender, race, tenure, English proficiency, and geo-location [16,17,66,67]. Through multiple discussions with our PD partners spanning 6 meetings, we finalized the diversity aspects of interest (i.e., gender, and organization affiliation attributes).

Our PD partner, the vice president of the DEI committee, recommended using the just concluded ASF survey on “the state of diversity in ASF” as the starting point for our discussions [68]. The survey ($N > 400$) included contributors’ demographics and various challenges that contributors faced. The underrepresented demographic attributes from this survey included gender, compensation (getting paid for OSS contributions), tenure, geo-location, and English proficiency. Note that the survey did not collect race data as the ASF is an international OSS community, and the survey was designed for all ASF contributors. How race is defined varies widely across nations, and which races are discriminated against depends on the race of the majority population of a country [69]. Based on similar logic, our PD partners decided not to consider race as a diversity aspect.

The survey results showed that contributors from minority groups encountered a host of challenges categorized into technical, social, and process challenges [70]. This corroborated the need to surface the state of diversity as one of the aspects of project health in an easy-to-digest manner.

Gender was the first diversity aspect of interest since (1) OSS is known to have gender imbalance [4,13,14], (2) peer-parity (similar individual one can compare oneself to) is an important criterion for women to continue in OSS [71], and (3) The ASF survey identified that women contributors face more challenges compared to others [72]. *Affiliation diversity* was the second diversity aspect of interest. Paid vs. volunteer contributors have varying levels of engagement and motivation to continue in a project. In the changed landscape of OSS today, company participation plays a critical role in the sustainability of an OSS project [19]. Because Beam included paid support from multiple organizations, our ASF PD partners were particularly eager to monitor the diversity of organizational affiliation in their project. In fact, affiliation diversity can help mitigate the risks associated with dependency on a single company and foster a richer, more resilient environment.

Geo-location and English proficiency were omitted aspects due to technical constraints in the infrastructure. With geo-location, it is not always guaranteed that contributors will list this information on their GitHub profiles. Tracking IP addresses could be a potential data source, but our IRB (Institutional Review Board) did not allow this. The same limitation applies to English proficiency. Past works have indicated that geo-location could be used to analyze cultural barriers and English proficiency [17,73]. However, even accurate region inferences, which we lack here, do not guarantee precise English proficiency inferences.

Next, over the course of two meetings, we finalized the project metrics (outcome variables) to use for Community Tapestry. Our CHAOSS PD partners led this discussion, starting us with the list of CHAOSS metrics they identified for project health. We prioritized the data that the ASF PD partners were interested in and that Bitergia Analytics could collect. The first set of metrics included those that could be collected from Github, the development platform that Beam used. These metrics included details about issues, pull requests (PRs), PR comments, and commits. The ASF PD partners requested details about StackOverflow activity related to BEAM, as they have contributors on duty to answer questions and aim to capture contributions across platforms to systematically assess retention health comprehensively.

They were also interested in social media activity (e.g., X). However, Bitergia Analytics did not have the capability to track such activity, so the ASF PD partners agreed to drop this request.

3.2. Collaborative prototyping

In the *collaborative prototyping* step, PD partners express ideas and provide insight on how users may interact with the tool through interactive mockups [61]. The collaborative prototyping process lasted four months and spanned eight meetings, as shown in Fig. 1.

The first four meetings were used as conceptual design sessions with stakeholders, during which they shared insights on their needs, preferences, and expectations for the tool [74]. This phase involved steps such as understanding their requirements, exploring use cases, and prioritizing features. Based on these inputs, we used Figma[2] to design several versions of different dashboards showcasing different types of information. When designing these mock-ups we followed the dashboard design principles laid out in Few [75].

[2] https://www.figma.com/prototyping/

PD partners then reviewed and provided iterative feedback on these prototypes. During feedback sessions, PD partners freely explored the prototypes, thinking aloud and sharing their screens. We asked our Beam PD partners to think about the tasks they wanted to perform as community managers and/or how they envisioned someone using the data.

The PD partners provided feedback verbally or annotated on the screen share. The feedback ranged from enhancement of existing features to requests for new features. For instance, the PD partners requested that the PR communication network (network diagram based on PR review comments, Fig. 3, be updated such that the thickness of edges between the contributors (nodes) reflected the number of PRs instead of that of comments. The collaborative prototyping step spanned eight meetings. The prototype was updated after each meeting (We asked our Beam PD partners to think about the tasks they want to perform as community managers and/or how they foresee someone using the data). (Refer to the supplementary material for early outcome prototypes from PD phases [76]).

3.3. Infrastructuring

Infrastructuring is the process where the design decisions are implemented into the tool. “No design process, even PD, can fully predict the changes ahead” [77]. Thus, during *infrastructuring*, it is important to understand and align any differences between design and technical configurations. We continued another eight bi-weekly meetings with our PD partners for implementation.

System Architecture. A key request from our PD partners was the necessity for up-to-date data, which could support daily-updated monitoring and evaluation of turnover and diversity aspects. We then implemented Community Tapestry in GrimoireLab from Bitergia Analytics (an OSS analytics toolkit and a part of the CHAOSS project), which allowed us to integrate daily-updated data from GitHub repositories and StackOverflow. Fig. 2 demonstrates the architectural design of our system. GrimoireLab leverages a Python API for fetching data from repositories, which enables us to access all retrieved items from the repositories as dictionaries (JSON documents). The Perceval library [78] then streamlines the process of obtaining daily updates from the repository, capturing only incremental changes through a 24-hour data collection cycle.

Identity analysis. A first step in our analysis was to separate bot-generated project activity from real user data. Bots are increasingly common in contemporary software development for tasks such as automated code review [79–81]. We used SortingHat (as a component of Bitergia Analytics Platform), a relational database-powered tool used by GrimoireLab [82], to filter out bot activities. SortingHat also helped connect contributors with their affiliations by keeping track of corporate and non-corporate email domains (e.g., google.com vs. gmail.com). We classify non-corporate email domains as “affiliation unknown”. The next concern was merging data from the same person. In OSS, individuals can use their private email or their corporate email when making contributions; additionally, people switch jobs. Thus, simply combining all affiliations into a single affiliation is not right [83]. Therefore, when we merged user IDs into a single account, we kept their list of affiliations mapped to their specific contributions.

In Sorting Hat, we merged user IDs into a single account but listed all their affiliations. Ultimately, this approach allows CommunityTapestry to accurately map and display the timeline of the user’s activities based on their affiliation status. For example, User A has two email addresses with two different affiliations: @companyA.com and @companyB.com. SortingHat merges these two identities for UserA and recognizes their affiliations with Company A and Company B. Without a specific date, SortingHat defaults to a recent time order. Assuming User A worked for Company A from January 1, 2020, to June 1, 2021, and Company B from July 1, 2021, onwards, SortingHat maps their activities associated with each company during the relevant periods.

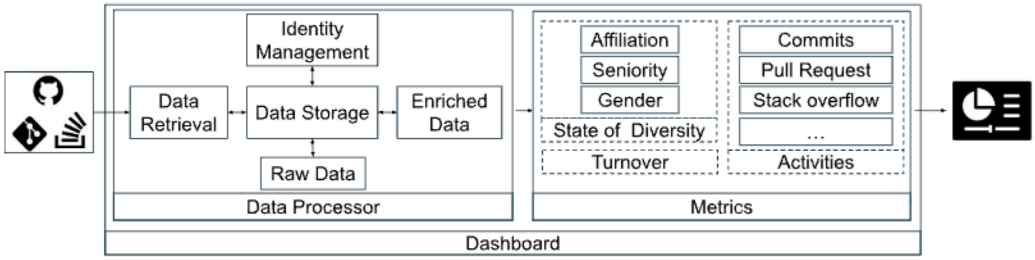

Fig. 2. Overview of the architecture of Community Tapestry. Data was collected from different sources (e.g., GitHub) through a data retrieval process and enriched using GrimoireLab [38]. The metrics module computes analysis results on diversity, activity, and turnover, which are then visualized on a dashboard to provide actionable insights.

Gender classification. To infer the gender of contributors in our dataset, we used the Namsor API [84]. This name recognition tool estimates the gender of a full name on a −1 to +1 probability scale. To improve accuracy, we implemented a two-step process. First, we used Namsor to predict the origin of contributors' names, to estimate the likely cultural or ethnic origin of the name, which can enhance gender inference. For example, when an Asian name (e.g., Jixiang) can be better categorized into gender when the name origin (China) is considered, as it allows Namsor to infer cultural signals from the country of (name) origin to infer gender. To minimize prediction errors, we applied a threshold and filtered out gender predictions with probabilities lower than 90%, following best practices from prior literature [85–87]. Namsor is just one method for gender inference; Community Tapestry is designed to be flexible and incorporate improved tools or self-identified gender data when available. In cases where maintainers or PMC members spot errors in the inferred data, they can provide corrections.

Additionally, past works have found that Namsor does not perform well for Chinese names since they originate from Chinese characters, and using only the "Pinyin" format causes errors [88,89]. Korean and Japanese names may also produce similar errors as Chinese names due to their unique writing systems [90]. Therefore, we performed an additional level of manual analysis performed by two researchers; one of Chinese origin and the other an American of Japanese and Korean descent. The manual verification entailed cross-referencing the contributors' GitHub profiles and LinkedIn pages to infer their gender. We first checked contributors' profile page(s) to see if they stated their pronouns; if not, then we checked for a clear photo. If the photo was unclear, we marked the gender as unknown. We acknowledge that gender identity extends beyond a binary classification, and we have designed our system to be flexible, allowing for future updates and modifications to the database. In total, we manually validated 991 names, of which 559 could not be inferred from their profiles due to the absence of a profile photo on their GitHub or LinkedIn pages. 191 underwent double-validation and 45 required corrections from NamSor prediction results (3 from male to female and 42 from female to male).[3]

For new contributors joining the project, we applied the same gender classification process that was used on existing contributors. When our confidence was low, we reverted to classifying the gender as "unknown." Furthermore, we allowed the PMC to update gender classifications based on their contextual knowledge of the project and its cultural norms.

Next, we introduce Community Tapestry, followed by the three types of evaluations we performed, one of which is evaluating the tool with future users—the final step in PD.

4. Community tapestry

Let us consider a hypothetical scenario in which Riley, a new PMC chair in Beam, believes it important to be aware of the turnover rate and improve gender and affiliation diversity within their project. Fig. 3 presents the dashboard pages of Community Tapestry. The letters in purple circles in Fig. 3 identify the information we explain below. Refer to supplementary material [76] for details of each dashboard.

Riley begins by navigating to the dashboard "Contribution Retention Trends" in Community Tapestry A. She notices the visualization of the trends of newcomers, contributors who left (i.e., inactive in the last six months), and might be leaving (i.e., inactive in the last three months), and the retention trend disaggregated by gender B. Using this information, Riley notices that while there is similar retention between men and women in the project (black line under contributor retention trends in B), there are 20 men to 5 women contributors, which indicates a significant difference.

Riley decides to dive deeper. She filters the data to the last year. She reviews the detailed list of women contributors who might be leaving (the list contains names, number of contributions, affiliation, and last contribution date) to reach out to them and help if possible. Community Tapestry allows filtering the visualizations by different criteria: time C, affiliation name, and gender D. These filters can be applied by clicking on the visualization or typing. Riley wonders if the women are inactive because they were affiliated with a company that left. She then looks at the list of contributors who might be leaving. Riley notices two women from the company "A3" (Affiliation 3)[4] leaving the project and decides to contact them to find out if the reason was a personal or company-related decision.

4.1. Additional dashboards

Community Tapestry has two additional dashboards that can be disaggregated by gender or affiliation. Each dashboard contains different visualizations, which are described below.

Communication. The communication dashboard (Fig. 3 bottom) depicts the PR interactions between contributors. This dashboard has four visualizations. *(1) PRs overview* E, which shows the number of PRs by group (i.e., gender, affiliation), comments, and likes. *(2) Time to merge PRs* F displays the average time (in days) a group (i.e., gender, affiliation) has to wait to get their PR merged in descending order. *(3) PRs communication network* G is a graph where nodes are contributors, and the edges link contributors interacting on the same PR. The size of the nodes reflects the number of PRs authored by a contributor, and the thickness of an edge reflects the number of PRs between two nodes (i.e., contributors). The colors differentiate the disaggregated groups (i.e., gender, affiliation) and are displayed in a legend. *(4) PRs that need attention* lists PRs that did not receive any comments and their

[3] We used biological sex categories (male/female) in accordance with the terminology used by Namsor; these categories may not accurately reflect individuals' gender identity.

[4] We have anonymized the names and affiliations here for confidentiality reasons.

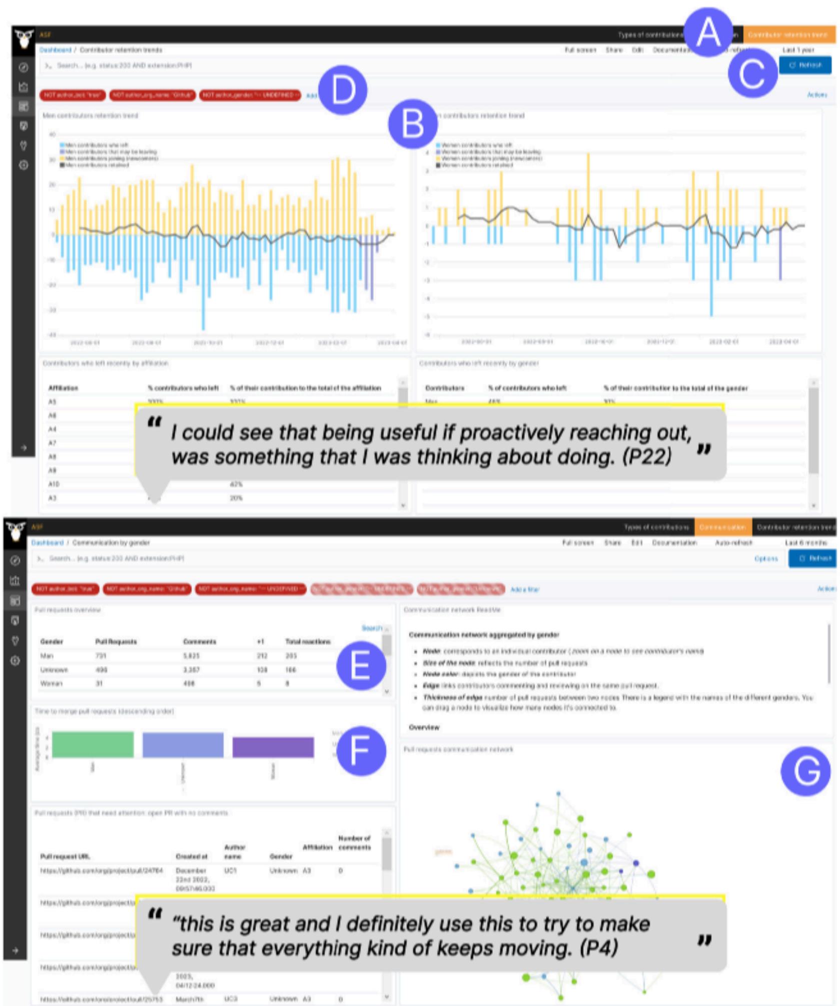

Fig. 3. Top: Community Tapestry contributor retention trend. Contributors' names and affiliations have been pseudo-anonymized. Bottom: Community Tapestry community network. A: Contribution Retention Trends; B: retention trend disaggregated by gender; C: Time filter; D: Affiliation and gender filters; E: PRs overview; F: Time to merge PRs; G: PRs communication network.

details (See supplementary materials [76]). This visualization contains the links to the PRs and the contributor information (name, affiliation, gender).

Types of Contributions. The types of contributions dashboard (as shown in Fig. 4) details the different kinds of contributions in an OSS project using the gender and affiliation aspects. This dashboard has four sets of visualizations. 1 *Contributors*, this set contains two visualizations presenting the percentage of contributors and the total number of contributors, both broken down by affiliation or gender. 2 *PRs*: [i] PRs proportion over time, which shows the percentage of PRs during a certain period of time disaggregated by gender or affiliation. [ii] PRs' evolution over time displays the evolution of PRs' counts by group (i.e., gender, affiliation). [iii] PRs count shows the total count of PRs within a group (i.e., gender, affiliation). [iv] PRs days to merge display the number of days before merging a PR disaggregated by gender or affiliation. 3 *Issues*, this set consists of four visualizations that are similar to the ones mentioned in (2) but present information related to the issue's contributions. 4 *StackOverflow*, this last set of four visualizations contains information on StackOverflow's questions and answers. This set is similar to information in (2) and (3). All dashboard pages allow filtering by different criteria, such as time, gender, and affiliation.

5. Evaluating for inclusivity

Because of the focus of Community Tapestry on diversity, we ensured that the dashboard supports and accommodates different approaches to how individuals process information (i.e. diverse cognitive styles of learning). To identify and fix usability and inclusivity bugs in this context, we used the Why/Where/Fix systematic debugging process [25], which is based on the GenderMag method [24]. This systematic process allows the designers to identify inclusivity bugs (instances where cognitive styles are unsupported), why they arise (which styles are unsupported), where they arise (UI element), and how to fix these bugs. The Why/Where/Fix process relies on personas with a customizable background to reflect the background of users along with a set of five cognitive style values. During the Why/Where/Fix process, the participants are requested to channel a persona and reflect on its ability to accomplish specific tasks using the software, e.g., to find information about the affiliation diversity of the persona's project.

We selected the Abi persona as the cognitive styles embedded in this persona tend to be overlooked by software [91,92]. Abi-like individuals have the following cognitive styles: (i) *Task-oriented motivation:* use technology for what they can accomplish with it, and not for enjoyment per se [93–95]; (ii) *Comprehensive information processing styles:* gather fairly complete information before proceeding [96,97]; (iii) *Lower computer self-efficacy* as compared to their peers; Computer self-efficacy relates with a person's confidence about succeeding at a

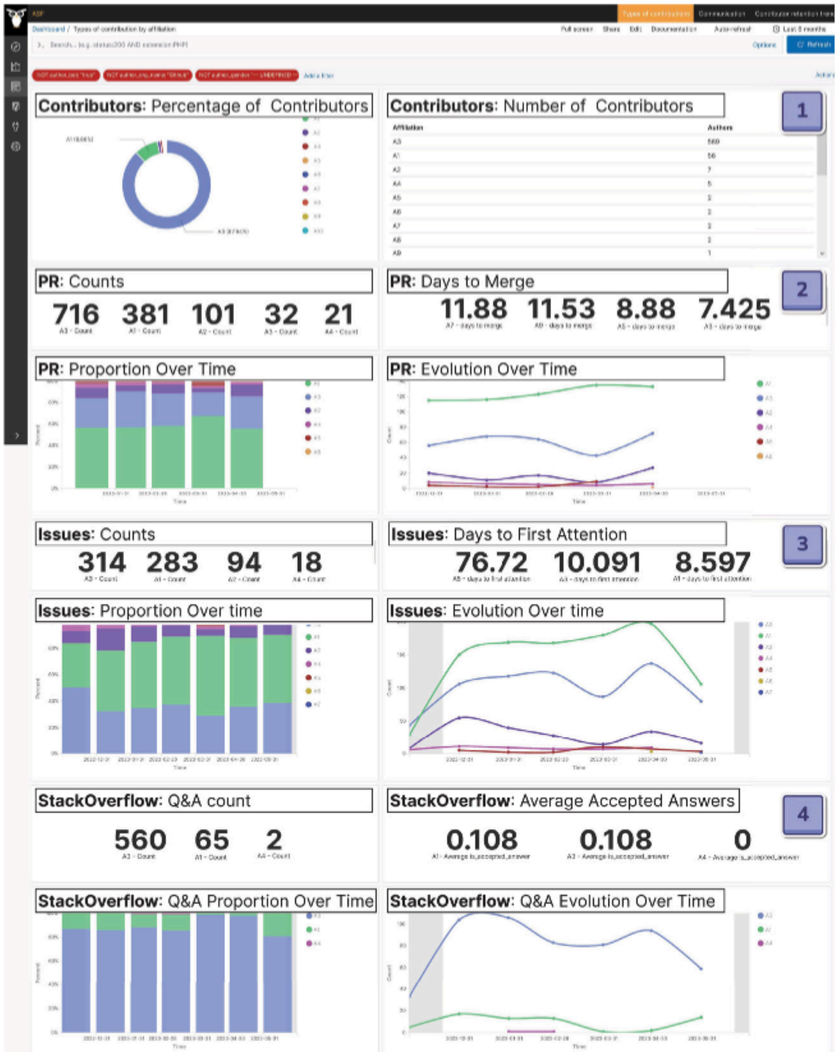

Fig. 4. Types of Contributions Dashboard. This dashboard visualizes OSS contributions across four areas: 1 contributor demographics, 2 pull requests, 3 issues, and 4 StackOverflow activity.

specific task, which influences their use of cognitive strategies and persistence; (iv) *Higher risk aversion:* when trying out new features [98, 99], which impact their decisions about the feature they use; and (v) *Learning by Process:* instead of playfully experimenting ("tinker") with new software features [93,100,101]. We customized the Abi persona to reflect the background of a PMC member.

We selected five use cases that yielded 41 evaluation questions related to users' goals and interface actions. These questions spanned each of the dashboards in Community Tapestry: 20 questions related to the types-of-contribution dashboard, 14 questions regarding the communication dashboard, and nine questions related to the trends dashboard. Three of the authors, designers of the system, performed this evaluation. Answering these 41 evaluation questions using the Why/Where/Fix approach led us to find ten usability bugs (24%), 9 of which were inclusivity bugs (22%), that is, bugs arising because Abi's cognitive styles were not supported.

We then designed and implemented fixes to Community Tapestry. An experienced researcher with the GenderMag method then reevaluated the fixes using GenderMag moments [102], a fragment of a GenderMag session where the evaluation targets specific features just in time. Our redesign reduced the number of usability bugs from 24% to 5% and inclusivity bugs from 22% to 2%. Note that the two usability bugs and the one inclusivity bug we did not fix were related to the UI of GrimoireLab we used. For example, a usability bug we could not fix is the size and position of the clock used to select the period to display information.

We selected five use cases that yielded 41 evaluation questions related to users' goals and interface actions. These questions spanned each of the dashboards in Community Tapestry: 20 questions related to the types-of-contribution dashboard, 14 questions regarding the communication dashboard, and nine questions related to the trends dashboard. Three of the authors, designers of the system, performed this evaluation using the Why/Where/Fix method [25], which is part of the well established GenderMag framework [91]. This method applies a structured cognitive walkthrough to assess how well the interface supports users with diverse cognitive styles. Each use case was broken down into subgoals and actions, and the evaluators – adopting the perspective of the Abi persona – answered three questions at each step: (1) whether <Persona> would have the subgoal the dashboard owners hoped for and why, (2) whether <Persona> would take the action the dashboard owners hoped for and why, and (3) if <Persona> did take the hoped-for action, would they know they did the right thing and were making progress toward their goal, and why. A "no" answer to any of these questions indicated a potential bug and its corresponding UI element(s); when the rationale included a specific cognitive style, it was flagged as an inclusivity bug. Using this method, we identified ten usability bugs (24%), nine of which were inclusivity bugs (22%)—i.e., bugs stemming from unsupported cognitive styles.

We then designed and implemented fixes to Community Tapestry using both the Why and the Where element of the Why/Where/Fix systematic process. The first author, an experienced GenderMag researcher, reevaluated the revised dashboards using GenderMag moments [102], which are targeted just-in-time assessments of specific features. This redesign reduced the number of usability bugs from 24% to 5% and inclusivity bugs from 22% to 2%. Note that the two remaining usability bugs and the one remaining inclusivity bug were due to limitations in GrimoireLab's UI. For example, we were unable to address the size and position of the clock used to select time periods. All forms used in the evaluation of all use cases across all dashboards are included in the supplemental material [76].

6. Evaluating with future users: ASF beam

This project was developed to design a dashboard intended as an empirically-evidenced pilot tool for ASF communities to help community leaders monitor turnover and diversity within their projects. The practicality of this dashboard for monitoring turnover and diversity within the OSS community was evaluated by future users (i.e., ASF community leaders and contributors who will adopt the dashboard to manage and analyze their OSS projects). The findings from these evaluations are being used by the ASF to design a framework that could be adopted by other OSS projects.

Evaluation design. The evaluation with future users in their natural environment is a key aspect of assessing design prototypes through PD [61,103]. Thus, we evaluate Community Tapestry with the *future users* of Beam with data extracted from their *own project* through *observations* of their unguided explorations of the dashboards.

Following empirical guidelines established by Lam et al. [104], we collaboratively designed our evaluation with PD partners. The first three authors: (1) iterated over the study questions and tasks, holding weekly meetings with the entire research team to refine the study; (2) discussed the study design with the PMC Chair of Project-B; (3) iteratively refined the study protocol for evaluating the ASF Beam dashboard through seven rounds of sandboxing with researchers at Oregon State University, until no further changes were required to materials such as the study scripts and questionnaires; and (4) piloted the study with the PMC chair.

Evaluation Scenario: To the goal of the project, our PD partners suggested we focus the evaluation on the **user experiences**. The project's primary goal was to assess user experiences with the dashboard to determine its potential adoption across other ASF projects to understand user engagement and gauge the dashboard's efficacy in enhancing awareness, utilities, and proactive actions toward turnover and diversity.

Evaluation question: The main question for user experiences is "What do my target users think of the visualization?" To this end and because awareness is the first step toward action, we decided to answer two evaluation questions:

- **EQ1:** how do the dashboards affect participants' awareness of turnover and diversity and their plans to take action?
- **EQ2:** how participants use the turnover and diversity information?

Evaluation method: Following the guidance from Lam et al. [104], we adopted a cross-sectional evaluation approach to assess the current and potential usage of the dashboard. This method involves collecting data at a time to provide a snapshot of user interactions, behaviors, and perceptions [105,106]. We chose this approach for three reasons: (1) Cross-sectional evaluations are proven effective for understanding immediate, actionable insights and predicting future engagement [105, 106]; (2) This timeframe aligns with our goal to demonstrate the dashboard's utility in helping community leaders monitor health and potential action based on their insights; (3) Longer-term methods like diary studies could be complicated by the dynamic nature of ASF projects and the multiple variables affecting outcomes. For example, ASF consistently utilizes other approaches to promote project health, such as public events [107], making it hard to determine what specifically caused changes in the outcomes.

Evaluation protocol: We used informal evaluation through observation to address our research questions. This approach does not utilize predefined task lists and is aimed at assessing "intuitiveness and functionality" [108], "probing for utility and usability" [109], and "identifying users' subjective preferences" [110]. The study included three parts (see Supplemental document for details [76]) and was approved by our university IRB. After the study, participants were compensated with a $50 gift card as a token of appreciation [111].

The first part was a pre-study questionnaire to collect participants' demographic information, their awareness about their project diversity, and their actions to improve diversity. We used this data as a baseline for comparison with post-study data. To familiarize the participants with the dashboard features, we provided them with a link and credentials to Community Tapestry. We then guided their exploration through the dashboard (See [76] for study script.)

We then asked the participants to explore each dashboard while thinking aloud. After exploring a dashboard, participants answered Likert-scale questions on their likelihood of using the information presented in that dashboard. They also answered two open-ended questions: (1) Is there any other way you would use the [dashboard] that we did not cover? and (2) Is there any other information you want to see in this [dashboard]? For each question, participants were asked to explain their answers.

We wrapped up the evaluation session with a post-study questionnaire, including the questions from the pre-study questionnaire and questions related to the dashboard's usefulness and usability. Additional evaluation questions were integrated to enhance our understanding of user experiences. These questions include: "What features are seen as useful?", "What features are missing?", "How can features be reworked to improve the supported work processes?", "Are there limitations of the current system that would hinder its adoption?" and "Is the tool understandable, and can it be learned?" All the pre- and post-study questionnaires are included in the supplemental material [76].

Recruitment. Our PD partners suggested evaluating the dashboard with community leaders, such as members from the PMC, since they are essential in managing the project, from setting community goals to taking immediate actions such as reviewing or accepting a "languishing" PR. The PMC chair first shared our study advertisement with the project committee through the project's mailing list. To recruit additional participants, the PMC chair recommended that we attend the conference held by the organization. We recruited 15 participants, 9 of whom were recruited from the conference, and the study was conducted in person.

Out of 15 community leader participants, 14 identified as men and one as a woman. This reflects the gender disparity within the community itself, as identified by the PMC chair. Their experience as community leaders ranged from < 1 year to 6–10 years. For 10 of our 15 participants, contributing to Beam was part of their primary employment.

Data analysis. We used descriptive statistics and qualitative coding [112] to analyze the data. Two researchers independently inductively-coded [113] the screen recordings and transcripts of participants using the dashboard. We used the screen recording to analyze the specific dashboard information used by participants and for what purpose. Then, the two researchers performed two rounds of negotiated agreement [114] on the three top-level categories; they then added subcategories, resulting in a final set of eight codes. After this, one researcher returned to the transcripts and re-coded them as needed.

The codebook consists of three top-level categories namely insight, explain, and action, reflecting how participants verbalized the use of

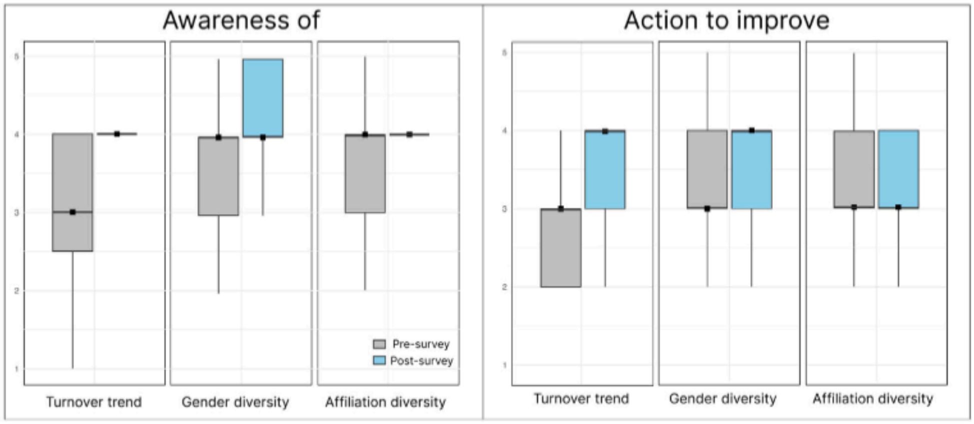

Fig. 5. Likert Scale responses for Pre-Survey and Post-Survey. Black dots: median responses. Outliers are not depicted to emphasize the overall trends.

the information found in the dashboards during their free-form explorations. For example, when participants verbalized synthesizing information from the dashboard (e.g., *"I am surprised by this one, [name of company] because I don't know it... "(P3)*), we coded such information as an INSIGHT. In cases where the participant verbalized a reason or explanation (e.g., *"...[company name] is consulting company. So they'll do some projects, and you know they will do project end to end...and they will kind of loop out..." (P1)*), we coded it as an EXPLAIN. In some cases, Community Tapestry served as a springboard for participants to take and/or think of taking future actions (e.g., *"...almost certainly that would be a good time to reach out to somebody because that's somebody who is given a lot to the project and understanding what is going on..."(P4)*), to which we assigned the code ACTION.

These three top-level categories were then enriched with 8 subcategories. First, INSIGHT and EXPLAIN were subdivided into PROJECT and CASE related. For instance, an explanation could be about a project (e.g., its culture or practices) or about a specific case that a participant was exploring in the dashboard (e.g., specific contributor who left). Second, we subdivided ACTION into three types of future actions that participants verbalized they would take (i.e., ACTION-COMMUNITY, ACTION-PERSONAL, ACTION-AFFILIATION). A fourth code, ACTION-EXPLORE, was used to identify scenarios where participants navigated outside of the Community Tapestry (e.g., GitHub profile) for additional information.

6.1. Findings

Below, we discuss the overall usefulness, turnover, and diversity awareness before and after the introduction of Community Tapestry. We then detail how participants used the information from Community Tapestry and the likelihood of using this information in the future.

6.1.1. EQ1: Overall usefulness and awareness

Awareness refers to participants' recognition of diversity and turnover issues. Usefulness, in contrast, captures participants' perception of the dashboard's utility for guiding practical decision-making and action. We analyzed the responses from two 5-point Likert scale questions on the dashboard usefulness and plan for continued use from Qiu et al. [32]. Overall, participants reported that the dashboard was useful to them (Q_usefulness: Mean = 4, SD = 0.97) and that they would continue to use it (Q_continue to use: Mean = 4.1, SD = 0.64). Community Tapestry's potential to inspire action was reflected by the post-survey responses, where most participants reported that the dashboard helped identify aspects to improve (Mean = 4.1, SD = 0.74).

Fig. 5 depicts the distributions of Likert scale responses in the pre- and post-evaluation questions about their *awareness* of turnover and diversity aspects (e.g., awareness of gender diversity) and *plan to take action* questions (e.g., I plan to take action to improve gender diversity). The distribution of responses (Fig. 5) for awareness questions (turnover, gender, affiliation) show an upward distribution shift between the pre- and post-evaluation questions, although the median (Med=4) stays the same for turnover and affiliation. Self-selection bias might be the reason behind the unchanged medians; participants who volunteered might be those who care about and are already aware of diversity in their project. There is a 1-point improvement in awareness of turnover. The increase in median awareness about turnover could be due to its unexpected nature as explained by P5, *"So that [turnover] definitely is generally like high across the boards for percent who left I wouldn't have expected it to be that high".(P5)*

When it comes to planning to take action to improve, we see distribution shift for turnover, and median shifts for gender and turnover. We interpret these behavioral intentions as evidence of the dashboard's usefulness, that is, its ability to not only raise awareness but also prompt users to consider concrete next steps. A reason for both distribution and median shifts for turnover could be its importance to project sustainability, as P4 explains *"if you notice trends of either a specific gender or a specific affiliation leaving that indicates that we might need to act in some way"*. Overall, while Community Tapestry did not have an effect on action to improve gender and affiliation diversity, it helped move the awareness distribution upward for gender, affiliation, and turnover and participants reported they planned future action to improve the turnover trend within their project.

We also conducted Wilcoxon signed-rank tests on the responses to the Pre-Survey and Post-Survey. The results showed statistically significant increases with large effect sizes in awareness of turnover trends (*Mean difference*[5] *= 0.93,p = 0.01, effect size = 0.90*) and action to address turnover (*Mean difference = 0.53,p = 0.01, effect size = 0.83*), indicating that Community Tapestry has the potential to enhance participants' recognition of and readiness to act on turnover issues. Additionally, we observed an increase in awareness of gender diversity (*Mean difference = 0.80,p = 0.03, effect size = 0.61*). While other measures did not reach statistical significance, mean differences for awareness of affiliation diversity, action on gender diversity, and action on affiliation diversity all showed a positive shift. Note, our limited sample size (N = 15) along with tied ranks can impact the results (See the supplementary material for details of the statistical analysis [76]).

6.1.2. (EQ2): Information use

We first look at how the information was used (EQ2A) during the think-aloud sessions and then the likelihood of participants using this information based on the post-study questionnaire (EQ2B).

EQ2A - Information usage. In their free form explorations, participants used the dashboard information to (1) get insights, (2) get an explanation for the insight, and (3) take action at different levels, as

[5] Mean difference = Post-Survey Responses - Pre-Survey Responses

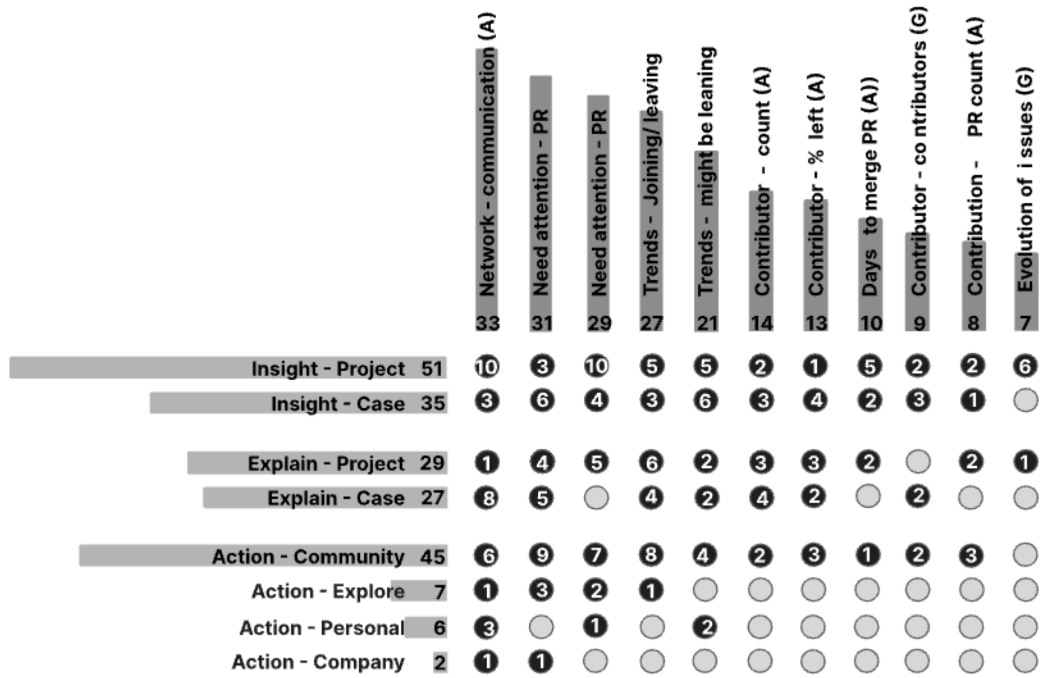

Fig. 6. Information usage. The number in the circles shows the number of participants who used [specific] information for INSIGHT, EXPLAIN, or ACTION. The vertical bar charts reflect the frequency of usage of an information source and the horizontal bar charts denote the frequency of a particular information use (e.g., INSIGHT-PROJECT) across all participants. Information sources are annotated with an (A) or (G) when that information source was disaggregated only by that aspect (Affiliation or Gender).

reflected in our codeset in Section 6. Fig. 6 depicts how participants used the dashboard information. The first dimension (vertical bar chart) shows the frequency of insights, explanations, and actions generated from a particular piece of information. The second dimension (horizontal bar chart) shows the frequency of which information was used (i.e. insights, explanations, actions).

Participants most frequently used Community Tapestry to get insights about the project (Fig. 6, rows 1 and 2) whether at the project level or related to a specific case (e.g., a specific affiliation leaving the project). This highlights one of the most important uses of Community Tapestry where aggregated data helps uncover patterns and trends that project leaders may not initially recognize. By providing a clearer, data-driven perspective, Community Tapestry facilitates a shift from an "intuition-based" approaches to more data-driven decision-making. Additionally, this data serves as a basis for deeper reflections on underlying trends and dynamics.

Actions in general and actions related to the community, were the second most frequent information usage.

Next, we present specific examples of information usage, by selecting the top three information sources in Fig. 6: (1) PR communication network (2) PRs that need attention, and (3) overall contributors joining/ leaving.

PR communication network: While navigating the dashboard, participants verbalized insights they gained about their projects. For instance, Participants (P1, P4, P6, P9, P12) realized that groups of contributors formed due to their project review culture *"it does stand out that there are these clusters of people with same affiliations reviewing each other's code" (P1)* and *"[nodes] contributing to [project], but they are not attached to the main network" (P12)*. They had a possible explanation for it: *"the main contributors to [Beam] are people inside [main affiliation]" (P12)*. Seven participants (P2, P3, P4, P5, P6, P7, P9, P14) articulated explanations for specific situations they observed. For example, P5 noted a possible reason for the clusters in the network: *"[they are] working on maybe this at the same company on the same features, and therefore you have you know, yo'are communicating more because you have more context thats shared" (P5)*.

PR that need attention: When reviewing the table of PRs that need attention, participants felt the overall table to be *"very helpful for the maintainer" (P8)*. The project context was instrumental in helping individual contributors use this information. P1, looking at a particular PR in the table, explains: *"he's [PR author] kind of actually a fish in the water in the community... He doesn't need help just because I know him. And so in this case, if this PR has not received comments, it's because he's not interested in moving it forward" (P1)*. On the other hand, when observing the same information about another contributor, P1 shared: *"out of the people that I don't know, or that I know, that are not inside the community very much, I might look at their PRs in detail" (P1)*.

The information provided by Community Tapestry inspired some participants (P1, P2, P3, P6, P12) to explore outside of the tool and get additional information such as GitHub PR. For instance, when participants noticed an active affiliation or newcomers that they did not know (P1, P3, P2, P6), or when they noticed a PR that needed attention (P1, P3, P12): *"can I check this one? I think there are comments [in the PR body in GitHub]... Oh, it is the same person commenting on himself...that doesn't count" (P12)*.

Trends of contributors joining/leaving: Participants (P4, P5, P13) shared project insights and possible explanations about their community composition in terms of gender where *"it definitely seems like statistically, there are more men that join per month the project than women" (P13)* and the trends of joining and leaving suggest the presence of one time contributors *"here the newcomer's line and the people who left line are both very long. It makes me think that there might be people that make one single contribution" (P1)*.

These insights inspired participants to reflect on how to increase retention, especially of women contributors. For instance, participants shared the need to *"connect them [newcomers] to people who have been retained previously" (P2)* and *"trying to be extra prompt on PR reviews and kind of flagging PR reviews from you know, under-represented groups, can be helpful" (P4)*, especially when *"the number of women is small enough that you can look at individual cases" (P1)*.

EQ2B - Likelihood of information usage. After using each dashboard participants provided their opinion on the likelihood (5-point Likert scale) of using the information provided in that dashboard. Overall there were 34 information sources. We grouped related information sources together into *five information sets* (Fig. 7). For example, the NEED

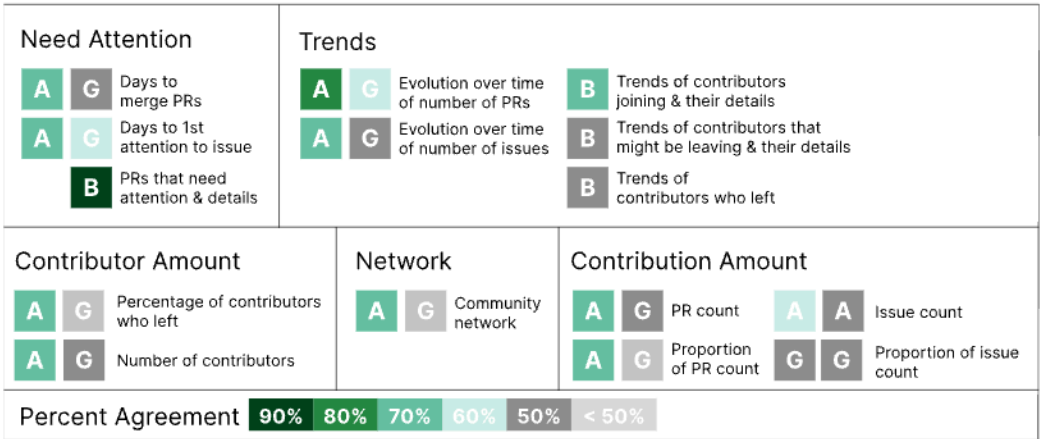

Fig. 7. Dashboard information sets ranked by usage likelihood. The likelihood is a response to Likert Scale questions about participants using different pieces of information. Each square represents an information source within Community Tapestry. "A" or "G" represent information sources disaggregated by affiliation or gender, respectively. "B" denotes both diversity aspects.

Attention set includes information such as, "days to merge PR", "days to first attention to issues" and "PR that needs attention". In Fig. 7 each box represents an information source and is annotated by "A", "G" or "B" if that information source was disaggregated by Affiliation-only, Gender-only, or both. Using participants' Likert scale responses, we calculated the percentage of participants at least "likely" to use said information (see Fig. 7).

Information that participants reported they were most likely to use was in the Need attention set. For example, 90% of participants were at least likely to use the "PRs that did not receive any comments", both for gender and affiliation disaggregations (see Fig. 7, row 3, col 1). This is probably because the information in this set allows participants to *take a remediation action.* P15 shared: *"I might have unconsciously prioritized contributions from advanced users of Beam, and I think for the project to be successful we should increase the state of affiliation diversity" (P15).* Others reported the dashboard's help in their current effort of seeking items that needed attention. P4 said:*"We already have patterns of rotating through and looking for PRs that are out of date, and this would make it a lot easier".*

trends was the next most likely to be used information. Participants used trends to identify "worrisome" patterns that would prompt them to take action: *"these numbers, of course, are interesting to see like a strong trend that in some ways is a sign to talk to people that represent such groups" (P9).* Trends were also useful to understand the impact of a change or decision. P4 said: *"overtime view is helpful for evaluating tools [migrating from Jira to Github issues]...did that have an impact on specific contributors...did that help more people from these backgrounds enter into the project" (P4).*

Participants preferred the*affiliation aspect over the gender aspect* across the information sets (Fig. 7). This might be due to the existence of some level of awareness about gender diversity in the project, as explained by P1 *"there's not that many women, so we can just look at individuals directly" (P1).* Further, multiple companies use Beam and have dedicated contributing teams where *"if there was sort of an unexplained mass exodus from a given affiliation...that might have a really big impact on the overall health of the project" (P5).*

7. Confirmatory evaluation: CNCF OpenShift

To investigate the applicability of Community Tapestry in communities beyond the one designed through our participatory design, we evaluated it with another OSS community: OpenShift (data extracted from OpenShift since January 1st, 2011) from the CNCF [115]. Such confirmatory studies help validate the findings and in our case assess to what extent a dashboard designed with Beam stakeholders could transfer to another OSS project.

We chose OpenShift for two reasons: (i) OpenShift is a larger community with more contributors than Beam, enabling us to test our dashboard on a different scale, and (ii) OpenShift is from a different foundation with a different governance structure compared to ASF. Evaluating Community Tapestry in different scenarios helps generate an in-depth understanding of its adaptability and operational capacity, giving insight into its strengths and areas for improvement in varied contexts [116].

We evaluated Community Tapestry with OpenShift using the same evaluation approach as Beam (Section 6). We piloted the study with a director of community development in OpenShift (this position is equivalent to the PMC chair position in Beam) to ensure that the study settings apply to OpenShift. Unlike BEAM, where on-duty contributors actively monitor StackOverflow activity, OpenShift's director indicated that they do not have contributors to engage with StackOverflow, and their community is not active there; thus this data is not relevant to monitor their project health management. Therefore, we excluded this data from OpenShift's analysis. However, this capability can be "turned on" for OpenShift if their needs change. All other processes, including data mining, gender inference, and affiliations classifications, followed the same approach as Beam.

Recruitment. To recruit participants with equivalent positions or roles in OpenShift, the director suggested recruiting members from community development. The director then directly contacted members of the community development team. In total, we recruited eight participants. We stopped our recruitment at eight because we began to see saturation from participants 6 onwards. That is, we analyzed the data from each study right after it was conducted and we did not find any new "codes" from the last two observation. Evaluations were conducted online via Zoom and lasted an hour.

Out of eight participants, one identified as a woman, and the other seven—as men; the gender distribution of participants is similar to Beam. Three participants are committers, and five serve as community managers and co-chairs. Their experience in the current roles ranges from less than one year to 6 to 10 years. For seven of them, contributing to OpenShift was part of their primary employment.

Data analysis. To analyze the study results from OpenShift, we followed the same approach as for Beam. When qualitatively coding data from OpenShift, no new codes emerged. While we observed minor differences in how information was used, the overall trends were similar. Since the overall results are similar, we provide the details on information usage for OpenShift (similar to Fig. 6) in the supplementary material [76]. Next, we discuss the results from OpenShift, showing the usefulness, diversity awareness, and likelihood of using Community Tapestry.

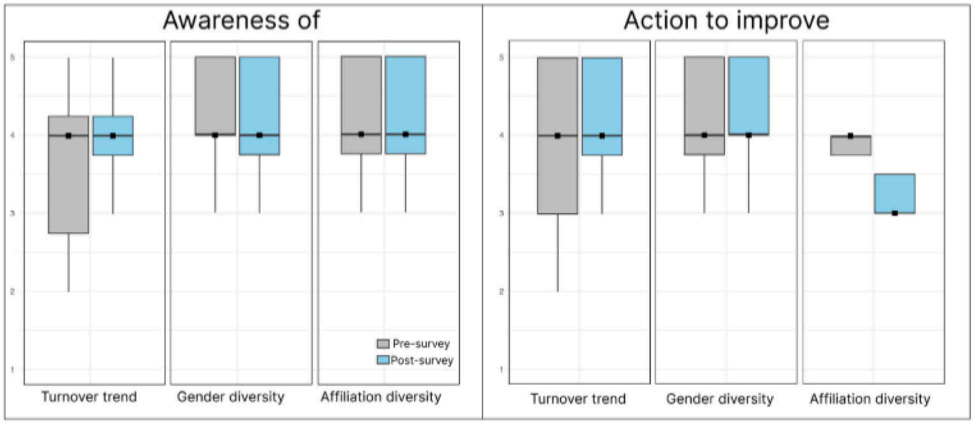

Fig. 8. Responses to the Pre-Survey and Post-Survey from OpenShift.

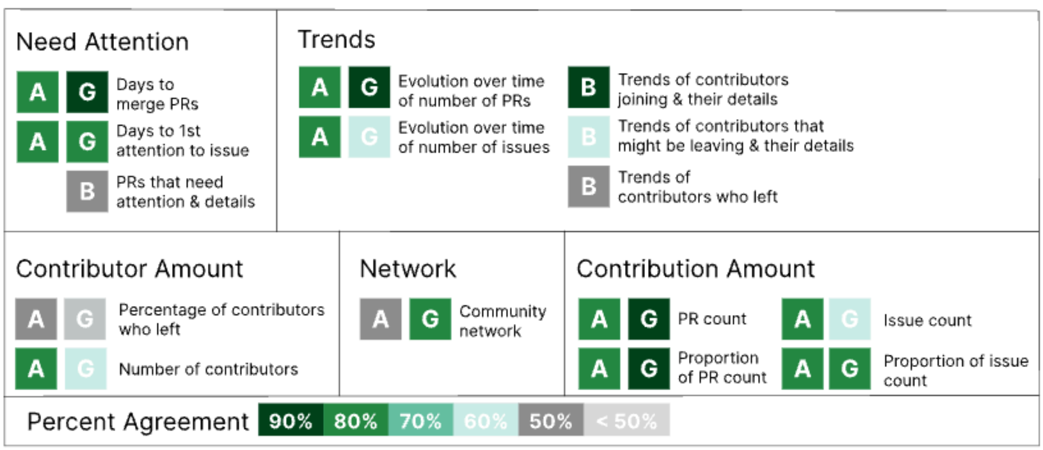

Fig. 9. Dashboard information sets ranked by usage likelihood (OpenShift).

Findings. *Overall usefulness and awareness.* In summary, participants from OpenShift agree on the overall usefulness and satisfaction of using Community Tapestry. They reported that the dashboard was useful to them (Q_usefulness: Mean = 5, SD = 0.76), helped them identify areas for improvement (Mean = 4.5, SD = 0.53), and that they would continue using it (Q_continue to use: Mean = 4.5, SD = 0.53).

Fig. 8 shows the results from OpenShift comparing the pre- and post-evaluation questions about their *awareness* of turnover and diversity aspects. Overall, the distributions of responses for awareness questions (turnover, gender, affiliation) show consistency between the pre-and post-evaluation questions; the medians remain the same at 4 (agree) for all awareness questions. OpenShift, as with many CNCF projects, is invested in creating diverse, inclusive environments. This might be the reason behind OpenShift participants being already aware of the diversity in their project. The shift in turnover awareness distribution could be because of the unexpected nature of this information: *"around [date], we have like a huge number of people leaving and still pretty much the same number of new contributors. This is interesting". (P20)* (Similar to what we saw in Project B.)

Regarding taking action to improve turnover and gender diversity, while the medians stayed the same, the distributions shifted up post-evaluation. The action to improve affiliation, however, shifted down to 3 (neutral). Participants mentioned that as OSS community members there was not much they could do to entice companies to contribute: *"...the company affiliation one, while interesting, is probably at least for me...probably less useful for me. If I'm a company person [examples of company]...that would be interesting (P17)".*

Likelihood of information usage. Fig. 9 shows the information sources. Participants were most likely to use information in the Need attention and Contribution amount, especially for gender aspect. Multiple participants expressed interest in observing the contribution amount by *"gender: men, women and unknown. Nice. Super nice Pull request count" (P20)*; *"So I would probably check this one frequently" (P16)*. In the set of trends, we observed similar trends from Need attention for gender, as "evaluation over time of a number of the PR". Over 90% of the participants were at least likely to use the "trends of contributors joining and their details": *"newcomers is always very interesting because recently I was trying to find out this data. Yeah, I had to write my own script, some tooling... I create a post on the social media accounts... I say, 'here are this months' newcomers... Thank you for your contribution.' " (P20)*

In summary, through another confirmatory study of OpenShift, we have observed agreement on the usefulness of Community Tapestry across all analysis dimensions. Just as P19 mentioned *"Yes!! That's the data I've requested couple of times for us to trump internally". (P19)*. Participants (P16, P17) suggested including geo-location as another diversity aspect (similar to the request by ASF PD partners). As we are limited by Grimoirelab infrastructure, we postpone this investigation for future work. Participants also compared Community Tapestry to existing tools they are using in their community, such as Devstats [35], acknowledging the unique value of Community Tapestry *"definitely, retention isn't that I've been looking at in the past.* and expressed their interest and satisfaction *"this is fascinating data" (P16)*, *"this is definitely useful" (P20)*, *"this is helpful for sure, to get a sense"(P19)*, *"I'm surprised...Yeah, it's very interesting". (P18)*

8. Discussions

Reflection from our findings. One of the most interesting findings from these results is the potential of Community Tapestry to inspire data-driven decision making. For example, in Project Beam participants gained considerable insights (often deemed surprising) from the dashboard and used the information to reflect on underlying trends, such as clusters in the PR communication network or high turnover rates within specific groups. In another instance, Beam participants were surprised to discover unexpectedly high percentages of contributors leaving (see Section 6.1.1, see Fig. 5), which prompted discussions about project sustainability and equity. Additionally, the dashboard's flexibility was showcased in its dashboard's ability to disaggregate data by gender and affiliation, supporting participants in focusing on issues most relevant to their community's priorities and goals. For instance, participants from Beam were able to delve into affiliation diversity and turnover trends (key to maintaining cross-company collaboration) while participants from OpenShift concentrated on gender diversity and newcomer integration.

When aligned with a community's priorities, an insight often prompts participants to provide explanations, typically by drawing on their contextual knowledge and understanding. In cases where an immediate explanation is unavailable, insights tied to community priorities tend to transition into actionable responses. These actions may include expressions of the need to address the issue or exploratory steps to gather additional data, with the ultimate goal of uncovering potential explanations. This dynamic highlights how Community Tapestry effectively bridges the gap between raw data and intention to take action by aligning insights with project priorities. Examples include prioritizing PR reviews for contributors from underrepresented groups and fostering connections between newcomers and existing contributors, which is a direct outcome of community and inclusivity efforts. The tool's ability to adapt seamlessly across two structurally and operationally distinct OSS communities (Apache and CNCF) further demonstrates its robustness, offering valuable support in varying governance and operational contexts while maintaining its focus on equity and sustainability [117].

This work significantly advances the field of collaborative software development by demonstrating how data-driven tools like Community Tapestry empower collaborative communities to address structural challenges and enhance equity and sustainability. The fine-grained analysis of how OSS project leaders use the dashboard—emphasizing actionable insights and reactions to surprising trends—offers valuable guidance on designing solutions that drive meaningful action. For example, participants used the tool to uncover patterns like high turnover or uneven communication, translating these insights into tailored interventions.

The research highlights the adaptability of Community Tapestry across diverse OSS ecosystems, such as Apache and CNCF, illustrating how a single tool can be leveraged effectively by communities with different structures and priorities. This adaptability underscores its value in enhancing software engineering practices, contributing to the broader goal of improving software development processes, quality, and sustainability.

Reflections from participatory design. The key principles of PD include mutual learning between users and designers and reflecting on the product's development process [39]. To support our understanding of the organization's needs, we selected our PD partners to include a wider set of stakeholders than just members of Beam. Each of our PD partners had different priorities and concerns, and our discussions helped us reach a collaborative decision. At the same time, we had to consider the constraints of our infrastructure provider, Bitergia Analytics, who had a (paid) contract with the ASF. For example, geo-locations as an information source were particularly interesting to community managers. However, integrating them into GrimoireLab's infrastructure proved unfeasible.

As researchers, we gained a deeper understanding of the conflicting priorities between paid and volunteer contributors in a hybrid project. For example, Beam's PMC chair had to navigate the "politics" of reviewer cliques and their reflection on the project, especially to those outside the company. There was a concern about whether collaboration patterns (e.g., whose PR gets reviewed by whom and how quickly) now visible through the dashboard could hurt morale and sponsorship, both of which impact retention. We had long conversations about what types of demographic information could be inferred, if they should be surfaced in the dashboard, and whether the dashboard should be public. Ultimately, the dashboard visibility was restricted to the PMC, potentially opening the dashboard to those with commit access. Whether the dashboard will be public will depend on how the community adopts it.

Finally, our conversations made us realize that the information extracted from the dashboard, its signals, and its fit within the future users' process was more important than the dashboard itself. Thus, we focused a large portion of our evaluation on this topic.

The ASF is continuing to promote project health. Through user studies conducted on two projects within prominent OSS communities, we gained valuable insights into the effectiveness of Community Tapestry. This tool enhances community awareness and facilitates strategic actions for the project involved in PD and projects operating under different community structures. The ASF community has not only endorsed the use of Community Tapestry in Beam but is also moving to its adoption across various projects. This decision underscores a dedication to leveraging Community Tapestry to enhance project health and suggests its potential for broader application. As one community leader remarked, "I have been looking for this for a long time" (P17).

In Section 6 on evaluation design, we discuss the complexity of assessing the Community Tapestry's effectiveness in fostering a healthy, sustainable community. This endeavor necessitates navigating various challenges and considering multiple confounding variables inherent to these community environments. Ongoing discussions with PD partners are focused on exploring the feasibility of conducting future impact evaluation through a counterbalance study.

Limitations. Like all studies, ours carries risks. However, we have taken reasonable steps to minimize these potential threats. Our decision to use PD may affect the generalizability of our findings. Other communities with different demographic distributions may prioritize different project health aspects and find different information to be helpful. However, we believe that PD allowed us to deeply understand the organizational context and needs of Beam, which outweighed PD's drawbacks. Moreover, similar issues may also arise concerning contribution metrics, particularly as our data collections did not include contributions that are not traceable on GitHub, such as organizing events and answering questions on StackOverflow, which is often unacknowledged [47,118]. To address these challenges, both the contribution metrics and project health aspects were developed based on discussions with community leaders. For example, it was highlighted that contributions to StackOverflow, which are examples of non-traditional contributions, are frequently overlooked for Beam. This insight shows the importance of adapting Community Tapestry to the different cultures of different projects. We also evaluated Community Tapestry with prospective users from two projects within two OSS communities. This approach strengthened our findings by highlighting the tool's adaptability and providing an understanding of different community structures.

To mitigate potential reliability issues, we used validated questions and questions from prior literature when possible. We sandboxed and piloted the evaluation instruments. We acknowledge that the only reliable way to know one's gender identity is to ask. Since such an approach does not scale, we used algorithmic tools such as NamSor. We reduced potential errors to the greatest extent possible (see Section 3). Similar reliability issues may be raised with affiliation. We manually checked 384 contributors' affiliation profiles (95% confidence interval, 5% margin error of 25,000 contributors across two communities) and ran a reliability test; our accuracy is 89%.

We acknowledge that gender classification is a double-edged sword. On one hand, gender identity is a complex and multifaceted social construct as both automated and manual classification methods are inherently limited, often relying on heuristics such as names or regions. To balance out these considerations, our approach provides flexibility for future updates, allowing the PMC to refine classifications based on their contextual knowledge of the project and its cultural norms. Additionally, our system is designed with privacy and sensitivity in mind. The dashboard is accessible only to the PMC and is not publicly visible. This ensures gender classification aligns with ethical standards and IRB requirements. For future deployments in different communities, we acknowledge that regulations and cultural norms regarding gender and location monitoring may vary.

Moreover, interventions like Community Tapestry, which support community leaders in monitoring OSS project health and sustainability, should consider regulations such as the Cyber Resilience Act (CRA), which emphasizes maintenance and security for the supply chain [119]. Our study, approved by IRB, aims to minimize privacy risks by using only public data. Additionally, for aggregated data, we implemented secure login systems [120], restricting access to PMC members.

In this study, we focused on contributor turnover and diversity (gender and organizational affiliation) as priority metrics to help community leaders understand project health, which are prioritized through PD discussions with all stakeholders. However, we acknowledge that many other metrics can be applied to understand a project's health [121], as diversity in OSS is a multifaceted concept that arises from differences in contributors' gender, seniority, language, region, and other characteristics [72], which could also be explored. Our PD approach provides insights for future studies to innovate and support OSS community's health.

9. Conclusion

In this paper, we report our experiences and outcomes in designing, developing, and evaluating a dynamic, daily-updated dashboard called Community Tapestry. Community Tapestry prioritizes three project health aspects: contributor turnover, gender diversity, and organizational affiliation diversity. Our evaluation of Community Tapestry with future users from the two most active and large communities demonstrates its usefulness and its capability to enhance proactive actions. *"This is nice, to kinda go and reach out and say, "hey, what's going on?" That'll be nice. And newcomers are always very interesting because recently I was trying to find out this data".(P20)*. Participants recognized the dashboards' potential not only in the short term but also in the long run, as highlighted by P7: *"I think this is like the best dashboard for overall project help that I would look at and again gives me the chance to see the results of targeted interventions most clearly" (P7).*

CRediT authorship contribution statement

Mariam Guizani: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Zixuan Feng:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Emily Arteaga:** Formal analysis, Data curation. **Katie Kimura:** Writing – review & editing, Formal analysis, Data curation. **Diane Mueller:** Data curation. **Luis Cañas Díaz:** Software, Resources, Project administration. **Alexander Serebrenik:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Investigation. **Anita Sarma:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank all participants in our study, with special gratitude to the ASF BEAM and CNCF OpenShift project teams for their participation and support. This work was partially supported by NSF Grant No. 2303043 and NSERC Discovery Grant RGPIN-2024-06511.

Data availability

The authors do not have permission to share data.

References

[1] F. Ferreira, L.L. Silva, M.T. Valente, Turnover in open-source projects: The case of core developers, in: Proceedings of the XXXIV SBES, 2020, pp. 447–456.

[2] C. Miller, D.G. Widder, C. Kästner, B. Vasilescu, Why do people give up flossing? a study of contributor disengagement in open source, in: Open Source Systems: 15th IFIP WG 2.13 International Conference, Montreal, QC, Canada, Proceedings 15, Springer, 2019, pp. 116–129.

[3] P.N. Sharma, J. Hulland, S. Daniel, Examining turnover in open source software projects using logistic hierarchical linear modeling approach, in: Open Source Systems: Long-Term Sustainability: 8th IFIP WG 2.13 International Conference, Hammamet, Tunisia. Proceedings 8, Springer, 2012, pp. 331–337.

[4] A. Bosu, K.Z. Sultana, Diversity and inclusion in open source software (OSS) projects: Where do we stand? in: ACM/IEEE ESEM, IEEE, 2019, pp. 1–11.

[5] B. Vasilescu, D. Posnett, B. Ray, M.G. van den Brand, A. Serebrenik, P. Devanbu, V. Filkov, Gender and tenure diversity in GitHub teams, in: Proceedings of the 33rd Annual ACM CHI, 2015, pp. 3789–3798.

[6] G. Robles, L.A. Reina, J.M. González-Barahona, S.D. Domínguez, Women in free/libre/open source software: The situation in the 2010s, in: IFIP International Conference on Open Source Systems, Springer, 2016, pp. 163–173.

[7] H.S. Qiu, Z.H. Zhao, T.K. Yu, L. Dabbish, B. Vasilescu, How much do women build open source infrastructure? in: D. Damian, K. Blincoe, D. Ford, A. Serebrenik, Z. Masood (Eds.), Equity, Diversity, and Inclusion in Software Engineering: Best Practices and Insights, A Press, 2024, pp. 231–253.

[8] D. Izquierdo-Cortazar, G. Robles, F. Ortega, J.M. Gonzalez-Barahona, Using software archaeology to measure knowledge loss in software projects due to developer turnover, in: 42nd HICSS, IEEE, 2009, pp. 1–10.

[9] M. Zhou, A. Mockus, X. Ma, L. Zhang, H. Mei, Inflow and retention in OSS communities with commercial involvement: A case study of three hybrid projects, ACM TOSEM 25 (2) (2016) 1–29.

[10] P. Hynninen, A. Piri, T. Niinimäki, Off-site commitment and voluntary turnover in GSD projects, in: 5th IEEE International Conference on Global Software Engineering, IEEE, 2010, pp. 145–154.

[11] Y. Yu, A. Benlian, T. Hess, An empirical study of volunteer members' perceived turnover in open source software projects, in: 45th HICSS, IEEE, 2012, pp. 3396–3405.

[12] A. Mockus, Organizational volatility and its effects on software defects, in: Proceedings of the 18th ACM SIGSOFT FSE, 2010, pp. 117–126.

[13] M. Ortu, G. Destefanis, S. Counsell, S. Swift, R. Tonelli, M. Marchesi, How diverse is your team? Investigating gender and nationality diversity in GitHub teams, J. Softw. Eng. Res. Dev. 5 (1) (2017) 1–18.

[14] B. Trinkenreich, I. Wiese, A. Sarma, M. Gerosa, I. Steinmacher, Women's participation in open source software: A survey of the literature, ACM TOSEM 31 (4) (2022) 1–37.

[15] M. Guizani, A. Chatterjee, B. Trinkenreich, M.E. May, G.J. Noa-Guevara, L.J. Russell, G.G. Cuevas Zambrano, D. Izquierdo-Cortazar, I. Steinmacher, M.A. Gerosa, A. Sarma, The long road ahead: Ongoing challenges in contributing to large OSS organizations and what to do, Proc. ACM Hum.-Comput. Interact. 5 (CSCW2) (2021) http://dx.doi.org/10.1145/3479551.

[16] G.A.A. Prana, D. Ford, A. Rastogi, D. Lo, R. Purandare, N. Nagappan, Including everyone, everywhere: Understanding opportunities and challenges of geographic gender-inclusion in oss, IEEE Trans. Softw. Eng. 48 (9) (2021) 3394–3409.

[17] S. Balali, I. Steinmacher, U. Annamalai, A. Sarma, M.A. Gerosa, Newcomers' barriers... is that all? an analysis of mentors' and newcomers' barriers in OSS projects, Comput. Support. Coop. Work (CSCW) 27 (3–6) (2018) 679–714.

[18] H.S. Qiu, A. Nolte, A. Brown, A. Serebrenik, B. Vasilescu, Going farther together: The impact of social capital on sustained participation in open source, in: ACM/IEEE ICSE, ICSE '19, IEEE Press, Piscataway, NJ, USA, 2019, pp. 688–699.

[19] M. Guizani, A.A. Castro-Guzman, A. Sarma, I. Steinmacher, Rules of engagement: Why and how companies participate in OSS, in: 45th IEEE/ACM ICSE, Melbourne, Australia, IEEE, 2023, pp. 2617–2629.

[20] Y. Zhang, M. Zhou, A. Mockus, Z. Jin, Companies' participation in OSS development–an empirical study of openstack, IEEE Trans. Softw. Eng. 47 (10) (2019) 2242–2259.

[21] E. Murphy-Hill, J. Dicker, M.M. Hodges, C.D. Egelman, C. Jaspan, L. Cheng, E. Kammer, B. Holtz, M.A. Jorde, A.K. Dolan, et al., Engineering impacts of anonymous author code review: A field experiment, IEEE Trans. Softw. Eng. 48 (7) (2021) 2495–2509.

[22] T. Ahmed, A. Bosu, A. Iqbal, S. Rahimi, Senticr: A customized sentiment analysis tool for code review interactions, in: 32nd IEEE/ACM ASE, IEEE, 2017, pp. 106–111.

[23] E. Murphy-Hill, J. Dicker, D. Carlson, M. Harbach, A. Murillo, T. Zhou, Did Gerrit's respectful code review reminders reduce comment toxicity? in: D. Damian, K. Blincoe, D. Ford, A. Serebrenik, Z. Masood (Eds.), Equity, Diversity, and Inclusion in Software Engineering: Best Practices and Insights, A Press, 2024, pp. 309–321.

[24] M. Burnett, S. Stumpf, J. Macbeth, S. Makri, L. Beckwith, I. Kwan, A. Peters, W. Jernigan, GenderMag: A method for evaluating software's gender inclusiveness, Interact. Comput. 28 (6) (2016) 760–787.

[25] M. Guizani, I. Steinmacher, J. Emard, A. Fallatah, M. Burnett, A. Sarma, How to debug inclusivity bugs? A debugging process with information architecture, in: Proceedings of the ACM/IEEE 44th ICSE-SEIS, in: ICSE-SEIS '22, Association for Computing Machinery, New York, NY, USA, 2022, pp. 90–101, http://dx.doi.org/10.1145/3510458.3513009.

[26] I. Santos, J.F. Pimentel, I. Wiese, I. Steinmacher, A. Sarma, M.A. Gerosa, Designing for cognitive diversity: Improving the github experience for newcomers, in: IEEE/ACM 45th ICSE-SEIS, IEEE, 2023, pp. 1–12.

[27] E. Murphy-Hill, A. Elizondo, A. Murillo, M. Harbach, B. Vasilescu, D. Carlson, F. Dessloch, GenderMag improves discoverability in the field, especially for women, in: IEEE/ACM 46th ICSE, IEEE Computer Society, 2024, 973–973.

[28] M.M. Hamid, A. Chatterjee, M. Guizani, A. Anderson, F. Moussaoui, S. Yang, I. Escobar, A. Sarma, M. Burnett, How to measure diversity actionably in technology, 2023, Equity, Diversity, and Inclusion in Software Engineering: Best Practices and Insights.

[29] D.A. Tamburri, F. Palomba, A. Serebrenik, A. Zaidman, Discovering community patterns in open-source: a systematic approach and its evaluation, Empir. Softw. Eng. 24 (2019) 1369–1417.

[30] M. Guizani, B. Trinkenreich, A.A. Castro-Guzman, I. Steinmacher, M. Gerosa, A. Sarma, Perceptions of the state of D&I and D&I initiative in the ASF, in: Proceedings of the ACM/IEEE 44th ICSE-SEIS, 2022, pp. 130–142.

[31] CHAOSS Community, Metrics for event organizers - CHAOSS, 2021, CHAOSS URL https://chaoss.community/metrics-for-event-organizers/.

[32] H.S. Qiu, A. Lieb, J. Chou, M. Carneal, J. Mok, E. Amspoker, B. Vasilescu, L. Dabbish, Climate coach: A dashboard for open-source maintainers to overview community dynamics, in: Proceedings of the CHI, 2023, pp. 1–18.

[33] M. Guizani, T. Zimmermann, A. Sarma, D. Ford, Attracting and retaining OSS contributors with a maintainer dashboard, in: Proceedings of the ACM/IEEE 44th ICSE-SEIS, 2022, pp. 36–40.

[34] A. Ramchandran, L. Yin, V. Filkov, Exploring apache incubator project trajectories with APEX, in: Proceedings of the 19th International Conference on Mining Software Repositories, 2022, pp. 333–337.

[35] Cloud Native Computing Foundation, DevStats: CNCF's development analytics, 2023, Available online at https://devstats.cncf.io/.

[36] Bitergia, Bitergia: Analytics for open source development, 2024, https://bitergia.com/. (Accessed 18 November 2024).

[37] C. Spinuzzi, The methodology of participatory design, Tech. Commun. 52 (2) (2005) 163–174.

[38] S. Dueñas, V. Cosentino, J.M. Gonzalez-Barahona, A.d.C. San Felix, D. Izquierdo-Cortazar, L. Cañas-Díaz, A.P. García-Plaza, GrimoireLab: A toolset for software development analytics, PeerJ Comput. Sci. 7 (2021) e601.

[39] D. Schuler, A. Namioka, Participatory Design: Principles and Practices, CRC Press, 1993.

[40] G. Rodríguez-Pérez, R. Nadri, M. Nagappan, Perceived diversity in software engineering: a systematic literature review, Empir. Softw. Eng. 26 (5) (2021) 102, http://dx.doi.org/10.1007/s10664-021-09992-2.

[41] B. Lin, A. Serebrenik, Recognizing gender of stack overflow users, in: Proceedings of the 13th International Conference on Mining Software Repositories, 2016, pp. 425–429.

[42] S. van Breukelen, A. Barcomb, S. Baltes, A. Serebrenik, "Still around": Experiences and survival strategies of veteran women software developers, in: 45th IEEE/ACM International Conference on Software Engineering, ICSE 2023, Melbourne, Australia, May 14-20, 2023, IEEE, 2023, pp. 1148–1160, http://dx.doi.org/10.1109/ICSE48619.2023.00103.

[43] M.B. Miranda, R. Prikladnicki, The challenges of ethnic-racial diversity within the IT sector, in: D. Damian, K. Blincoe, D. Ford, A. Serebrenik, Z. Masood (Eds.), Equity, Diversity, and Inclusion in Software Engineering: Best Practices and Insights, A Press, 2024, pp. 37–54.

[44] N. Munga, T. Fogwill, Q. Williams, The adoption of open source software in business models: a Red Hat and IBM case study, in: Proceedings of the SAICSIT, 2009, pp. 112–121.

[45] K. Mouakhar, A. Tellier, How do Open Source software companies respond to institutional pressures? A business model perspective, J. Enterp. Inf. Manag. (2017).

[46] Y. Zhang, H. Liu, X. Tan, M. Zhou, Z. Jin, J. Zhu, Turnover of companies in OpenStack: Prevalence and rationale, ACM TOSEM 31 (4) (2022) 1–24.

[47] Z. Feng, A. Chatterjee, A. Sarma, I. Ahmed, A case study of implicit mentoring, its prevalence, and impact in apache, in: Proceedings of the 30th ACM ESEC/FSE, 2022, pp. 797–809.

[48] S.P. Goggins, M. Germonprez, K. Lumbard, Making open source project health transparent, Computer 54 (8) (2021) 104–111.

[49] I. Steinmacher, T. Conte, M.A. Gerosa, D. Redmiles, 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, 2015, pp. 1379–1392.

[50] I. Steinmacher, A.P. Chaves, T. Conte, M.A. Gerosa, Preliminary empirical identification of barriers faced by newcomers to open source software projects, in: Proceedings of the 28th SBES, SBES '14, IEEE Computer Society, 2014, pp. 51–60.

[51] I. Steinmacher, I. Wiese, A.P. Chaves, M.A. Gerosa, Why do newcomers abandon open source software projects? in: 6th International Workshop on Cooperative and Human Aspects of Software Engineering, CHASE, IEEE, 2013, pp. 25–32.

[52] G. Pinto, I. Steinmacher, M.A. Gerosa, More common than you think: An in-depth study of casual contributors, in: IEEE International Conference on Software Analysis, Evolution, and Reengineering, SANER, in: SANER 2016, vol. 1, IEEE, 2016, pp. 112–123.

[53] I. Steinmacher, G. Pinto, I.S. Wiese, M.A. Gerosa, Almost there: A study on quasi-contributors in open-source software projects, in: 2018 IEEE/ACM 40th ICSE, IEEE, 2018, pp. 256–266.

[54] J.M. Levine, H.-S. Choi, Impact of personnel turnover on team performance and cognition, in: E. Salas, S. Fioren (Eds.), Team Cognition: Understanding the Factors that Drive Process and Performance, American Psychological Association, 2004, pp. 153—-176.

[55] J.M. Levine, R.L. Moreland, L. Argote, K.M. Carley, Personnel Turnover and Team Performance, Technical Report, PITTSBURGH UNIV PA, 2005.

[56] A. Chatterjee, M. Guizani, C. Stevens, J. Emard, M.E. May, M. Burnett, I. Ahmed, AID: An automated detector for gender-inclusivity bugs in OSS project pages, in: IEEE/ACM 43rd ICSE, IEEE, 2021, pp. 1423–1435.

[57] I. Steinmacher, T.U. Conte, C. Treude, M.A. Gerosa, Overcoming open source project entry barriers with a portal for newcomers, in: Proceedings of the 38th ICSE, 2016, pp. 273–284.

[58] J. Austgen, A. Fábrega, S. Allen, K. Babel, M. Kelkar, A. Juels, Dao decentralization: Voting-bloc entropy, bribery, and dark daos, 2023, arXiv preprint arXiv:2311.03530.

[59] T. Sharma, Y. Potter, K. Pongmala, H. Wang, A. Miller, D. Song, Y. Wang, Unpacking how decentralized autonomous organizations (daos) work in practice, in: 2024 IEEE International Conference on Blockchain and Cryptocurrency, ICBC, IEEE, 2024, pp. 416–424.

[60] DeepDAO, Discovery, listing and analytics for all DAOs, 2025, URL https://deepdao.io/organizations. (Accessed 27 May 2025).

[61] S. Bødker, C. Dindler, O.S. Iversen, R.C. Smith, What is participatory design? in: Participatory Design, Springer, 2022, pp. 5–13.

[62] M. Zhou, A. Mockus, What make long term contributors: Willingness and opportunity in OSS community, in: ACM/IEEE ICSE, ICSE '12, IEEE Press, 2012, pp. 518–528.

[63] K. Yamashita, Y. Kamei, S. McIntosh, A.E. Hassan, N. Ubayashi, Magnet or sticky? Measuring project characteristics from the perspective of developer attraction and retention, J. Inf. Process. 24 (2) (2016) 339–348.

[64] I. Steinmacher, M.A.G. Silva, M.A. Gerosa, D.F. Redmiles, A systematic literature review on the barriers faced by newcomers to open source software projects, Inf. Softw. Technol. 59 (2015) 67–85.

[65] CHAOSS Project, GrimoireLab sigils: Gerrit retention and newcomers panel, 2024, URL https://chaoss.github.io/grimoirelab-sigils/panels/gerrit-retention-newcomers/. (Accessed 10 June 2025).

[66] A. Rastogi, N. Nagappan, G. Gousios, Geographical Bias in GitHub: Perceptions and Reality, Technical Report, IIIT-Delhi, 2016.

[67] H. Carter, J. Groopman, Diversity, equity, and inclusion in open source: Exploring the challenges and opportunities to create equity and agency across open source ecosystems, 2021, The Linux Foundation.

[68] The Apache Software Foundation, The apache software foundation diversity, equity, and inclusion (DEI) report 2023, 2023, URL https://www.apache.org/foundation/docs/2023DEIReport.pdf. (Accessed 20 March 2025).

[69] L. Quillian, A. Heath, D. Pager, A.H. Midtbøen, F. Fleischmann, O. Hexel, Do some countries discriminate more than others? Evidence from 97 field experiments of racial discrimination in hiring, Sociol. Sci. 6 (2019) 467–496.

[70] Apache Software Foundation, 2023 Diversity, Equity, and Inclusion Report, 2023, https://apache.org/foundation/docs/2023DEIReport.pdf. (Online; accessed 20 March 2024).

[71] D. Ford, A. Harkins, C. Parnin, Someone like me: How does peer parity influence participation of women on stack overflow? in: IEEE VL/HCC, IEEE, 2017, pp. 239–243.

[72] Z. Feng, M. Guizani, M.A. Gerosa, A. Sarma, The state of diversity and inclusion in apache: A pulse check, 2023, arXiv preprint arXiv:2303.16344.

[73] Education First, EF english proficiency index (EPI), 2024, https://www.ef.com/wwen/epi/. (Accessed 18 November 2024).

[74] M.M. Andreasen, C.T. Hansen, P. Cash, Conceptual Design, Springer, Cham,Switzerland, 2015.

[75] S. Few, Information Dashboard Design: The Effective Visual Communication of Data, O'Reilly Media, Inc., 2006.

[76] M. Guizani, Z. Feng, E. Arteaga, K. Kimura, D. Mueller, L.C. Díaz, A. Serebrenik, A. Sarma, Community Tapestry: An Actionable Tool to Track Diversity and Turnover in OSS, Zenodo, 2025, http://dx.doi.org/10.5281/zenodo.15061207.

[77] V. Pipek, V. Wulf, Infrastructuring: Toward an integrated perspective on the design and use of information technology, J. Assoc. Inf. Syst. 10 (5) (2009) 1.

[78] GrimoireLab, Perceval: Fetch data from software repositories, 2021, Accessed 8 May 2023.

[79] M.S. Wessel, A. Serebrenik, I. Wiese, I. Steinmacher, M.A. Gerosa, Effects of adopting code review bots on pull requests to OSS projects, in: ICSME, IEEE, 2020, pp. 1–11, http://dx.doi.org/10.1109/ICSME46990.2020.00011.

[80] T. Dey, S. Mousavi, E. Ponce, T. Fry, B. Vasilescu, A. Filippova, A. Mockus, Detecting and characterizing bots that commit code, in: Proceedings of the 17th International Conference on Mining Software Repositories, 2020, pp. 209–219.

[81] N. Chidambaram, A. Decan, T. Mens, Distinguishing bots from human developers based on their GitHub activity types, in: CEUR Workshop Proceedings, Presented at the Seminar on Advanced Techniques and Tools for Software Evolution (SATToSE), 2023.

[82] CHAOSS, GrimoireLab: Open-source software development analytics toolset, 2025, URL https://github.com/chaoss/grimoirelab. (Accessed 27 May 2025).

[83] I.S. Wiese, J.T. Da Silva, I. Steinmacher, C. Treude, M.A. Gerosa, Who is who in the mailing list? comparing six disambiguation heuristics to identify multiple addresses of a participant, in: 2016 IEEE ICSME, IEEE, 2016, pp. 345–355.

[84] A. Serebrenik, How to ask about gender identity of software engineers and "guess" it from the archival data, in: Equity, Diversity, and Inclusion in Software Engineering: Best Practices and Insights, Apress Berkeley, CA, 2024, pp. 487–505.

[85] E. Carsenat, Inferring gender from names in any region, language, or alphabet, 2019, http://dx.doi.org/10.13140/RG.2.2.11516.90247.

[86] P. Sebo, Performance of gender detection tools: a comparative study of name-to-gender inference services, J. Med. Libr. Assoc.: JMLA 109 (3) (2021) 414.

[87] A. Serebrenik, How to ask about gender identity of software engineers and "guess" it from the archival data, in: D. Damian, K. Blincoe, D. Ford, A. Serebrenik, Z. Masood (Eds.), Equity, Diversity, and Inclusion in Software Engineering: Best Practices and Insights, A Press, 2024, pp. 487–505.

[88] P. Sebo, How accurate are gender detection tools in predicting the gender for Chinese names? A study with 20,000 given names in Pinyin format, J. Med. Libr. Assoc.: JMLA 110 (2) (2022) 205.

[89] P. Sebo, NamSor's performance in predicting the country of origin and ethnicity of 90, 000 researchers based on their first and last names, Sci. Data 9 (1) (2022) 1–7, http://dx.doi.org/10.1038/s41597-022-01166-4.

[90] M.M. Taylor, I. Taylor, Writing and Literacy in Chinese, Korean and Japanese, John Benjamins Publishing Company, 2014.

[91] M. Burnett, A. Peters, C. Hill, N. Elarief, Finding gender-inclusiveness software issues with GenderMag: A field investigation, in: CHI, ACM, 2016, pp. 2586–2598.

[92] M. Burnett, R. Counts, R. Lawrence, H. Hanson, Gender HCI and microsoft: Highlights from a longitudinal study, in: IEEE Symposium on Visual Languages and Human-Centric Computing, IEEE, 2017, pp. 139–143.

[93] M. Burnett, S.D. Fleming, S. Iqbal, G. Venolia, V. Rajaram, U. Farooq, V. Grigoreanu, M. Czerwinski, Gender differences and programming environments: Across programming populations, in: Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM, ACM, Bolzano-Bozen, Italy, 2010, pp. 1–10, http://dx.doi.org/10.1145/1852786.1852796.

[94] J. Margolis, A. Fisher, Unlocking the Clubhouse: Women in Computing, MIT Press, 2002.

[95] M.M. Burnett, L. Beckwith, S. Wiedenbeck, S.D. Fleming, J. Cao, T.H. Park, V. Grigoreanu, K. Rector, Gender pluralism in problem-solving software, Interact. Comput. (2011).

[96] R. Riedl, M. Hubert, P. Kenning, Are there neural gender differences in online trust? an fMRI study on the perceived trustworthiness of eBay offers, MIS Q. (2010).

[97] J. Meyers-Levy, B. Loken, Revisiting gender differences: what we know and what lies ahead, J. Consum. Psychol. (2015).

[98] T. Dohmen, A. Falk, D. Huffman, U. Sunde, J. Schupp, G.G. Wagner, Individual risk attitudes: measurement, determinants, and behavioral consequences, J. Eur. Econ. Assoc. (2011).

[99] G. Charness, U. Gneezy, Strong evidence for gender differences in risk taking, J. Econ. Behav. Organ. (2012).

[100] L. Beckwith, C. Kissinger, M. Burnett, S. Wiedenbeck, J. Lawrance, A. Blackwell, C. Cook, Tinkering and gender in end-user programmers' debugging, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, Montréal, Québec, Canada, 2006, pp. 231–240, http://dx.doi.org/10.1145/1124772.1124809.

[101] J. Cao, K. Rector, T.H. Park, S.D. Fleming, M. Burnett, S. Wiedenbeck, A debugging perspective on end-user mashup programming, in: Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing, VL/HCC, IEEE, 2010, pp. 149–156.

[102] C. Hilderbrand, C. Perdriau, L. Letaw, J. Emard, Z. Steine-Hanson, M. Burnett, A. Sarma, Engineering gender-inclusivity into software: ten teams' tales from the trenches, in: Proceedings of the ACM/IEEE 42nd ICSE, 2020, pp. 433–444.

[103] C. Bossen, C. Dindler, O.S. Iversen, Evaluation in participatory design: a literature survey, in: Proceedings of the 14th Participatory Design Conference: Full Papers-Volume 1, 2016, pp. 151–160.

[104] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, S. Carpendale, Empirical studies in information visualization: Seven scenarios, IEEE Trans. Vis. Comput. Graphics 18 (9) (2011) 1520–1536.

[105] X. Wang, Z. Cheng, Cross-sectional studies: strengths, weaknesses, and recommendations, Chest 158 (1) (2020) S65–S71.

[106] S. Kujala, T. Miron-Shatz, J.J. Jokinen, The cross-sequential approach: a short-term method for studying long-term user experience, J. Usability Stud. 14 (2) (2019) 105–116.

[107] The Apache Software Foundation, Apache software foundation events, 2024, https://events.apache.org/. (Accessed 18 November 2024).

[108] F. Janoos, S. Singh, O. Irfanoglu, R. Machiraju, R. Parent, Activity analysis using spatio-temporal trajectory volumes in surveillance applications, in: 2007 IEEE Symposium on Visual Analytics Science and Technology, IEEE, 2007, pp. 3–10.

[109] R. Eccles, T. Kapler, R. Harper, W. Wright, Stories in geotime, Inf. Vis. 7 (1) (2008) 3–17.

[110] H. Song, E.P. Curran, R. Sterritt, Multiple foci visualisation of large hierarchies with FlexTree, Inf. Vis. 3 (1) (2004) 19–35.

[111] J. Pater, A. Coupe, R. Pfafman, C. Phelan, T. Toscos, M. Jacobs, Standardizing reporting of participant compensation in HCI: A systematic literature review and recommendations for the field, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, Association for Computing Machinery, Yokohama, Japan, 2021, pp. 1–16.

[112] G. Gibbs, Analyzing Qualitative Data, London, England, 2007.

[113] S.H. Khandkar, Open coding, Univ. Calg. 23 (2009) (2009).

[114] N. McDonald, S. Schoenebeck, A. Forte, Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice, Proc. the ACM Human-Computer Interact. 3 (CSCW) (2019) 1–23.

[115] Cloud Native Computing Foundation, Cloud native computing foundation (CNCF), 2024, https://www.cncf.io/. (Accessed 18 November 2024).

[116] P. Runeson, M. Host, A. Rainer, B. Regnell, Case Study Research in Software Engineering: Guidelines and Examples, John Wiley & Sons, 2012.

[117] P. Lago, N. Condori Fernandez, I. Fatima, M. Funke, I. Malavolta, The sustainability assessment framework toolkit: a decade of modeling experience, Softw. Syst. Model. (2024) 1–23.

[118] B. Trinkenreich, M. Guizani, I. Wiese, A. Sarma, I. Steinmacher, Hidden figures: Roles and pathways of successful OSS contributors, Proc. the ACM Human-Computer Interact. 4 (CSCW2) (2020) 1–22.

[119] P.G. Chiara, The cyber resilience act: the EU commission's proposal for a horizontal regulation on cybersecurity for products with digital elements: An introduction, Int. Cybersecur. Law Rev. 3 (2) (2022) 255–272.

[120] C. Payne, On the security of open source software, Inf. Syst. J. 12 (1) (2002) 61–78.

[121] CHAOSS Project, Diversity and inclusion metrics, 2017, https://github.com/chaoss/wg-diversity-inclusion/tree/master/focus-areas. (Online; Accessed 30 December 2020).