Conseil de recherches en sciences naturelles et en génie du Canada

FORM 100						
Personal Data Form						
PART I						

Date

	PART I				2015/05/23				
Family name				Initial(s) of all given names Pe					
Rigby			Peter			PC	Vali	d 3	09207
	a faculty positi lete Appendic	ion at an eligible Can es B1 and C)	adian college						
I do not or will not hold an academic appointment at a Canadian postsecondary institution Place of employment other t							stseconda	ary	
ADDOINTME	NT AT A DC	STSECONDARY	INSTITUTION	Institution	(give addres	s in Appendix A	.)		
Title of position		DOTOLCONDANT	INSTITUTION						
Assistant 1	Professor			Tenured or tenure-track Yes X No academic appointment					
Department Computer S	Science and	d Software Eng	ineering	Part-time appointment Full-time appointment X					
Campus				For all non-tenured or non tenure-track academic appointment and Emeritus Professors, complete Appendices B & C					
Canadian posts Concordia	secondary inst	itution		For life-time Emeritus Professor and part-time positions, complete Appendix C				complete	
ACADEMIC	BACKGROU	IND							
Degree	Name (of discipline	Institution			Country			Date yyyy/mm
Bachelor's	Software Engineering with Management Option		Ottawa			CANADA		2004 / 05	
Doctorate	Computer Science		Victoria			CANADA		2011 / 4	
TRAINING O	F HIGHLY C	QUALIFIED PERSO	ONNEL		,				'
Indicate the nu	mber of studer	nts, fellows and other	research personnel that	you:			_		
		С	urrently		Over the past six years (excluding the current year)		
		Supervised	Co-supervised	Supe	rvised	Co-superv	rised	7	Total
Undergraduate		2			3	1	6		
Master's		2			4		(6
Doctoral		1					1		1
Postdoctoral					1				1
Others									
Total		5			8	1			14

Personal identification no. (PIN)

Valid 309207

Family name

Rigby

ACADEMIC, RESEARCH AND INDUST	TRIAL EXPERIENCE (use one additional p	age if necessary)	
Position held (begin with current)	Organization	Department	Period (yyyy/mm to yyyy/mm)
Assistant Professor	Concordia	Computer Science and Software Engineering	2012/08
Researcher	University of Victoria	Computer Science	2011/09 to 2011/10
Instructor of SENG 310 Human Computer Interaction	University of Victoria	Computer Science	2011/05 to 2011/08
Postdoctoral researcher	McGill	Computer Science	2011/11 to 2012/10
Instructor of SENG 371 Software Evolution	University of Victoria	Computer Science	2010/01 to 2010/04
Visiting Researcher	University of California at Davis	Computer Science	2008/09 to 2009/02
PhD Student	University of Victoria	Computer Science	2006/05 to 2011/04

Personal identification no. (PIN) Family name

Valid 309207

Rigby

DECEMBON CURRORT							
Family name and initial(s) of applicant	Title of proposal, funding source and program, and time commitment (hours/month)	Amount per year	Years of tenure (yyyy)				
List all sources of support (including NSERC grants and university start-up funds) held as an applicant or a co-applicant: a) support held in the past four (4) years but now completed; b) support currently held, and c) support applied for. For group grants, indicate the percentage of the funding directly applicable to your research. Use additional pages as required.							
a) Support held in the past 4 years	ears						
Peter C Rigby	Empirical Examinations of Industrial Peer Review and Distributed Version Control Work Practices Concordia University Startup grant 25 hours/month	25,000 25,000	2012 2013				
b) Support currently held							
Peter C Rigby	Contemporary Software Peer Review: Modern practices, fault prediction, and extraction of design decisions NSERC Discovery Grant (deferred for 2015) 0 hours/month	25,000 25,000 25,000 25,000 25,000	2013 2014 2016 2017 2018				
Peter C Rigby	The Impact of Disruptive Events on Software Systems NSERC, DRDC, KDM Analytics DND/NSERC Research Partnership Grants 35 hours/month	63,696 63,696 63,696	2013 2014 2015				
Peter C Rigby	Statistical Machine Translation and Summarization of Software Library Identifier Names from English to French FRQNT Établissement de nouveaux chercheurs universitaires 20 hours/month	20,320 20,320	2015 2016				
	+	+					

Highly Qualified Personnel (HQP)

Provide personal data about the HQP that you currently, or over the past six years, have supervised or co-supervised.

			Personal identification no. (PIN)	Family name
			Valid 309207	Rigby
Name	Type of HQP Training and Status	Years Supervised or Co-supervised	Title of Project or Thesis	Present Position
Andrey Krutauz	Master's (In Progress)	Supervised 2015 -	Test prioritization on Chrome	Masters student
Mark Karanfil	Res. Associate (In Progress)	Supervised 2015 -	Simple translations of English code	to USRA
Louis Querel	Master's (In Progress)	Supervised 2014 -	Using statistical bug prediction improve static analysis	to Masters student
Yue Cai Zhu	Res. Associate (In Progress)	Supervised 2014 -	Comparing test and code review effectiveness	w USRA and continues as an RA
Tajmilur Rahman	Doctoral (In Progress)	Supervised 2013 -	Release engineering and live feature toggles	PhD student
Amish Gala	Master's (Completed)	Supervised 2014 - 2015	Disruptions to developers cause by blocker bugs	ed Quality assurance at Nuance
Matthias Martineau	Res. Associate (Completed)	Supervised 2014 - 2015	Developer problem solving vs problem fixing	Software developer
Latifa Guerrouj	Postdoctoral (Completed)	Supervised 2013 - 2015	Summarizing code elements an prediction API changes	d Researcher
Samuel Donadelli	Master's (Completed)	Supervised 2013 - 2015	The impact of knowledge loss of software projects	on IBM EMM Consultant at Client Spectrum
Shams Azad	Master's (Completed)	Supervised 2013 - 2015	Predicting API changes and lic violations on Apps	ense Big data analyst for product recommendation at SAP
David Bourque	Res. Associate (Completed)	Supervised 2014 - 2014	Summarizing code elements in English	Masters student
Tavish Armstrong	Res. Associate (Completed)	Supervised 2013 - 2014	Mining and modeling review a bug data on Chrome	nd Data miner at Stripe in Bay Area
Mukadam, Murtuza	Master's (Completed)	Supervised 2012 - 2014	git-reviewed: a tool and user st	udy Software tool developer at Ericsson Ottawa
Bhageerutty, Dwijesh	Res. Associate (Completed)	Co-supervised 2012 - 2012	Ultimate Documentation Datab	pase Unknown
Form 100 (2009 W	 nage 4 of 4 Per 	sonal information c	ollected on this form and appendices wi	Il be Version française disponible

Form 100 (2009 W), page 4 of 4

Personal information collected on this form and appendices will be stored in the Personal Information Bank for the appropriate program.

Version française disponible



PROTECTED WHEN COMPLETED

1 Most Significant Contributions to Research and Practical Applications

Peer Review on Open Source Software (OSS) Projects: Parameters, Statistical Models, and Theory

This work represents the first set of studies that systematically and comparatively quantified how peer reviews are performed in a successful, mature OSS project. The first paper to come out of this work was accepted to ICSE in 2008. Prior to this paper, there were only anecdotal descriptions and limited quantitative analyses of OSS peer review practices. The main contribution was the creation of measures from email and version control archives that mirrored the measures created over the last 35 years to assess peer review in industrial development. This paper was the starting point and first case study in my dissertation and led to a preliminary theory of peer review in OSS. I validated this theory with replications, which appeared in the Transactions on Software Engineering and Methodology (TOSEM), the ACM's premier Software Engineering journal.

The measures and statistical models above provide convincing evidence that OSS peer review is both efficient and effective. They do not, however, describe the mechanisms and behaviours that facilitate broadcast based peer review in an OSS context. To gain a richer understanding and to triangulate my previous findings, I manually examined 500 instances of peer review and interviewed 10 core developers across the same six OSS projects. The results validated my quantitative findings and was published at ICSE 2011.

Modern Peer Review in Software Firms

Modern peer review techniques have become a hot topic over the last two years, with sessions on code review occurring at SANER 2015 and MSR 2015.

On the basis of my systematic analysis of OSS peer review, the Department of National Defense (DND) contacted me to evaluate the peer review processes used by the Land Software Engineering Centre (LSEC). The LSEC provided a document describing their current practices. The review process was an inefficient mix of formal inspection techniques combined with informal shortcuts. During the meetings, I described my findings from OSS review as well as the literature on software inspection. Using their requirements, I wrote a report that developed possible lightweight alternatives that would suit the LSEC development environment. The DND is committed to working with me to implement and validate a contemporary review process using a development team at LSEC. I also have a DND, NSERC, Industry grant to investigate disruptive events at DND.

Out of the DND work, I published a feature article in IEEE Software, 2012, "Contemporary Peer Review in Action: Lessons from Open Source Development". This article proved of interest to commercial firms. For example, I was invited to give a talk to Microsoft on peer review. I also meet on a weekly basis with Dr. Bird, a researcher at Microsoft. After sending the paper to a developer at AMD, Ratcliffe, he gave me their review data to analyze. My collaboration with Dr. Bird resulted in a joint publication, "Convergent Contemporary Software Peer Review Practices" at the FSE in 2013.

Statistical Translations and Summaries of English and Code

An increasing amount of knowledge about software is shared and archived in informal documents, such as forum posts and mailing lists. Unfortunately, while the informal structure promotes more sharing and rapid development, it also makes the discussion more difficult to index and search. A particular problem is searching for discussions related to the code elements (*e.g.* a class, method, field) that make up a software system. By locating these code elements a developer can find usage patterns, bug workarounds, or alternatives elements.

I developed a novel automated code element extractor that works without a pre-defined set of known elements. Based on the Java language specification, I wrote an island parser that can identify Java code elements in freeform text and code fragments that do not compile. The list of valid code elements is extracted based on elements frequency instead of a set of pre-defined valid elements. Using a benchmark,

I attained an average precision and recall at or above 90%. This work appeared at ICSE 2013.

In work with my postdoc and a research assistant, we developed statistical language models from StackOverflow data that described the purpose of each Android API class and method. The work was published this year at ICSE's NIER. With this language model we can create English queries and return related code elements. We are collaborating with Nguyen who has developed techniques to autocomplete blocks of code based on the surrounding context. In this collaboration, we plan to transform these sets code elements into the most likely combination of actual code based on a large corpus of applications.

2 Research Contributions and Practical Applications

Articles in refereed publications

J1: Md Tajmilur Rahman, and **Peter C. Rigby**. Release stabilization on linux and chrome. *Software, IEEE* 32, 2, Mar 2015, 81–88

Work funded by DND/NSERC.

J2: **Peter C. Rigby**, Daniel M. German, Laura Cowen, and Margaret-Anne Storey. Peer Review on Open-Source Software Projects: Parameters, Statistical Models, and Theory. *ACM Trans. Softw. Eng. Methodol.* 23, 4, 2014, 35:1–35:33

Work funded by NSERC CGSD and NSERC Discovery. I did the writing and research. Cowen guided my choice of statistical techniques that allowed for a model that included all the 6 OSS projects (instead of one per project). The remaining authors advised.

J3: **Peter Rigby**, Brendan Cleary, Frederic Painchaud, Margaret-Anne Storey, and Daniel German. Contemporary Peer Review in Action: Lessons from Open Source Development. *IEEE Software* 29, 6, Nov. 2012, 56–61

Work funded by NSERC CGSD. I did the writing and distilling of ideas. Cleary provided interesting insights from industry and meetings with Painchaud and the DND provided the impetus for presenting the concepts as lessons and recommendations. The last two authors were my advisers.

Other refereed contributions

C1: <u>Latifa Guerrouj</u>, <u>David Bourque</u>, and **Peter C. Rigby**. Leveraging Informal Documentation to Summarize Classes and Methods in Context. In *37th IEEE New Ideas Emerging Research track of the International Conference on Software Engineering (to appear)*, 2015, NIER '15, IEEE Press. (Acceptance rate: 25/135 or 18%)

Work funded by DND/NSERC/Industry grant.

C2: <u>Samuel M. Donadelli</u>, <u>Yue Cai Zhou</u>, and **Peter C. Rigby**. Organizational volatility and post-release defects: A replication case study using data from Google Chrome. In *Proceedings of the 12th Working Conference on Mining Software Repositories (to appear)*, 2015

Work funded by NSERC/DND/Industry grant.

B1: **Peter C. Rigby**, Alberto Bacchelli, Georgios Gousios, and <u>Murtuza Mukadam</u>. A Mixed Methods Approach to Mining Code Review Data: Examples and a replication study of multi-commit reviews. In *Analyzing Software Data*, C. Bird, T. Menzies, and T. Zimmermann, Eds. Morgan Kaufmann, 2015. To appear

Work funded by NSERC Discovery

- C3: <u>Latifa Guerrouj</u>, <u>Shams Azad</u>, and **Peter C. Rigby**. The Influence of App Churn on App Success and StackOverflow Discussions. In *22nd IEEE International Conference on Software Analysis, Evolution, and Reengineering*, 2015, SANER '15, IEEE Press. (Acceptance rate: 46/144 or 32%) Work funded by NSERC/DND/Industry grant.
- C4: Md Tajmilur Rahman, and **Peter C. Rigby**. Contrasting Development and Release Stabilization Work on the Linux Kernel. In *International Workshop on Release Engineering*, 2014 Work funded by NSERC/DND/Industry grant.
- C5: **Peter C. Rigby**, and Christian Bird. Convergent Contemporary Software Peer Review Practices. In *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering*, 2013, ESEC/FSE 2013, ACM, pp. 202–212. (Acceptance rate: 51/251 or 20%)

Work funded by a Concordia University startup grant and NSERC Discovery

C6: **Peter C. Rigby**, and Martin P. Robillard. Discovering Essential Code Elements in Informal Documentation. In *Proceedings of the 2013 International Conference on Software Engineering*, 2013, ICSE '13, IEEE Press, pp. 832–841. (Acceptance rate: 85/461 or 18.5%)

I did the writing and research. Robillard advised and framed the work as it is part of his larger research program.

- C7: **Peter C. Rigby**, Earl T Barr, Christian Bird, Prem Devanbu, and Daniel M German. What Effect does Distributed Version Control have on OSS Project Organization? In *International Workshop on Release Engineering*, 2013
- C8: <u>Murtuza Mukadam</u>, Christian Bird, and **Peter C. Rigby**. Gerrit software code review data from Android. In *Proceedings of the 10th Working Conference on Mining Software Repositories*, 2013, MSR '13, IEEE Press, pp. 45–48

Work funded by my Concordia University startup grant.

C9: Earl T. Barr, Christian Bird, **Peter C. Rigby**, Abram Hindle, Daniel M. German, and Premkumar Devanbu. Cohesive and isolated development with branches. In *Proceedings of the 15th International Conference on Fundamental Approaches to Software Engineering (FASE'12)*, 2012

I wrote the simulation, measures, and first draft of this paper. Pushing to finish my dissertation meant that I did little of the writing for the final version.

C10: **Peter C. Rigby**, and Margaret-Anne Storey. Understanding Broadcast Based Peer Review on Open Source Software Projects. In *Proceeding of the 33rd International Conference on Software Engineering*, 2011, ICSE '11, ACM, pp. 541–550. (Acceptance rate: 62/441 or 14%)

Work funded by CGSD. I did the writing and research. Storey advised and helped in the selection of an appropriate methodology.

C11: Christian Bird, **Peter C. Rigby**, Earl T. Barr, David J. Hamilton, Daniel M. German, and Prem Devanbu. The Promises and Perils of Mining git. In *MSR '09: Proceedings of the 2009 6th IEEE International Working Conference on Mining Software Repositories*, 2009, IEEE Computer Society, p. 10

I developed the original mining infrastructure. Bird and I discussed this infrastructure and he did the majority of the writing of this paper with Barr.

C12: **Peter C. Rigby**, Daniel M. German, and Margaret-Anne Storey. Open Source Software Peer Review Practices: A Case Study of the Apache Server. In *ICSE '08: Proceedings of the 30th International Conference on Software engineering*, 2008, ACM, pp. 541–550. (Acceptance rate: 56/371 or 15%)

Work funded by CGSD. I did the writing and research. German pushed me to develop appropriate measures. Storey advised.

NOTE: The International Conference on Software Engineering (ICSE) has a higher h-index than the top journal in the area, IEEE Transactions on Software Engineering (TSE). I have three papers at ICSE. I also have one paper at the Foundations of Software Engineering, which is ranked more highly than the top Empirical Software Engineering Journal (ESE) (http://scholar.google.ca/citations?view_op=top_venues&hl=en&vq=eng_softwaresystems).

3 Other Evidence of Impact and Contributions

In a keynote talk at Empirical Software Engineering and Measurement (ESEM), Bertrand Meyer held up my work on software peer reviews as an illustration of the style of impactful empirical research he would like to see more of in the future. Both SANER and MSR 2015 had a session on peer review, recognizing the importance of a topic I have been studying for 9 years.

Invited Talks and Presentations:

I have given invited talks on "Transferring Open Source Review Practices to Industrial Projects" to DND and Avaya and on "Broadcast Based Review" at Microsoft research. I gave an invited talk to the Consortium for Software Engineering Research (CSER) on current research program. I was on a panel at the 2nd Workshop on User evaluations for Software Engineering Researchers, which is an ICSE workshop.

I have also presented my work at numerous conferences and workshops, including ICSE.

Research Fellowships and Awards:

In 2014, I won the Concordia University's Emerging Researcher for cluster of 'Technology, Industry and the Environment." The award came with \$5 000 research stipend.

I was awarded an the top NSERC award during my PhD: an NSERC CGSD3.

Service

I have reviewed for IEEE Transactions on Software Engineering, Transactions on Software Engineering and Methodology, Journal of Software Systems, Journal of Information and Software Technology, the Journal of Empirical Software Engineering, and IEEE Software.

I have co-reviewed submissions with my advisers for a number of conferences, including the International Conference on Software Engineering (ICSE), and Foundations of Software Engineering (FSE).

The program committees I have been on include the International Working Conference on Mining Software Repositories (MSR), the MSR Mining Challenge Track International Workshop on Emerging Trends in FLOSS Research and Development, the International Workshop on Empirical Software Engineering in Practice, the international Workshop on Release Engineering.

I have co-organized the Doctorial Symposium at SANER 2015. I co-organized CSER summer 2013.

4 Contributions to the Training of HQP

I have trained 14 HQPs in total. I teach them a software analytics pipeline, which has the following stages. First, we develop research questions, which involves a literature review and preliminary qualitative analysis. Second, we scrape raw data. Third, we extract a minimal set of interesting attributes into a database. Third, we link and process the data in the database. Fourth, we summarize the data and output it to data mining or statistical modelling tools, such as R. Fifth, we interpret the results and models. Sixth, often we return to the raw data to extract further attributes and triangulate our finding with qualitative evidence. This training has helped my students get interesting data mining jobs, which I describe below.

Mukadam graduated in April 2014 and now works at Ericsson Ottawa. His masters thesis is entitled, "git-reviewed: A Distributed Peer Review Tool & User Study." In this work, we developed a tool that allowed peer reviews to be integrated into git in a distributed manner. This contrasts with all other peer review tools, that require a central repository for review discussion. Mukadam's knowledge of git infrastructure made him an ideal fit for Ericsson Ottawa's Development Environment team. Griffiths's was so impressed by Mukadam's training that he contacted me about a possible collaboration, which became the impetus for this Engage proposal.

As a USRA in 2013, I taught Armstrong how to mine and clean data. He proved to be a very proficient student who developed a mining infrastructure for Chrome peer reviews, bug reports, and vulnerabilities. This infrastructure has been helpful for other students in the lab including Zhu and Donadelli. Armstrong works for Stripe in the San Francisco Bay Area, where he mines software engineering data to monitor project progress and increase developer efficiency.

Guerrouj worked for me as a postdoctoral research from November 2013 to February 2015. She helped co-supervise Bourque and Azad. Guerrouj and I shared ideas and I helped her to develop as a writer as well as how to plan and structure her work. Unfortunately, health issues required Guerrouj to take a leave of absence from her position.

Azad graduated with a masters in April 2015. His thesis is entitled, "Empirical Studies of Android APIs: Prediction of API Changes and License Violation among Applications." My postdoc Guerrouj cosupervised Azad. We mined a community of Android Apps to predict the changes that developers would make to Apps. We also mined StackOverflow for combinations of API calls. Azad worked diligently to mine the ASTs from hundreds of Apps. After some initial difficulties and further training, Azad became proficient in mining data and storing it in a database. Azad now works for SAP as a big data analyst on the Product Recommendation team.

Bourque started as a USRA in May 2014 and finished as a research assistant in December 2014. He was co-supervised by my postdoc Guerrouj. I hired Bourque because he was doing a double major in Math and Computer Science. He built language models to summarize the purpose and use of Android Library classes and methods in English.

Karanfil started as a USRA in May 2015. He is building on Bourque's work as we are trying to develop language models that will allow us to perform basic translations from English to Code. This work is in collaboration with Devanbu and Nguyen.

Donadelli graduated with a masters in April 2015. His thesis is entitled, "The impact of knowledge loss on software projects: turnover, customer found defects, and dormant files." In this work, we studied the impact that leaving developers have on a project. We created statistical models and estimated the "bus-factor". Donadelli works as an IBM management consultant for Client Spectrum.

Zhu started as a USRA in May 2014 and continues to work for me as a research assistant. He has a degree in Economics and is pursuing a second undergraduate degree in Computer Science. I trained Zhu in mining software repositories and he has been able to use some of his knowledge of statistical models to help Donadelli in studies of knowledge loss on software projects. Zhu also works on extracting test information from the Chrome project.

Querel started working with me in September 2014. His masters thesis involves using statistical bug models to prioritize and reduce the number of static analysis warnings presented to developers. This work is done in collaboration with DND and KDM Analytics.

Rahman started working with me in September 2013. His PhD thesis is examining release engineering. So far we have walked the git DAG to differentiate between development work and stabilization work on Chrome and Linux. He is currently examining the use of feature flags to isolate features without using branches. This involves interviewing practitioners to understand the process and to extract flags from source code. We are also apply his release engineering insights to DND projects.

APPENDIX A Personal Data (Form 100)



Date

Complete this appendix (i) if you are an applicant or co-applicant applying for the first time; (ii) if you need to update information submitted with a previous application; or (iii) if you do not hold an appointment at a Canadian postsecondary institution. For updates, include only the revised information in addition to the date, your name and your PIN.

This information will be used by NSERC p	rimarily to contact applicants and	award holders. It may also be	Date			
used to identify prospective reviewers and seen or used in the adjudication process.	committee members, and to gen	erate statistics. It will not be	201:	5/05/23		
Family name	Given name	Initial(s) of all given nan	nes Personal iden	Personal identification no. (PIN		
Rigby	Peter	PC	Valid	309207		
Position and complete mailing address if postsecondary institution or if your currer		nt is not a Canadian	If address is indicate:	temporary,		
1515 St. Catherine W						
EV 3.118						
Montreal QC H3G2W1						
CANADA						
			Starting date			
			Leaving date			
			Loaving date			
Telephone number	Facsimile number	E-mail address				
1 (514) 8482424 3167		peter.rigby@concord	lia.ca			
Telephone number (alternate)	Give an alternate telep	Give an alternate telephone number only if you can				
		nber during business hours.	X Male	Femal		
LANGUAGE CAPABILITY			•			
English Read X	Write	X	Speak X	peak X		
French Read	Write	Write Sp				
	L		· 🗀			
I wish to receive my correspondence:	in English	X in	French			
AREA(S) OF EXPERTISE						
Provide a maximum of 10 key words that to separate them. If you have expertise which one(s).			esearch subject co	de(s)		
Empirical Software Engineering	rimary					
Studies, Professional/Industrial Developers, Statistical Machine						
Translations, Test Prioritization, Mining Software Repositories, Release						
Engineering, Software Measures and Analytics, Statistical Models, Software Peer Review and Inspection				ondary		
				2705		

Form 100, Appendix A (2009 W)

PROTECTED WHEN COMPLETED

Version française disponible

