Natural Sciences and Engineering Research Council of Canada | Conseil de recherches en sciences naturelles et en génie du Canada

FORM 100
Personal Data Form
PART I

Date: 2018/09/17

Family name	Given name	Initial(s) of all given names	Personal identification no. (PIN)
Seltzer	Margo	MI	542742

☐ I hold a faculty position at an eligible Canadian college (complete Appendices B1 and C)

☐ I do not or will not hold an academic appointment at a Canadian postsecondary institution

Place of employment other than a Canadian postsecondary Institution (give address in Appendix A)

APPOINTMENT AT A POSTSECONDARY INSTITUTION

Title of position: Professor

Department: Computer Science

Campus:

Canadian postsecondary institution: British Columbia

Tenured or tenure-track academic appointment — Yes [X] No [ ]

Part-time appointment [ ] Full-time appointment [X]

- For all non-tenured or non tenure-track academic appointment and Emeritus Professors, complete Appendices B & C
- For life-time Emeritus Professor and part-time positions, complete Appendix C

ACADEMIC BACKGROUND

Degree	Name of discipline	Institution	Country	Date yyyy/mm
Bachelor's	Applied Mathematics	Harvard University	UNITED STATES	1983 / 6
Doctorate	Computer Science	University of California - Berkeley	UNITED STATES	1992 / 12

TRAINING OF HIGHLY QUALIFIED PERSONNEL

Indicate the number of students, fellows and other research personnel that you:

	Currently		Over the past six years (excluding the current year)		
	Supervised	Co-supervised	Supervised	Co-supervised	Total
Undergraduate			10	7	17
Master's					
Doctoral		4	6	2	12
Postdoctoral		2	2		4
Others	1		2		3
Total	1	6	20	9	36

ACADEMIC, RESEARCH AND INDUSTRIAL EXPERIENCE (use one additional page if necessary)

Position held (begin with current)	Organization	Department	Period (yyyy/mm to yyyy/mm)
Professor	British Columbia	Computer Science	2018/09
Architect	Oracle Corporation	Oracle Labs	2012/01
Architect	Oracle Corporation	Embedded Systems	2006/03 to 2011/12
Harvard College Professor	Harvard University	Computer Science	2005/07 to 2010/06
Associate Dean of Computer Science and Engineering	Harvard University	Div. of Engineering and Applied Sciences	2002/07 to 2006/07
Professor	Harvard University	Computer Science	2000/07 to 2018/08
Associate Professor	Harvard University	Computer Science	2000/06 to 1997/07
Assistant Professor	Harvard University	Computer Science	1993/01 to 1997/06

Personal identification no. (PIN)	Family name
542742	Seltzer

RESEARCH SUPPORT

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.

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)
a) Support held in the past 4 years			
Margo Seltzer and two others	CSR: Medium: Collaborative Research: Workload-Aware Storage Architectures for Optimal Performance and Energy Efficiency National science foundation (US) Computer Systems Research 10 hours/month	124,375(100%) 142,478(100%) 65,023(100%) 65,023(100%)	2014 2015 2016 2017
Margo Seltzer and two others	Citation++: Data citation, provenance, and documentation National Science Foundation (US) 12 hours/month	137,988(100%) 126,025(100%) 126,025(100%)	2015 2016 2017
Margo Seltzer	New Approaches for Ranking in Machine Learnin Duke University (subcontract) IIS Career 8 hours/month	64,697(100%)	2018
Margo Seltzer with one other	Towards a FAIR Digital Ecosystem in the Cloud National Institutes of Health (NIH) 1 hours/month	841,387 (0%)	2018

Personal identification no. (PIN)	Family name
542742	Seltzer

RESEARCH SUPPORT

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

Margo Seltzer	Computer Systems Laboratory	70,000(100%)	2019
	Startup	70,000(100%)	2020
	Startup	70,000(100%)	2021
	160 hours/month	70,000(100%)	2022
		70,000(100%)	2023

b) Support currently held

Margo Seltzer and two others	XPS: FULL: CCA: Collaborative Research:	149,500(100%)	2015
	Automatically Scalable Computation	133,926 (90%)	2016
	National Science Foundation (US)	294,151 (50%)	2017
	Exploiting Parallelism and Scalability	127,211 (20%)	2018
	4 hours/month	127,211 (20%)	2019
Margo Seltzer and one other	PRINCESS : Probabilistic Representation of	248,586 (75%)	2016
	Intent Commitments to Ensure Software Survival	323,960 (75%)	2017
	Defense Advanced Research Projects Agency	537,986 (75%)	2018
	(DARPA) (USA)	553,667 (75%)	2019
	BRASS		
	8 hours/month		
Margo Seltzer and three others	CISE-Provenance : SI2-SSI: Collaborative	560,473(100%)	2016
	Research: Bringing End-to-End Provenance to	633,064(100%)	2017
	Scientists	328,004(100%)	2018
	National Science Foundation (US)	328,004(100%)	2019
	Sustainable Software Infrastructure		
	16 hours/month		

PROTECTED WHEN COMPLETED

Personal identification no. (PIN)	Family name
542742	Seltzer

RESEARCH SUPPORT

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.			
b) Support currently held			
Margo Seltzer	Computer Systems	1,000,000(100%)	2019
	Canada 150 Research Chair Program	1,000,000(100%)	2020
	Canada 150	1,000,000(100%)	2021
	120 hours/month	1,000,000(100%)	2022
		1,000,000(100%)	2023

PROTECTED WHEN COMPLETED

Canada

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
542742	Seltzer

Name	Type of HQP Training and Status	Years Supervised or Co-supervised	Title of Project or Thesis	Present Position
(Name withheld)	Res. Associate	Supervised 2017 -	Detecting and displaying controversy in wikipedia articles	Research associate
(Name withheld)	Postdoctoral	Co-supervised 2017 -	Optimal linear models	Post doctoral fellow
(Name withheld)	Doctoral (In Progress)	Co-supervised 2016 -	Provenance-based intrusion detection	Intern, University of Cambridge
(Name withheld)	Doctoral (In Progress)	Co-supervised 2016 -	Provenance-based OS synthesis	PhD student
(Name withheld)	Postdoctoral	Co-supervised 2016 -	Domain specific languages for OS synthesis	Post-doctoral researcher
(Name withheld)	Doctoral (In Progress)	Co-supervised 2015 -	Synthesizing assemblers and linkers	PhD Student
(Name withheld)	Doctoral (In Progress)	Co-supervised 2013 -	Automatic program repair	PhD Student
(Name withheld)	Undergraduate (Completed)	Co-supervised 2017 - 2018	Coding BetteR: Assessing and Improving Reproducibility	Engineer, Uber
(Name withheld)	Res. Associate	Supervised 2017 - 2018	OS Code synthesis	PhD Student, NYU
(Name withheld)	Doctoral (In Progress)	Supervised 2017 - 2018	OS Code synthesis	PhD Student, Harvard
(Name withheld)	Undergraduate (Completed)	Co-supervised 2017 - 2018	Generating Tests to Verify Machine Descriptions	Engineer, Google
(Name withheld)	Res. Associate (Completed)	Supervised 2016 - 2018	Incorporating provenance in dataverse	Engineer, Wave
(Name withheld)	Undergraduate (Completed)	Co-supervised 2016 - 2018	Simulation of open quantum systems	PhD student, Stanford
(Name withheld)	Doctoral (Not Completed)	Supervised 2011 - 2018	Automatically Scalable Computation	On leave
(Name withheld)	Undergraduate (Not Completed)	Supervised 2017 - 2017	OS Code Synthesis	deceased
(Name withheld)	Doctoral (In Progress)	Supervised 2016 - 2017	Certifiably optimal rule lists	PhD student (changed areas)
(Name withheld)	Undergraduate (Completed)	Supervised 2016 - 2017	Bias in Wikipedia articles of computer scientists	Engineer, Facebook
(Name withheld)	Undergraduate (Completed)	Supervised 2016 - 2017	Unsandboxing	Engineer, Microsoft
(Name withheld)	Undergraduate (Completed)	Co-supervised 2016 - 2017	Gender and Computer science at Harvard	Product Manager, Google
(Name withheld)	Undergraduate (Completed)	Co-supervised 2016 - 2017	Creating an Accessible Network of Life-Saving Devices	Engineer

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
542742	Seltzer

Name	Type of HQP Training and Status	Years Supervised or Co-supervised	Title of Project or Thesis	Present Position
(Name withheld)	Undergraduate (Completed)	Supervised 2016 - 2017	More Automatically Scalable Computation	PhD Student, Stanford
(Name withheld)	Undergraduate (Completed)	Supervised 2016 - 2017	A Case for Discrete Optimization in the 21st Century	Engineer, Benevolent AI
(Name withheld)	Postdoctoral	Supervised 2016 - 2017	Whole system provenance	Lecturer, University of Bristol
(Name withheld)	Undergraduate (In Progress)	Supervised 2016 - 2017	Incorporating provenance into Dataverse	Undergraduate student (senior)
(Name withheld)	Doctoral (Completed)	Supervised 2008 - 2017	Sorting shapes the performance of graph-structured systems	Senior engineer, Google
(Name withheld)	Undergraduate (Completed)	Co-supervised 2015 - 2016	Community Attribute Models via Dynamic Graph Evolution	Quantitative trader, Jane Street Capital
(Name withheld)	Undergraduate (Completed)	Supervised 2015 - 2016	Using Gaussian Processes to find predictable points in progs	Software engineer, Facebook
(Name withheld)	Undergraduate (Completed)	Supervised 2013 - 2015	Data structures for fast lookup on large, sparse bit vectors	Software engineer, Microsoft
(Name withheld)	Undergraduate (Completed)	Supervised 2013 - 2015	Data structures for fast lookup on large, sparse bit vectors	PhD Student, Harvard
(Name withheld)	Doctoral (Completed)	Supervised 2008 - 2015	LLAMA: A Persistent, Mutable Representation for Graphs	Member of the technical staff, NetApp
(Name withheld)	Undergraduate (Completed)	Supervised 2013 - 2014	Evaluating the efficacy of a flipped classroom for OS course	Software engineer, Facebook
(Name withheld)	Doctoral (Completed)	Co-supervised 2011 - 2014	Accelerating Markov chain Monte Carlo via parallel predictiv	Independent
(Name withheld)	Doctoral (Completed)	Supervised 2004 - 2014	Limiting Disclosure in Annotated Graphs	CEO, Sybil Security
(Name withheld)	Undergraduate (Completed)	Co-supervised 2013 - 2013	Evaluating the efficacy of style evaluation in a Mooc	Software engineer, Asana
(Name withheld)	Postdoctoral	Supervised 2010 - 2012	Hypervisor based provenance collection	Senior Performance Engineer, Mark Logic
(Name withheld)	Doctoral (Completed)	Co-supervised 2010 - 2012	Coordinated Resource Management in Network	Assistant Professor, Vassar

Margo Seltzer

1. Most Significant Contributions to Research and/or Practical Applications

Data Provenance is the formal record of how a data item came to be in its current state. It includes the identification of input data sources and the transformations of those sources into an artifact. It is widely accepted that data provenance lies at the heart of reproducibility.

Seltzer is a pioneer in this area. Her three most significant contributions in this area are: 1) the categorization and definition of the two modes of provenance collection: observation and disclosure [1], 2) the concept and implementation of whole-system provenance [2], and 3) the concept and implementation of semantic integration of provenance via layering [3]. First generation research in data provenance focused on mechanisms for capturing provenance, however, adoption requires applications that use provenance. Seltzer's recent work in this area focuses on applications of provenance: intrusion detection [6,16], other security applications, such as data loss prevention and attack attribution [5], computational reproducibility [7], and regulatory compliance [12].

Automatically Scalable Computation: Since the creation of the first multiprocessor computer system, a grand challenge has been the quest to "magically" transform a single-threaded program into a parallel one. The Automatically Scalable Computation (ASC) [27] project is a collaboration between Harvard and Boston University to achieve this vision, by transforming the problem of parallelization into a problem of prediction and speculation.

We assume a computational model based upon dynamical systems. The registers and memory of a conventional single-threaded processor comprise a (very large) state space. Instruction corresponds to a step in the state space, and program execution out a trajectory in that space. We find points in that trajectory from which we can build models that will accurately predict future states that the program will encounter. We then use excess cores to speculatively execute from those predicted states. If our predictions were accurate, the speculative executions can be used when we later encounter the predicted state from which they started, by fast forwarding to the end state of the speculation thereby achieving speedup.

Seltzer's group has produced two systems that embody this approach. The first realization of this architecture used a software virtual machine that emulated a subset of the Intel x86-32 architecture [31]. It demonstrated linear scalability up to hundreds of cores on a small collection of computational kernels. The second realization runs natively on Intel hardware, producing real-time speedup on a 44-core machine. On the kernels that it can effectively predict (the vast majority of the Polybench kernels), its performance is limited only by the overhead of the tools that monitor program execution. The team has proposals for new hardware to make monitoring more efficient.

This project spawned work in a number of disciplines. Early in the project's history, one of Seltzer's students (Elaine Angelino) used the idea of computational reuse and speculation for predictive prefetching for Markov Chain Monte Carlo simulation [24]. Using similar techniques, Seltzer's student, Amos Waterland, teamed up with colleagues in Applied Physics to apply ASC-like techniques to long timescale simulation of the atomistic evolution of materials, producing a paper in the Journal of Chemical Theory and Computation [4]. At Boston University, theorist Steve Homer has been developing a framework that defines the program properties that make an implementation amenable to ASC parallelization.

Efficient and Interpretable Models: In joint work with Cynthia Rudin, Seltzer has been developing high-performance implementations for interpretable models, e.g., rule lists, decision trees. In a series of conference and longer papers on constructing rule lists, Seltzer and her team have demonstrated how combining systems' techniques with strong mathematical bounds allows for solution of previously intractable problems.

Building on the idea of computational reuse, introduced in the ASC project, Seltzer developed a way to reuse computation in the Bayesian Rule List setting. Bayesian Rule Lists use Markov Chain Monte Carlo. However, any two consecutive rule list evaluations are quite similar. Storing incremental results for these evaluations produces an order of magnitude speedup. Tighter mathematical bounds on the objective function also improve performance significantly. To support these computations, Seltzer developed a high-performance library for rule list evaluations; this resulted in another order of magnitude performance improvement. This work, Scalable Rule Lists, won a best student paper award at the 2016 AISTATS conference and was then published in the 2017 ICML conference [14].

Building on this combination of computational reuse and clever data structures, Seltzer and Rudin moved on to the challenge of constructing a rule list for which they could produce a certificate of optimality. CORELS (Certifiably Optimal RulE ListS), uses a branch and bound algorithm to produce on optimal rule list with respect to a regularized objective function [10,15]. Once again, mathematical bounds played a significant role in reducing the search space, and computational reuse makes the implementation more efficient. Additionally, CORELS uses a collection of clever and efficient data structures to manage the algorithm. One of the more important such additions, conceived of by Seltzer, is the symmetry aware map. The key insight is that all orderings of a particular set of rules will *capture* the same samples from the training data. (A rule captures a sample if it is the first rule in the list for which the sample evaluates to true. The set of all samples captured by a sequence of rules is said to be captured by that sequence.) Thus, only the best performing permutation of a set of rules needs to be kept; all other permutations can be pruned. This again greatly reduces the search space.

The code for both projects is available under an open source license. SBRL has already been used by hundreds of researchers; CORELS was released in August 2017.

Seltzer's references prior to 2013 appear here as do references to others' work; Seltzer's post-2013 publications appear in the following section.

1. Braun, U., Garfinkel, S., Holland, D., Muniswamy-Reddy, K., Seltzer, M., "Issues in Automatic Provenance Collection," *Proceedings of the 2006 International Provenance and Annotation Workshop (IPAW)*, Chicago, IL, May 2006.
2. Muniswamy-Reddy, K., Holland, D., Braun, U., Seltzer, M., "Provenance-Aware Storage Systems," *Proceedings of the 2006 USENIX Annual Technical Conference*, Boston, MA, June 2006.
3. Muniswamy-Reddy, K., Braun, U., Holland, D., Macko, P., Maclean, D., Margo, D., Seltzer, M., Smogor, R., "Layering in Provenance Systems," *Proceedings of the 2009 USENIX Annual Technical Conference*, San Diego, CA, June 2009.
4. Perez, D., Cubuk, E., Waterland, A., Kaxiras, E. Voter, A., "Long-time dynamics through parallel trajectory splicing," *Journal of Chemical Theory and Computation (JCTC)* (2016).

2. Research Contributions and Practical Applications - Since January 1, 2013

Note: Unless otherwise specified, Seltzer was lead advisor on the projects associated with each publication. In these cases, Seltzer met with students weekly to guide their research, but they were responsible for conducting the work. Papers were then written collaboratively, typically with Seltzer and the lead student splitting the work approximately in half with minor contributions from other students.

Published Refereed Contributions

5. Pasquier, T., **Han, X**., Moyer, T., Bates, A., Hermant, O., Eyers, D., Bacon, J., Seltzer, M., "Runtime Analysis of Whole-System Provenance," *Proceedings of the 2018 Conference on Computer and Communications Security*, (CCS '18) Toronto CA, October 2018. (16 pages).
6. **Han, X**., Pasquier, T., Seltzer, M., "Provenance-based Intrustion Detection: Opportunities and Challenges," Proceedings of the Workshop on the Theory and Practice of Provenance, (Tapp 2018), London UK, July 2018. (4 pages).
7. Pasquier, T., Lau, M., **Han, X**., Fong, E., Lerner, B., Boose, E., Crosas, M., Ellison, A., Seltzer, M., "Sharing and Preserving Computational Analyses for Posterity with encapsulator," IEEE Computing in Science and Engineering 20(111), July 2018. doi: 10.1109/MCSE.2018.042781334 [Seltzer collaborated with Pasquier and Lau on development and edited paper; contribution 15%]
8. **Huang, Y**., **Pavlovic, M**., Marathe, V., Seltzer, M., Harris, T., Byan, S., "Closing the Performance Gap Between Volatile and Persistent Key-Value Stores Using Cross-Referencing Logs," Proceedings of the 2018 USENIX Annual Technical Conference, Boston MA, June 2018. (13 pages) [Seltzer co-advised Huang, helped develop the evaluation, and edited the paper; contribution 10%]
9. Rao, M., Bacon, D.F., Parkes, D., and Seltzer, M., "Incentivizing Deep Fixes in Software Economies," *IEEE Transactions on Software Engineering*. (June 2018, 21 pages) doi: 10.1109/TSE.2018.2842188 [Seltzer guided development of model and edited paper; contribution 10%]
10. Angelino, E., **Larus-Stone, N**., **Alabi, D**., Seltzer, M., Rudin, C., "Learning Certifiably Optimal Rule Lists for Categorical Data," *Journal of Machine Learning Research*, 18(234):1—78, 2018.
11. Pasquier, T., Lau, M., Trisovic, A., Boose, E., Couturier, B., Crosas, M., Ellison, A., Gibson, V., Jones, C., Seltzer, M., "If these data could talk," to appear in *Nature Scientific*. (7/24/17, 5 pages) [Pasquier, Lau, and Trisovic led this project; Seltzer was responsible for paper editing; 10% contribution.]
12. Pasquier, T., Singh, J., Powles, J., Eyers, D., Seltzer, M., Bacon, J., "Data Provenance to Audit Compliance with Privacy Policy in the Internet of Things," to appear in the *Journal of Personal and Ubiquitous Computing*. (8/15/17, 13 pages) [Pasquier was responsible for approximately 50% of the work, with each of the other co-authors each contributing approximately 10%.]
13. Pasquier, T., **Han, X**., **Goldstein, M**., Moyer, T., Eyers, D., Seltzer, M., Bacon, J., "Practical Whole-System Provenance Capture," *Proceedings of the ACM Symposium on Cloud Computing (SOCC 2017),* Santa Clara, CA, September 2017. (accepted 7/18/17; 14 pages) [Seltzer advised Han, Goldstein, and Pasquier; wrote/edited significant portions of the paper; contribution 20%]
14. **Yang, H**., Rudin, C., Seltzer, M., "Scalable Bayesian Rule Lists," *Proceedings of the International Conference on Machine Learning (ICML2017),*" Sydney, Australia, August 2017. [Seltzer wrote the underlying rulelist library, conducted code reviews of all of Yang's code, contribution 33%]
15. Angelino, E., **Larus-Stone, N**., **Alabi, D**., Seltzer, M., Rudin, C., "Learning Certifiably Optimal Rule Lists for Categorical Data," *Proceedings of the 23rd ACM Conference on Knowledge Discovery and Data Mining* (KDD2017)." Accepted for oral presentation. Halifax, Nova Scotia, Canada, August 2017.
16. **Han, X**., Pasquier, T., Ranjan, T., **Goldstein, M**., Seltzer, M., "FRAPpuccino: Fault-detection through Runtime Analysis of Provenance," *Proceedings of the Workshop on Hot Topics in Cloud Computing (HotCloud'17)*, USENIX Association, Santa Clara, CA, July 2017.

17. Marathe, V., Seltzer, M., Byan, S., Harris, T., "Persistent Memcached: Bringing Legacy Code to Byte-Addressable Persistent Memory." *Proceedings of USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 17),* USENIX Association, Santa Clara, CA, July 2017. [Marathe did the implementation; Seltzer advised on implementation and co-wrote the paper; 20% contribution.]
18. Whitehill, J., Seltzer, M., "A Crowdsourcing Approach to Collecting Tutorial Videos -- Toward Personalized Learning-at-Scale(L@S 2017); doi: 10.1145/3051457.3053973.
19. **Margo, D**., Seltzer, M., "A Scalable Distributed Graph Partitioner," *Proceedings of the 41st International Conference on Very Large Databases (VLDB)*, Kohola Coast, HA, September 2015.
20. **Eldridge, S.,** Appavoo, J., Joshi, A., **Waterland, A.**, Seltzer, M., "Towards General-Purpose Neural Network Computing," *Proceedings of the 13th International Conference on Parallel Architectures and Compilation Techniques (PACT 2015)*, Petrozavodsk Russia, September 2015. [Seltzer co-wrote and edited paper; 10% effort.]
21. Balakrishnan, N., Bytheway, T., Carata, L., Chick, O., Snee, J., Akoush, S., Sohan, R., Seltzer, M., Hopper, A., "Recent Advances in Computer Architecture: The Opportunities and Challenges for Provenance," *Proceedings of the 7th Workshop on the Theory and Practice of Provenance (TaPP 2015)*, Edinburgh Scotland, July 2015. [Seltzer did minor writing and editing; contribution 5%]
22. **Macko, P., Margo, D.,** Marathe, V., Seltzer, M., "LLAMA: Efficient Graph Analytics Using Large Multiversioned Arrays," *Proceedings of the 31st IEEE International Conference on Data Engineering (ICDE 2015)*, Seoul Korea, April 2015.
23. Appavoo, J., **Waterland, A., Eldridge, S**., **Zhao, K,.** Joshi, A, Homer, S, Seltzer, M, "Programmable Smart Machines: A Hybrid Neuromorphic approach to General Purpose Computation," *Proceedings of the first workshop on Neuromorphic Architectures (NeuroArch)* collocated with ISCA 2014. [Seltzer co-wrote and edited paper; 10% effort.]
24. **Angelino, E.,** Kohler, E., **Waterland, A**., Seltzer, M., Adams, M., "Accelerating MCMC via parallel predictive prefetching," *Proceedings of the 2014 Conference on Uncertainty in Artificial Intelligence (UAI '14)*, Quebec City, Quebec Canada, July 2014. [Angelino was co-advised by Seltzer and Adams; Adams appears last because we published in his (ML) community.]
25. **Rao, M.,** Parkes, D., Seltzer, M., Bacon, D., "A Framework for Incentivizing Deep Fixes," *Proceedings of the AAAI 2014 Workshop on Incentives and Trust in E-Communities (WIT-EC'14*), Quebec City, Quebec Canada, July 2014. [Seltzer guided development of model and edited paper; contribution 10%]
26. Carata, L., Akoush, S., Balakrishnan, N., Bytheway, T., Sohan, R., Seltzer, M., Hopper, A., 2014. "A primer on provenance." *Communications of the ACM* 57, 5 (May 2014), 52-60. DOI=10.1145/2596628 http://doi.acm.org/10.1145/2596628. [Highly collaborative effort; 15% contribution.]
27. **Waterland, A., Angelino, E**., Adams, R., Appavoo, J., Seltzer, M., "ASC: Automatically Scalable Computation" *Proceedings of the 2014 Conference on Architectural Support for Programming Languages and Operating Systems,* Salt Lake City Utah, March 2014.
28. Herzlinger, R., Seltzer, M., Gaynor, M., "Applying KISS to Healthcare Information Technology," *Computer* 46(11), IEEE, November 2013, 72--74. [Seltzer produced most of the final prose for the paper; 50% contribution.]
29. Borkin, M., **Yeh, C., Boyd, M., Macko, P.,** Gajos, K., Seltzer, M., Pfister, H., "Evaluation of Filesystem Provenance Visualization Tools," *Proceedings of the 2013 Conference on Information Visualization (Vis 2013)*, Atlanta Georgia, October 2013. [Yeh and Boyd were

co-advised by Seltzer and Borkin(postdoc)/Pfister(faculty); Pfister appears last because we published in a visualization venue, and that is his community.]

30. **Macko, P., Margo, D**., Seltzer, M., "Local Clustering in Provenance Graphs," *Proceedings of the 2013 International Conference on Information and Knowledge Management (CIKM 2013),* Burlingame CA, October 2013.
31. **Waterland, A., Angelino, E.,** Cubuk, E., Kaziras, E., Adams, R., Appavoo, J., Seltzer, M., "Computational Caches," *Proc. of the 2013 SYSTOR Conference*, Haifa Israel, July 2013.
32. **Macko, P., Margo, D.,** Seltzer, M., "Performance Introspection of Graph Databases," *Proceedings of the 2013 SYSTOR Conference*, Haifa Israel, July 2013.
33. **Holland, D., Angelino, E., Wald, G**., Seltzer, M., "Flash Caching on the Storage Client," *Proceedings of the 2013 USENIX Annual Technical Conference*, San Jose, CA, June 2013.

Other Refereed Contributions: None

Non-Refereed Contributions: None

Contributions to practical applications of knowledge

CamFlow

- Provenance capture software: http://www.camflow.org
- FRAPpucino Intrusion Detection system: https://github.com/michael-hahn/frap

Certifiably Optimal Rule Lists (CORELS)

- Website: https://corels.eecs.harvard.edu/
- Source code repository: https://github.com/nlarusstone/corels

Scalable Bayesian Rule List Software

- R Package on CRAN: https://cran.r-project.org/web/packages/sbrl/index.html
- Github Repository for C library and R package: https://github.com/Hongyuy/sbrlmod

Scalable Host-tree Embeddings for Efficient Partitioning

- Software: https://github.com/dmargo/sheep

Linked-node analytics using Large Multiversioned Arrays

- Software: https://github.com/goatdb/llama

Patent #9734607, "Graph processing using a mutable multilevel graph representation."
Granted: August 15, 2017 with Peter Macko and Virendra Marathe.

3. Other evidence of Impact and Contributions

A. USENIX Association: Board President (2012-2014); Vice-President (2008-2012); Acting Executive Director USENIX (2011-2012); Director at Large (1996-1998, 2006-2008)
B. National level boards
   - DARPA Information Science and Technology Study Group (ISAT) (2017-2020)
   - National Academy of Sciences Computer Science and Telecommunications Board (2016-present)
   - CRA Board of Directors (2014-2017); CRA Computing Community Consortium 2010-2013; Executive committee 2012-2013
C. Conference Steering Committees: Operating System Design and Implementation (OSDI) 2000-present; File and Storage Technologies (FAST) 2002-present; Hot Topics in Operating Systems (HotOS) 2007-present; Workshop on the Theory and Practice of Provenance (TaPP) 2009-2016
D. Architect, Oracle Corporation (1 day per week consulting, 2006-present)

4. Delays in Research Activity: None

Natural Sciences and Engineering Research Council of Canada | Conseil de recherches en sciences naturelles et en génie du Canada

SEND ONE ORIGINAL ONLY DO NOT PHOTOCOPY

APPENDIX A
Personal Data
(Form 100)

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 primarily to contact applicants and award holders. It may also be used to identify prospective reviewers and committee members, and to generate statistics. It will not be seen or used in the adjudication process.

Date: 2018/09/17

Family name	Given name	Initial(s) of all given names	Personal identification no. (PIN)
Seltzer	Margo	MI	542742

Position and complete mailing address if your primary place of employment is not a Canadian postsecondary institution or if your current mailing address is temporary

If address is temporary, indicate:

Starting date

Leaving date

Telephone number	Facsimile number	E-mail address
(604) 8226642		mseltzer@cs.ubc.ca

Telephone number (alternate): 1 (617) 3084872

Give an alternate telephone number only if you can be reached at that number during business hours.

Gender (completion optional): [ ] Male [X] Female

LANGUAGE CAPABILITY

English	Read [X]	Write [X]	Speak [X]
French	Read [ ]	Write [ ]	Speak [ ]

I wish to receive my correspondence: in English [X] in French [ ]

AREA(S) OF EXPERTISE

Provide a maximum of 10 key words that describe your area(s) of expertise. Use commas to separate them. If you have expertise with particular instruments and techniques, specify which one(s).

operating systems, database systems, transaction processing, data provenance, intrusion detection, file systems

Research subject code(s)

Primary: 2720

Secondary: 2711

Appendix D (Form 100)
Consent to Provide Limited Personal Information About Highly Qualified Personnel (HQP) to NSERC

NSERC applicants are required to describe their contributions to the training or supervision of highly qualified personnel (HQP) by providing certain details about the individuals they have trained or supervised during the six years prior to their current application. HQP information must be entered on the Personal Data Form (Form 100). This information includes the trainee's name, type of HQP training (e.g., undergraduate, master's, technical etc.) and status (completed, in-progress, incomplete), years supervised or co-supervised, title of the project or thesis, and the individual's present position.

Based on the federal *Privacy Act* rules governing the collection of personal information, applicants are asked to obtain consent from the individuals they have supervised before providing personal data about them to NSERC. In seeking this consent, the NSERC applicant must inform these individuals what data will be supplied, and assure them that it will only be used by NSERC for the purpose of assessing the applicant's contribution to HQP training. To reduce seeking consent for multiple applications, applicants will only need to seek consent one time for a six-year period. If the trainee provides consent by e-mail, the response must include confirmation that they have read and agree to the text of the consent form.

When consent cannot be obtained, applicants are asked to not provide names, or other combinations of data, that would identify those supervised. However, they may still provide the type of HQP training and status, years supervised or co-supervised, a general description of the project or thesis, and a general indication of the individual's present position if known.

An example of entering HQP information on Form 100 (with and without consent):

Name	Type of HQP Training and Status	Years Supervised or Co-supervised	Title of Project or Thesis	Present Position
Consent Received from Marie Roy				
Roy, Marie	Undergraduate (Completed)	Supervised 1994 - 1997	Isotope geochemistry in petroleum engineering	V-P (Research), Earth Analytics Inc., Calgary, Alberta
Consent Not Obtained from Marie Roy				
(name withheld)	Undergraduate (Completed)	Supervised 1994 - 1997	Isotope geochemistry	research executive in petroleum industry - western Canada

Consent Form

Name of Trainee

Applicant Information

Name Seltzer, Margo MI

Department	Postsecondary Institution
Computer Science	British Columbia

I hereby allow the above-named applicant to include limited personal data about me in grant applications submitted for consideration to NSERC for the next six years. This limited data will only include my name, type of HQP training and status, years supervised or co-supervised, title of the project or thesis and, to the best of the applicant's knowledge, my position title and company or organization at the time the application is submitted. I understand that NSERC will protect this data in accordance with the *Privacy Act*, and that it will only be used in processes that assess the applicant's contributions to the training of highly qualified personnel (HQP), including confidential peer review.

Trainee's signature _____ Date _____

Note: This form must be retained by the applicant and made available to NSERC upon request.