

2023-12-19

Evaluation of Volunteering Capabilities in an Open-Source Software Community

Hariharan, Aadharsh

Hariharan, A. (2023). Evaluation of volunteering capabilities in an open-source software community (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from <https://prism.ucalgary.ca.https://hdl.handle.net/1880/117791>

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Evaluation of Volunteering Capabilities in an Open-Source Software Community

by

Aadharsh Hariharan

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN ELECTRICAL ENGINEERING

CALGARY, ALBERTA

DECEMBER, 2023

© Aadharsh Hariharan 2023

Abstract

Open-source software is a cornerstone of modern technology. Embodying principles of transparency, collaboration, and innovation, it nurtures a vibrant ecosystem that empowers individuals, businesses, and communities. Open-source software has impacted software development significantly; the longevity of open-source projects is essential to the entire field of software development. Challenges faced by open-source software communities include the management of contributors, effective utilization of them, retention of existing contributors, and recruitment of new contributors. For projects where most contributors are volunteers – which remains the case for several projects such as Gnome, Perl, and Python – attracting and retaining volunteers becomes crucial to success. Crowston (2011) argued that because of the high mobility of knowledge workers, even paid employees require personal motivation to participate in projects. In this sense, they should also be viewed as volunteers.

Numerous studies explore the dynamics of open-source communities and volunteer contributions. This research has yielded models to assess the volunteering prowess of open-source software communities, and proposed solutions to address challenges. However, most studies have taken a collective approach, encompassing multiple open-source software communities, which presents a generalized perspective. Utilizing a fusion of quantitative and qualitative techniques, this research project gauges the degree of relevance and applicability of existing theories, models, and solutions within the unique context of the Perl and Raku community. This case study offers valuable insights into the community's existing skills, capabilities, and resources available for constructive contributions to growth and

development. These insights are instrumental in identifying and implementing strategies to attract and retain volunteers within the community. Conflict within communities can be a significant factor in retaining volunteers, and the latter portion of the thesis emphasizes identifying techniques to address these challenges.

Preface

This thesis is original, unpublished, independent work by the author, A. Hariharan. The experiments reported in Chapters 4-6 were covered by Ethics Certificate REB20-2135_REN2, issued by the University of Calgary Conjoint Health Ethics Board for the project “Evaluation of Volunteering Capabilities in an Open-Source Software Community” on February 25, 2023.

Declaration

I solemnly affirm that I am the sole author of this thesis, and it has been crafted independently, devoid of any unauthorized external aid or recourse to materials beyond those explicitly specified. Any concepts or content obtained directly or indirectly from external sources have been duly acknowledged and appropriately attributed within this work.

The research presented in this thesis spans 2021 to 2023, conducted under the mentorship of Dr. Ann Barcomb at the University of Calgary. Dr. Barcomb also served as the second coder for qualitative data analysis in one of the interviews.

Segments of this research project were conducted in collaboration with external entities, most notably The Perl and Raku Foundation, which played a pivotal role in data collection and developing adapted techniques to benefit the Perl and Raku community. I received invaluable assistance from Andrea Gail Tampus, who acted as the secondary coder for the remaining interviews.

Auburn Phillips provided indispensable support by meticulously proofreading my thesis and offering constructive feedback, instrumental in refining and formatting the document into its final stage.

Acknowledgements

I would like to extend my heartfelt appreciation to my supervisor, Dr. Ann Barcomb, for affording me the invaluable opportunity to expand my knowledge in a nurturing academic environment, and for her unwavering belief in my capabilities. Her guidance and support have been indispensable throughout my academic journey at the University of Calgary.

I am deeply thankful to The Perl and Raku Foundation for their close collaboration with our research project, generously providing essential data and resources that played a pivotal role in the success of our endeavour.

Special gratitude is also owed to Andrea Gail Tampus, whose role as the secondary coder for our qualitative analysis of interviews significantly expedited the data analysis process. Her contribution is sincerely appreciated.

I want to express my profound gratitude to my family members: my mother Usha, my father Hariharan, and my sister Arya, who have been a steadfast source of emotional support, particularly during challenging times.

Lastly, I thank all my lab mates and friends for their unwavering camaraderie and support throughout my academic journey. Your presence has made this journey all the more enriching.

Table of Contents

ABSTRACT	II
PREFACE	IV
DECLARATION	V
ACKNOWLEDGEMENTS.....	VI
TABLE OF CONTENTS	VII
LIST OF TABLES.....	IX
LIST OF FIGURES.....	X
LIST OF ABBREVIATIONS	XI
CHAPTER 1: INTRODUCTION	1
1.1 MOTIVATION	1
1.2 RESEARCH OBJECTIVES.....	3
CHAPTER 2: BACKGROUND	6
2.1 PERL AND RAKU COMMUNITY.....	6
2.2 MODEL OF RETENTION	8
CHAPTER 3: LITERATURE REVIEW	11
3.1 OPEN-SOURCE SOFTWARE COMMUNITY LIFECYCLE AND SUSTAINABILITY	11
3.2 OPEN-SOURCE SOFTWARE COMMUNITY AND MANAGEMENT	15
3.3 DYING AND REVITALIZED COMMUNITIES	17
CHAPTER 4: METHODOLOGY.....	19
4.1 METHODOLOGICAL APPROACH.....	19
4.2 SAMPLING STRATEGY.....	20
4.3 DATA COLLECTION	20
4.4 DATA ANALYSIS	25
CHAPTER 5: QUANTITATIVE DATA ANALYSIS AND RESULTS	32
5.1 DATA QUALITY	32
5.2 REPRESENTATIVENESS.....	34
5.3 PLS-PM	37
5.4 PLS-PM MODEL EVALUATION.....	40
5.5 PLS-PM HYPOTHESES TESTING AND EXPLORATION.....	43
5.6 SUMMARY	47
CHAPTER 6: QUALITATIVE DATA ANALYSIS AND RESULTS.....	49
6.1 CODEBOOK	49
6.2 CONTRIBUTOR BENEFIT MOTIVATION	50
6.3 SOCIAL NORMS	52
6.4 PSYCHOLOGICAL SENSE OF COMMUNITY	53
6.5 SATISFACTION	55
6.6 COMMUNITY COMMITMENT	58

6.7 SUMMARY	59
CHAPTER 7: IDENTIFICATION OF EXISTING TECHNIQUES	61
7.1 POPULARITY OF THE COMMUNITY	62
7.2 AVAILABILITY OF VOLUNTEERS (RETENTION OF VOLUNTEERS)	63
7.3 AVAILABILITY OF VOLUNTEERS (RECRUITMENT OF VOLUNTEERS)	64
7.4 CHALLENGES FACED BY BEING A PART OF THE COMMUNITY	66
CHAPTER 8: ADAPTATION OF TECHNIQUES FOR PERL AND RAKU	70
8.1 MARKETING AND PROMOTIONS.....	70
8.2 FORMAL COMMUNITY STRUCTURE.....	71
8.3 ABANDONING OF MODULES	71
8.4 RECRUITMENT AND RETENTION OF VOLUNTEERS.....	72
8.5 FUNDING	73
CHAPTER 9: DISCUSSION.....	74
9.1 COMPARISON TO PREVIOUS STUDIES.....	74
9.2 LIMITATIONS.....	80
9.3 FUTURE WORK.....	86
CHAPTER 10: CONCLUSION	88
REFERENCES.....	90
APPENDICES.....	95
APPENDIX A: SURVEY QUESTIONS.....	95
APPENDIX B: INTERVIEW QUESTIONS	100
APPENDIX C: ETHICS APPROVAL CERTIFICATE.....	105
APPENDIX D: INTERVIEW CONSENT FORM.....	106
APPENDIX E: SURVEY RECRUITMENT	110
APPENDIX F: INTERVIEW RECRUITMENT	110
APPENDIX G: CODEBOOK	111

List of Tables

<i>Table 3.1: Survey Questions for Each Construct.....</i>	<i>22</i>
<i>Table 5.1: Representativeness of the members of the community.....</i>	<i>35</i>
<i>Table 5.2: Convergent Validity</i>	<i>40</i>
<i>Table 5.3: Internal Consistency Reliability</i>	<i>42</i>
<i>Table 5.4: Effects Between Latent Variables.....</i>	<i>43</i>
<i>Table 5.5: Bootstrapping result.....</i>	<i>44</i>
<i>Table 5.6: Cluster Segmentation</i>	<i>46</i>

List of Figures

<i>Fig 3.1: Pictorial Representation of the Model of Retention.....</i>	<i>10</i>
<i>Fig 4.1: Flowchart depiction of the qualitative coding process.....</i>	<i>29</i>
<i>Fig 4.2: Illustration of the Qualitative Coding Process.....</i>	<i>31</i>
<i>Fig 5.1: Pictorial Representation of the PLSPM model.....</i>	<i>38</i>

List of Abbreviations

AVE	Average Variance Extracted
GOF	Goodness of Fit
REBUS	Response Based Unit Segmentation
RO1	Research Objective 1
RO2	Research Objective 2
RO3	Research Objective 3
TPRF	The Perl and Raku Foundation

Chapter 1: Introduction

1.1 Motivation

It is widely recognized that software has revolutionized technology, profoundly influencing the technological landscape. This has led to numerous innovations once considered unattainable. As a cornerstone of Engineering, Science, and Technology, it is worth exploring what supports and sustains software. Among the various critical verticals representing essential focus areas, one stands out for its pivotal role in establishing software as a global technological foundation: open-source software. Open-source software is the fundamental framework of modern software development, underpinning its vitality and innovation. A 2008 report by the Standish Group revealed that open-source models resulted in substantial annual savings, amounting to nearly \$60 billion for consumers. More recently, the European Commission reported that open-source software had contributed to direct cost savings of approximately €114 billion annually within the European economy (Schleife & Niemann, 2015). Considering the continued growth of open-source solutions, it is plausible this figure has increased in the current context.

The Open-Source Initiative¹ defines *open-source software* as software that can be freely accessed, used, changed, and shared (in modified or unmodified form) by anyone and is made by many people and distributed under licenses that comply with the Open-Source Definition. Other key characteristics of open-source software include:

¹ <https://opensource.org/osd/>

- Transparency - the source code of open-source software is accessible to anyone, allowing for increased transparency in how the software works and operates.
- Customization – open-source software can be tailored to meet specific needs, as the source code is available for users to modify and improve.
- Collaboration - open-source software development relies on a large, global community of developers working together to build, maintain, and improve the software.
- Innovation - the open-source development model encourages experimentation, as developers can suggest and implement new features and improvements.

Open-source software is often self-managed by communities. This research project focused on examining the self-managed Perl and Raku community, notable for its absence of a single dominant corporate entity. To facilitate collaboration and engagement with the community, this study partnered with The Perl and Raku Foundation (TPRF), as it oversees and supports the Perl and Raku community.

Volunteers constitute a significant part of such self-managed communities and looking at all community volunteers (not just core contributors) helps increase understanding, because episodic and periphery volunteers are vital to communities (Hyde et al., 2016). While Hyde emphasized the significance of episodic and periphery volunteers in a study conducted beyond the realm of open-source software, Barcomb et al. (2018) reinforced this observation within the open-source software context.

Episodic volunteering is when a person contributes infrequently or irregularly. Contributors often make a single contribution, then become inactive. Hyde et al. (2016) devised distinct metrics for studying episodic volunteering, which served as the foundation for the Model of

Retention used by Barcomb et al. (2017) in the context of episodic volunteering in open-source software. Acknowledging the shared motivations between habitual and episodic volunteers, this research project applied the model comprehensively to the volunteering landscape of the Perl and Raku community. The goal was to develop strategies addressing issues such as the limited availability of volunteers and funding shortages identified in the process.

The contributions of this research study are:

- Validating and exploring the extent to which general findings specifically apply to the Perl and Raku community.
- Offering a nuanced examination of both episodic and habitual volunteers' participation dynamics and behaviours within this distinct context.
- Examining the adaptability and generalizability of existing solutions tailored to address issues within open-source software communities, shedding light on the extent to which these solutions can be effectively applied to a specific community, and identifies necessary adaptations for compatibility.
- Adding to the literature on declining communities, an area with limited research. While the contribution may be modest in scope, it addresses a gap in the existing literature.

1.2 Research Objectives

This research project's objective is to assess the applicability of existing literature findings, analytical techniques, and methodologies within a specific community. This encompasses a comprehensive understanding of the entire volunteering landscape, including episodic and habitual participation, within a community. The research identifies strategies for tackling the

challenges within the Perl and Raku community, including adapting specific techniques to align more closely with the unique context of this community.

The research project was broadly classified into three phases, each aimed at addressing a specific research objective. The first phase involved data collection through surveys, interviews, and public information. Data was analyzed based on past theories and methodologies to evaluate the current situation of the open-source community, pinpoint the exact issue plaguing it, and understand the applicability of these theories when studying a specific open-source software community. This formulated the first research objective:

RO1: Understanding the applicability of current theories for gaining insight into the volunteering problems faced by an open-source software community.

The second phase of the research involves determining techniques to help solve the issues identified in the first phase, and the relevance of such practices in the context of a specific open-source software community. This primarily includes past-developed techniques and new ones specifically required for the community. This gives the second research objective:

RO2: Understanding the relevance of recommended solutions in the existing literature to the problems identified for an open-source software community.

The third phase considers implementation of the techniques derived in phase two and adapting and modifying the techniques to suit the Perl and Raku community. This could be considered the

final aspect of RO2, but to provide clearer structure, it became the third and final research objective:

RO3: Understand the adaptations that might be required on the identified techniques from existing literature to fit an open-source software community specifically.

To fulfil RO1, a survey was undertaken to collect data, allowing for the implementation of existing models and theories from the literature. Survey analysis was crucial to this phase. Interviews were conducted to explore insights from survey data, identifying pertinent techniques and solutions from existing literature, thus addressing RO2. Finally, as part of RO3, a series of discussions were conducted with a community representative to tailor the techniques specifically to the community's needs.

These three research objectives help understand the validity of past theories and methodologies when applied to a specific community and its entire volunteering situation (episodic and habitual volunteers). It also helps understand the Perl and Raku community better and solve some of their significant issues.

Chapter 2: Background

This chapter provides an overview of certain aspects of the research. This includes additional information about the Perl and Raku community and the implemented model, which is valuable for comprehending the research study.

2.1 Perl and Raku community

The Perl and Raku community is a vibrant and diverse community of developers and enthusiasts who share a passion for the Perl and Raku programming languages. While Perl and Raku are distinct languages, they share a common heritage and have an overlapping community due to their historical connection. Perl was developed by Larry Wall in 1987 as a general-purpose Unix scripting language to make report processing easier (Wall, 1994). Over the years, Perl has transformed into a language that currently serves the needs of many businesses. Raku, formerly known as Perl 6, is a separate language that evolved from Perl but with a distinct syntax and features (Lenz, 2020).

This community represents a volunteer-driven ecosystem without major corporate sponsorship, making it a noteworthy case study. Its self-governance by dedicated volunteers aligns with the study's focus on decentralized decision-making, collaborative dynamics, conflict resolution, and community governance within specific open-source software communities. Additionally, the close affiliation of the principal investigator, Dr. Ann Barcomb, with the Perl and Raku community granted invaluable access and insight.

While the Perl and Raku community overlap, there has been conflict and divergence between the two communities. The conflict primarily arose in early 2000 when the decision was

taken to develop a newer version of Perl, (i.e., Perl 6) and this was meant to be a rewrite of the codebase to address some of the significant issues identified in the previous version (i.e., Perl 5). The development continued for an extended period. It became apparent that the new codebase would not be backward-compatible with Perl 5 but would become a new language. In 2019, the version of Perl that was supposed to be Perl 6 was renamed as a new language called “Raku” (Lenz, 2020). This led to conflict and differences of opinion within the Perl community regarding the direction and future of the language. While some members embraced Raku as a natural progression and improvement upon Perl, others expressed concerns about the fragmentation of the community and Perl in its current version being outdated due to the long development period of Raku. Division emerged between the Perl and Raku community. Some Perl enthusiasts remained focused on Perl 5 and continued to develop and promote the language. Meanwhile, the Raku community formed its own separate identity, actively developing and promoting Raku as a distinct programming language. Nevertheless, TPRF manages both Perl and Raku. Since the conflict between Perl and Raku began, the community has experienced a decline, which was observed during data collection.

The Perl and Raku Foundation² is a non-profit organization dedicated to advancing the Perl and Raku programming languages. It provides support and resources to the Perl and Raku community, including developers, users, and enthusiasts. The primary goals of TPRF include promoting the use of Perl and Raku, supporting development of Perl and Raku projects and tools,

² <https://www.perlfoundation.org/>

and ensuring long-term viability of the Perl and Raku language. They are central to community management, serving as an elected board of directors.

2.2 Model of Retention

A crucial part of the analysis was structuring the survey data using a model and finding correlations. The Model of Retention (Barcomb et al., 2020) was developed to fit a similar type of data and study. This model was derived from the study conducted by Hyde et al. (2016), where a model for episodic volunteer retention looked at non-profit organizations, and Barcomb et al. (2020) successfully applied it to episodic volunteering in open-source software. Numerous models exist for assessing volunteering in open-source software communities, including Stewardship Theory (Davis et al., 1997), Social Exchange Theory (Emerson, 1976), and the Community of Practice Framework (Wenger, 1998). However, none of these models are tailor-made for the specific needs of volunteer retention, intervention, and retention strategy development as comprehensively as the Model of Retention. It is important to note that the superiority of a model hinges on the specific context of its application. For future studies considering the Model of Retention, it is recommended that this application context be adequately reviewed before deciding. In the context of this research study, which strongly emphasizes volunteer retention, the Model of Retention emerged as the more fitting choice due to its specialized focus on this aspect.

The Model of Retention maps five constructs to the **Intention to Remain (IR)**. Hyde et al. (2016) developed the constructs based on understanding volunteer motivations and their nature. The five constructs are:

- **Contributor Benefit Motivation (CBM):** The personal benefits a volunteer might gain or expect to gain when contributing to the community. This construct is primarily evident before joining the community.
- **Social Norms (SN):** The pressure the participant faces in their social environment outside of the community being studied. This construct can be evident at any point in joining or participating in the community.
- **Psychological Sense of Community (PSC):** The feeling of similarity and comfortability the individual experiences when joining the community. It is the emotional attachment and connection a person feels towards a community. This construct is primarily evident when entering the community.
- **Satisfaction (S):** The sense of fulfilment or pleasure one gets from participating in the community. This construct is primarily evident after joining the community.
- **Community Commitment (CC):** The extent to which the individual identifies with the community and sees themselves as a member of the community. This construct is primarily evident after joining the community.

The survey utilized the Likert scale to measure the different constructs of the model of retention. A Likert scale is a commonly used psychometric tool for measuring individuals' attitudes, opinions, and perceptions. It provides a structured way to assess subjective experiences and opinions by asking respondents to rate their level of agreement or disagreement with a set of statements or questions (Likert, 1932).

The survey sections (Appendix A) aimed to capture information related to the constructs of the Model of Retention. These sections were structured based on the 5-item Likert scale, in which

respondents are presented with five response options and are tasked with selecting the one that best aligns with their perspective. Response options encompassed "*Strongly Agree*," "*Agree*," "*Neither Disagree Nor Agree*," "*Disagree*," and "*Strongly Disagree*". Employing the Likert scale, the responses could be quantified. A more in-depth review of the survey questions is explained in Chapter 4 (section 4.3).

The model of retention is pictorially represented in Fig 3.1:

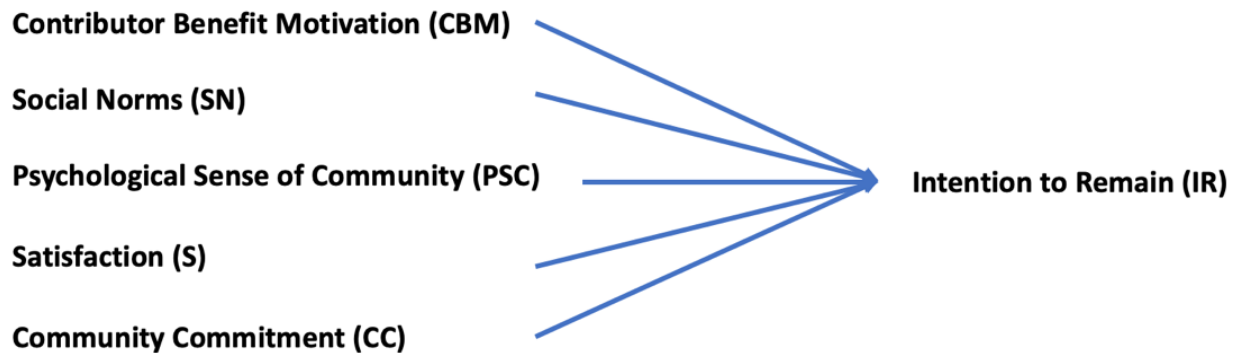


Fig 3.1: Pictorial Representation of the Model of Retention

Chapter 3: Literature Review

Three areas of literature inform this study: the lifecycle and health of open-source software projects and communities, the recruitment, management, and retention of contributors in open-source software, and the decline and revitalization of communities. Initially robust, the Perl and Raku community has been declining over the years. Comprehending these topics helps illuminate the issues faced in the community concerning volunteers and provides context for applying the practices to the Perl and Raku community.

3.1 Open-Source Software Community Lifecycle and Sustainability

Like any organization, open-source software communities experience various phases and transitions influenced by multiple factors. Guimarães et al. (2013) discuss how the lifecycle of an open-source community can be represented as an “Inverted U-shape”, showing five stages of the lifecycle: startup, growth, maturity, decline, and death. As the community grows from its initial startup phase, the coordination of activities must become more effective. It is vital as a community expands for the community coordinators to identify essential issues in the domain, improve communication strategies, document, foster relationships, maintain the knowledge base, and assess the community’s overall health. Failing to do this can result in the community eventually reaching the downward slope of the inverted U-shaped lifecycle.

Guimarães et al. (2013) also explore an alternative curve for tracking the community lifecycle based on activity levels. Their research illustrates a distinctive “Inverted S-shape” curve that reflects fluctuations in activity throughout the different stages of a community’s lifecycle. In the initial startup phase, activity levels are notably high as members invest significant effort in

launching their projects. As the community matures, activity levels tend to decrease. However, during the decline stage, activity is resurgent, possibly due to revitalization efforts. If these efforts fail, activity rapidly declines, forming the characteristic inverted S-shape curve. While their research introduces distinct states in the open-source software community lifecycle, it does not assert that these states are predetermined or irreversible. In fact, the paper articulates that when both curves are considered, the analysis portrays a dynamic perspective where communities can transition between lifecycle phases, oscillating back and forth. This ability allows them to regain significance, draw in fresh contributors and users, and effectively rejuvenate themselves. This dynamic view contrasts with the notion of irreversibility often associated with organismic lifecycle models.

There is a notable gap in the existing literature regarding techniques for transitioning from decline back to previous states, such as growth or maturity. The Perl and Raku community has been in a prolonged state of decline, but delving into the underlying causes of this decline may offer insights that could lead to its reversal or reinvigoration. Rather than evaluate activity levels of the Perl and Raku community's lifecycle stage, this study focuses on comprehending the current state of the community and delving into the underlying reasons for this decline.

Robles and González-Barahona (2012) also examine how open-source software may come into existence, which can be closely linked to the lifecycle of a community. Initially, people come together on specific projects, which attract curious minds from different backgrounds, who finally become contributors to the project. As the project becomes more popular, different adaptations arise, causing further expansion. As community popularity increases, a governing structure and

collaborative environment must be implemented. Finally, the attraction of volunteers from diverse backgrounds fuels sustainability.

Foulonneau et al. (2013) propose studying the lifecycle of an open-source software community by analyzing communication data of that community. Activity levels on the established communication channel can be used to discover different stages of the community lifecycle such as growth, maturity, and decline. Activity levels in the communication channels provide a visual assessment of the date of decline when measuring the community lifecycle. This approach could be used for analyzing the chat logs from the Perl and Raku community's official communication channels, such as Slack or Facebook groups. This was not conducted as it seemed likely to simply reaffirm that the community is in a state of decline.

Wynn (2004) considers open-source software community lifecycles in the framework of a project lifecycle. Wynn evaluates the framework by including the concepts of labour, a hierarchy of authority, decentralization, and management processes; and studying how these vary through the various stages of the community lifecycle. For example, during the introduction and growth stage, labour is more prominent to expand the scope of the initial project. As the community expands, the hierarchy of authority and management becomes more critical. During the maturity stage, decentralization is most prominent. During the stage of decline or revitalization, labour becomes the most critical factor. When applying Wynn's theory or framework to the context of the Perl and Raku community, it becomes evident that the community is experiencing a decline phase. This conclusion arises from the initial data analysis, emphasizing a significant challenge faced by the community, notably the scarcity of available volunteers. Additionally, the ranking of

the language indicates it is in a declining phase. Perl, once a prevalent language, now ranks 27th on the Tiobe Index³ for top programming languages in the world.

Nyman and Lindman (2013) summarize the sustainability of open-source software in two phrases: "A strong community offers better insurance of sustainability of the software level," but at the same time, "better software can more easily attract a bigger community". A poorly managed community or a poorly constructed project can be the bane to the sustainability of open-source software. Foster (2015) explores different factors that contribute to open-source software community sustainability, such as governance and leadership, community engagement and participation, document and knowledge management, finance and resource sustainability, long-term planning and vision, and diversity and inclusivity. Foster proposes a model that provides a framework for understanding open-source software sustenance and contributing factors, while Curto-Millet and Jiménez (2022) look at sustainability as multifaceted and interrelated. They propose five themes of sustainability:

- Participation
- Focus on certain actors
- Time
- Dimensions (through which problems of sustainability are made explicitly)
- Logic

While participation has been a critical factor for various other studies, the other four indicators are unique to this framework. The presence of influential actors strongly indicates community sustainability: the community assumes a pivotal role by creating an environment that nurtures

³ <https://www.tiobe.com/tiobe-index>.

collaboration, while individuals give back to its vitality. Time is an indicator in that, as long as communication remains positive and stable, it does not harm community sustainability.

Curto-Millet and Jiménez describe three dimensions of community sustenance: economic, existential, and physical. Threats to any one of these can affect community sustainability. Finally, Curto-Millet and Jiménez propose two logics for open-source software: competition and collaboration. Collaboration is when the individuals of the community value social good over other factors, and drive development towards high-quality software; this is generally considered the boon of sustainability. Competition occurs when volunteer actions may be erratic, driven by prospect of challenging other projects; this is the bane of sustainability.

On viewing the Perl and Raku community through the frameworks discussed, it is evident that volunteer participation, finance and resource availability, and economic and existential dimensionality are negatively affecting the sustainability of the community. Time indicates the decline of the community over the years, which is also evident from the fall in rankings of the language. However, upon examining the logics put forth by Curto-Millet and Jiménez (2022), it is clear that this community thrives on collaboration rather than competition to fuel its development, a factor that positively contributes to sustainability. This may explain why the Perl and Raku community has not faded away entirely.

3.2 Open-Source Software Community and Management

Certain open-source software development is facilitated and managed by communities. Often, these communities consist of groups of people who band together and volunteer time and expertise (Bonaccorsi & Rossi, 2003). Volunteers can be geographically spread out and convene through virtual methods (Crowston et al., 2005; Dabbish et al., 2012).

Open-source software communities have been described as structures formed through a knowledge-intensive and non-commercial conducive, resulting in "exemplars of knowledge sharing" (Iskoujini & Roberts, 2015) or a social network based on the connections between individuals with an initial common interest (Xu et al., 2005). For such organizations, volunteer management is essential. This is evident when Shah (2006) discusses the significance of effective management practices that foster motivation and collaboration among volunteers.

Barcomb et al. (2020) bifurcate volunteers into two categories based on the frequency of their contributions, i.e., habitual, and episodic volunteers. Habitual volunteers contribute more consistently, and episodic volunteers offer sporadic or solitary contributions. While habitual volunteers may form a core part of the community, episodic volunteers are also vital (but more challenging to manage). Cnaan & Goldberg-Glen (1991) describe motivation and loyalty, which can be translated to commitment, as the main reasons for the contributors becoming habitual. Cnaan et al. (2021) describe the main areas where episodic volunteers are more prominent: significant events, periodically rotating groups, and limited-scope activities. Although the paper acknowledges that the impact episodic volunteering can leave on the volunteer itself is uncertain, it is generally estimated that any type of volunteering positively impacts the volunteer. Another bifurcation seen in open-source software communities, as described by Crowston et al. (2012), is based on the level of involvement in the community: Core and Periphery volunteers. Core volunteers contribute to the central workings of the community and are highly involved, while periphery volunteers are not deeply involved in the community and may contribute sporadically.

Barcomb et al. (2020) conducted a Delphi Study similar to this project, on a larger scale of many open-source software communities while concentrating more on episodic volunteering.

The Delphi Study is a research method used to gather and synthesize expert opinions and insights on a particular topic or issue. It is typically employed in situations where there is uncertainty or a lack of consensus. Barcomb et al. were able to adapt and implement generic volunteering theories and methodologies on open-source software communities, identify the significant concerns with episodic volunteering and propose practices to overcome these issues.

This research project mainly uses Barcomb et al.'s (2020) Delphi Study as a reference while conducting the research project. While the Delphi Study was at a much broader scale, this study uses the same methodologies and theories to validate the applicability when implemented on the specific Perl and Raku community. The Delphi study concentrated on episodic volunteers, while this project studies both episodic and habitual volunteers.

3.3 Dying and Revitalized Communities

Very little literature is available about dying or revitalized communities, regardless of whether one looks at open-source software or volunteerism in general. Available literature indicates that a major problem in open-source software communities is that it is easy for the volunteer base to shrink and eventually die out unless retention techniques are used. The reverse is also possible, with communities rapidly expanding. Eckert & Muller (2017) state “software isn’t sustainable without users” and describe the vital relationship between open-source software communities and their users.

Another significant issue is sustainability, especially after a fork, when a project divides into two streams, which then evolve into something completely different (Gamalielsson and Lundell, 2012). This is relevant even though the Perl and Raku community has not experienced a fork but has witnessed the emergence of factions due to conflicts between Perl and Raku. While forking

has always been considered troubling to the sustainability of open-source software communities, some communities have succeeded (Ven and Maennert, 2008). Such a fork occurred in 2010 when a faction splintered from the OpenOffice Community, establishing The Document Foundation, which subsequently embarked on the development of LibreOffice as a fork derived from the OpenOffice codebase. While OpenOffice continues to be maintained and developed, the fork that created LibreOffice has been thriving ever since, gaining prominence and community support in recent years (Gamalielsson & Lundell, 2014).

The literature review yielded insights that significantly enhanced understanding of open-source software and communities. At the outset of the research, there was little to no prior knowledge of the Perl and Raku community. However, the literature review proved instrumental, offering valuable guidance on how to approach and assess the community. One crucial observation emerged: preliminary indications suggest that the community may be experiencing a decline. This revelation was pivotal in shaping the research design and its execution. While scholarly resources on communities that have successfully revitalized after experiencing a decline remain scarce, these research endeavours hold promise in shedding light on this relatively underexplored area.

Chapter 4: Methodology

This chapter discusses methods used in each phase of the research project.

4.1 Methodological Approach

In this study, a mixed methodological approach was used. Mixed methods research involves combining both qualitative (interview) and quantitative (survey) research methods within a single study. This was conducted to gain a more comprehensive understanding by combining the strengths of both methods. Surveys and interviews complement each other and allow for triangulation, such as cross-verifying results from different data sources. This increases the reliability and validity of the findings. Surveys provide an understanding of *what* is amiss within the community or the constructs influencing it, while interviews offer insights into *why* these issues are occurring and their impact on the community. The integration of *what* and *why* provides profound insight into the volunteering landscape in the community, enhancing the effectiveness of proposed techniques and solutions to address challenges.

Surveys were conducted to understand on a collective basis the trends and patterns of the community. Surveys provide a systematic way of gathering information and understanding the perspectives and behaviours of a specific group or population. Interviews were conducted to gain more personal insight into the community. Interviews are valuable for collecting in-depth qualitative information from individuals, and a more nuanced understanding of people's experiences, perspectives, and opinions. This gives a more profound understanding of the issues affecting the community and helps in making the techniques to resolve these issues more effective.

Survey questions were designed to gather specific insights based on the five constructs of the Model of Retention. The survey was launched on 15th July 2021 and was kept live until 30th September 2021. Following the survey, interviews were conducted to collect qualitative data. While the interview script was initially prepared based on the Model of Retention, a quick analysis of the survey data gave insights on adapting some questions or asking follow-up questions to derive more information.

The primary engagement with the community took place through the Perl and Raku Foundation, including facilitating the distribution of the survey and recruiting interview participants from the community. After identifying techniques from existing literature, a report was compiled with a predominant focus on the results rather than the methodology. This report was subsequently submitted to TPRF and later disseminated throughout the entire community. The adaptation of techniques outlined in Chapter 8 also involved collaboration with a representative from the Perl and Raku Foundation.

4.2 Sampling Strategy

Survey and interview data were collected through the Perl and Raku community, targeting current members. Both the survey and the interview recruitment message were marketed through the TPRF website and official Slack and Facebook communication channels, and any member of the Perl and Raku community was welcome to participate.

4.3 Data Collection

The data collected in surveys and interviews give powerful insights into the volunteering situation of the Perl and Raku community. The following details the metrics of the data collected:

4.3.1 Surveys

A survey was shared with community members (see Appendix E for survey recruitment message). The survey was designed to capture information about parameters people consider when volunteering and whether they decide to continue volunteering. The Perl and Raku Foundation promoted the survey through their official website, various dedicated community Slack channels, and Facebook groups. Of the 142 responses received, 88 complete responses could be used. The survey was completely anonymized unless the participant chose to disclose certain information, such as their website or GitHub information.

Survey questions (Appendix A) were carefully structured in alignment with the Model of Retention discussed in Chapter 2 (Section 2.2) and were adapted from the EV'17 survey conducted by Barcomb et al. (2018), which drew from earlier literature (e.g., Hyde et al., 2016). Questions were customized to the specific context of open-source software.

The first two sections of the survey gathered general information about participants and their volunteering activities, such as the duration of their involvement as open-source software contributions, the nature of their contributions (whether episodic or habitual), and their specific type of contribution (ranging from code contributions to documentation and translation, or even a combination of these). The next section captured demographic information (gender identification, educational background, and age) to assess the community's representativeness, as reflected in Table 5.1.

Subsequent segments of the survey were tailored to collect information specific to each construct within the retention model. A series of questions explored different aspects related to the community, as shown in Table 3.1. The section commenced with an exploration of

Psychological Sense of Community, questions gauging community members' psychological perceptions, such as *"If I feel like talking, I can generally find someone in the community to talk to right away."* The inquiry shifted to Social Norms, investigating whether contributors experienced societal pressures to engage with the community. For instance, participants were asked questions like *"Other people think that contributing is important to me."*

The survey transitioned to Satisfaction, where participants expressed their sentiments by agreeing or disagreeing with statements like *"I enjoy my contribution experience."* Next, Community Commitment utilized inquiries such as *"I am proud to tell others that I am part of this community,"* measuring the loyalty participants felt toward the community.

The final segment delved into the Contributor Benefit Motivation construct. Questions such as *"I participate to make money"* probed primary motivations behind participants' contributions. This structured questionnaire, coupled with inquiries targeting Intention to Remain (e.g., *"I plan to participate in this community in the future"*), systematically gathered vital data pertaining to the study's constructs.

Table 3.1: Survey Questions for Each Construct

Psychological Sense of Community
I am quite similar to most people in my community
If I feel like talking, I can generally find someone in the community to talk to right away
If there was a serious problem in the community, the people could get together to solve it
Social Norms
Other people think that contributing is important to me
It is important to my friends and relatives that I continue contributing
Many of the people that I know expect me to continue as a contributor
No one would really be surprised if I just stopped contributing

Satisfaction
I enjoy my contribution experience
My contribution experience is personally fulfilling
My contribution experience is worthwhile
I am likely to continue to contribute to this project
Community Commitment
I feel very little loyalty to this community
I am proud to tell others that I am part of this community
It would take very little change in my present circumstances to cause me to leave the community
There's not too much to be gained by sticking with this community indefinitely
I really care about the fate of this community
For me this is the best of all possible communities to participate in
Contributor Benefit Motivation
I want to be recognized for my contributions.
I want to receive a tangible acknowledgment of my contributions
I participate to get a reputation in the free/open-source developers' scene, or within the Perl/Raku community
I participate to improve my job opportunities
I participate to make money
I participate to learn and develop new skills
Intention to Remain
I plan to participate in this community in the future
I will recommend that others participate in this community
I will tell others about the positive experiences that I had participating in this community
I hope that participation in this community is a part of my life for years to come
I am more motivated to participate because of my recent contribution experience with this community
I care about the Perl/Raku community

4.3.2 Interviews

Interviews were designed to get a more personal view of the community members, including contributions they have made, interactions with other volunteers, and issues faced by them and the community. Interview recruitment followed the same approach as the survey, utilizing TPRF website, various Slack channels, and Facebook groups (See Appendix F for recruitment message). Interviews were recorded and transcribed.

A total of six interviews were conducted, including interviewees from geographical locations spread across three continents. Four interviewees worked mostly with Perl, and two worked with both Perl and Raku. Hence, most responses were more Perl-centric than Raku. An important question that arises is whether the interviews genuinely reflect the diversity of the Perl and Raku community, considering only six interviews were conducted, all involving male participants. While the interview sample may not comprehensively represent the community's diversity, the survey data, which draws from a broader and more diverse pool of participants, extensively captures the community's diversity. This is evident in the representativeness in Table 5.1. Additionally, interview participants were geographically dispersed, spanning various regions across the globe.

In the interest of maintaining anonymity, interviewees will be referred to as Interviewee1 to 6. The number assignment is based on the chronological order of the date the interview was taken, i.e., Interviewee1 was the first interview taken, Interviewee2 the second and so on. (See Appendix B for interview questions).

As previously mentioned, surveys and interviews complement each other. Surveys, characterized by structured questionnaires with standardized questions and response options,

efficiently capture information. In contrast, interviews exhibit a high degree of flexibility, enabling tailored questions for each participant, fostering probing inquiries, follow-ups, and a thorough exploration of individual responses. Furthermore, surveys and interviews synergize effectively in terms of sample size. Surveys accommodate a larger participant pool, compensating for the smaller interview sample size. Simultaneously, interviews yield deeper insights compared to surveys, enriching the overall research perspective. In this research study, the combination of surveys and interviews yielded a profound understanding of the community's volunteering dynamics and contributed valuable insights.

4.4 Data Analysis

Once the data collection process was completed, analysis proceeded, utilizing two primary methods: Quantitative and Qualitative data analysis. A combination of quantitative qualitative analyses was employed in response to the presence of two distinct data types: surveys, which were quantifiable and analyzed using quantitative techniques, and interviews, which were subjected to qualitative data analysis. Qualitative and quantitative data analysis go hand in hand, especially when working with different types of data.

4.4.1 Quantitative Data Analysis

Quantitative data analysis is rooted in mathematical modelling and measurement techniques employed to gain insights into specific behaviours (Albers, 2017). It is particularly suitable for numeric data or data that can be converted into a numeric format. In this research project, the survey data was coded into a numeric format to facilitate quantitative data analysis. The Model of Retention, with its different constructs to survey data, aided in the exploration of key patterns and relationships within the dataset. The term "constructs" typically refers to abstract or

conceptual ideas, concepts, or variables that are created to represent specific, often complex, phenomena. Constructs give researchers a way to measure or quantify something that may not have a straightforward, observable form.

The constructs investigated in this research study encompass Community Commitment, Psychological Sense of Community, Social Norms, Contributor Benefit Motivation, Satisfaction, and Intention to Remain. By converting the collected survey data into a numerical format, researchers can use statistical methods to analyze data related to these constructs, helping to draw conclusions and make sense of the phenomena being studied. For accurate measurement of the constructs, it is imperative to develop specific indicators that effectively capture these elements. In the context of surveys, indicators manifest as distinct questions, variables, or metrics meticulously crafted to gauge particular facets or concepts of interest. These purposeful indicators serve as instruments that aid researchers and survey designers in comprehending and evaluating specific phenomena. Indicators are thoughtfully selected to mirror the critical variables or dimensions being explored in a survey, enabling the acquisition of quantifiable and actionable insights. Their role is pivotal in the process of quantifying, analyzing, and deriving meaningful conclusions from the responses collected. Within the scope of this thesis, these indicators are referred to as "*questions*", simplifying their representation.

Quantitative data analysis includes data preprocessing, data validation and finally, data analysis (Duncan Kramer, 2003). Data preprocessing involves converting data into a numeric format, while data validation applies various techniques to ensure data quality. A deeper exploration of these techniques will be provided in Chapter 5. Subsequently, the model of

retention was applied to the quantitative dataset utilizing PLSPM a statistical method capable of implementing the model.

4.4.2 Qualitative Data Analysis

Qualitative data analysis is a process of examining non-numerical data, such as text, images, audio recordings, or videos, to identify patterns, themes, and meanings. Qualitative research often involves collecting data through interviews, focus groups, observations, or surveys that ask open-ended questions. The analysis of such data involves several stages, including data preparation, data coding, theme identification, and interpretation (Seers, 2012).

Data preparation involves organizing and cleaning the data, transcribing audio or video recordings, and reviewing the data to identify patterns and themes. Data coding involves assigning labels or tags to segments of data that represent specific ideas, concepts, or themes (Zhang et al., 2003). This process can be done manually or with specialized software tools.

After coding, the analyst examines the coded data to identify patterns and themes. Themes or patterns are recurring ideas that emerge from the data, and they can be identified by looking for similarities and differences across the data set. The analyst may use various techniques, such as content analysis, discourse analysis, or grounded theory (Seers, 2012). Content analysis involves three key steps: the development of code schemes, their application, and the subsequent interpretation of the findings. This method was selected for its structured approach to content examination (Krippendorff, 2018), offering a systematic means to analyze the content under investigation. Conversely, discourse analysis delves into the intricate aspects of language, communication, and meaning construction through spoken or written discourse. Due to limitations in qualitative data availability, it was not a suitable choice for this study. On the other

hand, grounded theory aims to derive theories or conceptual frameworks directly from qualitative data. This method prioritizes theories that naturally emerge from the data itself, avoiding the imposition of predefined constructs. In the context of this research study, where the application of a predefined construct (such as the model of retention) is necessary, grounded theory was not a viable option.

Finally, the analyst interprets the data to generate insights and conclusions. This involves drawing connections between themes and identifying the broader implications of the findings. Qualitative data analysis is an iterative process, meaning the analyst may need to revisit earlier stages as new insights emerge from the data.

In this project, interviews served as the primary source of qualitative data for analysis. Dr. Ann Barcomb, the project supervisor, acted as the second coder for one of the interviews, while Andrea Tampus, an undergraduate student at the University of Calgary, served as the second coder for the remaining interviews. Engaging a second coder in qualitative data analysis is a widely adopted practice aimed at bolstering the validity, reliability, and robustness of research findings. This approach safeguards against undue influence from individual biases or perspectives, ultimately enhancing the overall quality of the analysis (Miles et al., 2014).

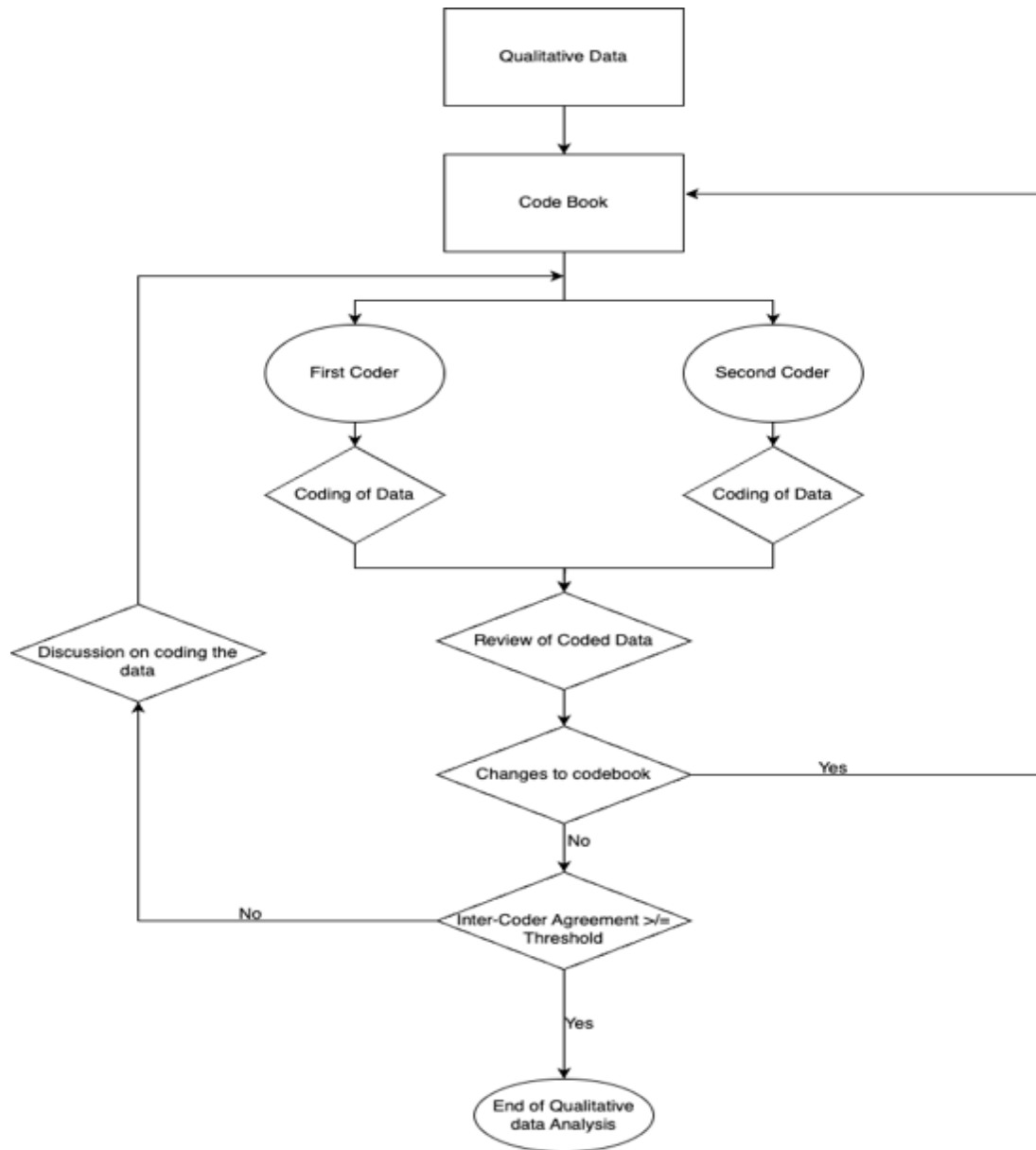


Fig 4.1: Flowchart depiction of the qualitative coding process

Figure 4.1 demonstrates that the qualitative data analysis process commenced with the creation of a codebook, which serves as a comprehensive reference containing predefined labels or "codes" along with precise definitions and illustrative case scenarios to guide their application (Seers, 2012). Subsequently, the primary coder utilizes the initial version of the codebook to code

the data, a process mirrored by the second coder. Upon completing the initial coding phase, both coders engage in a collaborative review, which could yield various outcomes, including the addition, removal, consolidation, or division of codes within the codebook. This iterative process continues with the updated codebook until a substantial level of similarity between the coders is attained.

At the end of each iteration, a measure of similarity is calculated using a metric known as Inter-Coder Reliability. Inter-Coder Reliability quantifies the degree of concurrence or consistency among different coders when coding the same data or event. It assesses the reliability of judgments made by multiple coders. One of the ways of calculating Inter-Coder Reliability is using Cohen's Kappa (κ), which accounts for the possibility of agreement occurring by chance and provides a measure of agreement beyond what would be expected by random chance alone (Brennan & Prediger, 1981). Cohen's Kappa (κ) provides a value between -1 and 1, with '1' being that the coders agree all the time perfectly, '0' being if the coders agree just as much they would by random chance and '-1' if the coders disagree perfectly all the time. Ideally, a Kappa value should be between 0 and 1, the higher the value, the better. To validate the coding in this project, a predefined threshold of 80% ($\kappa = 0.8$) was set for the Inter-Coder Reliability. This threshold was set based on the research objectives, methodological considerations, and desired level of agreement for the study. In instances where this threshold was not met, coders engaged in Inter-Coder Agreement discussions to refine the coding guidelines and procedures, aiming to reach a shared understanding and enhance the coding's reliability and accuracy.

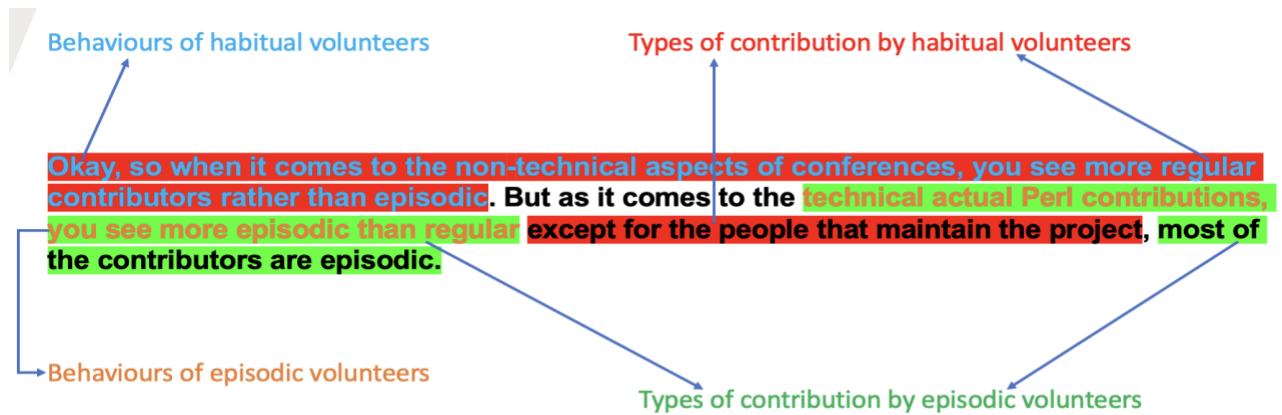


Fig 4.2: Illustration of the Qualitative Coding Process

In Figure 4.2, certain lines of text are highlighted in various colours to illustrate their corresponding labels or codes. Some lines of text correspond to more than one code. The MAXQDA⁴ tool was used for its wide variety of applications, including the calculation of the Inter-Coder Reliability.

After completing the coding process, the textual information was summarized according to their respective labels. This entails organizing information based on the assigned labels, allowing for a cohesive grouping that aligns with the underlying context. This approach facilitates a clear and accurate comprehension of the significance of each piece of information. Subsequently, these grouped data sets can effectively correlate with the insights derived from the quantitative data analysis, enhancing holistic understanding of the research findings.

⁴ <https://www.maxqda.com/interview-transcription-analysis>

Chapter 5: Quantitative Data Analysis and Results

The methodology chapter (Chapter 4) overviewed the methods used to analyze the data in this project. This chapter provides deeper insight into specific techniques used in the quantitative analysis as well as discuss the results obtained.

5.1 Data Quality

The first data validation step is to check for the quality of the data to see how suitable it was for use in this analysis. This was first done by checking for Common Method Bias, a systematic error or bias that can occur in research when the same method or source (i.e., the survey) is used to collect both the predictor (independent variable – Intention to Remain) and the criterion (dependent variable – constructs of the Model of Retention) data in a study. It can result in inflated or spurious relationships between constructs and compromise the validity of research findings. Common method bias is typically considered a form of measurement error (Doty & Glick, 1998). Addressing common method bias is essential for maintaining the integrity and validity of research findings, ensuring accurate data interpretation, and making meaningful contributions to the body of knowledge in a given field.

Efforts were made to reduce the influence of Common Method Bias, aiming to improve the research's overall quality and credibility. Harman's one-factor analysis was used to perform factor analysis on the dataset, with all the survey questions related to each construct included as variables. Factor analysis is a statistical technique used to identify underlying factors (or constructs) that explain the correlations among variables (survey questions). If a single factor emerges that accounts for a substantial portion of the variance in the dataset (usually more than

50%), it suggests that common method bias might be an issue. After conducting the test, it is clear that a maximum variance of 23% is achieved, underscoring that no individual construct explains more than 23% of the total variance. This indicates that no individual construct exerts substantial influence, as none surpasses the critical 50% threshold that would suggest common method bias. Instead, a collective contribution of multiple constructs is necessary to explain more than 50% of the variance. It can be confidently concluded that common method bias is not a significant concern in this research data.

Next, Bartlett's test was used to assess potential intercorrelations among the constructs to discern any underlying structural patterns. Bartlett's test scrutinizes intercorrelations among constructs using statistical significance (Tobias & Carlson, 1969), providing valuable insights into the suitability of a model to elucidate relationships among these constructs. Bartlett's test yields a p-value which ideally should be less than the threshold of 0.05. The p-value serves as a statistical gauge of the strength of evidence against the null hypothesis during a hypothesis test. It quantifies the likelihood of obtaining results as extreme as those observed, assuming the null hypothesis to be true. When the p-value is deemed statistically significant, the observed data is highly improbable to occur by chance alone, providing substantial grounds for rejecting the null hypothesis. The null hypothesis posits that the constructs are uncorrelated and function independently. Conversely, the alternative statistically significant hypothesis suggests the existence of correlations, indicating that a model may be warranted.

Upon conducting the test, results yielded a remarkably low p-value of 1.408e-88, significantly below the conventional threshold of 0.05. Based on this outcome, it can be confidently concluded that the test has successfully passed, indicating statistically significant

correlations among the constructs and supporting the suitability of employing a model to explain their relationships.

The Kaiser-Meyer-Olkin (KMO) Test (Shrestha, 2021) is similar in purpose to Bartlett's test but focuses on assessing whether the dataset possesses sufficient common variance and correlations among variables to support factor analysis. The KMO Test yields a single KMO value, ranging from 0 to 1. Higher KMO values indicate greater suitability for factor analysis, with a desirable threshold typically set at 0.5 or higher. Upon conducting the KMO Test, the obtained KMO value stands at 0.704, surpassing the minimum acceptable threshold of 0.5 and affirming that the dataset meets the requirements of the KMO Test.

The Bartlett's test and the Kaiser-Meyer-Olkin test are typically administered in conjunction to evaluate both the intercorrelations among constructs and the suitability of the data for factor analysis. The outcomes of these tests unequivocally confirm that the dataset is well-suited for factor analysis and the subsequent application of the Model of Retention. The data has successfully passed the quality assessments, affirming its readiness for analysis.

5.2 Representativeness

Prior to conducting the modelling, an assessment of data representativeness was performed to evaluate the community's similarities or differences from others. Demographic information was compared to data from four additional studies.

- EV'14: - This research was conducted on episodic volunteering in the public health and non-profit organization sector. Only summary information is available. (Hyde et al., 2014)
- EV'17: - This survey was conducted by Barcomb et al. This survey covers multiple open-source software communities and focuses on episodic participants. The data were

collected between 2016 and 2017. Only summary information is available. (Barcomb et al., 2019)

- OS'17: - This was a survey conducted by GitHub to understand open-source software in a detailed aspect, and one of the survey topics was volunteering and contributions. This is a public dataset. (Geiger, 2017)
- FLOSS '13: - This survey was conducted on over 2000 open-source software contributors; many aspects were covered, though this was specific to open-source software communities. This is a public dataset. (Robles et al., 2013)

The overall representativeness is summarised in Table 5.1:

Table 5.1: Representativeness of the members of the community

	This Study	EV'17	FLOSS '13	OS '17	EV '14
Mean Birth Year	1971	1977	1979	1987	1971
Gender					
Male	87.209%	84.44%	87.40%	95.38%	10.39%
Female	4.65%	13.33%	11.12%	3.52%	89.61%
Other	8.139%	2.22%	1.48%	1.10%	n/a
Education					
University	84.88%	78.57%	77.69%	65.70%	39.64%
Trade School	9.302%	8.16%	3.27%	3.44%	28.11%
High School	5.81%	11.22%	17.85%	27.45%	0.47%
Contribution					
Code, Programming	40.90%	33.66%	48.63%	n/a	n/a
Other contributions	25.0%	28.71%	26.77%	n/a	n/a
Both	34.09%	37.62%	24.59%	n/a	n/a

Since this study implements theories and concepts developed from and for a broader set of open-source software communities to a specific community, it is essential to compare the demographics of the other studies with this one, as some theories may require adaptation to suit this study's demographic. For example, it can be seen that the participants in this study are generally older but, at the same time, more educated, which might affect the general applicability of the theories.

Statistical analysis tests were conducted on the demographic information of the various studies compared in Table 5.1. The Welch two-sample T-test (Cressie & Whitford, 1986) was conducted on the age of respondents from various studies. The results indicate a statistically significant difference in age between the compared datasets, with a p-value of less than 0.05 when compared to the FLOSS '13 and OS '17 datasets. This is also seen in the difference in the mean birth years. Subsequent Chi-square tests were conducted to compare gender, education, and contribution levels between the dataset used in this study and the FLOSS '13 and OS '17 datasets. A p-value below 0.05 indicates statistical distinctions between the dataset of this study and the FLOSS '13 and OS '17 datasets concerning gender and education. A p-value of 0.1226 was obtained for contribution, indicating that the populations of the two studies are not statistically different from each other, as it is higher than the significance level of 0.05. Most tests indicated that the data exhibits statistical differences compared to the FLOSS'13 study, implying that the theories cannot be readily applied to this study.

Correlation tests were conducted between internal parameters of the dataset, including contribution type, age, hours contributed, episodic projects contributed, regular projects contributed, and gender. The primary motivation behind these tests was to explore potential

influences among parameters that could impact the implementation of established theories in the study. In all tests, the p-value exceeds 0.05, indicating statistical insignificance. Consequently, the hypotheses suggesting the absence of influence between the parameters cannot be rejected (or null hypothesis).

5.3 PLS-PM

To implement the model of retention, the PLSPM model of evaluation was employed. This choice was motivated by the dataset's limited size and the effectiveness demonstrated in a prior study using this technique (Barcomb et al., 2020). PLSPM stands for "*Partial Least Squares Path Modeling*", which is a multivariate statistical technique that combines the features of partial least squares regression (PLSR) and structural equation modelling (SEM). PLSPM is used for analyzing the relationships between multiple independent variables and multiple dependent variables in a complex system. This is particularly useful with highly correlated variables, a characteristic applicable to the constructs in this project. PLSPM was implemented based on the Model of Retention. The five constructs were utilized to map them to the Intention to Remain (see Fig 5.1). The model implementation using PLSPM is stored in a GitHub⁵ repository.

⁵ <https://github.com/aadharsh1997/EV-Research-Project>

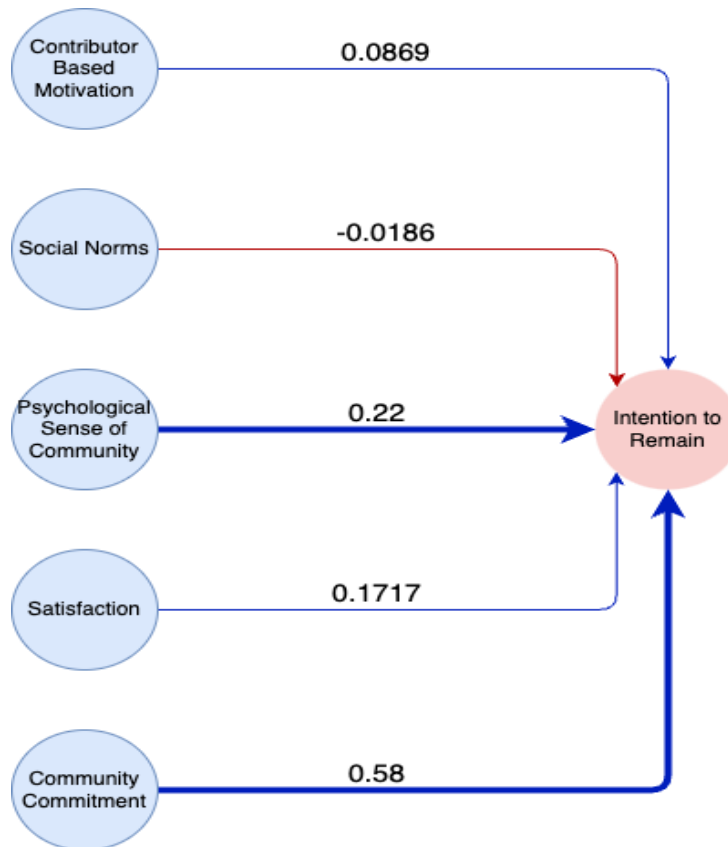


Fig 5.1: Pictorial Representation of the PLSPM model

Certain characteristics and information can be derived from the diagram of the PLSPM model. On each line, the path coefficients or weights that attribute to the strength and direction of each construct toward the Intention to Remain are shown.

In addition, a significance assessment for each construct within the implemented retention model was conducted, as outlined in Section 5.5.1 (Bootstrapping). This assessment utilized a 95% confidence interval, which signifies a range within which there exists 95% confidence that the parameter falls. In this context, if the entire interval resides in either the positive or negative domain (indicating the same sign at both ends), it is considered statistically significant. Conversely, if the interval spans both positive and negative values, it lacks statistical significance,

as it fails to establish the correlation direction with the Intention to Remain definitively. The constructs are summarized below:

- Contributor Benefit Motivations are not a significant construct for volunteers to join the community or remain within the community. This construct has a low correlation, meaning that it does not help explain the Intention to Remain in the community and is statistically insignificant.
- Social Norms volunteers may face is not a construct for them to remain or join the community. The construct has a low (negative) correlation and is statistically insignificant. (Represented in red in Fig 5.1)
- Satisfaction of volunteers within the community is not a demonstrated construct for volunteers to remain in the community as it is also statistically insignificant. However, it explains approximately 17% of the Intention to Remain. It may be prudent to consider this as a possible partial explanation behind the intention of volunteers to remain.
- Psychological Sense of Community plays a strong role in the Intention to Remain among volunteers. Psychological Sense of Community comprises 22% of the Intention to Remain. It is statistically significant, indicating an environment that fosters a sense of welcome for newcomers and potentially nurtures this psychological aspect among volunteers due to the presence of like-minded individuals. (Represented in thicker blue in Fig 5.1)
- Community Commitment has the highest path coefficient among constructs, making up 58% of the Intention to Remain and is statistically significant. This means that Community Commitment plays a vital role in the Intention to Remain among contributors. This shows that the devotion and loyalty shown by volunteers who have been contributing to the

community for a long time is strong and is what motivates them to continue contributing to the community. (Represented in a thicker blue in Fig 5.1)

5.4 PLS-PM Model Evaluation

The PLSPM packages in R come equipped with certain inbuilt functionalities that can be used to evaluate the model.

5.4.1 Convergent Validity

The initial test (Table 5.2) is the Convergent Validity Assessment (Carlson & Herdman, 2012) to determine the degree of convergence among the constructs of a specific model. Since the constructs are all correlated to the same factor, the Intention to Remain, much of the variance should be shared among the constructs and they should converge. In this test, the Average Variance Extracted (AVE) determines the convergence of the constructs. The typical rule of thumb is that if the AVE is greater than 0.5, then more than half of the variance is shared across the different constructs. In the model, every construct, except for Contributor Benefit Motivation, exhibits an AVE value exceeding 0.5. This may imply a weaker correlation between Contributor Benefit Motivation and the other constructs. This weaker association could be attributed to the statistical insignificance of Contributor Benefit Motivation.

Table 5.2: Convergent Validity

Construct	AVE
CBM	0.357579
CC	0.514443
PSC	0.519762
S	0.695700
SN	0.518717

5.4.2 Internal Consistency Reliability

The Internal Consistency Reliability test checks for the extent to which constructs are consistent with each other. It evaluates whether all constructs in a set are measuring the same thing and provide similar results. Internal Consistency Reliability is used to ensure that the items in the measurement instruments are dependable and effectively capture the concept or construct they intend to assess. Usually, a high degree of consistency is desired to show it is referring to the same construct or concept (Bentler, 2009). Three parameters are usually used to understand the consistency between the constructs, which are Cronbach's alpha, Dillon-Goldstein's rho, and 1st and 2nd Eigenvalues. Cronbach's alpha should be greater than 0.7, and Dillon-Goldstein's rho should be above 0.7 and below 0.95, showing the strength of the consistency towards the construct. The 1st Eigenvalue should be greater than one, and the 2nd Eigenvalue should be less than one, indicating statistical significance of the internal consistency of each construct. Upon conducting the test, Cronbach's alpha value exceeds 0.7 for all constructs except for Social Norms and Psychological Sense of Community. Except for Social Norms, all constructs are within the range for Dillon-Goldstein's rho value. The Eigenvalues are suitable for all constructs except Contributor Benefit Motivation.

In summary, the impact of internal consistency reliability plays a pivotal role in assessing the trustworthiness and quality of the measurement scales represented by each construct. Constructs such as Community Commitment and Satisfaction demonstrate strong reliability, bolstered by statistically significant coefficients. This indicates that the items within these scales consistently and reliably measure the intended constructs, thereby enhancing the validity of research findings. However, Social Norms, while statistically significant, exhibits weaker

consistency in measuring the constructs. Contributor Benefit Motivation, on the other hand, shows strong internal consistency but is statistically insignificant. These factors may be attributed to the statistical insignificance of those constructs, as further detailed in Section 5.5.1.

Psychological Sense of Community displays moderate internal consistency and statistical significance. This could be attributed to either an insufficient number of survey data points for Psychological Sense of Community to achieve the desired level of internal consistency reliability or the limited scope of the survey questions (Section C), which may not fully capture the multifaceted nature of Psychological Sense of Community. Further investigation and potential scale refinement may be needed in future studies for these cases to enhance their reliability and relevance in the research context.

The result is represented in Table 5.3:

Table 5.3: Internal Consistency Reliability

Construct	Cronbach's Alpha	Dillon-Goldstein's Rho	1 ST EV	2 nd EV
CC	0.812	0.864790	3.100548	0.771951
S	0.852046	0.901485	2.791580	0.633882
SN	0.065949	0.543144	2.112936	0.737596
PSC	0.546376	0.767834	1.574137	0.757316
CBM	0.761853	0.836752	2.826157	1.155115

5.4.3 Effects Between Latent Variables

The final test checks for the effects of latency between constructs (Table 5.4). The desired outcome is that constructs show a direct relation to the Intention to Remain and not to any other constructs. Upon conducting the test, only direct effects are observed, and there are no

undesired effects among latent constructs. It can be inferred that aside from a few weak correlations, the established constructs demonstrate validity, allowing for the evaluation of the model results.

Table 5.4: Effects Between Latent Variables

Construct → IR	Direct	Indirect	Total
CC → IR	0.575767	0.0	0.575767
S → IR	0.171743	0.0	0.171743
SN → IR	-0.018606	0.0	-0.018606
PSC → IR	0.219978	0.0	0.219978
CBM → IR	0.086861	0.0	0.086861

5.5 PLS-PM Hypotheses Testing and Exploration

The PLSPM model comes with certain functionalities that allow evaluation of the implemented model. These functionalities can be used to test the proposed model as well as gain more information from the data.

5.5.1 Bootstrapping

PLSPM on its own does not tell about the statistical significance of the different constructs, but knowing this will help draw more meaningful conclusions. Bootstrapping is a statistical method used to estimate the sampling distribution of a statistic by repeatedly sampling with replacement from a given dataset (Sanchez et al., 2013). It involves creating multiple "*bootstrapped*" datasets that are similar to the original dataset, but each is slightly different due to the random sampling process. The technique is useful when the underlying distribution of the data is unknown or when a small sample size makes it difficult to draw accurate conclusions about the population. Overall,

bootstrapping is a powerful statistical technique that can help provide more accurate and reliable estimates and inferences from a given dataset.

In this technique, the resamples are first combined to calculate the mean value of each construct along with the standard error and the 95% confidence interval. If it is significant, the two extremes of the 95% interval should have the same sign (positive or negative). From this analysis, it becomes evident that three construct hypotheses lack support through bootstrapping, namely, Social Norms, Satisfaction, and Contributor-Based Motivation, all demonstrating statistical insignificance (Table 5.5).

Table 5.5: Bootstrapping result

Construct	Mean	Std. Error	95% Interval
CC	0.566115	0.073084	(0.417622, 0.700622)
S	0.165509	0.094129	(-0.011874, 0.353735)
SN	0.005280	0.076640	(-0.147426, 0.152490)
PSC	0.217199	0.076448	(0.067902, 0.365630)
CBM	0.084889	0.095720	(-0.169281, 0.226682)

5.5.2 Moderating Factors

Moderating factors are crucial to model evaluation, as they assess the presence of external factors that could potentially impact the model (Sanchez et al., 2013). External factors refer to any influences that exist outside the defined boundaries of the model and its constructs. These factors are not part of the model's core components but have the potential to affect or modify the relationships and outcomes described by the model. It is important to look at moderating factors because they highlight the conditions or characteristics that determine when and how the relationship varies and how it is influenced. The ability to tailor changes based on individual

differences is improved by considering moderating factors, which also contributes to a more thorough and nuanced understanding of the relationship between constructs. The specific factors considered were Age, Tenure, Contribution, and Gender.

For Tenure, other than Community Commitment, all other constructs show statistical insignificance. The relationship between tenure and Community Commitment was positive. It can be concluded that volunteers who remain due to their commitment to the community are typically those with longer tenures or a significant amount of active involvement in the community. This goes hand in hand with the definition of Community Commitment as the extent to which the individual identifies with the community and sees themselves as a member of the community, is an expected moderation effect Tenure has on Community Commitment.

The impact of Age on the constructs is generally negligible, with the exception of Community Commitment, where age is the only statistically significant factor. The relationship between age and Community Commitment was positive, indicating that volunteers who remain within each community due to their commitment are typically older than average. This aligns with the finding that volunteers with longer tenure also exhibit higher Intentions to Remain as a result of their Community Commitment.

In Contribution, apart from Psychological Sense of Community, all other constructs exhibit statistical insignificance. This phenomenon manifests as a robust moderating effect, particularly concerning code contributions made by volunteers, indicating that volunteers contributing code derive satisfaction from their community or environment. It is again expected that contribution would have a moderating effect on Psychological Sense of Community as it corroborates the very definition of it.

It was not possible to determine if gender was a factor due to the higher proportion of men and people who identify with genders other than male and female.

5.5.3 Cluster Analysis

It is helpful to examine the unknown factors influencing the construct within the PLSPM model. Cluster analysis uncovers patterns, structures, and relationships within a dataset by identifying clusters or groups of data points that share common characteristics. This aids in pinpointing the specific groups within the community whose attributes align most closely with the developed techniques and solutions, making them the most suitable candidates to implement these strategies. A functionality of PLSPM was used known as Response Based Unit Segmentation or REBUS (Sanchez et al., 2013). REBUS adapts global and local models by assigning observations to groups based on distance. REBUS initially recommended a 4-cluster segmentation, which was deemed optimal when compared to running 2, 3 and 5-cluster segmentation based on the Goodness of Fit (GOF). The GOF of REBUS clusters was 0.714 as compared to the GOF of the original PLSPM model of 0.59. The clusters are as shown in Table 5.6, and the bolded values are statistically significant values at 95% confidence.

Table 5.6: Cluster Segmentation

Group	CBM	SN	PSC	S	CC	GoF
Satisfied	0.2184	-0.1267	0.2852	0.7226	-0.0227	0.7726144
Individualistic	0.6648	0.0232	-0.1993	0.1616	0.6371	0.6724868
Obligated	-0.5072	0.0588	0.2178	-0.0709	0.7663	0.6880414
Optimistic	-0.1059	-0.1984	0.6530	0.1890	0.3701	0.7429961

A group name was assigned to each cluster based on the highest correlated constructs. Analyzing each construct and its correlation with the cluster determines potential associated attributes.

- Satisfaction: typically exhibit specific traits that reflect contentment and fulfillment. This cluster associates with attributes such as positive attitudes, elevated morale, and cooperation. CBM also contributes to this cluster, suggesting that satisfaction may result from individuals deriving personal benefits within the community.
- Individualistic: typically exhibits specific characteristics that emphasize personal independence, self-reliance, and a focus on individual needs and goals.
- Obligated: typically exhibits a sense of duty, responsibility, and commitment to fulfilling obligations and responsibilities, often driven by personal expectations.
- Optimistic: typically exhibits traits that emphasize a positive outlook on the community and a hopeful attitude toward the future. An optimistic outlook also stems from a sense of loyalty towards the community, believing in a positive future.

5.6 Summary

The quantitative data analysis yields significant insights into the model of retention's key constructs. Among the five constructs from the model, only Community Commitment and Psychological Sense of Community emerge with statistical significance as contributors to the Intention to Remain. Conversely, the remaining constructs do not exhibit statistical significance, suggesting that they do not play a significant role in influencing the intention to remain.

Before applying the model of retention to the survey data, a comprehensive data quality check was conducted, including tests for common method bias, all of which were successfully passed. Following the model's implementation, additional statistical tests were employed to

assess validity and reliability, including tests for convergent validity and internal consistency reliability. Social Norms and Contributor Benefit Motivation were found to be statistically insignificant, potentially contributing to their lack of internal consistency reliability. Contributor Benefit Motivation also demonstrated weak convergent validity. In contrast, Psychological Sense of Community exhibited moderate internal consistency reliability, which could be attributed to the limited data points available for this construct to demonstrate strong reliability. Cluster analysis was applied to identify clusters based on specific attributes.

The quantitative data analysis revealed three statistically insignificant constructs, which constitutes the primary limitation as it does not offer a definitive answer or conclusion regarding the studied relationship or effect. Statistical insignificance should not be equated with the absence of a relationship or effect; it simply means that, based on the data and the chosen construct model, there is not enough evidence to assert the presence of the relationship or effect confidently. To address this challenge, one of the most reliable methods is to augment the findings by collecting data from an additional source, thereby enhancing the results' reliability through triangulation. Getting data from another source helps determine if the same results are consistently observed across different datasets, strengthening the reliability of the conclusions. The interview data proved invaluable in addressing the uncertainty stemming from the statistical insignificance of the three constructs in relation to Intention to Remain. Through qualitative data analysis, it becomes apparent that these three constructs indeed exhibit a less pronounced relationship with the Intention to Remain, as elaborated upon in the following chapter.

Chapter 6: Qualitative Data Analysis and Results

This chapter details the qualitative data analysis and findings. When examining the results, intriguing inferences and potential explanations emerge, providing deeper insights into the research objective of understanding the applicability of current theories to the volunteering challenges encountered by the Perl and Raku community. A codebook overview is presented in this chapter, and the qualitative data results are again grouped based on the constructs for better understanding, as each issue identified is relevant to that particular construct.

6.1 Codebook

Section 4.1.2 in Chapter 4 discusses using a codebook (Appendix G) as part of qualitative data analysis. The codebook gives various relations and definitions to different information in qualitative data. A total of 38 codes were present in the codebook and used for coding the qualitative data. Using these codes, a total of 896 instances were labelled in the six interviews. All codes in the codebook are divided into two categories: *community* and *volunteers*. Of the 38 codes, 15 fell under community, and 23 under volunteers. Within each subcategory, nested recursive codes can be found, each addressing a specific topic related to the main category. In some instances, this hierarchy extends to two levels of recursive codes for greater detail.

The primary code in the Community category was the "*Community*" label itself, which was used to label all the generic information about the Perl and Raku community and was used in 76 instances. The following two most used codes under the same category were *Practices to improve the community* and *Problems faced by the community*, which usually went in tandem to label a problem faced by the community and a proposed solution. They were used at 57 and 56

instances, respectively. The next most prevalent code was *Contribution*, which described the role the interviewee played in the community and their contributions and was used in 48 instances.

The primary code in the Volunteer subcategory was *Interaction with volunteers*, which labelled experiences the interviewee had with other volunteers in the community. It was used in 51 instances in the data. The next most used label in the same subcategory was *Practices on volunteers*, which specified techniques to resolve volunteer issues, and was used in 46 instances. Another notable code was *Availability of volunteers*, which highlighted cases related to volunteers' presence and their access when working on projects. It was used in 37 instances in the data. The codes that more closely relate to the research topic were used the most.

6.2 Contributor Benefit Motivation

Quantitative data analysis evidenced that volunteers' reluctance to join or remain within the community is often tied to their perception of a lack of personal benefits. To delve deeper into the reasons behind this phenomenon, an analysis of the qualitative data, specifically the interviews, was conducted and yielded the following results.

6.2.1 The Perl community is declining

In general, the community is different from a couple of years ago. It has been declining ever since causing modules to be abandoned, resulting in future projects not being supported. Volunteers are taking their projects to other languages where more support is available. This is the case for both episodic and habitual volunteers. Interviewee2 illustrates this:

“...it is sometimes a challenge because some modules get abandoned by the authors, the authors change fields, or they are not interested in maintaining it anymore, or they just do not want to anymore.” - Interviewee2

This shows that the abandonment of modules is a significant issue within the community, and the fact that these modules no longer support any technology advancements is a reason why volunteers no longer feel there is any benefit to contributing to the community. Also, it was illustrated by Interviewee4 that:

“The less work that goes into Perl, the harder it is to use it” - Interviewee4

“Very often, I find myself in a situation where we have to do things, not the language, because the SDK (software development kit) toolkit, which another service provider using this service requires, has not got Perl support.” - Interviewee4

These statements indicate that the community is declining as more users find it challenging to implement technologies using it.

6.2.2 Fewer Perl Jobs

There are fewer Perl jobs compared to other languages, and the Perl community has not been able generate more Perl jobs, which might be important for volunteers not joining the community as they do not see any employment benefits. The migration of people from Perl to other languages because of the lack of jobs in Perl is highlighted in Interviewee1’s statement:

“Meanwhile, Perl people we are losing our jobs, and everyone is leaving Perl to go to Python and Java” - Interviewee1

Perl is not widely recognized within the realm of job opportunities, which may result in limited resources for in-depth Perl learning compared to more popular programming languages. The statement by Interviewee4 gauges this:

"I trained developers from other languages into Perl because Perl is not a very popular programming language" - Interviewee4

Few Perl-related careers is a significant reason why volunteers are not motivated to contribute.

6.2.3 Perl 5 and Perl 6 Confusion

The confusion created by the simultaneous availability of Perl 5 and Raku has resulted in the perception that Perl is dead, outside of the Perl and Raku community. Potential volunteers may feel that there is no benefit in contributing to the Perl community. This is evident from the statements made by Interviewee1:

"Perl was going to be the number one language; we were number three, and we were going up to become number one. We are now below twenty. Okay, we have dropped to the bottom of the stack because of Raku. That is the only reason, and there is no other reason. Perl would actually be the number one right now if it wasn't for Raku, and that's a very, very sad and disturbing thing." - Interviewee1

"Most people believe that Perl is dead, and we're fighting to now reverse that inaccuracy. So even just getting past that first sentence that will come out of people's mouth requires such a deep knowledge of how messed up the Perl and Raku situation is" - Interviewee1

The community is struggling with the splitting of Perl and Raku, which has led to factions forming among members. The lack of resolution of this matter has affected the development of the language. This is a primary reason why the volunteer population is declining, as they feel this conflict hinders their involvement in the community.

6.3 Social Norms

Surveys indicates that social norms do not drive volunteer contributions, as detailed below.

6.3.1 Popularity of the Community

A major reason why Social Norms do not play a significant role is because popularity of the language is low. As long as popularity keeps declining, social constructs (such as increasing competency to secure a job) will not influence volunteers to join the community,

“I mean, interestingly, it has gone down over time. A decade ago, it was super popular. And now, it’s less popular. So, that’s gone down a lot.”- Interviewee3

6.3.2 Social Norms in Open-Source

In general, participation in open-source software communities is not something people feel pressured about, positively or negatively, by friends, family, community, nor industry. In fact, all interviewees indicate that social norms did not influence their participation.

6.4 Psychological Sense of Community

Psychological Sense of Community exhibits a noteworthy and statistically significant positive correlation with volunteers' Intention to Remain. This suggests that the community, particularly the Perl faction, fosters an environment that offers volunteers a sense of comfort and motivation to sustain their involvement. It is likely that this sense of comfort and the presence of like-minded individuals play a pivotal role in retaining volunteers.

The community boasts a rich diversity of contributors hailing from various backgrounds, including geography, culture, and gender. Table 5.1 details the community's inclusivity, with a notably higher percentage of individuals identifying with genders beyond male or female compared to similar studies. Interviews underscore the geographical diversity, as respondents were dispersed across different regions worldwide.

While the Psychological Sense of Community demonstrated strong significance in the quantitative data analysis, certain characteristics identified in the qualitative data analysis contradicted this finding, as elaborated upon below.

6.4.1 Unwelcoming and Abrasive

Some community members are unwelcoming, displaying abrasive personalities and friction against new ideas and methodologies within the community. Interviewee2 elaborates:

"The people who developed Perl, the people who contributed to Perl, they're a very bright bunch. And sometimes with bright people, you get abrasive personalities." - Interviewee2

"I guess another part of the challenge is that some people have been a part of the community for quite a while, and they've got, oh, we've always done it this way. They're kind of stuck in one way of doing it. And I think it is important to keep an open mind about approaches that we take in solving problems because the way we've always done it that way is not necessarily the best way." - Interviewee2

Interviewee4 observes similar issues:

"One of the issues that came up is how do we manage misbehaviour and that kind of stuff because Perl has a reputation— at least I think Perl among other languages— Perl has been very loud in this respect for some people being complacent to others. And particularly, in the case of Perl, I think people have been doing it for a long time being unwelcoming to beginners." - Interviewee4

The display of abrasive personalities might be attributed to Institutional Inertia (Osterloh & Rota, 2007), which refers to the inclination of organizations or institutions to resist change and uphold existing practices, structures, or policies, even when it might be more advantageous or necessary to adapt and evolve. This tendency can stem from various factors, including a commitment to

tradition, aversion to risk, and a yearning for stability. Consequently, it can impede an organization's capacity to effectively address new challenges, seize opportunities, or adapt to shifts in its environment. The quantitative data analysis reveals that higher tenure and age are characteristic traits of the Perl and Raku community, which could account for this institutional inertia.

Although Psychological Sense of Community emerges as a strong construct of the Intention to Remain from survey data, interviews indicate that newer members might perceive abrasive personalities among their older peers. This incongruity may be because the survey primarily garnered responses from longstanding community members rather than newcomers.

6.4.2 Social Standards

Another effect of Psychological Sense of Community is a significant dip in the social standards, and often, the Perl community resists new ideas and methodologies. This aligns with earlier observations about abrasive and unwelcoming behaviours and is a probable consequence of institutional inertia. This is quite evident from Interviewee4's statements:

"It's more people within the community I think should be adhering to certain social standards, which will make the community inviting to other people and very good if that will sort itself out if the vibe within the community is well-founded." - Interviewee4

6.5 Satisfaction

Quantitative data analysis indicates that satisfaction does not significantly indicate Intention to Remain, and it is statistically insignificant. Qualitative data analysis provides several discernible reasons why, as outlined below.

6.5.1 Conflict Between Perl and Raku

Interview analysis provides numerous reasons for the persistence of this tension, including leadership being unable to resolve the problem, the belief that funding is not effectively distributed between the two factions, and poor communication. Interviewee3 states:

“I wish that people would communicate more. I wish that people would talk more. People think that participating is just writing code and a lot of it is communicating. That’s great that you can write code. But if you don’t come in and talk to us before you start writing code, then that’s gonna be very helpful. You need to talk to people before you go and write code so that you know what’s going on.” - Interviewee3

Interviewee1 describes the ineffective distribution of funding and unsatisfactory leadership:

“We’re so far below that right now that we’re in nobody’s budget. And they’re still trying to take the Perl money and give it to the Raku people. Someone should be sent to jail for that, I believe it’s actually not only unethical I believe it’s illegal what they’re doing.” – Interviewee1

“Unfortunately, the Perl Foundation has never had sufficient leadership because Larry Wall the creator of Perl was never the leader of the Perl Foundation and he’s not a strong leader anyway. So, the community has been permanently crippled by never having strong leadership, never having strong corporate backing, and never having a roadmap or community plan of any kind.” – Interviewee1

6.5.2 Community Size

With the Perl community declining, members of the community feel that failing to release a new version of Perl (e.g., Perl 6 or 7) may be significant to volunteer (dis)satisfaction. Some interviewees believe that the lack of a new major Perl release shows that the Perl community is

not ready to move on from Perl 5 despite the split of Perl and Raku into separate languages.

Interviewee1, Interviewee4 and Interviewee6 discuss the shrinking community:

“Essentially what’s happening is Perl is continuing to decline, Perl is not increasing, it’s decreasing.” - Interviewee1

“it’s shrinking in itself” - Interviewee4

“Firstly, the size of the community. I’m becoming more concerned that it’s not a sustainable size.” - Interviewee6

Interviewee1 also notes the impact of not releasing a major version of Perl:

“We had a little increase for two or three months a year ago when they announced that Perl 7 would be released soon because we were finally deleting Perl 6, we were just going to skip over 6 and go to 7” - Interviewee1

6.5.3 Availability of Volunteers

With few volunteers available, contributing become more difficult. Episodic volunteers cannot contribute effectively and do not follow protocols, leading to disruptions in habitual volunteer’s contributions. Specific modules not being available to use for projects as their maintainers have abandoned them may also contribute in affecting the satisfaction levels of volunteers in the community. The abandonment of modules has been discussed under Contributor Benefit Motivation. When it comes to the contributions of episodic volunteers, Interviewee6 states:

“Contributions from the episodic volunteers is typically very focused on one particular aspect of the project. They’ll try to fix one particular issue or add functionality without really understanding or testing the impact on everything else whereas the developers who have been around for a while tend to be more aware of that sort of thing. Yeah, the quality of the

contributions from episodic contributors to significantly less. There's more work in tidying things up so it doesn't break everything else.” - Interviewee6

6.5.4 Funding

Often, satisfaction levels drop because there are too few exciting projects. Within the Perl and Raku Community this lack may be due to little corporate funding and backing, unsatisfactory leadership, protocols, and community roadmaps. The perceived tension between Perl and Raku also plays a role, causing the general impression that Perl is no longer advancing. While the leadership and conflicts within the community were extensively discussed, funding has also been a weak aspect for the community. Interviewee1 explains:

“We need money, we need financial backing, we need big corporations. The only thing the Perl Foundation should ever have been trying to do is go court the big companies and get them to write fat checks to us. That’s it, there should be nothing else, but they’ve never done that. They’ve completely failed to do the one thing they should have done.” - Interviewee1

“So, the community has been permanently crippled by never having strong leadership, never having strong corporate backing, and never having a roadmap or community plan of any kind.” - Interviewee1

6.6 Community Commitment

Quantitative showed Community Commitment was the construct with the highest path coefficient and was the most significant factor towards the Intention to Remain, along with being statistically significant. This is confirmed in the interviews, and in that the volunteers who remain in the community are Perl loyalists. Considerable effort is being made to revive the community. Community observation indicated that volunteers, especially habitual ones, engage in a healthy

exchange of ideas and discussions on advancement of the community and projects. This is evident from Interviewee1's statements:

"Most people believe that Perl is dead and we're fighting to now reverse that inaccuracy." - Interviewee1

"And that's even more disturbing because those of us that are left as hard-core Perl loyalists like I spent my whole life and built the compiler, I can't quit Perl, it's very disturbing to see that happen." - Interviewee1

6.7 Summary

Findings from quantitative data analysis in Chapter 5 revealed Community Commitment and Psychological Sense of Community as the two primary constructs significantly associated with the Intention to Remain. In contrast, the remaining constructs exhibited weaker correlations with Intention to Remain, rendering them statistically insignificant in the quantitative analysis. The accompanying qualitative data analysis was pivotal to providing richer insights into these constructs.

For constructs exhibiting a weaker relationship with the Intention to Remain, it became evident that the identified issues held greater severity compared to other constructs, signifying their more substantial impact on the community. For instance, funding emerged as a significant contributing factor to Satisfaction, which carries greater significance than issues affecting Psychological Sense of Community, where social standards played central roles. The comparison of issue severity shed light on why certain constructs had a more substantial influence on the Intention to Remain. The more significant the severity of the issue, the weaker the correlation with the Intention to Remain.

Notably, a recurrent challenge identified across multiple constructs in the qualitative data analysis was the conflict between Perl and Raku. This conflict emerged as a significant and overarching issue within the Perl and Raku community, and addressing this conflict was recognized as a potential catalyst for substantial positive change within the community.

Chapter 7: Identification of Existing Techniques

Following the comprehensive analysis of both interviews and surveys, valuable insights were gleaned, as discussed in the preceding analysis section. This segment delves into potential strategies aimed at mitigating the issues unearthed during the analysis phase.

These strategies draw from established practices within open-source software communities and volunteer organizations. Most of these techniques have been sourced from the paper "*Managing episodic volunteers in free/libre/open-source software communities*" (Barcomb et al., 2020), which itself originated from the Delphi study conducted by Barcomb et al. in 2017 (referred to as EV'17). Additionally, some techniques were tailored to address specific challenges, drawing inspiration from various literature, interview data, and insights obtained through analysis. Besides those emanating from Barcomb's paper, the origin of each technique is explicitly indicated. It is important to note that these techniques are proposed in response to identified issues without considering whether they have been previously attempted. In the subsequent research phase, special attention was directed toward adapting techniques to address the community's more pressing concerns. The problems encountered by the community and the corresponding techniques for potential resolution have been categorized into six overarching themes. These categories were selected to align with the areas under which the issues identified for each construct in Chapter 6 are grouped.

In Section 5.5.3, clusters were identified that may be particularly well-suited for carrying out the implementation of specific techniques. Notably, the "Optimistic" cluster has not been used for any specific techniques since all proposed techniques align with the goal of improving the

community, which centers around a belief in a brighter community future. Consequently, any approach that fosters positivity and optimism is advantageous for this cluster's unique characteristics and aspirations. The "Satisfied" cluster might not significantly impact implementation of these techniques because they may already be content with the current status quo in the community. Consequently, their motivation to adopt any of these techniques might not be as strong as that of the other clusters.

7.1 Popularity of the Community

The analysis reveals a decline in the size of the Perl community. In order to help improve this, some of these techniques would be helpful:

- Recruit people who enjoy social media specifically for the task of promoting and marketing. Those can be social media influencers and bloggers in the tech industry.
- Marketing and promoting the languages and projects by offering demos in universities, companies, and at technical events. University and corporate outreach will help more people become familiar with Perl and its usage. This will hopefully eventually increase the number of users once they know the benefits and advantages of Perl. This technique might suit the "Obligated" cluster as the sense of duty, responsibility, or commitment toward the community may help fulfil this task.
- Send ambassadors to attend smaller events to enable personal interactions with potential participants. This technique might be suited for the "Individualistic" cluster as they are often the holders of knowledge within the community and are best suited to give information to potential participants.

- Educate sponsoring organizations about open-source software projects, including topics such as the necessity of maintenance and the open model of production.
- Marketing major releases or releases in general. This will let the technical world know that the languages are progressing, which is especially useful for Perl, as the general opinion is that it has become stagnant. This can include social media, tech journalists, and bloggers. (derived from Interviewee1)

7.2 Availability of Volunteers (Retention of Volunteers)

Analysis shows the number of volunteers in the Perl community is reducing every year, which is a significant issue. Surveys indicate that volunteers with longer tenure demonstrate higher rates of retention, and lower rates among newer volunteers. To help increase retention, the following techniques can be tried:

- Guiding individuals who are still determining where to begin, particularly by directing them toward junior roles. Mentorship, supplemented by orientation documents for new volunteers, significantly enhances their onboarding experience. This is more suited for the "Individualistic" cluster as they know more about the development done in the community and will know the more manageable tasks for new joiners.
- Educate contributors about what happens to a contribution after it is included in the project. Explain the benefits to the project if they maintain their contribution.
- Ensure that communication channels, both online and offline, are monitored, that queries are directed appropriately, and responses are given. Examples include forums and FAQ documents. Consider automation to help people work through the early processes, such as a chatbot or step-by-step interactive site. This will help newer

volunteers contribute. An integrated system enabling newcomers to connect with experienced volunteers for assistance would also be advantageous.

- Provide mentoring when contributions are rejected due to insufficient quality. This might include tools to help people meet quality requirements. Ensure contributors can always reach out to mentors for guidance. This is suited for the “Individualistic” cluster.
- Exciting volunteers with new, more ambitious projects is one of the best ways to retain volunteers. This particular activity may require more funding, which means techniques to acquire the funding should be considered. (derived from Interviewee3)

7.3 Availability of Volunteers (Recruitment of Volunteers)

Few volunteers are joining the community. To increase recruitment, the following can be tried:

- Learn about the experience, preferences, and time constraints of participants. This will facilitate the creation of well-defined tasks and projects, making it easier for new volunteers to get involved. This can be implemented using surveys and questionnaires during outreach programs. A generic web form could be available on the website, which potential volunteers fill out to show their interests and availability.
- Document the community's working practices, placing particular emphasis on those areas most relevant to new contributors and most appreciated in the community. This, along with an orientation document, will help new volunteers understand precisely what sort of contributions can be made. This, coupled with showcasing contribution benefits, will help recruit volunteers. This technique is suited for the "Individualistic" cluster who hold the most knowledge about development.

- At events, offer walk-through tutorials on getting started as a contributor, culminating in a hackathon working on a specific beginner problem. This includes demos and user experience opportunities. This might be more suited for the “Obligated” cluster.
- Showcase tangible rewards for participation, such as an organizer's dinner, swag, recommendation letters, certificates, or online recommendations.
- Translate documents to attract diverse volunteers from different geographical locations. (derived from Interviewee5)
- Market and promote the community and projects through social media. This can be done by recruiting people skilled in social media.
- Send ambassadors to smaller events to enable personal interactions with potential participants, showcase projects, and promote of the community.
- Invite creators of unofficial initiatives to incorporate them into the main project if they are successful and of high quality. If the project is stand-alone, recognize successes within the project. Provide resources for project implementation and development.
- Foster the expansion of Perl job opportunities to attract volunteers to the community, thereby facilitating its growth. The pivotal role of corporate support in achieving this goal cannot be overstated. Engaging in collaborative partnerships with corporate entities not only secures vital funding but also paves the way for the advancement of essential Perl modules. As these modules evolve, becoming increasingly indispensable, the demand for Perl expertise naturally surges, resulting in the creation of more job opportunities within the Perl ecosystem (Lerner and Tirole, 2002).

- University and Corporate outreach programs are important in locating young talents who are looking for opportunities. Analysis indicates that while corporate outreach was pursued (although subdued), university outreach programs were very much missing. This would benefit the community, as students are seeking opportunities to contribute and put their skills to the test. They also want to experience working on industry projects. This can be a resource to gain volunteers and expand the community.

7.4 Challenges faced by being a part of the community

Interview analysis indicates volunteers encounter numerous challenges while contributing to the community. The lack of available volunteers for projects affects others' contributions. Techniques for addressing this are listed in section 7.3. Other challenges are noted below.

7.4.1 Abandoning of modules

Often, when maintainers abandon modules, they become outdated and cannot support future projects, causing volunteers to seek other solutions, often outside the community. It would be helpful to:

- Educate contributors about what happens to a contribution after it is included in the project and explain the benefits to the project if they maintain their contribution. This goes hand in hand with Retention of Volunteers (section 7.2).
- Give sustained participants access to more critical positions such as maintainers. The only thing required for this is to educate the new volunteer about the module and how to maintain it. (derived from Interviewee2)
- Create a bot for tracking abandoned modules to immediately notify the community so that modules stay updated and to enable quicker transfer of maintainers. There is

currently no formal communication or handover practice, so the community remains unaware of abandoned and outdate modules until someone tries to use them. Automatic tracking will help ensure modules are always maintained and never become outdated.

7.4.2 Implementing Expected Behaviours

Volunteers often do not exhibit certain behaviours expected by the community. To help, the following techniques may be used:

- Define what successful engagement of episodic contributors looks like. Describe how to measure the impact.
- Document the community's working practices, placing particular emphasis on areas most relevant to new and episodic contributors, and where contributions will be most appreciated. This technique is more suited for the "Individualistic" cluster.
- Document the module's workflow architecture and use a project container to allow people to build a local system quickly. Decide upon one recommended way to set up a development environment and focus on this in the documentation. This technique is more suited for the "Individualistic" cluster.
- Utilize advances in continuous integration/continuous delivery to automate routine evaluation.
- Require people to sufficiently document their submissions before they are accepted.
- Help people understand the cooperative values that underlie free and open-source software. This is best done by leading through example.

- Apply project management techniques to the volunteers, such as Work Breakdown Structure (WBS) and Scrum. These will help in better volunteer management and coordination. This was derived from the analysis because it was noticed that, at present, there are no such management techniques implemented in the working practices of the community, which has resulted in communication issues, resource mismanagement and poor-quality control. Employing project management techniques will significantly help tackle these issues.

7.4.3 Tensions Between Perl and Raku Community

One significant issue is the tension between the Perl and Raku community. Interviewees stated that conflicts between the Perl and the Raku factions may have also been the root cause for many issues within the community. To help resolve this conflict, the following techniques are proposed:

- Third-party intervention in open-source software communities. There are two main types of third-party intervention, namely mediation and arbitration. Both play the role of listening to all sides of the dispute and helping the parties to reach a resolution. The main difference is that an arbitrator has the authority to impose the solution. Analysis indicates a mediator would be more helpful in this conflict than an arbitrator because the issue is essentially tension between the two groups, so a collective hearing and proposal of solutions would be most appropriate. The mediator must maintain an unbiased view. Since open-source software is highly collaborative, picking a solution based on majority opinion is preferable to imposing a solution. (RW De Joode, 2004)

- Establish a formal community structure. In general, open-source software communities have very informal structures, but utilizing a formal structure can keep most conflicts at bay. This will help create a communication pathway which will decrease issues escalating caused by poor communication. (Derived from analysis)
- Separate Perl and Raku into two distinct entities with different support systems and organizational structures. This would include determining how to divide the existing funding pools and other shared resources between the two factions and establishing separate funding sources for each community. (Derived from Interviewee1)
- Hire an external service provider to serve as an intermediary in providing sponsorships. In the context of an open-source software community, an external service provider typically refers to a third-party entity that offers specialized services or support to open-source projects or communities. These service providers may offer a range of services tailored to the needs of open-source projects and the communities that develop and maintain them. Look for a stable sponsor to ensure event continuity.

In conclusion, problems were identified based on in-depth analysis of data obtained through the survey and interviews. The proposed solutions are derived from existing literature on thriving open-source software communities for addressing similar problems. However, no evaluation has been conducted regarding implementation of the solutions at this phase of the research, nor the extent to which similar solutions may have already been implemented. To achieve this, a series of meetings were conducted with a Perl and Raku representative to discuss the proposed techniques and any necessary adjustments explicitly tailored to the Perl and Raku community. Details on adaptations of the techniques are in Chapter 8.

Chapter 8: Adaptation of Techniques for Perl and Raku

This chapter considers how the proposed solutions can be modified and expanded to suit the current environment in the Perl and Raku community. A set of the most pressing issues the community is currently facing was chosen and adaptations were developed through discussions and meetings with a representative of TPRF who was knowledgeable in the current issues the community is facing and the best way to transform them to suit the community's current working practices. Going forward, the community could create subcommittees to take work on implementing sections of techniques. The specific techniques have been grouped based on the type of activity or domain they are intended to address.

8.1 Marketing and Promotions

Through the discussions It was decided that the most effective marketing strategy is social media advertisements such as Facebook and Google ads. These platforms can be used for promotion of new releases in Perl and Raku, conferences, businesses that use Perl and Raku, and university outreach programs, triggering options for potential collaborations and community expansion.

Hosting events is a great way to promote the community. Previously release parties, cruises, and hackathons were all conducted more frequently; these sorts of events need to be brought back for community members as well as collaborators. Another common event in the community is Perl Mongers⁶ (2023), a group where everyone meets monthly in their respective regions and collaborates on projects and community management. Over the years, this has reduced but

⁶ <https://www.pm.org/>

rejuvenation would help create a continuous environment and a space for people to collaborate more frequently than annual events such as conferences.

Other options include regular publication of advancements in the technical space, radio and TV commercials, websites, podcasts, merchandising, brochures, and prospectuses.

8.2 Formal Community Structure

Establishing a formal structure, including committees beyond a management board, is very important for any organization. A steering committee could help with strategic direction, risk management, monitoring and evaluation, and accountability. This will help management be more effective in its decision-making. A steering committee typically consists of key stakeholders and domain experts who can provide guidance.

A community affairs committee can enhance relationships within and outside the community. This helps foster positive relationships, addressing community needs as well as maintaining the code of ethics within the community.

A committee specific to the Perl and Raku community would be for the Perl mongers' group. Having a committee oversee the monthly collaborations between volunteers and outside parties will help ensure its longevity. Since this is important for community promotion and fostering collaborations, it will be beneficial for the committee to be comprised typically of those who head the events in their respective locations, to help with funding and logistics.

8.3 Abandoning of Modules

Module maintenance abandonment plagues the Perl and Raku community, hindering the development and relevancy of Perl and Raku. A repository digger (bot) could identify modules that have not progressed over time and locate potential maintainers of the modules.

The creation of a central list of all modules needing adoption, and guidance material on those modules would help new volunteers take up the task of being a maintainer. The guidance material will also contain updated documentation showing which specific areas require upgrades and bug fixes.

Other methods include promotion of these modules through mailing lists and social media groups, and events such as conferences and Perl mongers' meetings.

8.4 Recruitment and Retention of Volunteers

Creating a central list of all the activities and projects, with a description of required contributions categorized based on priority would be a first step toward getting volunteers to work on them. This also helps gauge the volunteers required at varying priority levels. Once finished, different marketing and promotion platforms can be used to make this widespread.

Since the community is declining, it would be helpful to have outreach programs to more prominent open-source software communities, such as Python, Ruby, and Java. Collaborating with these communities will give the opportunity to promote Perl and Raku along with tapping into bigger volunteer pools. This will help retain existing Perl and Raku volunteers as they get a chance to work on other technologies. The other communities also have a chance to experience a different type of technology and collaborate on projects that include both pieces of technology. This may speed development as they might fill each other's shortcomings.

Other techniques include inclusivity workshops; a platform for anonymous questions, feedback, and suggestions; recognition through awards and prizes; bug bounties; and micro-grants for Pull Requests and Merge Requests.

8.5 Funding

Businesses and organizations still use Perl and Raku to implement their technologies, so it is in the best interest of these companies to see Perl and Raku advance. Showcasing new advancements in Perl and Raku helps convince potential industry partners to collaborate and fund new Perl and Raku projects. Some exciting new technologies include ML/AI support packages and BioPerl. Displaying live demos at conferences and job fairs will help promote and engage companies further.

The creation of new jobs will also help increase dependency on the Perl and Raku community for paid training and other types of support. Providing these types of support will help foster better relationships with companies, which can eventually lead to more extensive collaborations in the future. In order to create more jobs, usage of the language must increase in companies, which can be done through extensive marketing and promotion as well as more language advancements. Another advantage of this is that as more jobs are created, more people will be looking to join the community to volunteer and increase their competency in the language. This helps in volunteer recruitment and the expansion of the Perl and Raku community.

Chapter 9: Discussion

9.1 Comparison to Previous Studies

This research project aimed to validate theories previously developed in studies examining various open-source software communities, with a particular focus on the Perl and Raku community. This study introduces a novel perspective by encompassing not only episodic volunteers but also habitual volunteers, offering a comprehensive evaluation of the community's volunteering capabilities.

The most striking distinctions become evident when examining the representativeness table in Chapter 5 (Table 5.1). This table underscores the demographic disparities between this study and prior research efforts. Notably, the average age of community volunteers in this study surpasses that of previous studies. A significant contrast emerges in terms of gender identification, with a higher proportion of volunteers identifying outside the traditional male or female categories. This underscores the uniqueness of the community and emphasizes how the broader conclusions drawn from past studies may not necessarily apply universally.

The Model of Retention's performance falls short compared to the EV'17 study. Of the five constructs examined, only Community Commitment and Psychological Sense of Community exhibited statistical significance. This contrasts the EV'17 study, where Social Norms, Community Commitment, and Satisfaction were statistically significant. In the EV'17 study, since Psychological Sense of Community and Contributor Benefit Motivation proved statistically insignificant, specific questions pertaining to these constructs were dropped from the survey, as they could not be adapted to the context of open-source software. This drop bolstered the

convergent validity of the constructs and contributed to an enhanced level of internal consistency reliability in the EV'17 study. Typically, questions are dropped when cross-loadings, which indicate the extent to which a question accurately measures its intended construct and demonstrates substantial associations with other constructs in the model, exhibit inconsistencies in the constructs.

However, in the current study, the cross-loadings for the questions were generally consistent, with loadings exceeding the 0.5 threshold, except for two questions. When this inconsistency is context-specific, there may be no necessity to eliminate the questions (Joseph et al., 2010). Another reason for retaining the questions related to statistically insignificant constructs in this study was to thoroughly assess the entire spectrum of the model of retention as applied to a specific community, encompassing both episodic and habitual volunteers. This approach holds promise for future endeavours aimed at refining the model and survey questions to optimize model fit. These refinements might permit correlated errors between questions that are influenced by common method bias or revise the measurement model according to theoretical principles. Correlated errors can account for relationships between questions affected by similar sources of bias, improving the model's accuracy.

The model did not perform as well as in the EV'17 study. Factors contributing to this discrepancy might include that the EV'17 study solely focused on episodic volunteers, whereas this study encompassed both episodic and habitual volunteers. This variation may have been due to the foundational principles of Hyde's (2016) Model of Retention, which was designed specifically for studying episodic volunteering. Notably, even though the model of retention

performed better in the EV'17 study with only episodic volunteers, the overall conclusion was that the model was still not suitable in the context of open-source software.

A crucial consideration for the weaker performance of the model in open-source software is that Hyde's model originated from the realm of cancer non-profit organizations, which may have influenced the selection of constructs. These constructs might not function optimally within the context of open-source software communities. Even under such circumstances, constructs like Social Norms may not accurately reflect the dynamics of open-source software communities. This aligns with the theory proposed by Barcomb et al. (2018), suggesting that volunteers contributing to open-source software often operate without significant societal pressures, as the general public typically lacks a vested interest in their activities. A plausible recommendation would involve the development of a new model with constructs tailored to the unique dynamics of open-source software communities. The revised model should incorporate constructs that cater to the needs of both episodic and habitual volunteers, as a diversity of motivations, skills, and availability exists within these two volunteer categories. The model then becomes adaptable to studying either episodic volunteers, habitual volunteers, or a combination of both, provided that they have access to adequate data.

The model must possess the versatility to adjust seamlessly to the evolving circumstances of volunteers. For instance, the model needs to flex to the shifting volunteering patterns of individuals. Habitual volunteers may, at times, transition into episodic volunteering due to shifts in their life circumstances. Therefore, the model must be equipped to account for this dynamic. In such cases, constructs like Community Commitment may assume paramount significance, as other constructs may not be apt within the current framework of the model of retention in the

context of open-source software. This will be particularly valuable for researchers examining and comprehending the long-term evolution of volunteering dynamics within communities, especially those that have traversed various phases over the years.

In addition to the changing situational dynamics of volunteers, the development of this model should consider theories addressing motivations driving people to contribute to an open-source software community and factors influencing contributors' decisions to switch to other communities. For example, this may occur when there is a shift in social standards, rendering it no longer compatible with their preferences (Von Krogh et al., 2012). These aspects are notable features of volunteer behaviour within the open-source software ecosystem, which contribute to the volunteering dynamics of the community.

In the EV'17 study, Social Norms, Community Commitment, and Satisfaction were all positively correlated to Intention to Remain, while Contributor Benefit Motivation and Psychological Sense of Community did not demonstrate a relationship with the Intention to Remain. In this study, only Community Commitment and Psychological Sense of Community had a positive relation to the Intention to Remain and were statistically significant, whereas the other constructs did not demonstrate a relationship with the Intention to Remain.

Previous findings did not heavily attribute Social Norms to the Intention to Remain in open-source software. In fact, it was negatively correlated with the Intention to Remain in this study. This negative correlation might be attributed to the inclusion of habitual volunteers in the study, who are typically not motivated by social constructs. Community Commitment, in general, was considered not significant when it comes to episodic volunteering but is actually the strongest

construct towards Intention to Remain in the previous studies (Barcomb et al., 2018) and is further strengthened in this study due to the inclusion of habitual volunteers.

Psychological Sense of Community was the only construct that differed from EV'17; it was statistically significant in this study, which again can be attributed to the inclusion of habitual volunteers in the study. When comparing, all different constructs exhibit varying relationships with the Intention to Remain, except for Community Commitment. This observation highlights differences in the impact on constructs when considering only episodic volunteers (EV'17) compared to volunteers as a whole (this study). Another notable difference was that the internal consistency measure was not as strong as the EV'17 study, which might have been because of the inclusion of both habitual and episodic volunteers together, or simply because the community is in decline. Any one or a combination of these factors could have contributed to why the internal consistency reliability was not as strong as the other study.

When examining moderating factors in EV'17, age and tenure did not affect the constructs, whereas in this study they did proportionately affect Community Commitment. This is consistent with the inclusion of habitual volunteers in the study because habitual volunteers tend to have more extended tenure periods as volunteers in the community.

One interesting finding in the EV'17 study was the identification of four distinct groups in episodic volunteering. A similar clustering was conducted in this study as well, and four distinct groups were also identified here (See Table 5.6). The Satisfaction group was well documented in EV'17 and in previous studies such as Wu et al. (2017), so it was not unexpected. The Obligated group piqued curiosity in EV'17 and was also found in this study. This is more closely associated with Community Commitment. In the context of episodic volunteering Community Commitment

might have been an interesting observation, but in this study it was not unexpected as it included habitual volunteers as well. Two clusters not seen in other literature and unique to the Perl and Raku community were the Individualistic and Optimistic groups. An Optimistic group was likely formed because of the decline of the community and the efforts to revitalize it. Since literature on dying and revitalizing communities was not available, this cluster may be unique to the Perl and Raku community. Individualistic group formation is likely due to the lack of volunteers available in the community, so most of the work and development is solely driven by individual members, leaving the locus of knowledge in the hands of individual volunteers who have been contributing to the community for an extended period. This is discussed in Von Krogh (2009). Also, compared to the EV'17 study, the clusters are different, suggesting there is no universality: It can vary from one community to another. This can depend on several different factors for that particular community, such as community health, stage in the lifecycle, and size of the community.

Finally, three research objectives were delineated at the outset of this thesis, serving as the guiding framework for this research endeavour. The first research objective aimed to explore the potential applicability of existing theories and methodologies pertaining to volunteering within the realm of open-source software, particularly within a specific community context. It became evident that the findings differed, with notable variations observed, such as factoring the Model of Retention in different ways. Consequently, a straightforward translation to accommodate this specific community and encompass the entire spectrum of volunteering resources (both episodic and habitual) proved challenging. Amidst these disparities, certain parallels emerged, permitting the techniques identified in prior studies to bear relevance to this research project. These

techniques could be adapted to better align with the unique dynamics of the Perl and Raku community, aligning with the second and third research objectives.

9.2 Limitations

Creswell's framework is a valuable tool for addressing limitations in both quantitative and qualitative data analysis (Creswell & Creswell, 2017). This framework includes the following categories: Sampling Limitations, Data Quality and Reliability, and Researcher Subjectivity.

Sampling limitations in quantitative data analysis encompass concerns related to sample size, selection bias, response bias, and self-selection bias, whereas in qualitative data analysis, it revolves around the potential for limited diversity in the sample and leads to lesser generalizability. Data Quality and Reliability considerations involve the assessment of measurement errors, missing data, and data validity in quantitative analysis, while in the qualitative context, they involve the reliability of coding and analysis processes. The issue of Researcher Subjectivity, characterized by confirmation bias, emerges as a common thread, affecting both quantitative and qualitative data analysis.

9.2.1 Sampling Limitation

When examining sampling limitations within quantitative data analysis, sample size plays a pivotal role in any data analysis project. Inadequate sample sizes can introduce significant limitations, such as the potential for a sample to inaccurately reflect the broader population or the specific group under study, as well as reduced statistical power, making it challenging to detect genuine effects or differences between groups.

In this research study, the survey consisted of a pool of responses that captured different aspects of the community. However, only 88 of the 142 total responses could be used with

complete and usable data, which may potentially introduce bias into the survey, such as sampling bias, response bias, and self-selection bias. Selection bias arises when the selected participants do not accurately represent the larger population (Groves et al., 2009). To assess this, external validation was employed to gauge the study's representativeness compared to other studies, as demonstrated in Table 5.1. It was anticipated that some degree of selection bias might occur due to the study's use of self-selection as well as the challenge of accurately gauging the surveyed population's representativeness due to a lack of knowledge about its composition. This approach meant that participants who were more deeply committed to the community were more likely to partake in the survey.

The next aspect of sampling limitation is response bias. Response bias occurs when participants do not provide answers that genuinely reflect their beliefs, potentially leading to inaccurate results (Dillman et al., 2014). To mitigate this concern, the survey was thoughtfully designed to ensure complete anonymity, unless participants voluntarily disclosed some identifying information. This approach minimized the likelihood of participants offering inaccurate information.

Lastly, self-selection bias arises when participants voluntarily choose to engage in the survey, potentially yielding data that may not fully represent the entire population, as opposed to randomized or carefully curated selections (Couper et al., 2008). In light of this study's use of self-selection for the online survey, quantifying a specific response rate proved challenging due to uncertainty of how many Perl and Raku community members were aware of the survey's existence. Notably, extensive efforts were made to promote the survey across various communication channels, including Slack, Facebook, and community website news articles.

When considering sampling limitations for qualitative data analysis, limited diversity in the sample is concerning. This may be applicable for the interview data collected in this study. Since only six interviews were conducted, it may not be representative of the diversity of the entire population of the Perl and Raku community. Since this was done in tandem with the survey, which draws from a broader and more diverse pool of participants, the data as a whole captures the community's diversity. The presence of self-selection in the interview data is noteworthy, as it reflects a bias toward individuals with greater community engagement and investment participating in the interviews. This bias is evident in the composition of the six participants, with four showing stronger affiliations with Perl rather than Raku. This pattern may stem from the perception that those aligned with the Perl faction in the Perl-Raku conflict sense a more significant effect of the decline in the community because the community was once originally associated solely with Perl. They might be more inclined to take part in the interviews. This aspect may also carry positive implications for future endeavours, particularly in implementing techniques within the community. Interview participants, given their active involvement and willingness to volunteer, could represent valuable contributors to future initiatives. Therefore, obtaining insights from these engaged participants, even if it introduces a degree of self-selection bias, can ultimately enhance the overall outcome of the study.

9.2.2 Data Quality and Reliability

The second category for assessing research limitations centers on data quality and reliability. This category pertains to challenges stemming from potential measurement errors or biases that can undermine survey data quality and the efficacy of the analytical model employed. To ensure data quality, a set of statistical tests (section 5.1), including assessments for common-

method bias, was conducted, with all tests yielding successful outcomes. The implementation of the Model of Retention shows that the internal consistency reliability (section 5.4.2) of three constructs—Contributor Benefit Motivation, Social Norms, and Psychological Sense of Community—does not meet the desired level of reliability. This poses a challenge to the overall model.

Several factors may contribute to this outcome. Firstly, Social Norms and Contributor Benefit Motivation exhibit statistical insignificance, meaning that their lower reliability could be attributed to chance or random variability. For Psychological Sense of Community, it is possible that insufficient data points were available to establish strong internal consistency reliability (Brace et al., 2016). Brace et al. noted that with limited sample size, the internal consistency reliability of their questionnaire was compromised, hindering the ability to detect consistent patterns. A larger sample size helps address this concern and is advisable to explore diverse methods for encouraging volunteers to engage in the study, including offering incentives. Pilot tests with a small group of volunteers may help identify and rectify any issues with the survey.

One crucial aspect regarding social norms as a construct in the Model of Retention is its contextual relevance. Initially, it was incorporated into the model while examining volunteerism within cancer control non-profit organizations, where individuals often grapple with societal expectations and pressures that motivate their involvement (Hyde et al., 2014). However, when applying this construct to open-source software communities, Barcomb et al. (2020) suggest that the general public tends to have fewer opinions and exert less societal pressure regarding open-source software compared to areas like cancer control non-profits. The deliberate decision to include Social Norms as a construct persisted, even when acknowledging the potential absence

of a significant impact. This decision was informed by insights from Barcomb et al.'s (2020) Delphi study, which indicated that Social Norms did indeed influence non-code volunteers within open-source software communities. Nevertheless, in the specific context of the Perl and Raku community, this particular construct did not exhibit the same influential effect on non-code contributors. This underscores the notion that each open-source software community is unique, and generalizing findings from past studies to all such communities may not hold true.

In qualitative data analysis, this category pertains to the coding and analysis process. The main limitation here is the potential for interpretation bias, which can be particularly pronounced when a single researcher handles the qualitative data. Interpretation bias refers to the cognitive bias where researchers, consciously or unconsciously, interpret information or research findings in a way that aligns with their pre-existing beliefs, expectations, or hypotheses (Rosenthal, 1976). To address this, a secondary coder was enlisted. The coding and analysis were conducted independently by both coders. Subsequently, a meticulous comparison of their work ensued, and Inter-Coder Reliability was calculated to gauge agreement. If the predefined threshold was not attained, the coders engaged in collaborative discussions to enhance their mutual understanding of the research topic. This iterative process was repeated until the requisite threshold was met. Through this rigorous method, the data analysis incorporated more than one perspective, significantly reducing the risk of interpretation bias.

9.2.3 Researcher Subjectivity

Researcher subjectivity refers to the influence of a researcher's personal beliefs, values, experiences, and perspectives on the research process and its outcomes. It is the inherent subjectivity that researchers bring to their work, which can shape how they design studies,

collect data, analyze findings, and interpret results. This is often termed confirmation bias, which is the inclination of individuals to prioritize information aligning with their pre-existing beliefs or values while diminishing or dismissing contradictory evidence (Klayman & Ha, 1987). Confirmation bias can significantly impact decision-making, problem-solving, and information interpretation. This concern was pre-emptively addressed in the study, as the primary researcher possessed no prior knowledge of the Perl and Raku community before the survey was conducted. All information pertaining to the community was exclusively derived from the research study itself, effectively mitigating the risk of confirmation bias in the survey. Furthermore, as detailed in Chapter 4 (section 4.3), the survey questions were adapted from the previously validated survey questions used in the EV'17 study (Barcomb et al., 2018).

Similarly, the interview questions were modelled after the survey questions, albeit designed to elicit deeper insights. Additionally, previous literature, including the EV'17 study, served as a reference point when crafting the interview questions for this study. This approach minimizes the potential for confirmation bias to influence the results.

To validate the representativeness of the analysis results within the community, a post-analysis member validation (member checking) was conducted. Member checking typically involves presenting participants with summaries of the data collected or the researcher's initial analysis. Participants are then asked to provide feedback, corrections, or additional insights to validate or refine the findings. This acknowledges participants as co-constructors of knowledge and ensures their voices are accurately represented in the research (Morse et al., 2002). Member checking is valuable for qualitative researchers as it allows for participant input, helps identify potential researcher bias or misinterpretations, and enhances credibility of findings. A

presentation on this research project was delivered at the Perl and Raku Conference held in Toronto in June 2023. It was strikingly apparent that a majority of the analysis outcomes and techniques strongly resonated with audience members. Genuine enthusiasm was expressed for implementation of the recommended techniques to address the community's challenges and foster positive change.

9.3 Future Work

Future work worth considering includes implementing the proposed techniques in collaboration with the community and conducting action research, which is a collaborative and iterative process that combines a phenomenon and its implementation (or an *“action”*) along with reflective evaluation (*“research”*). Action research would provide more insights into how techniques were implemented. After the techniques were implemented and a certain period had elapsed, another analysis would be conducted, which would include both quantitative and qualitative data analysis (surveys and interviews) to evaluate the outcome of the implemented techniques.

Another future work direction could be the extrapolation of the methodology, results, and techniques obtained in this research to larger open-source software communities that also face issues in volunteer management. Because other studies on open-source software have concentrated on episodic volunteering and not on the volunteering situation as a whole, extrapolating this study onto other communities can help address volunteering problems in the community as well as better management of all volunteers in the community, and not just Episodic Volunteers. It will be interesting to see adaptations that might be required to implement these theories and solutions in other communities with different volunteer clusters (as shown in

section 5.5.3). This can give insights into how similar and different open-source communities are from one another.

Given the dearth of existing literature concerning the revitalization of declining open-source software communities, this research has the potential to serve as a valuable resource for future studies in this domain. By shedding light on the identifiable characteristics of a declining community and the associated challenges, this study paves the way for a better understanding of this issue. Such insights can be instrumental in formulating effective strategies to rejuvenate and support these communities.

Chapter 10: Conclusion

This research sought to validate existing theories and practices that were originally developed for a broader set of open-source software communities and focused primarily on episodic volunteers. The aim was to apply and assess their relevance within the specific context of the Perl and Raku community, considering all types of volunteers, not exclusively episodic ones. This validation was conducted through an in-depth case study that examined the volunteering dynamics within the Perl and Raku community. The study demonstrated that the Model of Retention did not perform well when applied to a single community like Perl and Raku, and it also highlighted the challenge of directly translating techniques and solutions from past studies to fit the Perl and Raku community specifically. Nevertheless, the research identified areas of relevance, enabling the adaptation of techniques and methods to align better with the Perl and Raku community's distinct characteristics and requirements.

The conclusions drawn from this study hold the potential for broader applicability to other open-source software communities experiencing a similar stage of decline. Specifically, constructs such as Community Commitment and Psychological Sense of Community may emerge as prominent constructs influencing the Intention to Remain, as individuals who exhibit loyalty and find a sense of comfort within the community are more likely to stay engaged and continue their participation. Finally, this research study provides a valuable contribution as a case study of a declining community, an area characterized by limited research. While the contribution may be modest in scope, it addresses a gap in the existing literature.

This research project also examined the volunteering situation of an open-source software community (Perl and Raku) to find out the current volunteering problems faced by the community. Finally, it proposed generic and adapted techniques to overcome these problems.

Data collected in surveys and interviews was analyzed using quantitative and qualitative data analysis methods. The Model of Retention was employed on the data using PLSPM to understand which constructs of the model are working for the community and which are not. Community Commitment and Psychological Sense of Community were the two biggest constructs toward Intention to Remain and were statistically significant. In contrast, Social Norms were negatively correlated to the Intention to Remain, and all the others were statistically insignificant. Qualitative data analysis was conducted to learn why each factor of the model is the way it is, and a set of issues was identified for each construct. The most significant and common issues were the decline of the community in terms of popularity as well as the number of active volunteers in the community. The conflict between Perl and Raku was also one of the significant issues identified.

Based on the issues identified in the analysis, a total of 35 techniques were proposed, which were mostly from past literature, with some explicitly developed to address specific issues in the community. Based on the proposed solutions, some techniques for the most pressing issues faced by the communities were modified into an implementable action item.

References

- Albers, M. J. (2017). Quantitative data analysis—In the graduate curriculum. *Journal of Technical Writing and Communication*, 47(2), 215-233.
- Barcomb, A., Kaufmann, A., Riehle, D., Stol, K. J., & Fitzgerald, B. (2018). Uncovering the periphery: A qualitative survey of episodic volunteering in free/libre and open-source software communities. *IEEE Transactions on Software Engineering*, 46(9), 962-980.
- Barcomb, A., Stol, K. J., Fitzgerald, B., & Riehle, D. (2020). Managing episodic volunteers in free/libre/open-source software communities. *IEEE Transactions on Software Engineering*, 48(1), 260-277.
- Barcomb, A., Stol, K. J., Riehle, D., & Fitzgerald, B. (2019, May). Why do episodic volunteers stay in FLOSS communities? *2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)* (pp. 948-959). IEEE.
- Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, 74(1), 137-143.
- Bonaccorsi, A., & Rossi, C. (2003). Why open-source software can succeed. *Research policy*, 32(7), 1243-1258.
- Brace, N., Kemp, R., & Snelgar, R. (2006). *SPSS for psychologists: A guide to data analysis using SPSS for Windows*. Lawrence Erlbaum Associates Publishers.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement*, 41(3), 687-699.
- Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods*, 15(1), 17-32.
- Couper, M. P., & Miller, P. V. (2008). Web survey methods: Introduction. *Public Opinion Quarterly*, 72(5), 831-835.
- Cramer, D. (2003). *Advanced Quantitative Data Analysis*. McGraw-Hill Education (UK).
- Cressie, N. A. C., & Whitford, H. J. (1986). How to use the two-sample t-test. *Biometrical Journal*, 28(2), 131-148.
- Creswell, J. W., & Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage publications.

- Crowston, K. (2011, June). Lessons from volunteering and free/libre open-source software development for the future of work. *Researching the Future in Information Systems: IFIP WG 8.2 Working Conference*, Turku, Finland, June 6-8, 2011. Proceedings (pp. 215-229). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Crowston, K., & Howison, J. (2005). The social structure of free and open-source software development. *First Monday*, Volume 10, Number 2 – 7.
- Crowston, K., Wei, K., Li, Q., & Howison, J. (2012). Core and periphery in free/libre and open-source software team communications. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 501-510).
- Cnaan, R. A., & Goldberg-Glen, R. S. (1991). Measuring motivation to volunteer in human services. *The Journal of Applied Behavioral Science*, 27(3), 269-284.
- Cnaan, R. A., Meijs, L., Brudney, J. L., Hersberger-Langloh, S., Okada, A., & Abu-Rumman, S. (2021). You thought that this would be easy? Seeking an understanding of episodic volunteering. *VOLUNTAS: International Journal of Voluntary and Non-profit Organizations*, 1-13.
- Curto-Millet, D., & Corsin Jimenez, A. (2022). The sustainability of open-source commons. *European Journal of Information Systems*, 1-19.
- Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012, February). Social coding in GitHub: transparency and collaboration in an open software repository. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 1277-1286).
- Davis, J. H., Schoorman, F. D., & Donaldson, L. (2018). Toward a stewardship theory of management. *Business Ethics and Strategy*, Volumes I and II (pp. 473-500).
- De Joode, R. V. W. (2004). Managing conflicts in open-source communities. *Electron. Mark.*, 14(2), 104-113.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. John Wiley & Sons.
- Doty, D. H., & Glick, W. H. (1998). Common methods bias: does common methods variance really bias results? *Organizational Research Methods*, 1(4), 374-406.
- Eckert, R., & Mueller, A. (2017). Sustainability and Diversity of Open-Source Software Communities. *International Conference on Software Engineering Advances 2017*, 59-64.
- Foster, D. (2015). A Model of Open-Source Software Project Sustenance. *Information Systems Research*, 26(3), 631-649.

- Foulonneau, M., Pawelzik, R., Grégoire, B., & Donak, O. (2013). Analyzing the open-source communities' lifecycle with communication data. *Proceedings of the Fifth International Conference on Management of Emergent Digital EcoSystems* (pp. 340-344).
- Gamalielsson, J., & Lundell, B. (2012). Long-term sustainability of open-source software communities beyond a fork: A case study of LibreOffice. *IFIP International Conference on Open-Source Systems* (pp. 29-47). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Gamalielsson, J., & Lundell, B. (2014). Sustainability of Open-Source software communities beyond a fork: How and why has the LibreOffice project evolved? *Journal of Systems and Software*, 89, 128-145.
- Geiger, R. S. (2017). *Summary Analysis of the 2017 GitHub Open-Source Survey*. arXiv preprint arXiv:1706.02777.
- Guimarães, A. L. S., Korn, H. J., Shin, N., & Eisner, A. B. (2013). The life cycle of open-source software development communities. *Journal of Electronic Commerce Research*, 14(2), 167.
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey Methodology*. John Wiley & Sons.
- Hyde, M. K., Dunn, J., Scuffham, P. A., & Chambers, S. K. (2014). A systematic review of episodic volunteering in public health and other contexts. *BMC Public Health*, 14, 1-16.
- Hyde, M. K., Dunn, J., Bax, C., & Chambers, S. K. (2016). Episodic volunteering and retention: An integrated theoretical approach. *Nonprofit and Voluntary Sector Quarterly*, 45(1), 45-63.
- Iskoujina, Z., & Roberts, J. E. (2015). Knowledge sharing in open-source software communities: motivations and management. *Journal of Knowledge Management*, 19(4), 791–813.
- Joseph, F. H. J. R., Barry, J. B., Rolph, E. A., & Rolph, E. A. (2010). *Multivariate Data Analysis*. Pearson Prentice Hall.
- Krippendorff, K. (2018). *Content Analysis: An Introduction to its Methodology*. Sage publications.
- Lenz, M., & Lenz, M. (2020). What Is Raku?. *Raku Fundamentals: A Primer with Examples, Projects, and Case Studies*, 1-6.
- Lerner, J., & Tirole, J. (2002). Some simple economics of open source. *The Journal of Industrial Economics*, 50(2), 197-234.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.
- Meijs, L. C. P. M., & Brudney, J. L. (2007). Winning volunteer scenarios: The soul of a new machine. *International Journal of Volunteer Administration*, 24(6), 68-79.

- Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, 1(2), 13-22.
- Nyman, L., & Lindman, J. (2013). Code forking, governance, and sustainability in open-source software. *Technology Innovation Management Review*, 3(1), 7–12.
- Osterloh, M., & Rota, S. (2007). Open-source software development—Just another case of collective invention? *Research Policy*, 36(2), 157-171.
- Richard, E., & Emerson, R. M. (1976). Social exchange theory. *Annual Review of Sociology*, 2(1), 335-362.
- Robles, G., Arjona Reina, L., Serebrenik, A., Vasilescu, B., & González-Barahona, J. M. (2014, May). FLOSS 2013: A survey dataset about free software contributors: challenges for curating, sharing, and combining. *Proceedings of the 11th Working Conference on Mining Software Repositories* (pp. 396-399).
- Robles, G., Gonzalez-Barahona, J. M., & Herraiz, I. (2009, May). Evolution of the core team of developers in libre software projects. In *2009 6th IEEE International Working Conference on Mining Software Repositories* (pp. 167-170). IEEE.
- Rosenthal, R. (1976). Experimenter effects in behavioral research.
- Sanchez, G., Trinchera, L., Sanchez, M. G., & FactoMineR, S. (2013). Package ‘plsrm’. *Citeseer*: State College, PA, USA.
- Sanchez, G. (2013). *PLS Path Modeling with R*. Berkeley: Trowchez Editions, 383(2013), 551.
- Seers, K. (2012). Qualitative data analysis. *Evidence-based Nursing*, 15(1), 2-2.
- Shah, S. K. (2006). Motivation, governance, and the viability of hybrid forms in open-source software development. *Management Science*, 52(7), 1000-1014.
- Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, 9(1), 4-11.
- Tobias, S., & Carlson, J. E. (1969). Brief report: Bartlett's test of sphericity and chance findings in factor analysis. *Multivariate Behavioral Research*, 4(3), 375-377.
- Treude, C., & Robillard, M. P. (2016, May). Augmenting API documentation with insights from stack overflow. *Proceedings of the 38th International Conference on Software Engineering* (pp. 392-403).
- Ven, K., & Mannaert, H. (2008). Challenges and strategies in the use of open-source software by independent software vendors. *Information and Software Technology*, 50(9-10), 991-1002.

von Krogh, G. (2009). Individualist and collectivist perspectives on knowledge in organizations: Implications for information systems research. *The Journal of Strategic Information Systems*, 18(3), 119-129.

Von Krogh, G., Haefliger, S., Spaeth, S., & Wallin, M. W. (2012). Carrots and rainbows: Motivation and social practice in open-source software development. *MIS Quarterly*, 649-676.

Wall, L. (1994). *The Perl programming language*.

Wenger, E. (1999). Communities of Practice: Learning, Meaning, and Identity. *Cambridge University Press*.

Wynn Jr., Donald E. (2004). Organizational Structure of Open-Source Projects: A Life Cycle Approach. *SAIS 2004 Proceedings*. 47.

Xu, J., Christley, S., & Madey, G. (2005). The open-source software community structure. *NAACSOS2005*.

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence*, 17(5-6), 375-381.

Appendices

Appendix A: Survey Questions

Section A: General open-source participation

The following questions are used to understand your participation habits, which means any type of unpaid activity, including documentation, translation, bug reports, mentoring, programming, or any other activity you do for an open-source project. For this survey we use the following definitions:

Habitual participation: you contributed frequently (10 or more substantial contributions)

OR

you made 2 or more contributions of any size per month, for 6 consecutive months If your participation is not habitual, we refer to your participation as “episodic”.

⇒ A1. How many free/libre/open-source software projects have you participated in in the last 12 months?

[number]

⇒ A2. In how many free/libre/open-source software projects have you clearly been a habitual contributor in the last 12 months?

[number]

⇒ A3. In how many free/libre/open-source software projects have you clearly been an episodic contributor in the last 12 months?

[number]

⇒ A4. Have you contributed to the Perl/Raku community more episodically or habitually for the last 12 months? The rest of the survey pertains to your experiences in the Perl/Raku community.

1. episodic
2. habitual
3. neither

⇒ If neither, end survey with: We thank you for your willingness to participate in the survey, but we are only collecting responses about Perl/Raku contributors.

Section B: Participation Overview

⇒ B1. Please select the area where your main contributions are:

1. code, programming
2. other contributions (documentation, translations, tests, artwork...)
3. both

⇒ B2. In which year did you first contribute to the community?

[number]

⇒ B3. On average, how many hours per month did you contribute to this project over the last 12 months?

[number]

Section C: Participation

Please indicate the extent to which you agree with each statement.

⇒ C1. I am quite similar to most people in my community.

1. Strongly disagree

2. Disagree
3. Neither disagree nor agree
4. Agree
5. Strongly agree

[All questions in this section use the same scale for answers]

- ⇒ C2. If I feel like talking, I can generally find someone in the community to talk to right away.
- ⇒ C3. If there was a serious problem in the community, the people could get together to solve it.
- ⇒ C4. Other people think that contributing is important to me.
- ⇒ C5. It is important to my friends and relatives that I continue contributing.
- ⇒ C6. Many of the people that I know expect me to continue as a contributor.
- ⇒ C7. No one would really be surprised if I just stopped contributing.
- ⇒ C8. I enjoy my contribution experience.
- ⇒ C9. My contribution experience is personally fulfilling.
- ⇒ C10 My contribution experience is worthwhile.

Section D: Participation (cont.)

- ⇒ D1. I am likely to continue to contribute to this project.
- ⇒ D2. I feel very little loyalty to this community
- ⇒ D3. I am proud to tell others that I am part of this community.
- ⇒ D4. It would take very little change in my present circumstances to cause me to leave the community.

- ⇒ D5. There's not too much to be gained by sticking with this community indefinitely.
- ⇒ D6. I really care about the fate of this community.
- ⇒ D7. For me this is the best of all possible communities to participate in.
- ⇒ D8. I plan to participate in this community in the future.
- ⇒ D9I will recommend that others participate in this community.
- ⇒ D10. I will tell others about the positive experiences that I had participating in this community.

Section E: Participation (cont.)

- ⇒ E1. I hope that participation in this community is a part of my life for years to come.
- ⇒ E2. I am more motivated to participate because of my recent contribution experience with this community.
- ⇒ E3. I care about the Perl/Raku community.
- ⇒ E4. I want to be recognized for my contributions.
- ⇒ E5. I want to receive a tangible acknowledgment of my contributions.
- ⇒ E6. I participate to get a reputation in the free/open-source developers' scene, or within the Perl/Raku community.
- ⇒ E7. I participate to improve my job opportunities.
- ⇒ E8. I participate to make money.
- ⇒ E9. I participate to learn and develop new skills.

Section F: Demographics

Demographic information is collected to gauge the representativeness of the survey. Please answer the questions honestly. If you would prefer not to answer, please simply skip the question. These questions are all optional.

⇒ F1. What is your year of birth?

[year]

⇒ F2. Which of the following describes how you think of yourself?

1. male
2. female
3. other _____

⇒ F3. What is your highest education level?

1. University/college
2. Technical/trade
3. High school
4. Primary school

⇒ F4. GitHub offers us very valuable insights into contribution patterns of developers when paired with your survey answers. We will not include your GitHub identity in any reports/papers. However, by providing your GitHub identity, your survey responses will no longer be anonymous to the researchers. If you are willing, please enter your GitHub username:

[text]

Section G: Conclusion

⇒ G1. If you have any comments for the researchers about this survey, please state them here:

[text]

⇒ G2. By submitting this survey, you agree to participating in the research. If you have changed your mind, do not hit submit. Incomplete surveys are not included in the analysis.

[submit]

⇒ G3. Thank you for your participation.

Appendix B: Interview Questions

Section 1: Introduction

[Self-introduction and recording of consent]

Section 2: Experience

⇒ 2.1 Personal experience

1. How long have you been involved in the Perl/Raku community?
2. In which ways do you contribute to the project?

⇒ 2.2 Experience with volunteers

1. How many contributors would you estimate you regularly work with?
2. How many contributors would you estimate you work with who participate intermittently?
3. In what ways are people contributing habitually to the project, which you have observed?
4. In what ways have you observed people contributing episodically to the project?

5. In which ways do you work with other contributors?

Section 3: Current volunteering situation

⇒ 3.1. Existing practices

1. What are some of the ways in which volunteers contribute to the project's success?
2. What practices do you use to help volunteers contribute to the project's success?
3. What sort of behavior do you want to promote in volunteers?
4. What practices do you use to promote this behavior?
5. How important is it to you to retain volunteers?
6. What practices do you use to retain volunteers?
7. What other objectives, if any, do you have concerning volunteers?
8. What practices are you using to further those objectives?

⇒ 3.2. Episodic volunteers

1. What are some of the differences between working with habitual volunteers and episodic volunteers?
2. What are some of the challenges with involving episodic volunteers?
3. What are some of the challenges with making use of episodic volunteers?
4. What other objectives, if any, do you have concerning episodic volunteers?
5. What practices are you using specifically for episodic volunteers?
6. Are there activities that you think are particularly suited to episodic volunteering?
7. Are there activities that you think are particularly unsuited to episodic volunteering?

⇒ 3.3. Volunteer scenario

1. What would you define as high availability, in terms of contributing to the project?
2. What would you define as low availability?
3. What would you consider to be specialized knowledge or expertise which could be useful in your area of the project?
4. What kind of skills would you expect someone whom you did not consider specialized to nonetheless possess?
5. If you think of the volunteers you work with, and try to categorize them in terms of their availability, what proportion would you describe as having high availability, and what proportion low availability?
6. Thinking again of the volunteers you work with, what proportion would you categorize as having high levels of specialized knowledge or expertise, compared to low levels of specialized knowledge?
7. Could you give approximate percentages for each combination of availability and knowledge?
8. Specifically, high availability and high expertise, high availability and low expertise, low availability and high expertise, and low availability and low expertise?

Section 4: Desired outcomes

⇒ 4.1. Approach

1. What approaches do you use to identify and implement new practices in the community?

⇒ 4.2. Problems

1. What are the biggest challenges you are facing in the part of the community you work with?

⇒ 4.3. Imagining

1. What are some potential improvements you can envision in volunteering in the part of the Community you work with?
2. What are some potential improvements you can envision which specifically involve episodic volunteering?

⇒ 4.4. Assistance

1. Are there any changes that you would like to see in the part of the community you work in, but do not know how to bring about?
2. At the conclusion of this research, we will analyze the current volunteering situation and make recommendations for improving volunteer management and retention. Subsequently, I will work with community managers on implementation of practices and observe the outcome. Are you potentially interested in participating in the next phase of the research?

Section 5: Conclusion

⇒ 5.1. Open topic

1. This concludes my questions. Do you have any additional thoughts you'd like to share?
2. Do you have any questions for me about the research?

⇒ 5.2. Recognition

1. In the published papers and reports, there is the opportunity for research participants to be credited by name. Would you like to be credited by name or nickname, or would you prefer to be recognized anonymously?
2. If you change your mind at any point prior to publication about being recognized, please let me know.
3. Would you like to be informed when any research papers are published?

⇒ 5.3. Thank you

[Thank you speech]

Appendix C: Ethics Approval Certificate



Conjoint Faculties Research Ethics Board
Research Services Office
2500 University Drive, NW
Calgary AB T2N 1N4
Telephone: (403) 220-4283/6289
cfreb@ucalgary.ca

CERTIFICATION OF INSTITUTIONAL ETHICS APPROVAL

Ethics approval for the following research has been renewed by the Conjoint Faculties Research Ethics Board (CFREB) at the University of Calgary. The CFREB is constituted and operates in compliance with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS 2).

Ethics ID: REB20-2135_REN2
Principal Investigator: Ann Barcomb
Co-Investigator(s):
Student Co-Investigator(s): Aadharsh Hariharan
Andrea Gail Tampus
Study Title: Evaluation of existing volunteering capabilities of a
free/libre/open source software project
Sponsor: University of Calgary

Effective: 25-Feb-2023

Expires: 25-Feb-2024

Restrictions:

This Certification is subject to the following conditions:

1. Approval is granted only for the research and purposes described in the application.
2. Any modification to the approved research must be submitted to the CFREB for approval.
3. Reportable events (SAE's, new safety information, protocol deviations, audit findings, privacy breaches, and participant complaints) are to be submitted in accordance with the Board's reporting requirements.
4. An annual application for renewal of ethics certification must be submitted and approved by the above expiry date.
5. A closure request must be sent to the CFREB when the research is complete or terminated.

Approved By:

Jenny Godley, PhD, Chair, CFREB

Date:

3-Feb-2023 11:30 AM

Appendix D: Interview Consent Form



Name of Researcher, Faculty, Department, Telephone & Email:

Dr. Ann Barcomb
Department of Electrical and Computer Engineering
ann.barcomb@ucalgary.ca
+1 (403) 210-6631

Title of Project:

Evaluation of volunteering capabilities in open source software

This consent form, a copy of which has been given to you, is only part of the process of informed consent. If you want more details about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

The University of Calgary Conjoint Faculties Research Ethics Board has approved this research stu

Participation is completely voluntary, and will remain confidential unless you subsequently choose to disclose your participation by engaging in a later phase in the study.

Purpose of the Study

Recent scholarship has shown that episodic, or periphery, contributors can also benefit communities in a number of ways, and that, additionally, communities may need to adapt to take full advantage of this resource. In a recent study, we identified 65 practices which community managers believed might be effective in including episodic contributors and making use of their contributions. In order to examine the effect of these practices in situ, this phase of the study will evaluate the volunteer management practices currently in place in the Perl/Raku community.

What Will I Be Asked To Do?

You are asked to participate in an interview with a duration of approximately 1 hour. The interview will be recorded (audio) in order to create a transcript record for analysis.

You will be asked questions about your participation in and observations about the Perl/Raku community. Topics which will be included are your contributions, experiences with volunteers, volunteering practices, types of volunteers in the community, challenges in the community, and the desired state of volunteering.

In a subsequent phase of the study, we expect to work with members of the community to implement changes in proposed volunteering practices. You may be contacted for a later phase of the study in approximately six months if you express an interest now, or after viewing the report to The Perl Foundation. Participation in the next phase of the study is completely voluntary.

Participation in this study is completely voluntary and you may refuse to participate completely, or refuse to participate in parts of the study. You may decline to answer any and all questions. You may withdraw consent

1

within two weeks of being provided with a transcript of the interview. The transcript will be provided by email, unless you request a secure transfer. You may make corrections to the transcript up to two weeks after it has been provided to you.

What Type of Personal Information Will Be Collected?

Should you agree to participate, you will be asked about the length of time you have been involved in the Perl/Raku community, and the ways in which you contribute to the project. Anonymously attributed quotations may be included in the resulting papers and reports. It is possible that, based on your answers, you may be identifiable to other members of the community.

Recordings of the interview will be sent to a transcription service. Dr. Barcomb and students in her lab will have access to the recordings and transcripts.

There are several options for you to consider if you decide to take part in this research. You can choose all, some, or none of them. Please review each of these options and choose Yes or No:

I grant permission to be audio-taped: Yes: ☐ No: ☐

I grant permission to be video-taped: Yes: ☐ No: ☐

There are additional options for you to consider if you decide to take part in or none of them. Please review each of these options and choose Yes or No:

I grant permission for my transcript to be emailed to me: Yes: ☐ No: ☐

I wish to receive the transcript in a secure transfer: Yes: ☐ No: ☐

There are additional options for you to consider if you decide to take part in this research. You should choose one, or none of them. Please review each of these options and choose Yes or No:

I wish to remain anonymous: Yes: ☐ No: ☐

I wish to remain anonymous, but you may refer to me by an identifier such as "Interviewee 2": Yes: ☐ No: ☐

You may quote me and use my name: Yes: ☐ No: ☐

Are there Risks or Benefits if I Participate?

The topic of the interview is not expected to cause distress. The only identifiable risk of participation is of loss of anonymity.

In any publication/reports, only the aggregate analysis will be reported, although it will be illustrated through anonymous quotations from participants. There is some chance that others in the community will recognize speech patterns or phrases in quotations, and be able to identify the you. If you agree to be identified by name in the acknowledgements, the risk of loss of anonymity increases.

The expected benefits are to the Perl/Raku community, Free/Libre/Open/Source Software communities, and the scientific community.

The expected benefit to the Perl/Raku community is that an analysis of the current state of volunteering will enable the next phase of research, which involves implementing proposed changes.

The expected benefit to Free/Libre/Open/Source Software communities is that the research may be generalized to other communities.

The expected benefit to the scientific community is the validation of proposed best practices for managing episodic volunteers, through the observation of a real-life situation.

What Happens to the Information I Provide

Recordings of the interview will be sent to a transcription service. Dr. Barcomb and students in her lab will have access to the recordings and transcripts.

Transcripts will not be identified by name but by pseudonym. Transcripts and recordings will be stored on an encrypted computer disk for ten years following publication of research.

Email exchanges are stored unencrypted. A security breach could lead to the release of email contents, including transcripts.

Participants are free to withdraw up to two weeks after receiving a copy of the transcript. If you choose to withdraw, all original recordings and transcripts will be deleted. Your statement withdrawing consent will be retained in order to demonstrate research integrity in the event of any investigation into the study. It will be erased ten years following the publication of the research.

There are additional options for you to consider if you decide to take part in or none of them. Please review each of these options and choose Yes or No:

Would you like to receive a summary of the study's results?

Are you interested in being contacted about a follow-up interview, with the understanding that you can always decline the request? Yes: ____ No: ____

Do you wish to receive credit for your participation in the form of an acknowledgement in published papers and reports? Understand that receiving credit in the form of an acknowledgement may make it easier for people to identify you. Yes: ____ No: ____

Are you willing to be contacted concerning the reuse of your transcript in future research? If you agree to be contacted, you will be furnished with information about the new research project and have the opportunity to consent or decline to have your data used. Yes: ____ No: ____

Signatures

Your signature on this form indicates that 1) you understand to your satisfaction the information provided to you about your participation in this research project, and 2) you agree to participate in the research project.

In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from this research project at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name: (please print) _____

Participant's Signature: _____ Date: _____

Researcher's Name: (please print) _____

Researcher's Signature: _____ Date: _____

Questions/Concerns

If you have any further questions or want clarification regarding this research and/or your participation, please contact:

Dr. Ann Barcomb
Electrical and Software Engineering, Schulich School of Engineering
ann.barcomb@ucalgary.ca, +1 403.210.6631

If you have any concerns about the way you've been treated as a participant, please contact the Research Ethics Analyst, Research Services Office, University of Calgary at 403.220.6289 or 403.220.8640; email cfreb@ucalgary.ca. A copy of this consent form has been given to you to keep for your records and reference. The investigator has kept a copy of the consent form.

Appendix E: Survey Recruitment

Dr. Ann Barcomb (<http://barcomb.org>) of the University of Calgary is conducting research to understand episodic, or occasional, participation in the Perl/Raku community, in collaboration with The Perl Foundation. The results of her research will be provided as a TPF report and will assist the community in improving practices for managing episodic participation. Please consider assisting us by taking the survey about your participation in the Perl/Raku community.

The University of Calgary Conjoint Faculties Research Ethics Board has approved this study (REB20-2135).

Appendix F: Interview Recruitment

My name is Dr. Ann Barcomb, and I am an assistant professor at the University of Calgary. I am conducting a study on practices for managing episodic, or occasional, contributors to the Perl/Raku project, in collaboration with The Perl Foundation. I'm contacting you because you have been identified as a community manager or community leader, and are in the position of having an overview of the participation of numerous contributors.

I would like to interview about your experiences working with episodic contributors, and episodic contributions, in the Perl/Raku community. Additionally, I will ask if there is any difference between the current state and what you would like to see, in terms of episodic participation. If you agree to be interviewed, quotations from the interview may be included in reports/papers, although not attributed directly to you. You have the option of having your name included in a list of participants, if you would like to receive recognition for your effort.

The results of the interviews I am conducting will be delivered as part of a report to TPF, along with recommendations on practices (derived from previous research) which might be applied to improve the management of episodic participation. Subsequent research will involve working with interested community managers who participated in interviews, in order to implement the practices described in the proposal.

More information about the research is available in the overview of the proposed research:
[URL linking to research overview for participants.]

Please let me know if you are willing to be interviewed, and if you have any interest in the subsequent research. I really hope that this research will be relevant, and provide benefit to the Perl/Raku community, in addition to furthering understanding of practices for managing episodic participation in free/libre/open source software communities.

The University of Calgary Conjoint Faculties Research Ethics Board has approved this study (REB20-2135).

Appendix G: Codebook

Codes	Sub Code 1	Sub Code 2	Definition	When to Use	Where not to use
Community			Code relates to generic information about the community	This code is to be used whenever generic information about the community is given. It can be tools used in the community or processes or activities, etc. Use this when you feel the recursive codes are not able to cover the information	This code should not be used if any specific information about the community is given. There are recursive codes to use for specific information
Community	Getting started in the community		This code signifies ways one can get started in the community. This includes initial practices, workflows, setup, etc.	Use this code to describe ways an individual can get started in the community. Ex- Orientation documents, onboarding processes, etc.	Use this strictly only when describing the new joining of volunteers
Community	Diversity in the Community		This code signifies the diversity in the community. This includes, gender, geographic, sexual orientation, etc.	Use this code to signify the different demographical and cultural aspects of the community. Examples of Diversity factors are Gender, Race/Ethnicity, Age, Cultural Background, Level of Experience in the Professional Industry (Outside the Community), etc.	Not to be used in cases such as amount of time spent in the community as that would come under 'Tenure' and difference in the number of Contributions done.

Community	Popularity of the community		This code signifies information related to the popularity or PR of the community	Use this when describing about the popularity of the community and the projects done.	Not to be used in cases where it talks about past merits of the community or being a part of the community unless its ties to the popularity of the community.
Community	Practices to improve the community		This code signifies practices that can help improve the community	Use this code to describe techniques or practices that can help the community	Not to be used about retention or recruitment of volunteers. Other codes are available for that.
Community	Problems faced by the community		This code signifies the problems or challenges faced by the community	Use this code to describe any problems, challenges that harms the community. Use this codes in cases such as actions or practices that may damage the community. These can be internal or external to the community. An example of this can be financial constraints faced by the community. Another example is implementation of practices that may cause problems in the community. In some scenarios Issues faced in the community and this label goes hand in hand and both can be used.	Not to be used when talking about conflicts or difference of opinion between volunteers or even groups of volunteer. Can be used when it affects the community in a big way. For Minor conflicts it should not be used.

Community	Issues within the community		This code signifies any issues that might arise within the community	This code is to be used when describing issues, conflicts, etc. that may happen between volunteers or groups of volunteers within the community.	Not to be used in cases of describing problems faced by the community. There is a recursive code "Problems faced by the community" for such cases.
Community	Tenure		This code relates to the amount of time the individual has been involved in the community. Be it days, months, or years.	Use this code to describe the time spent by the individual in the community	Don't use this code to describe anything other than tenure such as age.
Community	Contribution		The different ways the interviewee has contributed to community. Describes the role of the person in the community.	Use it when it describes the contribution done by the interviewee or generic contributions done in the community. Use this when it is not possible to determine if it is by Episodic or Habitual Volunteers.	Don't use it when it describes something specific other than the individual's contribution or generic information on contributing. Recursive codes are available for such a scenario where we know if it is episodic or habitual or leading to a project's success or even category of contribution of the individual.

Community	Contribution	Contributions that lead to a project's success	This code signifies those contributions that usually lead a project's success	Use this to describe any contribution that leads to a project's success	Not to be used for anything else
Community	Contribution	Types of Contribution by Episodic Volunteers	Types of Contribution by Episodic Volunteers	This code is to be used whenever information on episodic volunteer's contribution is given	Don't use this code to describe about Habitual volunteer contributions or generic contributions within the community
Community	Contribution	Types of Contribution by Habitual Volunteers	This code signifies the types of contributions done by Habitual volunteers within the community	This code is to be used whenever information on Habitual volunteer's contribution is given	Don't use this code to describe about episodic volunteer contributions or generic contributions within the community
Community	Contribution	Category of Contribution (Non-Technical)	This code signifies the category or type of contribution that is Non-Technical done by the interviewee	Use this code contribution in instances where we describe the type of contribution done by the Interviewee in the Non-Technical Field.	Don't use this code when describing the type of contribution done by volunteers Episodic or habitual as there a recursive codes for that.
Community	Contribution	Category of Contribution (Technical)	This code signifies the category or type of contribution that is Technical done by the interviewee	Use this code contribution in instances where we describe the type of contribution done by the Interviewee in the Technical Field.	Don't use this code when describing the type of contribution done by volunteers Episodic or habitual as there a recursive codes for that.

Community	Challenges faced by being part of the community		This code signifies the challenges or problems faced by the interviewee by being a part of the community	This code is to be used whenever describing difficulties the interviewee faces in being a part of the community	This code is not to be used to describe any other factors of the community unless it relates to being a challenge of being in the community.
Volunteers			This code signifies the relationships and collaboration the individual has done with other contributors.	There are many recursive codes for specific topic, but this code can be used for any generic information about the community	Not to be used when talking about specific topics, there are specific codes available for it.
Volunteers	Issues faced by volunteers		This code signifies the issues face in the volunteers in the community	Use this label for generic information related to issues faced in the community by volunteers. An Example of this can be Documentation not being available to understand a piece of code or the community not being welcoming with enough information to get started in the community, etc. It should be strictly used when the interviewee talk about issues other volunteers faces in the community and not his/her own issues with the community.	Not to be used for specific types of issues such as issues within the community, or Problems faced by the community for which other labels already exist. Not to be used when describing the issues or challenger the Interviewee faces in the community. Use the "Challenges faced by being part of the community" label for that.

Volunteers	Interaction with volunteers		This code signifies the interaction the individual has had with volunteers	Use this label if the individual talks about experiences in a generic way that we cannot distinguish if it is about episodic or habitual volunteers. An example of this can be when he/she talks about a wide group of volunteers within the community or other experiences that constitutes a wider part of the community.	Not to be used when specific interactions for Habitual or Episodic volunteers are explained or mentioned
Volunteers	Interaction with volunteers	Interaction with Episodic Volunteers	Interaction of the individual with episodic volunteers in the community.	Use this code to when relationship or experiences between the individual and the episodic volunteers in the community are explained	Not to be used when talking about Habitual volunteers
Volunteers	Interaction with volunteers	Interaction with Habitual Volunteers	Interaction the of individual with Habitual contributors of the community	Use this code to when relationship or experiences between the individual and the Habitual volunteers in the community are explained	Not to be used when talking about episodic volunteers
Volunteers	Practices on volunteers		This code signifies the practices implemented by the individual to help create better volunteer experiences.	Use this code when describing practices or ideas implemented to help volunteers contribute better to the community.	Not to be used for anything else such as issues because of volunteers unless it leads to practices being formed to solve it.

Volunteers	Behaviors of Volunteers		This code signifies the behaviors and characteristics of volunteers.	Use this code when describing behaviors or characteristics of volunteers	Not to be used for practices used to help gain expected or better behaviors to make better contributions as there is a recursive code available for it.
Volunteers	Behaviors of Volunteers	Behaviors of Episodic volunteers	This code signifies the behaviors and characteristics of Episodic volunteers.	Use this code when describing behaviors or characteristics of Episodic volunteers. An example of Behaviors of Episodic volunteers would be "Episodic Volunteers are only interested in adding a feature that may be beneficial to them and they don't worry about the larger goal of the project"	Not to be used for practices used to help gain expected or better behaviors to make better contributions as there is a recursive code available for it. No tot be used on phrases that directly or indirectly define episodic volunteers. For example, "I contributed twice and then didn't contribute again".
Volunteers	Behaviors of Volunteers	Behaviors of Habitual volunteers	This code signifies the behaviors and characteristics of Habitual volunteers.	Use this code when describing behaviors or characteristics of Habitual volunteers.	Not to be used for practices used to help gain expected or better behaviors to make better contributions as there is a recursive code available for it.

Volunteers	Behaviors of Volunteers	Expected behaviors of volunteers	The code signifies the expected behaviors of volunteers within the community	Use this code when describing about the behaviors a volunteer should exhibit when contributing to the community	Not to be used when describing the behavior of the volunteer. There are recursive codes available for that
Volunteers	Behaviors of Volunteers	Practices to help implement the expected behaviors	This code signifies the practices that can be used to help bring out the behaviors in contributors	Use this code when describing practices or ideas implemented to help bring out the characteristics	Not used for describing the characteristics required to contribute better to the community.
Volunteers	Retention of volunteers		This code signifies information related to helping on retaining volunteers within the community	Use this code for information on retention of volunteers	Not to be used for practices that can help in doing so as a recursive code is available for it.
Volunteers	Retention of volunteers	Practices that can help retain volunteers	This code signifies the practices that be used to help retain volunteers	Use this code for practices that can help on retention of volunteers within the community	Other information on retention of volunteers
Volunteers	Skills required by Volunteers		This code signifies the skills or competency required by volunteers to be an active part of the community	Use this code in respect to skills required by volunteers to contribute	Not to be used for anything else
Volunteers	Availability of Volunteers		This code signifies the availability of volunteers for the community	Use this code in respect to availability of volunteers. This code can be used in instance such as talking about number of volunteers available for a certain project or task.	Not to be used for anything else. An example phrase is "it's good to figure out how much time they have available and how many tasks they should take on at once."

Volunteers	Availability of Volunteers	Availability based on knowledge	This code signifies the availability of volunteers for a certain knowledge or skillset.	Use this code to signify availability of volunteers based on skillset or knowledge	Not to be used for anything else
Volunteers	Habitual Volunteers		This code signifies generic information to Habitual volunteers	Use this code for generic information on Habitual volunteers	Not to be used for episodic volunteers or specific information. Use recursive codes for that
Volunteers	Habitual Volunteers	Suited Tasks for Habitual Volunteers	This code signifies the tasks that are more suited for Habitual Volunteers	Use this code when describing tasks that are suited for Habitual volunteers	Not to be used when describing for episodic volunteers
Volunteers	Episodic Volunteers		This code signifies generic information to episodic volunteers	Use this code for generic information on episodic volunteers	Not to be used for Habitual volunteers or specific information. Use recursive codes for that
Volunteers	Episodic Volunteers	Suited Task for Episodic Volunteers	This code signifies the tasks that are more suited for Episodic Volunteers	Use this code when describing tasks that are suited for episodic volunteers	Not to be used when describing for Habitual volunteers
Volunteers	Episodic Volunteers	Challenges faced because of Episodic Volunteering	This Code signifies the challenges faced by the community because of Episodic Volunteers.	This code is to be used only when describing about the difficulties the communities face when using episodic volunteers	Not to be used on habitual volunteers
Volunteers	Episodic Volunteers	Challenges faced by Episodic Volunteers	This code signifies the challenges that are faced by episodic volunteers within the community	This code has to be used when describing difficulties that are faced by the episodic volunteers in the community	Not to be used on habituais