

Proceedings

The 2022 Mining Software Repositories Conference

MSR 2022

18-20 May 2022, Virtual

23-24 May 2022, Pittsburgh, Pennsylvania

Proceedings

The 2022 Mining Software Repositories Conference

MSR 2022

18-20 May 2022, Virtual

23-24 May 2022, Pittsburgh, Pennsylvania



Association for
Computing Machinery



IEEE
COMPUTER
SOCIETY

The Association for Computing Machinery
2 Penn Plaza, Suite 701
New York New York 10121-0701

ACM COPYRIGHT NOTICE. Copyright © 2022 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Publications Dept., ACM, Inc., fax +1 (212) 869-0481, or permissions@acm.org.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, +1-978-750-8400, +1-978-750-4470 (fax).

Notice to Past Authors of ACM-Published Articles

ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM. If you have written a work that was previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform permissions@acm.org, stating the title of the work, the author(s), and where and when published.

ACM ISBN: 978-1-4503-9303-4

Editorial production by Lisa O'Conner
Cover art production by Annie Jiu



IEEE Computer Society
Conference Publishing Services (CPS)

<http://www.computer.org/cps>

The 2022 Mining Software Repositories Conference **MSR 2022**

Table of Contents

Message from the MSR 2022 General and Program Co-Chairs	xviii
Message from the MSR 2022 Data and Tool Showcase Track Co-Chairs	xxiii
Message from the MSR 2022 Hackathon Track Co-Chairs	xxiv
Message from the MSR 2022 Industry Track Co-Chairs	xxv
Message from the MSR 2022 Mining Challenge Track Co-Chairs	xxvi
Message from the MSR 2022 Registered Reports Track Co-Chairs	xxviii
Message from the MSR 2022 Shadow PC Track Co-Chairs	xxx
Message from the MSR 2022 Tutorials Track Co-Chairs	xxxi
Organizing Committee	xxxii
Program Committee	xxxv

Session 1: Communication (Documentation & Automation)

An Empirical Evaluation of GitHub Copilot's Code Suggestions	1
<i>Nhan Nguyen (University of Alberta, Canada) and Sarah Nadi (University of Alberta, Canada)</i>	
BotHunter: An Approach to Detect Software Bots in GitHub	6
<i>Ahmad Abdellatif (Concordia University, Canada), Mairieli Wessel (Delft University of Technology, Netherlands), Igor Steinmacher (Universidade Tecnológica Federal do Paraná, Brazil), Marco A. Gerosa (Northern Arizona University, USA), and Emad Shihab (Concordia University, Canada)</i>	
Comments on Comments: Where Code Review and Documentation Meet	18
<i>Nikitha Rao (Carnegie Mellon University, United States), Jason Tsay (IBM Research, United States), Martin Hirzel (IBM Research, United States), and Vincent J. Hellendoorn (Carnegie Mellon University, United States)</i>	
Does This Apply to Me? An Empirical Study of Technical Context in Stack Overflow	23
<i>Akalanka Galappaththi (University of Alberta, Canada), Sarah Nadi (University of Alberta, Canada), and Christoph Treude (University of Melbourne, Australia)</i>	
Towards Reliable Agile Iterative Planning via Predicting Documentation Changes of Work Items	35
<i>Jirat Pasuksmit (The University of Melbourne, Australia), Patanamon Thongtanunam (The University of Melbourne, Australia), and Shanika Karunasekera (The University of Melbourne, Australia)</i>	

Session 2: Maintenance (Issues & Smells)

Beyond Duplicates: Towards Understanding and Predicting Link Types in Issue Tracking Systems	48
<i>Clara Marie Lüders (University of Hamburg, Germany), Abir Bouraffa (University of Hamburg, Germany), and Walid Maalej (University of Hamburg, Germany)</i>	
Smelly Variables in Ansible Infrastructure Code: Detection, Prevalence, and Lifetime	61
<i>Ruben Opdebeeck (Vrije Universiteit Brussel, Belgium), Ahmed Zerouali (Vrije Universiteit Brussel, Belgium), and Coen De Roover (Vrije Universiteit Brussel, Belgium)</i>	
An Alternative Issue Tracking Dataset of Public Jira Repositories	73
<i>Lloyd Montgomery (University of Hamburg, Germany), Clara Lüders (University of Hamburg, Germany), and Walid Maalej (University of Hamburg, Germany)</i>	
Real-World Clone-Detection in Go	78
<i>Qinyun Wu (Bytedance Ltd., China), Huan Song (Bytedance Ltd., China), and Ping Yang (Bytedance Ltd., China)</i>	

Session 3: Introspection, Vision, and Human Aspects

Geographic Diversity in Public Code Contributions: An Exploratory Large-Scale Study Over 50 Years	80
<i>Davide Rossi (University of Bologna, Italy) and Stefano Zacchiroli (LTCI, Télécom Paris, Institut Polytechnique de Paris, France)</i>	
Operationalizing Threats to MSR Studies by Simulation-Based Testing	86
<i>Johannes Härtel (University of Koblenz, Germany) and Ralf Lämmel (University of Koblenz, Germany)</i>	
The General Index of Software Engineering Papers	98
<i>Zeinab Abou Khalil (Inria de Paris, France) and Stefano Zacchiroli (LTCI, Télécom Paris, Institut Polytechnique de Paris, France)</i>	
Challenges and Future Research Direction for Microtask Programming in Industry	103
<i>Masanari Kondo (Kyushu University, Japan), Shinobu Saito (NTT Computer and Data Science Laboratories, Japan), Yukako Imura (NTT Computer and Data Science Laboratories, Japan), Eunjong Choi (Kyoto Institute of Technology, Japan), Osamu Mizuno (Kyoto Institute of Technology, Japan), Yasutaka Kamei (Kyushu University, Japan), and Naoyasu Ubayashi (Kyushu University, Japan)</i>	
Starting the InnerSource Journey: Key Goals and Metrics to Measure Collaboration	105
<i>Daniel Izquierdo-Cortázar (Bitergia, Spain), Jesús Alonso-Gutiérrez (Santander Group, Spain), Alberto Pérez García-Plaza (Bitergia, Spain), Gregorio Robles (Universidad Rey Juan Carlos, Spain), and Jesús M. González-Barahona (Universidad Rey Juan Carlos, Spain)</i>	

Mining Challenge

An Exploratory Study on Refactoring Documentation in Issues Handling	107
<i>Eman Abdullah AlOmar (Stevens Institute of Technology, USA), Anthony Peruma (Rochester Institute of Technology, USA), Mohamed Wiem Mkaouer (Rochester Institute of Technology, USA), Christian D. Newman (Rochester Institute of Technology, USA), and Ali Ouni (University of Quebec, Canada)</i>	
Between JIRA and GitHub: ASFBot and its Influence on Human Comments in Issue Trackers	112
<i>Ambarish Moharil (Eindhoven University of Technology, The Netherlands), Dmitrii Orlov (Eindhoven University of Technology, The Netherlands), Samar Jameel (Eindhoven University of Technology, The Netherlands), Tristan Trouwen (Eindhoven University of Technology, The Netherlands), Nathan Cassee (Eindhoven University of Technology, The Netherlands), and Alexander Serebrenik (Eindhoven University of Technology, The Netherlands)</i>	
Is Refactoring Always a Good Egg? Exploring the Interconnection Between Bugs and Refactorings	117
<i>Amirreza Bagheri (University of Szeged, Hungary) and Peter Hegedus (University of Szeged, Hungary)</i>	
On the Co-Occurrence of Refactoring of Test and Source Code	122
<i>Nicholas Alexandre Nagy (Concordia University, Canada) and Rabe Abdalkareem (Carleton University, Canada)</i>	
Refactoring Debt: Myth or Reality? An Exploratory Study on the Relationship Between Technical Debt and Refactoring	127
<i>Anthony Peruma (Rochester Institute of Technology, USA), Eman Abdullah AlOmar (Stevens Institute of Technology, USA), Christian D. Newman (Rochester Institute of Technology, USA), Mohamed Wiem Mkaouer (Rochester Institute of Technology, USA), and Ali Ouni (University of Quebec, Canada)</i>	
Studying the Impact of Continuous Delivery Adoption on Bug-Fixing Time in Apache's Open-Source Projects	132
<i>Carlos Diego Andrade de Almeida (Federal University of Ceará, Brazil), Diego Nogueira Feijó (Federal University of Ceará, Brazil), and Lincoln Souza Rocha (Federal University of Ceará, Brazil)</i>	
Which Bugs are Missed in Code Reviews: An Empirical Study on SmartSHARK Dataset	137
<i>Fatemeh Khoshnoud (Shiraz University, Iran), Ali Rezaei Nasab (Shiraz University, Iran), Zahra Toudeji (Shiraz University, Iran), and Ashkan Sami (Shiraz University, Iran)</i>	

Tutorial

Empirical Standards for Repository Mining	142
<i>Preetha Chatterjee (Drexel University, USA), Tushar Sharma (Dalhousie University, Canada), and Paul Ralph (Dalhousie University, Canada)</i>	

Session 4: Software Quality (Bugs & Smells)

Dazzle: using Optimized Generative Adversarial Networks to Address Security Data Class Imbalance Issue	144
<i>Rui Shu (North Carolina State University, USA), Tianpei Xia (North Carolina State University, USA), Laurie Williams (North Carolina State University, USA), and Tim Menzies (North Carolina State University, USA)</i>	
How to Improve Deep Learning for Software Analytics (a Case Study with Code Smell Detection)	156
<i>Rahul Yedida (NC State University, USA) and Tim Menzies (NC State University, USA)</i>	
To What Extent do Deep Learning-Based Code Recommenders Generate Predictions by Cloning Code from the Training Set?	167
<i>Matteo Ciniselli (Università della Svizzera italiana, Switzerland), Luca Pascarella (Università della Svizzera italiana, Switzerland), and Gabriele Bavota (Università della Svizzera italiana, Switzerland)</i>	
Searching for High-Fidelity Builds using Active Learning	179
<i>Harshitha Menon (Lawrence Livermore National Laboratory, USA), Konstantinos Parasyris (Lawrence Livermore National Laboratory, USA), Tom Scogland (Lawrence Livermore National Laboratory, USA), and Todd Gamblin (Lawrence Livermore National Laboratory, USA)</i>	
ApacheJIT: A Large Dataset for Just-In-Time Defect Prediction	191
<i>Hossein Keshavarz (University of Waterloo, Canada) and Meiyappan Nagappan (University of Waterloo, Canada)</i>	
ReCover: a Curated Dataset for Regression Testing Research	196
<i>Francesco Altiero (Università degli Studi di Napoli Federico II, Italy), Anna Corazza (Università degli Studi di Napoli Federico II, Italy), Sergio Di Martino (Università degli Studi di Napoli Federico II, Italy), Adriano Peron (Università degli Studi di Napoli Federico II, Italy), and Luigi L. L. Starace (Università degli Studi di Napoli Federico II, Italy)</i>	

Tutorial

Mining the Ethereum Blockchain Platform: Best Practices and Pitfalls (MSR 2022 Tutorial)	201
<i>Gustavo A. Oliva (Queen's University, Canada)</i>	

Session 5: Communication & Domains

Mining the Usage of Reactive Programming APIs: A Study on GitHub and Stack Overflow	203
<i>Carlos Zimmerle (Federal University of Pernambuco, Brazil), Kiev Gama (Federal University of Pernambuco, Brazil), Fernando Castor (Utrecht University, The Netherlands), and José Murilo Mota Filho (Federal University of Pernambuco, Brazil)</i>	

Painting the Landscape of Automotive Software in GitHub	215
<i>Sangeeth Kochanthara (Eindhoven University of Technology, The Netherlands), Yanja Dajsuren (Eindhoven University of Technology, The Netherlands), Loek Cleophas (Eindhoven University of Technology, The Netherlands), and Mark van den Brand (Eindhoven University of Technology, The Netherlands)</i>	
DISCO: A Dataset of Discord Chat Conversations for Software Engineering Research	227
<i>Keerthana Muthu Subash (Carleton University, Canada), Lakshmi Prasanna Kumar (Carleton University, Canada), Sri Lakshmi Vadlamani (Carleton University, Canada), Preetha Chatterjee (Drexel University, United States), and Olga Baysal (Carleton University, Canada)</i>	
Inspect4py: A Knowledge Extraction Framework for Python Code Repositories	232
<i>Rosa Filgueira (University of St Andrews, UK) and Daniel Garijo (Universidad Politécnica de Madrid, Spain)</i>	
SLNET: A Redistributable Corpus of 3rd-Party Simulink Models	237
<i>Sohil Lal Shrestha (University of Texas at Arlington, USA), Shafiul Azam Chowdhury (University of Texas at Arlington, USA), and Christoph Csallner (University of Texas at Arlington, USA)</i>	
SoCCMiner: A Source Code-Comments and Comment-Context Miner	242
<i>Murali Sridharan (University of Oulu, Finland), Mika Mäntylä (University of Oulu, Finland), Maëlick Claes (University of Oulu, Finland), and Leevi Rantala (University of Oulu, Finland)</i>	
SOSum: A Dataset of Stack Overflow Post Summaries	247
<i>Bonan Kou (Purdue University, USA), Yifeng Di (Purdue University, USA), Muhao Chen (University of Southern California, USA), and Tianyi Zhang (Purdue University, USA)</i>	

Session 6: Maintenance & Testing

An Empirical Study on Maintainable Method Size in Java	252
<i>Shaiful Alam Chowdhury (University of British Columbia, Canada), Gias Uddin (University of Calgary, Canada), and Reid Holmes (University of British Columbia, Canada)</i>	
Characterizing High-Quality Test Methods: A First Empirical Study	265
<i>Victor Veloso (Universidade Federal de Minas Gerais, Brazil) and Andre Hora (Universidade Federal de Minas Gerais, Brazil)</i>	
CLIP Meets GamePhysics: Towards bug Identification in Gameplay Videos using Zero-Shot Transfer Learning	270
<i>Mohammad Reza Taesiri (University of Alberta, Canada), Finlay Macklon (University of Alberta, Canada), and Cor-Paul Bezemer (University of Alberta, Canada)</i>	
Complex Python Features in the Wild	282
<i>Yi Yang (Rensselaer Polytechnic Institute, USA), Ana Milanova (Rensselaer Polytechnic Institute, USA), and Martin Hirzel (IBM Research, USA)</i>	

ManyTypes4TypeScript: A Comprehensive TypeScript Dataset for Sequence-Based Type Inference.....	294
<i>Kevin Jesse (University of California, Davis, USA) and Premkumar T. Devanbu (University of California, Davis, USA)</i>	
Methods2Test: A Dataset of Focal Methods Mapped to Test Cases	299
<i>Michele Tufano (Microsoft, USA), Shao Kun Deng (Microsoft, USA), Neel Sundaresan (Microsoft, USA), and Alexey Svyatkovskiy (Microsoft, USA)</i>	
npm-Filter: Automating the Mining of Dynamic Information from npm Packages	304
<i>Ellen Arteca (Northeastern University, USA) and Alexi Turcotte (Northeastern University, USA)</i>	

Session 7: Developer Wellbeing & Project Communication

How Heated is it? Understanding GitHub Locked Issues	309
<i>Isabella Ferreira (Polytechnique Montréal, Canada), Bram Adams (Queen's University, Canada), and Jinghui Cheng (Polytechnique Montréal, Canada)</i>	
On the Violation of Honesty in Mobile Apps: Automated Detection and Categories	321
<i>Humphrey O. Obie (Monash University, Australia), Idowu Ileku (Data Science Nigeria, Nigeria), Hung Du (Deakin University, Australia), Mojtaba Shahin (RMIT University, Australia), John Grundy (Monash University, Australia), Li Li (Monash University, Australia), Jon Whittle (CSIRO's Data61, Australia), and Burak Turhan (University of Oulu, Finland)</i>	
Exploring Apache Incubator Project Trajectories with APEX	333
<i>Anirudh Ramchandran (University of California, Davis, USA), Likang Yin (University of California, Davis, USA), and Vladimir Filkov (University of California, Davis, USA)</i>	
The OCEAN Mailing List Data Set: Network Analysis Spanning Mailing Lists and Code Repositories	338
<i>Melanie Warrick (Google, Inc.), Samuel F. Rosenblatt (University of Vermont), Jean-Gabriel Young (University of Vermont), Amanda Casari (Google, Inc.), Laurent Hébert-Dufresne (University of Vermont), and James Bagrow (University of Vermont)</i>	
The Unexplored Treasure Trove of Phabricator Code Reviews	343
<i>Gunnar Kudrjavets (University of Groningen, Netherlands), Nachiappan Nagappan (Microsoft Research, USA), and Ayushi Rastogi (University of Groningen, Netherlands)</i>	
The Unsolvable Problem or the Unheard Answer? A Dataset of 24,669 Open-Source Software Conference Talks	348
<i>Kimberly Truong (Oregon State University, USA), Courtney Miller (Carnegie Mellon University, USA), Bogdan Vasilescu (Carnegie Mellon University, USA), and Christian Kästner (Carnegie Mellon University, USA)</i>	

Session 8: Large-Scale Mining & Software Ecosystems

A Large-Scale Comparison of Python Code in Jupyter Notebooks and Scripts	353
<i>Konstantin Grotov (JetBrains Research, ITMO University), Sergey Titov (JetBrains Research), Vladimir Sotnikov (JetBrains Research), Yaroslav Golubev (JetBrains Research), and Timofey Bryksin (JetBrains Research)</i>	
An Empirical Study on the Survival Rate of GitHub Projects	365
<i>Adem Ait (IN3 - UOC, Spain), Javier Luis Cánovas Izquierdo (IN3 - UOC, Spain), and Jordi Cabot (IN3 - UOC, ICREA, Spain)</i>	
Do Customized Android Frameworks Keep Pace with Android?	376
<i>Pei Liu (Monash University, Australia), Mattia Fazzini (University of Minnesota, United States), John Grundy (Monash University, Australia), and Li Li (Monash University, Australia)</i>	
DaSEA – A Dataset for Software Ecosystem Analysis	388
<i>Petya Buchkova (IT University of Copenhagen, Denmark), Joakim Hey Hinnerskov (IT University of Copenhagen, Denmark), Kasper Olsen (IT University of Copenhagen, Denmark), and Rolf-Helge Pfeiffer (IT University of Copenhagen, Denmark)</i>	
Dataset: Dependency Networks of Open Source Libraries Available Through CocoaPods, Carthage and Swift PM	393
<i>Kristiina Rahkema (University of Tartu, Estonia) and Dietmar Pfahl (University of Tartu, Estonia)</i>	
Lupa: A Framework for Large Scale Analysis of the Programming Language Usage	398
<i>Anna Vlasova (JetBrains Research), Maria Tigina (JetBrains Research, ITMO University), Ilya Vlasov (Saint Petersburg State University), Anastasiia Birillo (JetBrains Research), Yaroslav Golubev (JetBrains Research), and Timofey Bryksin (JetBrains Research)</i>	
GitDelver Enterprise Dataset (GDED): An Industrial Closed-Source Dataset for Socio-Technical Research	403
<i>Nicolas Riquet (University of Namur, Belgium), Xavier Devroey (University of Namur, Belgium), and Benoît Vanderose (University of Namur, Belgium)</i>	

Session 9: Scaling & Cloud

SniP: An Efficient Stack Tracing Framework for Multi-Threaded Programs	408
<i>Arun KP (Indian Institute of Technology, India), Saurabh Kumar (Indian Institute of Technology, India), Debadatta Mishra (Indian Institute of Technology, India), and Biswabandan Panda (Indian Institute of Technology, India)</i>	
Tooling for Time- and Space-Efficient git Repository Mining	413
<i>Fabian Hesinde (University of Potsdam, Germany), Willy Scheibel (University of Potsdam, Germany), and Jürgen Döllner (University of Potsdam, Germany)</i>	
TSSB-3M: Mining Single Statement Bugs at Massive Scale	418
<i>Cedric Richter (University of Oldenburg, Germany) and Heike Wehrheim (University of Oldenburg, Germany)</i>	

Session 10: Security

LibDB: An Effective and Efficient Framework for Detecting Third-Party Libraries in Binaries	423
<i>Wei Tang (Tsinghua University, China), Yanlin Wang (Microsoft Research, China), Hongyu Zhang (The University of Newcastle, Australia), Shi Han (Microsoft Research, China), Ping Luo (Tsinghua University, China), and Dongmei Zhang (Microsoft Research, China)</i>	
Noisy Label Learning for Security Defects	435
<i>Roland Croft (CREST - The Centre for Research on Engineering Software Technologies, University of Adelaide, Cyber Security Cooperative Research Centre, Australia), M. Ali Babar (CREST - The Centre for Research on Engineering Software Technologies, University of Adelaide, Cyber Security Cooperative Research Centre, Australia), and Huaming Chen (CREST - The Centre for Research on Engineering Software Technologies, University of Adelaide, Cyber Security Cooperative Research Centre, Australia)</i>	
WeakSATD: Detecting Weak Self-Admitted Technical Debt	448
<i>Barbara Russo (Free University of Bozen-Bolzano, Italy), Matteo Camilli (Free University of Bozen-Bolzano, Italy), and Moritz Mock (Free University of Bozen-Bolzano, Italy)</i>	
AndroOBFS: Time-Tagged Obfuscated Android Malware Dataset with Family Informationn	454
<i>Saurabh Kumar (Indian Institute of Technology Kanpur, India), Debadatta Mishra (Indian Institute of Technology Kanpur, India), Biswabandan Panda (Indian Institute of Technology Bombay, India), and Sandeep Kumar Shukla (Indian Institute of Technology Kanpur, India)</i>	
TriggerZoo: A Dataset of Android Applications Automatically Infected with Logic Bombs	459
<i>Jordan Samhi (University of Luxembourg, Luxembourg), Tegawendé F. Bissyandé (University of Luxembourg, Luxembourg), and Jacques Klein (University of Luxembourg, Luxembourg)</i>	
Vul4J: A Dataset of Reproducible Java Vulnerabilities Geared Towards the Study of Program Repair Techniques	464
<i>Quang-Cuong Bui (Hamburg University of Technology, Germany), Riccardo Scandariato (Hamburg University of Technology, Germany), and Nicolás E. Díaz Ferreyra (Hamburg University of Technology, Germany)</i>	

Session 11: Machine Learning & Information Retrieval

Challenges in Migrating Imperative Deep Learning Programs to Graph Execution: An Empirical Study	469
<i>Tatiana Castro Vélez (City University of New York (CUNY) Graduate Center, USA), Raffi Khatchadourian (City University of New York (CUNY) Hunter College, USA), Mehdi Bagherzadeh (Oakland University, USA), and Anita Raja (City University of New York (CUNY) Hunter College, USA)</i>	
Does Configuration Encoding Matter in Learning Software Performance? An Empirical Study on Encoding Schemes	482
<i>Jingzhi Gong (Loughborough University, UK) and Tao Chen (Loughborough University, UK)</i>	

Multimodal Recommendation of Messenger Channels	495
<i>Ekaterina Koshchenko (JetBrains Research, The Netherlands), Egor Klimov (JetBrains Research, Russia), and Vladimir Kovalenko (JetBrains Research, The Netherlands)</i>	
On the Naturalness of Fuzzer-Generated Code	506
<i>Rajeswari Hita Kambhamettu (Carnegie Mellon University, USA), John Billos (Wake Forest University, USA), Tomi Oluwaseun-Apo (Pennsylvania State University, USA), Benjamin Gafford (Carnegie Mellon University, USA), Rohan Padhye (Carnegie Mellon University, USA), and Vincent J. Hellendoorn (Carnegie Mellon University, USA)</i>	
Senatus - A Fast and Accurate Code-to-Code Recommendation Engine	511
<i>Fran Silavong (JPMorgan Chase, United Kingdom), Sean Moran (JPMorgan Chase, United Kingdom), Antonios Georgiadis (JPMorgan Chase, United Kingdom), Rohan Saphal (JPMorgan Chase, United Kingdom), and Rob Otter (JPMorgan Chase, United Kingdom)</i>	
GraphCode2Vec: Generic Code Embedding via Lexical and Program Dependence Analyses	524
<i>Wei Ma (University of Luxembourg), Mengjie Zhao (LMU Munich, Germany), Ezekiel Soremekun (University of Luxembourg), Qiang Hu (University of Luxembourg), Jie M. Zhang (University College London, United Kingdom), Mike Papadakis (University of Luxembourg), Maxime Cordy (University of Luxembourg), Xiaofei Xie (Singapore Management University, Singapore), and Yves Le Traon (University of Luxembourg)</i>	

Session 12: Integration & Large-Scale Mining

Do Small Code Changes Merge Faster? A Multi-Language Empirical Investigation	537
<i>Gunnar Kudrjavets (University of Groningen, Netherlands), Nachiappan Nagappan (Meta Platforms, Inc., USA), and Ayushi Rastogi (University of Groningen, Netherlands)</i>	
FaST: A Linear Time Stack Trace Alignment Heuristic for Crash Report Deduplication	549
<i>Irving Muller Rodrigues (Polytechnique Montreal, Canada), Daniel Aloise (Polytechnique Montreal, Canada), and Eraldo Rezende Fernandes (Leuphana University of Lüneburg, Germany)</i>	
Is Open Source Eating the World's Software? Measuring the Proportion of Open Source in Proprietary Software using Java Binaries	561
<i>Julius Musseau (Mergebase, Canada), John Speed Meyers (Chainguard, USA), George P. Sieniawski (IQT Labs, USA), C. Albert Thompson (Ford Motor Company, USA), and Daniel M. German (University of Victoria, Canada)</i>	
Methods for Stabilizing Models Across Large Samples of Projects (with Case Studies on Predicting Defect and Project Health)	566
<i>Suvodeep Majumder (North Carolina State University, USA), Tianpei Xia (North Carolina State University, USA), Rahul Krishna (North Carolina State University, USA), and Tim Menzies (North Carolina State University, USA)</i>	

Mining Code Review Data to Understand Waiting Times Between Acceptance and Merging: An Empirical Analysis	579
<i>Gunnar Kudrjavets (University of Groningen, Netherlands), Aditya Kumar (Snap, Inc., USA), Nachiappan Nagappan (Meta Platforms, Inc., USA), and Ayushi Rastogi (University of Groningen, Netherlands)</i>	
TwinDroid: A Dataset of Android app System Call Traces and Trace Generation Pipeline	591
<i>Asma Razgallah (Univeristé du Québec à Chicoutimi, Canada), Raphaël Khoury (Univeristé du Québec à Chicoutimi, Canada), and Jean-Baptiste Poulet (Univeristé du Québec à Chicoutimi, Canada)</i>	

Session 13: Security & Quality

LineVD: Statement-Level Vulnerability Detection using Graph Neural Networks	596
<i>David Hin (The University of Adelaide, Australia), Andrey Kan (AWS AI Labs, Australia), Huaming Chen (The University of Adelaide, Australia), and M. Ali Babar (The University of Adelaide, Australia)</i>	
LineVul: A Transformer-Based Line-Level Vulnerability Prediction	608
<i>Michael Fu (Monash University, Australia) and Chakkrit Tantithamthavorn (Monash University, Australia)</i>	
On the Use of Fine-Grained Vulnerable Code Statements for Software Vulnerability Assessment Models	621
<i>Triet Huynh Minh Le (The University of Adelaide, Australia) and M. Ali Babar (The University of Adelaide, Australia; Cyber Security Cooperative Research Centre, Australia)</i>	
ECench: An Energy Bug Benchmark of Ethereum Client Software	634
<i>Jinyoung Kim (Sungkyunkwan University, Republic of Korea), Misoo Kim (Sungkyunkwan University, Republic of Korea), and Eunseok Lee (Sungkyunkwan University, Republic of Korea)</i>	
Microsoft CloudMine: Data Mining for the Executive Order on Improving the Nation's Cybersecurity	639
<i>Kim Herzig (Microsoft Corporation, USA), Luke Ghostling (Microsoft Corporation, USA), Maximilian Grothusmann (Microsoft Corporation, Germany), Sascha Just (Microsoft Corporation, Germany), Nora Huang (Microsoft Corporation, Canada), Alan Klimowski (Microsoft Corporation, USA), Yashasvini Ramkumar (Microsoft Corporation, USA), Myles McLeroy (Microsoft Corporation, USA), Kivanc Muslu (Microsoft Corporation, USA), Hitesh Sajani (Microsoft Corporation, USA), and Varsha Vadaga (Microsoft Corporation, USA)</i>	

Session 14: Software Quality

Evaluating the Effectiveness of Local Explanation Methods on Source Code-Based Defect Prediction Models	640
<i>Gao Yuxiang (Jiangsu Normal University, China), Zhu Yi (Jiangsu Normal University, China), and Yu Qiao (Jiangsu Normal University, China)</i>	

Problems and Solutions in Applying Continuous Integration and Delivery to 20 Open-Source Cyber-Physical Systems	646
<i>Fiorella Zampetti (University of Sannio, Italy), Vittoria Nardone (University of Sannio, Italy), and Massimiliano Di Penta (University of Sannio, Italy)</i>	
To Type or Not to Type? A Systematic Comparison of the Software Quality of JavaScript and TypeScript Applications on GitHub	658
<i>Justus Bogner (University of Stuttgart, Germany) and Manuel Merkel (University of Stuttgart, Germany)</i>	
Using Bandit Algorithms for Selecting Feature Reduction Techniques in Software Defect Prediction	670
<i>Masateru Tsunoda (Kindai University, Japan), Akito Monden (Okayama University, Japan), Koji Toda (Fukuoka Institute of Technology, Japan), Amjed Tahir (Massey University, New Zealand), Kwabena Ebo Bennin (Wageningen University & Research, Netherlands), Keitaro Nakasai (National Institute of Technology, Kagoshima College, Japan), Masataka Nagura (Nanzan University, Japan), and Kenichi Matsumoto (Nara Institute of Science and Technology, Japan)</i>	
Constructing Dataset of Functionally Equivalent Java Methods using Automated Test Generation Techniques	682
<i>Yoshiki Higo (Osaka University, Japan), Shinsuke Matsumoto (Osaka University, Japan), Shinji Kusumoto (Osaka University, Japan), and Kazuuya Yasuda (Hitachi, Ltd., Japan)</i>	
Extracting Corrective Actions from Code Repositories	687
<i>Yegor Bugayenko (Huawei, Russia), Kirill Daniakin (Innopolis University, Russia), Mirko Farina (Innopolis University, Russia), Firas Jolha (Innopolis University, Russia), Artem Kruglov (Innopolis University, Russia), Giancarlo Succi (Innopolis University, Russia), and Witold Pedrycz (University of Alberta, Canada)</i>	

Session 15: Collaboration & Open Source

Code Review Practices for Refactoring Changes: An Empirical Study on OpenStack	689
<i>Eman Abdullah AlOmar (Stevens Institute of Technology, USA), Moataz Chouchen (University of Quebec, Canada), Mohamed Wiem Mkaouer (Rochester Institute of Technology, USA), and Ali Ouni (University of Quebec, Canada)</i>	
A Time Series-Based Dataset of Open-Source Software Evolution	702
<i>Bruno L. Sousa (Department of Computer Science - UFMG, Brazil), Mariza A. S. Bigonha (Department of Computer Science - UFMG, Brazil), Kecia A. M. Ferreira (Department of Computing - CEFET-MG, Brazil), and Glaura Franco (Department of Statistics - UFMG, Brazil)</i>	
A Versatile Dataset of Agile Open Source Software Projects	707
<i>Vali Tawosi (University College London, UK), Afnan Al-Subaihin (University College London, UK), Rebecca Moussa (University College London, UK), and Federica Sarro (University College London, UK)</i>	

FixJS: A Dataset of Bug-Fixing JavaScript Commits	712
<i>Viktor Csuwik (University of Szeged, Hungary) and László Vidács (University of Szeged, Hungary)</i>	
LAGOON: An Analysis Tool for Open Source Communities	717
<i>Sourya Dey (Galois, Inc., USA) and Walt Woods (Galois, Inc., USA)</i>	
Automatically Prioritizing and Assigning Tasks from Code Repositories in Puzzle Driven Development	722
<i>Yegor Bugayenko (Huawei, Russia), Ayomide Bakare (Innopolis University, Russia), Arina Cheverda (Innopolis University, Russia), Mirko Farina (Innopolis University, Russia), Artem Kruglov (Innopolis University, Russia), Yaroslav Plaksin (Innopolis University, Russia), Giancarlo Succi (Innopolis University, Russia), and Witold Pedrycz (University of Alberta, Canada)</i>	

Tutorial

Software Bots in Software Engineering: Benefits and Challenges	724
<i>Mairieli Wessel (Delft University of Technology, The Netherlands), Marco A. Gerosa (Northern Arizona University, USA), and Emad Shihab (Concordia University, Canada)</i>	

Hackathon

Bot Detection in GitHub Repositories	726
<i>Natarajan Chidambaram (University of Mons, Belgium) and Pooya Rostami Mazrae (University of Mons, Belgium)</i>	
GitRank: A Framework to Rank GitHub Repositories	729
<i>Niranjan Hasabnis (Intel Lab, USA)</i>	
Maintenance and Evolution: GrimoireLab Graal	732
<i>Willem Meijer (University of Groningen, The Netherlands), David Visscher (University of Groningen, The Netherlands), Erwin de Haan (University of Groningen, The Netherlands), Merijn Schröder (University of Groningen, The Netherlands), Leon Visscher (University of Groningen, The Netherlands), Andrea Capiluppi (University of Groningen, The Netherlands), and Ioan Botez (University of Groningen, The Netherlands)</i>	
OpenSSL 3.0.0: An Exploratory Case Study	735
<i>James Walden (Northern Kentucky University, USA)</i>	
Quid Pro Quo: An Exploration of Reciprocity in Code Review	738
<i>Gavidia-Calderon Carlos (The Open University, United Kingdom), Han DongGyun (Singapore Management University, Singapore), and Bennaceur Amel (The Open University, United Kingdom)</i>	
Replicating Data Pipelines with GrimoireLab	741
<i>Kalvin Eng (University of Alberta, Canada) and Hareem Sahar (University of Alberta, Canada)</i>	

Session 16: Non-functional Properties (Availability, Security, Legal Aspects)

A Deep Study of the Effects and Fixes of Server-Side Request Races in Web Applications	744
<i>Zhengyi Qiu (North Carolina State University, USA), Shudi Shao (North Carolina State University, USA), Qi Zhao (North Carolina State University, USA), Hassan Ali Khan (North Carolina State University, USA), Xinning Hui (North Carolina State University, USA), and Guoliang Jin (North Carolina State University, USA)</i>	
A Large-Scale Dataset of (Open Source) License Text Variants	757
<i>Stefano Zacchiroli (LTCI, Télécom Paris, Institut Polytechnique de Paris, France)</i>	
Detecting Privacy-Sensitive Code Changes with Language Modeling	762
<i>Gökalp Demirci (Meta Platforms, Inc., USA), Vijayaraghavan Murali (Meta Platforms, Inc., USA), Imad Ahmad (Meta Platforms, Inc., USA), Rajeev Rao (Meta Platforms, Inc., USA), and Gareth Ari Aye (Meta Platforms, Inc., USA)</i>	
SECOM: Towards a Convention for Security Commit Messages	764
<i>Sofia Reis (University of Lisbon, Portugal), Rui Abreu (University of Porto, Portugal), Hakan Erdogmus (Carnegie Mellon University, USA), and Corina Păsăreanu (Carnegie Mellon University, USA)</i>	
Varangian: A Git Bot for Augmented Static Analysis	766
<i>Saurabh Pujar (IBM Research, USA), Yunhui Zheng (IBM Research, USA), Luca Buratti (IBM Research, USA), Burn Lewis (IBM Research, USA), Alessandro Morari (IBM Research, USA), Jim Laredo (IBM Research, USA), Kevin Postlethwait (Red Hat, USA), and Christoph Görn (Red Hat, Germany)</i>	
Author Index	769

Message from the MSR 2022 General and Program Co-Chairs

Welcome to the 19th International Conference on Mining Software Repositories (MSR), co-located with the 44th International Conference on Software Engineering (ICSE 2022). MSR is a thriving research community that organizes a yearly conference that has gained a solid reputation amongst software engineering researchers. MSR 2022 features nine tracks – Technical, Data and Tool Showcase, Mining Challenge, Hackathon, Registered Report, Industry, Tutorials, Vision and Reflection, and Shadow PC – and prestigious awards – ACM SIGSOFT Distinguished Paper, Most Influential Paper, FOSS Impact Paper, MSR Ric Holt Early Career Achievement, and MSR Foundational Contribution.

Due to the COVID-19 pandemic, MSR 2022 has been organized following a hybrid format including a virtual event and an in-presence event. The virtual MSR 2022 of the program has been planned to be held on May 18-20 and will include all accepted paper presentations, keynotes, tutorials, and invited talks. An in-person MSR 2022 has also been organized on May 23-24 in Pittsburgh, USA, featuring interactive events.

This year, we received a record number of 235 submissions across the six paper tracks of the conference: the technical track (138), the data and tool showcase (49), the mining challenge (13), the hackathon (7), the registered reports track (14), and the industry track (14). For the sixth year in a row, MSR used double-anonymous reviewing in both the technical and mining challenge tracks to reduce reviewer bias and to ensure fairness in the review process. Due to the specific requirements of the tracks, the data and tool showcase, the hackathon, the registered reports, and the industry tracks followed a single-blind review model. Where possible, all MSR 2022 tracks encouraged Open Science policies to enable the sharing of tools and data for reviewers and fellow researchers.

The technical track received 138 submissions (111 full and 27 short). One full paper was desk rejected. The remaining 137 papers went through a thorough review process. We accepted a total of 47 out of 138 papers, with an overall acceptance rate of 34% across the technical track. By paper length, our acceptance rates are 39 out of 111 full papers (35%) and 8 out of 27 short papers (30%).

All technical track submissions were reviewed by at least three members of the program committee. To further ensure the quality of the reviews and fairness of the final decision, we also continued with the leading reviewer model this year, where a fourth reviewer moderated the discussion and summarized it into a meta-review. The reviews were released to the authors several weeks before final decisions on the manuscripts were made. The authors were then given an opportunity to respond to the reviews, and reviewers were required to respond to the authors' response if one was provided. The authors could also withdraw their manuscript after seeing the reviews if they desired.

We did not aspire towards meeting any quota on the number of papers to be accepted. Rather, the acceptance decisions were made based on the program committee's discussions on the content and quality of each paper individually. Following its tradition, MSR 2022 papers were accepted or rejected based on their submitted length, and we did not permit downgrading papers from full to short papers.

New to MSR 2022: We expanded the scope of the Data Showcase track to include tools (i.e., the track is now the Data and Tool Showcase). This expanded track not only encourages submissions of data sets that can be used to further advance research in the mining software repositories area, but also tools to benefit researchers and practitioners. This track is expanded to encourage more tools from the MSR community, as the number of tool papers has been low in prior editions of MSR. Our effort has been successful as evident from the large increase in the number of submissions to the newly expanded track (a record number of 49 submissions, as compared to 26 in the prior year). We have also created a new industry track, which has been highly successful, receiving 14 submissions despite being its inaugural edition. Moreover, we have introduced a Vision and Reflection track; this by-invitation track solicits reflections and discusses the future of MSR. Additionally, we have expanded our outreach efforts by having 4 Diversity, Inclusion, and Outreach Co-Chairs (as compared to 2 Diversity and Inclusion Co-Chairs in the prior year).

Continuing tracks: MSR 2022 continues all the tracks of MSR 2021:

- We continue the tradition of the Mining Challenge, where researchers from across the community apply their mining techniques to a common dataset. For the sixth year in a row, we issued an open call for challenge track proposals, which attracted three competitive submissions. From those, one was selected to provide the Mining Challenge for 2022: SmartSHARK, a dataset that combines detailed information from the version control system (commits, code metrics, code clones, PMD warnings, change types, refactorings) with issue tracking data from Jira, pull request data from GitHub and continuous integration data from Travis.
- For the third time, MSR 2022 featured a Registered Reports track, with the goals of (1) eliminating under-powered, selectively reported, or researcher-biased studies, and (2) providing early feedback to authors in their initial study design. With registration, authors can submit an experimental plan, including hypotheses and expected outcomes, and receive peer review feedback before data is collected. Registered reports could obtain in-principle acceptance at Springer’s Empirical Software Engineering Journal or continuity acceptance.
- For the second time, MSR 2022 featured the hackathon track. The goal of the hackathon was to provide the opportunity for participants to work with world-class researchers on relevant problems and research questions that explore problems and solutions in open source software development. Individuals came together virtually to define research questions, form teams, and scope problems. This year’s hackathon is based on GrimoireLab, which is a toolset designed to help retrieve, analyze, and visualize software development data stored in various supporting systems (issue tracking, source code management, etc.). The hackathon is supported by Bitergia, the company leading the development of GrimoireLab, and CHAOSS, the community in which GrimoireLab is maintained.
- For the second time, MSR 2022 featured a Shadow PC track. The goal of the Shadow PC was to train the next generation of program committee (PC) members and to expose early-career researchers (PhD students, postdocs, new faculty members, and industry practitioners) to the review process of the technical track.
- Finally, MSR 2022 will also have 4 exciting tutorials for both newcomers and seasoned MSR attendees.

Awards: As is tradition, a selection of the best technical track and data and tool showcase papers, will be invited to submit an extended version for consideration in a special issue of the Springer Journal on Empirical Software Engineering (EMSE). The invitees will be announced during the opening session of MSR 2022. MSR 2022 will also recognize technical papers with ACM SIGSOFT Distinguished Paper Awards as well as FOSS Awards, which are granted to papers that show outstanding contributions to the FOSS community. MSR also recognizes the best paper from the data and tool showcase track with a Best Data and Tool Showcase Paper Award. Finally, following the practice of the last editions, MSR 2022 will also present a Most Influential Paper award honoring a paper from MSR 2012 that stood the test of time and that has had a large impact on the MSR community. These awards will be announced at MSR 2022.

For the fifth time, the MSR community recognizes outstanding contributions in the MSR field through the establishment of two series of awards: the MSR Ric Holt Early Career Achievement Award, and the MSR Foundational Contribution Award. The MSR Ric Holt Early Career Achievement Award recognizes outstanding junior researchers who provided outstanding contributions in the MSR field. The MSR Foundational Contribution Award recognizes individuals, or groups of individuals, who produced fundamental contributions in the MSR field, where such contributions have helped many others (not limited to the MSR community) to advance the state of the art.

This year's Most Influential Paper award is given to the paper titled "GHTorrent: Github's Data from a Firehose" from MSR 2012 authored by Georgios Gousios, and Diomidis Spinellis. This year's committee also identified a runner-up, which is the paper titled "App Store Mining and Analysis: MSR for App Stores" authored by Mark Harman, Yue Jia, and Yuanyuan Zhang. We would like to congratulate the award winners and thank them for their influential contributions to the MSR community (and beyond)!

Finally, we are also very happy to have Christian Kästner as our keynote speaker this year. Christian is an Associate Professor and the Director of the Software Engineering Ph.D. program at the School of Computer Science at Carnegie Mellon University. In 2019, he started to co-teach a new course "Machine Learning in Production" at the intersection of software engineering and machine learning to better prepare a large number of students who, after graduation, start to work on software systems that integrate more and more machine learning (e.g., mobile apps, web applications, IoT devices). Since then, he also conducted research on collaboration, documentation, and quality assurance in teams where software engineers and data scientists interact. His 2022 keynote, entitled "From Models to Systems: Rethinking the Role of Software Engineering for Machine Learning" will present his experience at the intersection of software engineering and data science. His keynote will also encourage MSR attendees, who have a rare profile with expertise and a deep appreciation of both data science and software engineering, to improve the education and tooling of both software engineers and data scientists.

The diverse MSR 2022 program would not have been possible without the contributions of a large organizational team (a record number of 49 Organizing Committee members this year). We thank the Co-chairs of the Data and Tool Showcase (Chakkrit Tantithamthavorn and Xin Xia), the Co-chairs of the Mining Challenge Track (Steffen Herbold, Alexander Trautsch, and Fabian Trautsch), the co-chairs of the Registered Reports Track (Jin L.C. Guo and Raula Gaikovina Kula), the co-chairs of the Industry Track (Vladimir Kovalenko and Mei Nagappan), the co-chairs of the Hackathon Track (Maëlick Claes, Daniel Izquierdo Cortazar, Jesus M. Gonzalez-Barahona, Gregorio Robles, with support chairs Venu Vardhan

Reddy Tekula and Quan Zhou), the co-chairs of the Shadow PC Track (Eleni Constantinou and Sarah Nadi), and the coordinators of the Shadow PC Mentors (Ayushi Rastogi and Alexander Serebrenik). We also thank all members of the Program Committee as well as external reviewers for the Technical Track (125 PC members this year), and all the aforementioned tracks. We appreciate all the effort put into reviewing and discussing the submitted papers and providing constructive reviews to the authors. A special mention goes to the PC members who served on the rapid response reviewing team who provided emergency reviews to account for unforeseen circumstances (Diomidis Spinellis, Fabio Palomba, Gustavo Pinto, Kevin Moran, Marco Gerosa, Olga Baysal, Tse-Hsun Chen, and Weiyi Shang)!

We also thank the co-chairs of the Tutorials Track (Foutse Khomh and Andy Zaidman) and the co-chairs of the Vision and Reflection Track (Bram Adams and Shaowei Wang) for inviting great speakers for the tracks. The Tutorials Track features four invited tutorials including: “Empirical Standards for Repository Mining” (by Paul Ralph, Tushar Sharma, and Preetha Chatterjee), “Mining the Ethereum Blockchain Platform: Best Practices and Pitfalls” (by Gustavo A. Oliva), “Software Bots in Software Engineering: Benefits and Challenges” (by Mairieli Wessel, Marco Gerosa, Emad Shihab), and a tutorial about extending IntelliJ or Space with MSR-driven features by JetBrains. The Vision and Reflection Track will feature six speakers in two sessions: Vision and Reflection. Alexander Serebrenik, Denae Ford Robinson, and Foutse Khomh will present in the Vision Session, while Audris Mockus, Ahmed E. Hassan, and Daniela Damian will present in the Reflection Session.

In addition, special thanks to the MSR Awards co-chairs Alberto Bacchelli and Miryung Kim and their committee members; the FOSS Award co-chairs Julia Lawall, Fabio Palomba, and Stefano Zacchiroli and their committee members; and Massimiliano Di Penta and Tao Xie, who organized the selection of the Most Influential Paper from MSR 2012.

We are most grateful to the people who worked closely with us in the other organizing aspects of MSR 2022: Ting Zhang and Zhou Yang (Web Co-Chairs); Gemma Catolino, Tapajit Dey, Lingfeng Bao, and Masud Rahman (Publicity and Social Media Co-Chairs); Maxime Lamothe and Zhiyuan Wan (Proceedings Co-Chairs); Iftekhar Ahmed and Bogdan Vasilescu (Hybridization (Local) Co-Chairs); Felipe Ebert and Ferdian Thung (Hybridization (Virtual) Co-Chairs); and Tegawendé F. Bissyandé, Chaiyong Rakhitwetsagul, Yuan Tian, and Gias Uddin (Diversity, Inclusion, and Outreach Co-Chairs).

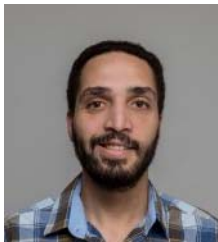
We also thank our sponsors: IEEE Computer Society, Association for Computing Machinery (ACM), Technical Council on Software Engineering (IEEE TCSE), and Special Interest Group on Software Engineering (ACM SIGSOFT). We also thank JetBrains for being our Platinum Sponsor, and our special thanks go to Vladimir Kovalenko who helped us to secure the fund. We also thank Ahmed E. Hassan and Huawei Canada for their support of the widening participation program, which provided funding for conference attendees who would have been unable to join otherwise. We definitely should not forget the ICSE 2022 General Chair Matthew B. Dwyer, and the Co-located Events Co-chairs Ivan Beschastnikh and Christine Julien, for their help with co-locating MSR with ICSE and coordinating arrangements. And of course, we thank the MSR steering committee, and especially its chair, Tom Zimmermann, for their advice and assistance in organizing the conference.

Last, but not definitely least, we want to thank all of the MSR community members who are dedicated to attending and participating in MSR either virtually or in-person (or both)! We thank all the authors for their submissions and the whole MSR community for their continuous contributions to the topic.

We hope you have a great time at MSR 2022 and enjoy the program!



David Lo, Singapore Management University, Singapore
MSR 2022 General Chair



Shane McIntosh, University of Waterloo, Canada
MSR 2022 Program Co-Chair



Nicole Novielli, University of Bari, Italy
MSR 2022 Program Co-Chair

Message from the MSR 2022 Data and Tool Showcase Track Co-Chairs

On behalf of the entire Program Committee, it is our pleasure to welcome you to the Data and Tool Showcase Track of MSR 2022! The goal of this track is to provide a forum for actively promoting and recognizing the creation of reusable datasets and tools that are designed and built not only for a specific research project, but for the MSR community as a whole.

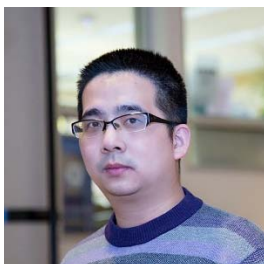
This year, the Data and Tool Showcase Track has attracted the highest record of 49 abstract submissions. Three of which were withdrawn, resulting in 46 paper submissions for review. The papers went through a rigorous review process with at least 3 reviewers. The papers were evaluated and discussed along with detailed evaluation criteria (value, usefulness, reusability, quality, clarity, and availability), according to our published Reviewer Guidelines. We actively asked reviewers to point out the merits of each paper and sought consistency in the application of the review criteria in all aspects of the decision process, as well as in summarizing the decision. We are pleased to report that the program committee has accepted 34 papers for publication and presentation at the conference, leading to an acceptance rate of 74%. Similar to previous years, we do not have a threshold of acceptance rate. This high acceptance rate has to do with the high-quality submissions from the MSR authors and our new review criteria. These papers offer and promote datasets, as well as tools developed to collect them, related to code review, commits, bug fixing, debugging, test cases, open-source libraries and licenses, Android malware, and among other interesting topics.

We are extremely grateful to all the authors who submitted high-quality work and program committee members for their thorough reviews and timely discussions on all the submitted papers. Finally, we would like to thank MSR 2022 GC, David Lo, the PC Co-Chairs, Shane McIntosh and Nicole Novielli, all the members of the organizing committee for supporting the Data/Tool Showcase track in many ways and making MSR 2022 happen by organizing and running a virtual conference this year.

MSR 2022 Data/Tool Showcase Program Co-Chairs



Dr. Chakkrit (Kla) Tantithamthavorn



Dr. Xin Xia

Message from the MSR 2022 Hackathon Track Co-Chairs

Today, software is developed thanks to many supporting systems, which provide help for source code management, code review, issue tracking, synchronous or asynchronous interpersonal communication, continuous integration, and many other tasks. Many of these systems store a wealth of data about how software is being developed, allowing for detailed studies and exploration tools that could be used to better understand software development.

GrimoireLab is a toolset designed to help retrieve, analyze, and visualize such data. It has components for gathering data from about 30 different software development supporting systems, and enriching, analyzing, and visualizing it, including modules for identity management. GrimoireLab is mainly written in Python, available for use as Python packages or with docker-compose. There is also some documentation: a paper, a tutorial, and detailed instructions in each of the source code repositories. To have a glimpse of the kind of analysis and visualizations that GrimoireLab provides out of the box, check the CHAOSS GrimoireLab dashboard (GrimoireLab is a CHAOSS project). A simplified version of the data in any collection of GitHub and GitLab repositories can also be obtained via Cauldron.io, which is powered by GrimoireLab.

Hackathons are effective ways to explore research and product ideas by teaming up with others on intense but limited duration tasks. We proposed a GrimoireLab online hackathon to explore problems and solutions in software development that require data collection from software development repositories.

The GrimoireLab MSR Virtual Hackathon provided activities typical of the in-person hackathon virtually. Participants formed teams, defined their own research questions and worked towards answering them utilizing GrimoireLab during the hackathons. Organizers provided advice on the best ways to address those research questions, conduct data processing and improve performance. During the hackathon the participants had the opportunity to work with and seek advice from world-class researchers who served as mentors and program committee members for the event.

This second hackathon at MSR was attended by 24 individuals who worked alone or in teams of 2 to 6 participants on 8 projects. Responses to this initiative from both participants and the program committee were very positive and encouraging. We would like to thank the MSR steering committee and the MSR general and program chairs for their encouragement and interest in this effort.

Special thanks to Venu Vardhan Reddy Tekula and Quan Zhou from Bitergia for acting as our Hackathon Support co-chairs!

Jesus M. Gonzalez-Barahona, Maelick Claës, Daniel Izquierdo, Gregorio Robles

MSR 2022 Hackathon Co-Chairs

Message from the MSR 2022 Industry Track Co-Chairs

The MSR 2022 Industry Track is the venue to present and learn about the opportunities, challenges, and cutting-edge technology related to using data from software repositories in practice. For a long time, academic researchers in software engineering have been looking to learn and collaborate with practitioners. Our goal for the new Industry Track is to be the space for a productive dialogue between software engineering researchers and practitioners, particularly those building tools for other software professionals.

In this first edition of the Industry Track, we were looking to minimize the known barriers to industry participation in software engineering research conferences. We invited practitioners to submit an abstract, maximum 1 page long, outlining a talk or a poster presentation.

We solicited abstracts in two categories:

- 1) Problems: the talk outlines a single problem in an industrial context that could be addressed by using data from software repositories.
- 2) Solutions: the talk reports on a data-driven tool or technique in practice, based on data from software repositories, that the authors have built or are working on.

We got 14 submissions from which we accepted 12, and out of that, 3 requested for a presentation only and 9 have both a presentation and publication in the proceedings.

Industry Track Co-Chairs



Vladimir Kovalenko, *JetBrains Research, Netherlands*



Mei Nagappan, *University of Waterloo, Canada*

Message from the MSR 2022 Mining Challenge Track Co-Chairs

The Mining Software Repositories (MSR) challenge is a long-standing tradition, dating back to 2006. It is open to all researchers in the field, and frequently participated in by young researchers and motivated students. MSR 2022 has been no exception, with the MSR conference holding the 17th edition of the challenge, which we have been honored to chair.

This year, the challenge was about the SmartSHARK data, a collection of rich and detailed information about the evolution of software projects. The data is unique in its diversity and contains information about each change, including issue tracking data, continuous integration data, as well as pull request and code review data. Moreover, the data does not only contain the raw data scraped from repositories, but also annotations in form of labels determined through a combination of manual analysis and heuristics, as well as links between the different parts of the data set. By now, the data set contains the following data:

- Data collected from Git, e.g., the commit messages, authors, dates, as well as the changed hunks. The clone of the Git repository at the time of collection is also stored to enable further analysis of the source code.
- Data about the source code for each commit focused on Java, e.g., software metrics (size, complexity, documentation, code clones), static analysis warnings from PMD, and the number of nodes of each type in the AST of a file.
- Data about code changes, i.e., the detection of change types with ChangeDistiller, as well as refactorings with RefDiff and RefactoringMiner.
- Data collected from Jira, i.e., the issues, comments, and changes to issues made.
- Data collected from GitHub, i.e., issues, pull requests, and code reviews as part of pull requests.
- Data collected from mailing lists, i.e., all emails from the developer and user mailing lists.
- Links commits and issues, as well as links between commits and pull requests.
- Manually validated links between commits and bug issues, as well as the type of issues labeled as bug for 38 projects.
- Manually validated line labels that mark which changes contributed to a bug fix for 23 projects as well as partial data for five additional projects.
- Annotations for commits and changes, i.e., bug fixing changes including their probable inducing changes, if changes modified Javadocs or inline comments, whether TODOs were added or removed, if test code changed or if we were able to detect refactorings.
- Travis CI build logs and build status information for all projects that use Travis CI.

The data contains only projects from the Apache Software Foundation that have Java as main language. The projects all have between 1,000 and 20,000 commits, i.e., the data does not contain very small or very

large projects. A detailed description of the data sources, collection tools, size, format, and database schema can be found on arXiv.¹

This year, the Mining Challenge Track attracted 13 paper submissions. The papers went through a rigorous review process. Every submission was reviewed by three members of the program committee, and an electronic discussion was held for all papers. Based on the reviews and discussion, the program chairs reached a decision. We are pleased to report the program committee has accepted 7 papers for publication and presentation at the conference.

MSR Mining Challenge 2022 Co-Chairs

Steffen Herbold, *TU Clausthal, Germany*

Alexander Trautsch, *University of Göttingen, Germany*

Fabian Trautsch, *Germany*

A. ¹Trautsch, F. Trautsch, S. Herbold: *MSR Mining Challenge: The SmartSHARK Repository Mining Data*, <https://doi.org/10.48550/arXiv.2102.11540>

Message from the MSR 2022 Registered Reports Track Co-Chairs

Empirical Software Engineering Journal (EMSE), in conjunction with the conference on Mining Software Repositories (MSR), is continuing the Registered Reports (RR) track. The RR track of MSR 2022 has two goals: (1) to prevent [HARKing](#) (hypothesizing after the results are known) for empirical studies; (2) to provide early feedback to authors in their initial study design. For papers submitted to the RR track, methods and proposed analyses are reviewed prior to execution. Pre-registered studies follow a two-step process:

- Stage 1: A report is submitted that describes the planned study. The submitted report is evaluated by the reviewers of the RR track of MSR 2022. Authors of accepted pre-registered studies will be given the opportunity to present their work at MSR.
- Stage 2: Once a report has passed Phase 1, the study will be conducted and actual data collection and analysis take place. The results may also be negative! The full paper is submitted for review to EMSE.

The RR track of MSR 2022 supports two types of papers:

Confirmatory: The researcher has a fixed hypothesis (or several fixed hypotheses) and the objective of the study is to find out whether the hypothesis is supported by the facts/data.

Exploratory: The researcher does not have a hypothesis (or has one that may change during the study). Often, the objective of such a study is to understand what is observed and answer questions such as WHY, HOW, WHAT, WHO, or WHEN. We include in this category registrations for which the researcher has an initial proposed solution for an automated approach (e.g., a new deep-learning-based defect prediction approach) that serves as a starting point for his/her exploration to reach an effective solution.

The outcome of the RR report review is one of the following:

- **In-Principal Acceptance (IPA):** The reviewers agree that the study is relevant. The authors can engage in the actual study for Stage 2. If the protocol is adhered to (or deviations are thoroughly justified), the study is published.
- **Continuity Acceptance (CA):** The reviewers agree that the study is relevant, that the (initial) methods appear to be appropriate. However, for exploratory studies, implementation details and post-experiment analyses or discussion (e.g., why the proposed automated approach does not work) may require follow-up checks.
- **Rejection** The reviewers do not agree on the relevance of the study or are not convinced that the study design is sufficiently mature.

This year, the registered reports track at MSR received 14 submissions, each reviewed by three PC members. Following the format in the previous year at MSR RR track, we allow authors to revise their initial submissions along with the rebuttal of Stage 1. We then invite work recommended by the PC members to proceed with the Stage 2 submissions at EMSE. Of the original 14 submissions, 10 were accepted to proceed with Stage 2 submissions (2 In-Principle Acceptance, and 8 Continuity Acceptance).

We would like to thank authors, PC members, EMSE EiCs, and the MSR steering committee who have contributed and made this edition of the RR Track successful. Our special thanks to prior chairs of MSR

2020, MSR 2021, and ICSME 2021 RR Tracks who have provided valuable inputs and shared their experience.

We look forward to seeing the final results!

Raula Gaikovina Kula and Jin L.C. Guo
MSR 2022 Registered Reports Track Co-Chairs

Message from the MSR 2022 Shadow PC Track Co-Chairs

The Shadow PC track was introduced in MSR 2021. It is a unique opportunity for PhD students and early career researchers to understand the peer review process first hand. MSR 2022 continued with the Shadow PC track while learning from the lessons of the previous year, mainly scaling the load of shadow PC members and mentors. We received 111 shadow PC member applications and accepted almost 50% of applicants. Specifically, we accepted 51 shadow PC members: 33% women, 67% men; 82% PhD students, 6% Postdoc, 10% junior faculty, 2% industry practitioners. There were 51 technical track papers that opted in for review by the shadow PC. We randomly selected 35 of these papers to control the size of the shadow PC track. Each shadow PC member reviewed 2-3 papers. 56 members of the main track program committee volunteered to act as mentors and provide feedback on the reviews and discussion; 27 of them have been asked to coordinate one or two discussions.

We hope that the Shadow PC was a useful experience to all its members. We also want to thank all the mentors who dedicated their time to provide feedback and train the next generation of MSR PC members.

Eleni Constantinou and Sarah Nadi
MSR 2022 Shadow PC Chairs

Ayushi Rastogi and Alexander Serebrenik
Coordinators of Shadow PC Mentors

Message from the MSR 2022 Tutorials Track Co-Chairs

For MSR 2022, we have invited seasoned software repository miners to give 4 exciting tutorials to our broad community, be it newcomers or mining experts. These tutorials will cover topics such as software bots, empirical standards, mining of blockchain platforms, and MSR features in the integrated development environment.

Tutorial Track Co-Chairs

Foutse Khomh, *Polytechnique Montréal, Canada*

Andy Zaidman, *Delft University of Technology, The Netherlands*

MSR 2022

Organizing Committee

General Chair

David Lo, *Singapore Management University, Singapore*

Program Co-chairs

Nicole Novielli, *University of Bari, Italy*

Shane McIntosh, *University of Waterloo, Canada*

Industry Track Co-chairs

Vladimir Kovalenko, *JetBrains Research, Netherlands*

Mei Nagappan, *University of Waterloo, Canada*

Mining Challenge Co-chairs

Steffen Herbold, *TU Clausthal, Germany*

Alexander Trautsch, *University of Göttingen, Germany*

Fabian Trautsch, *Germany*

Tutorials Co-chairs

Foutse Khomh, *Polytechnique Montréal, Canada*

Andy Zaidman, *Delft University of Technology, Netherlands*

Registered Reports Track Co-chairs

Jin L.C. Guo, *McGill University, Canada*

Raula Gaikovina Kula, *Nara Institute of Science and Technology, Japan*

Data and Tool Showcase Co-chairs

Chakkrit Tantithamthavorn, *Monash University, Australia*

Xin Xia, *Huawei Software Engineering Application Technology Lab, China*

Hackathon Track Co-chairs

Maëlick Claes, *University of Oulu, Finland*

Daniel Izquierdo Cortazar, *Bitergia, Spain*

Jesus M. Gonzalez-Barahona, *Universidad Rey Juan Carlos, Spain*

Gregorio Robles, *Universidad Rey Juan Carlos, Spain*

Hackathon Track Support Co-chairs

Venu Vardhan Reddy Tekula, *CHAOSS Community, India*

Quan Zhou, *Bitergia, Spain*

Vision and Reflection Track Co-chairs

Bram Adams, *Queen's University, Canada*
Shaowei Wang, *University of Manitoba, Canada*

Proceedings Co-chairs

Maxime Lamothe, *Polytechnique Montréal, Canada*
Zhiyuan Wan, *Zhejiang University, China*

Publicity and Social Media Co-chairs

Gemma Catolino, *Tilburg University & Jheronimus Academy of Data Science, Netherlands*
Tapajit Dey, *Lero – The Irish Software Research Centre and University of Limerick, Ireland*
Lingfeng Bao, *Zhejiang University, China*
Masud Rahman, *Dalhousie University, Canada*

Web Co-chair

Ting Zhang, *Singapore Management University, Singapore*
Zhou Yang, *Singapore Management University, Singapore*

Shadow PC Mentor Coordinators

Ayushi Rastogi, *University of Groningen, Netherlands*
Alexander Serebrenik, *Eindhoven University of Technology, Netherlands*

Shadow PC Co-chairs

Eleni Constantinou, *Eindhoven University of Technology, Netherlands*
Sarah Nadi, *University of Alberta, Canada*

Diversity, Inclusion, and Outreach Co-chairs

Tegawendé F. Bissyandé, *SnT, University of Luxembourg, Luxembourg*
Chaiyong Rakhitwetsagul, *Mahidol University, Thailand*
Yuan Tian, *Queen's University, Canada*
Gias Uddin, *University of Calgary, Canada*

MSR Awards Co-chairs

Alberto Bacchelli, *University of Zurich, Switzerland*
Miryung Kim, *University of California at Los Angeles, USA*

MIP Award Co-chairs

Massimiliano Di Penta, *University of Sannio, Italy*
Tao Xie, *Peking University, China*

FOSS Award Co-chairs

Julia Lawall, *Inria, France*
Fabio Palomba, *University of Salerno, Italy*
Stefano Zacchiroli, *Télécom Paris, Polytechnic Institute of Paris, France*

Hybridization (Local) Co-chairs

Iftekhhar Ahmed, *University of California at Irvine, USA*

Bogdan Vasilescu, *Carnegie Mellon University, USA*

Hybridization (Virtual) Co-chairs

Felipe Ebert, *Jheronimus Academy of Data Science, Netherlands*

Ferdian Thung, *Singapore Management University, Singapore*

MSR 2022

Program Committee

MSR Technical Papers

Program Co-Chairs

Shane McIntosh, *University of Waterloo, Canada*

Nicole Novielli, *University of Bari, Italy*

Program Committee Members

Abbas Heydarnoori, *Sharif University of Technology, Iran*

Ajay Kumar Jha, *University of Alberta, Canada*

Akond Rahman, *Tennessee Tech University, United States of America*

Alberto Bacchelli, *University of Zurich, Switzerland*

Alessio Ferrari, *ISTI-CNR, Italy*

Alexander Serebrenik, *Eindhoven University of Technology, Netherlands*

Amjed Tahir, *Massey University, Palmerston North, New Zealand*

Anand Ashok Sawant, *Siemens Corporate Technology, United States of America*

Andre Hora, *UFMG, Brazil*

Andrea Mocci, *Software Institute, Università della Svizzera italiana, Switzerland*

Andreea Vescan, *Babes-Bolyai University, Romania*

Andy Meneely, *Rochester Institute of Technology, United States of America*

Annie Ying, *Cisco AI Lab, Canada*

Aurora Ramírez, *University of Córdoba, Spain*

Barbara Russo, *University of Bolzano, Italy*

Bernhard Berger, *Hamburg University of Technology, Germany*

Bin Lin, *Università della Svizzera italiana (USI), Switzerland*

Bogdan Vasilescu, *Carnegie Mellon University, United States of America*

Breno Miranda, *Federal University of Pernambuco, Brazil*

Chetan Bansal, *Microsoft Research, United States of America*

Christian Bird, *Microsoft Research (MSR), United States of America*

Christoph Treude, *University of Melbourne, Australia*

Chunyang Chen, *Monash University, Australia*

Cor-Paul Bezemer, *University of Alberta, Canada*

Daniel Alencar da Costa, *University of Otago, New Zealand*

Danielle Gonzalez, *Microsoft, United States of America*

Dario Di Nucci, *University of Salerno, Italy*

Diego Elias Costa, *Concordia University, Canada*

Diomidis Spinellis, *Athens University of Economics and Business & Delft University of Technology, Greece*

Djamel Eddine Khelladi, *"CNRS, IRISA", France*

DongGyun Han, *Singapore Management University, Singapore*
Earl Barr, *University College London, United Kingdom*
Ewan Tempero, *University of Auckland, New Zealand*
Fabio Palomba, *University of Salerno, Italy*
Felipe Ebert, *Eindhoven University of Technology, Netherlands*
Fernando Castor, *Utrecht University, Netherlands*
Fiorella Zampetti, *University of Sannio, Italy*
Foyzul Hassan, *University of Michigan - Dearborn, United States of America*
Francesca Arcelli Fontana, *University of Milano Bicocca, Italy*
Gabriele Bavota, *Software Institute @ Università della Svizzera Italiana, Switzerland*
Gema Rodríguez-Pérez, *University of British Columbia, United States of America*
Gemma Catolino, *PostDoc at Tilburg University - Jheronimus Academy of Data Science, Netherlands*
Georgia Kapitsaki, *University of Cyprus, Cyprus*
Gias Uddin, *University of Calgary, Canada*
Gregorio Robles, *Universidad Rey Juan Carlos, Spain*
Gustavo Pinto, *Federal University of Pará, Brazil, Brazil*
Helge Pfeiffer, *IT University of Copenhagen, Denmark*
Hongyu Zhang, *The University of Newcastle, Australia*
Hourieh Khalajzadeh, *Monash University, Australia*
Igor Steinmacher, *Universidade Tecnológica Federal do Paraná, Brazil*
Iman Keivanloo, *Amazon, United States of America*
Ivan Machado, *Federal University of Bahia, Brazil*
Jason Tsay, *IBM Research, United States of America*
Jean Melo, *Dixa,*
Jesus M. Gonzalez-Barahona, *Rey Juan Carlos University, Spain*
Jinfu Chen, *Huawei,*
Jinqiu Yang, *Concordia University, Canada*
Jürgen Cito, *TU Wien and Facebook, Austria*
Katja Kevic, *Microsoft,*
Kevin Moran, *George Mason University, United States of America*
Kla Tantithamthavorn, *Monash University, Australia*
Kostadin Damevski, *Virginia Commonwealth University, United States of America*
Leandro Minku, *University of Birmingham, United Kingdom*
Li Li, *Monash University, Australia*
Luca Pascarella, *Università della Svizzera italiana, Switzerland*
Mahmoud Alfadel, *The University of Waterloo, Canada*
Marcelo Maia, *Federal University of Uberlandia, Brazil*
Marco Tulio Valente, *UFMG, Brazil*
Marco Gerosa, *Northern Arizona University, United States of America*
Maria Teresa Baldassarre, *University of Bari, Italy*
Maria Kechagia, *University College London, United Kingdom*
Maria Papoutsoglou, *Department of Computer Science - University of Cyprus, Greece*
Martin Pinzger, *University of Klagenfurt, Austria*
Massimiliano (Max) Di Penta, *University of Sannio, Italy*

Masud Rahman, *Dalhousie University, Canada*
Maxime Lamothe, *Polytechnique Montreal, Canada*
Md Tajmilur Rahman, *Gannon University, United States of America*
Melina Vidoni, *Australian National University, Australia*
Michel Albonico, *Federal University of Technology, Paraná (UTFPR), Brazil*
Mika Mäntylä, *University of Oulu, Finland*
Mohammed SAYAGH, ETS, *Quebec University, Canada*
Moritz Beller, *Facebook, Inc., Netherlands*
Mukul Prasad, *Fujitsu Research of America, United States of America*
Olga Baysal, *Carleton University, Canada*
Pasquale Salza, *University of Zurich, Switzerland*
Phuong Nguyen, *University of L'Aquila, Italy*
Preetha Chatterjee, *Drexel University, United States of America*
Rabe Abdalkareem, *Carleton University, Canada*
Ripon Saha, *Fujitsu Research of America, Inc., United States of America*
Robert Dyer, *University of Nebraska-Lincoln, United States of America*
Sarra Habchi, *University of Luxembourg, Luxembourg*
Sebastiano Panichella, *Zurich University of Applied Science (ZHAW), Switzerland*
Serge Demeyer, *University of Antwerp, Belgium*
Shaohua Wang, *New Jersey Institute of Technology, United States of America*
Shaowei Wang, *Department of Computer Science, University of Manitoba, Canada*
Shayan Akbar, *Amazon, United States of America*
Shinpei Hayashi, *Tokyo Institute of Technology, Japan*
Song Wang, *York University, Canada*
Stefano Zacchiroli, *LTCI, Télécom Paris, Institut Polytechnique de Paris, France*
Stephan Diehl, *University of Trier, Germany*
Tao Chen, *Loughborough University, United Kingdom*
Tao Zhang, *Macau University of Science and Technology, China*
Tapajit Dey, *Lero, University of Limerick, Ireland*
Thomas Durieux, *KTH, Sweden*
Tianyi Zhang, *Purdue University, United States of America*
Tien Nguyen, *University of Texas at Dallas, United States of America*
Timofey Bryksin, *JetBrains Research, Russia*
Toni Mattis, *Hasso Plattner Institute, University of Potsdam, Germany*
Triet Le, *The University of Adelaide, Australia*
Tse-Hsun (Peter) Chen, *Concordia University, Canada*
Valerio Cosentino, *Bonhams, Spain*
Vladimir Kovalenko, *JetBrains Research, Netherlands*
Wei Yang, *University of Texas at Dallas, United States of America*
Weiyi Shang, *Concordia University, Canada*
Wellington de Oliveira Júnior, *University of Lisbon, Brazil*
Wesley K. G. Assunção, *Johannes Kepler University Linz, Austria*
Xiaoyuan Xie, *School of Computer Science, Wuhan University, China*
Xin Xia, *Huawei, China*

Yang Liu, *Nanyang Technological University, Singapore*
Yaroslav Golubev, *JetBrains Research, Russia*
Ying Zou, *Queen's University, Canada*
Yoshiki Higo, *Osaka University, Japan*
Yves Le Traon, *University of Luxembourg, Luxembourg*
Zhe Yu, *Rochester Institute of Technology, United States of America*
Zsuzsanna Onet-Marian, *Babes-Bolyai University, Romania*

MSR Industry Track

Industry Track Co-chairs

Vladimir Kovalenko, *JetBrains Research, Netherlands*
Mei Nagappan, *University of Waterloo, Canada*

MSR Mining Challenge

Mining Challenge Co-Chairs

Steffen Herbold, *TU Clausthal, Germany*
Alexander Trautsch, *University of Göttingen, Germany*
Fabian Trautsch, *Germany*

Committee Members

Willian Oizumi, *PUC-Rio, Brazil*
Roberto Verdecchia, *Vrije Universiteit Amsterdam, Netherlands*
Anthony Peruma, *Rochester Institute of Technology, USA*
Valentina Lenarduzzi, *University of Oulu, Finland*
Ahmad Abdellatif, *Concordia University, Canada*
Ivan Pashchenko, *University of Trento, Italy*
Shangwen Wang, *National University of Defense Technology, China*
Maireli Wessel, *Delft University of Technology, Netherlands*
Huascar Sanchez, *SRI International, USA*
Alireza Aghamohammadi, *Sharif University of Technology, Iran*

Tutorials Co-chairs

Foutse Khomh, *Polytechnique Montréal, Canada*
Andy Zaidman, *Delft University of Technology, Netherlands*

MSR Registered Reports

Registered Reports Co-chairs

Raula Gaikovina Kula, *Nara Institute of Science and Technology, Japan*

Jin L.C. Guo, *McGill University, Canada*

Committee Members

Ali Ouni, *Ecole de Technologie Supérieure (ETS), Canada*

Chaiyong Ragkhitwetsagul, *Mahidol University, Thailand*

Christoph Treude, *University of Melbourne, Australia*

Dario Di Nucci, *University of Salerno, Italy*

Eunjong Choi, *Kyoto Institute of Technology, Japan*

Filipe Cogo, *Centre for Software Excellence - Huawei Canada, Canada*

Foutse Khomh, *Polytechnique Montreal, Canada*

Gemma Catolino, *Tilburg University - Jheronimus Academy of Data Science, Netherlands*

Kevin Moran, *George Mason University, United States of America*

Lili Wei, *The Hong Kong University of Science and Technology*

Lu Xiao, *Stevens Institute of Technology, United States of America*

Masanari Kondo, *Kyushu University, Japan*

Michael Vierhauser, *Johannes Kepler University, Austria*

Minghui Zhou, *Peking University, China*

Nikolaos Tsantalis, *Concordia University, Canada*

Patanamon (Pick) Thongtanunam, *The University of Melbourne, Australia*

Tse-Hsun (Peter) Chen, *Concordia University, Canada*

Yiling Lou, *Purdue University, United States of America*

Yusuf Sulisty Nugroho, *Universitas Muhammadiyah Surakarta, Indonesia*

Data and Tool Showcase

Data and Tool Showcase Co-Chairs

Chakkrit (Kla) Tantithamthavorn, *Monash University, Australia*

Xin Xia, *Huawei, China*

Committee Members

Amiangshu Bosu, *Wayne State University, USA*

Bach Le, *University of Melbourne, Australia*

Bodin Chinthanet, *Nara Institute of Science and Technology, Japan*

Bowen Xu, *Singapore Management University, Singapore*

Chaiyong Ragkhitwetsagul, *Mahidol University, Thailand*

Chunyang Chen, *Monash University, Australia*

Federica Sarro, *University College London, UK*

Filipe Cogo, *Huawei, Canada*

Gharib Gharibi, *TripleBlind, USA*

Gustavo Oliva, *Queen's University, Canada*

Heng Li, *Polytechnique Montréal, Canada*

Igor Steinmacher, *Universidade Tecnológica Federal do Paraná, Brazil*
Jinfu Chen, *Huawei, Canada*
Jirat Pasuksmit, *University of Melbourne, Australia*
Laksri Wijerathna, *Monash University, Australia*
Luca Pascarella, *Università della Svizzera italiana, Switzerland*
Mansoor Zahedi, *University of Adelaide, Australia*
Masud Rahman, *Dalhousie University, Canada*
Meng Yan, *Chongqing University, China*
Ming Wen, *Huazhong University of Science and Technology, China*
Morakot Choetkiertikul, *Mahidol University, Thailand*
Oscar Chaparro, *College of William & Mary, USA*
Shade Ruangwan, *Fujitsu Limited, Japan*
Song Wang, *York University, Canada*
Tianyi Zhang, *Purdue University, USA*
Timofey Bryksin, *JetBrains Research, Russia*
Triet Le, *University of Adelaide, Australia*
Tse-Hsun (Peter) Chen, *Concordia University, Canada*
Xiao Chen, *Monash University, Australia*
Xiaoxue Ren, *The Chinese University of Hong Kong, China*
Xing Hu, *Zhejiang University, China*
Yutaro Kashiwa, *Kyushu University, Japan*
Yuxia Zhang, *Beijing Institute of Technology, China*
Zhe Yu, *Rochester Institute of Technology, USA*

MSR Hackathon

Hackathon Track Co-chairs

Jesus M. Gonzalez-Barahona, *Universidad Rey Juan Carlos, Spain*
Maelick Claës, *University of Oulu, Finland*
Daniel Izquierdo, *Bitergia, Spain*
Gregorio Robles, *Universidad Rey Juan Carlos, Spain*

Hackathon Track Support Co-chairs

Venu Vardhan Reddy Tekula, *CHAOSS Community, India*
Quan Zhou, *Bitergia, Spain*

Committee Members

Ahmed Zerouali, *Vrije Universiteit Brussel, Belgium*
Andrea Capiluppi, *University of Groningen, The Netherlands*
Carol Alexandru, *University of Zurich, Switzerland*
Chaiyong Ragkhitwetsagul, *Mahidol University, Thailand*
Daniel German, *University of Victoria, Canada*
Emre Sülün, *Bilkent University, Turkey*
Gema Rodríguez-Pérez, *University of British Columbia, Canada*

Igor Wiese, *Federal University of Technology - Paraná, Brazil*
Michel Chaudron, *Eindhoven University of Technology, Netherlands*
Raula Gaikovina Kula, *Nara Institute of Science and Technology, Japan*
Simon Butler, *University of Skövde, Sweden*
Vahid Etemadi, *Shiraz University of Technology, Iran*
Valerio Cosentino, *Bonhams, Spain*

Shadow PC

Shadow PC Co-Chairs

Eleni Constantinou, *Eindhoven University of Technology, Netherlands*
Sarah Nadi, *University of Alberta, Canada*

Coordinators of Shadow PC Mentors

Ayushi Rastogi, *University of Groningen, Netherlands*
Alexander Serebrenik, *Eindhoven University of Technology, Netherlands*

MSR Shadow PC mentors

Shayan Akbar, *Amazon, United States*
Michel Albonico, *Federal University of Technology - Paraná (UTFPR), Brazil*
Mahmoud Alfadel, *University of Waterloo, Canada*
Wesley Assunção, *Johannes Kepler University Linz, Austria*
Earl T. Barr, *University College London, UK*
Moritz Beller, *Facebook, United States*
Fernando Castor, *Utrecht University & Federal University of Pernambuco, Netherlands & Brazil*
Gemma Catolino, *Tilburg University & Jheronimus Academy of Data Science, Netherlands*
Chunyang Chen, *Monash University, Australia*
Jinfu Chen, *Centre for Software Excellence, Huawei, Canada*
Tse-Hsun (Peter) Chen, *Concordia University, Canada*
Jürgen Cito, *TU Wien and Facebook, Austria*
Valerio Cosentino, *Bonhams, Spain*
Diego Costa, *Concordia University, Canada*
Serge Demeyer, *University of Antwerp, Belgium*
Tapajit Dey, *Lero - The Irish Software Research Centre and University of Limerick, Ireland*
Dario Di Nucci, *University of Salerno, Italy*
Thomas Durieux, *KTH, Sweden*
Robert Dyer, *University of Nebraska-Lincoln, United States*
Felipe Ebert, *Jheronimus Academy of Data Science, Netherlands*
Jesus M. Gonzalez-Barahona, *Universidad Rey Juan Carlos, Spain*
Hourieh Khalajzadeh, *Monash University, Australia*
Maxime Lamothe, *Polytechnique Montréal, Canada*
Triet Le Huynh Minh, *The University of Adelaide, Australia*
Bin Lin, *Università della Svizzera italiana (USI), Switzerland*
Ivan Machado, *Federal University of Bahia, Brazil*

Marcelo De Almeida Maia, *Federal University of Uberlandia, Brazil*
Jean Melo, *IT University of Copenhagen / Dixa, Denmark*
Andy Meneely, *Rochester Institute of Technology, United States*
Fabio Palomba, *University of Salerno, Italy*
Maria Papoutsoglou, *Aristotle University of Thessaloniki, Greece*
Luca Pascarella, *Università della Svizzera italiana (USI), Switzerland*
Martin Pinzger, *Alpen-Adria-Universität Klagenfurt, Austria*
Akond Rahman, *Tennessee Tech University, United States*
Masud Rahman, *Dalhousie University, Canada*
Md Tajmilur Rahman, *Gannon University, United States*
Aurora Ramírez, *University of Córdoba, Spain*
Gregorio Robles, *Universidad Rey Juan Carlos, Spain*
Gema Rodríguez-Pérez, *University of British Columbia (UBC), Canada*
Mohammed Sayagh, *Queen's University, Canada*
Ripon Saha, *Fujitsu, United States*
Pasquale Salza, *University of Zurich, Switzerland*
Anand Ashok Sawant, *University of California, Davis, United States*
Weiyi Shang, *Concordia University, Canada*
Gias Uddin, *University of Calgary, Canada*
Marco Tulio Valente, *Federal University of Minas Gerais, Brazil*
Bogdan Vasilescu, *Carnegie Mellon University, United States*
Andreea Vescan, *Babes-Bolyai University, Romania*
Melina Vidoni, *Australian National University, Australia*
Shaowei Wang, *University of Manitoba, Canada*
Song Wang, *York University, Canada*
Wei Yang, *University of Texas at Dallas, United States*
Zhe Yu, *Rochester Institute of Technology, United States*
Hongyu Zhang, *University of Newcastle, Australia*
Tao Zhang, *Macau University of Science and Technology (MUST), China*
Tianyi Zhang, *Purdue University, United States*

MSR Shadow PC members

Maram Assi, *Queen's University, Canada*
Aakash Bansal, *University of Notre Dame, United States*
Camila Mariane Costa Silva, *University of Canterbury, New Zealand*
Roland Croft, *University of Adelaide, Australia*
Léuson Da Silva, *Federal University of Pernambuco, Brazil*
Manuel De Stefano, *University of Salerno, Italy*
Mouna Dhaouadi, *University of Montreal, Canada*
Yangruibo Ding, *Columbia University in the City of New York, United States*
Saikat Dutta, *University of Illinois at Urbana-Champaign, United States*
Zachary Eberhart, *University of Notre Dame, United States*
Fahimeh Ebrahimi, *Louisiana State University, United States*
Madeline Endres, *University of Michigan, United States*

Vahid Etemadi, *Shiraz University of Technology, Iran*
Sarah Fakhoury, *Washington State University, United States*
Samuel Flint, *University of Nebraska - Lincoln, United States*
Armstrong Foundjem, *Queen's university, Canada*
Stefanos Georgiou, *Queen's University, Canada*
Mehdi Golzadeh, *University of Mons, Belgium*
Niranjan Hasabnis, *Intel, United States*
Jingzhu He, *ShanghaiTech University, China*
Max Hort, *University College London, England*
Emanuele Iannone, *University of Salerno, Italy*
Zoe Kotti, *Athens University of Economics and Business, Greece*
Rrezarta Krasniqi, *University of North Texas, United States*
Gunnar Kudrjavets, *University of Groningen, the Netherlands*
Gulsher Laghari, *University of Sindh, Jamshoro, Pakistan*
Michel Maes Bermejo, *Universidad Rey Juan Carlos, Spain*
Rungraj Maipradit, *Nara Institute of Science and Technology, Japan*
Antonio Mastropaolo, *Università della Svizzera Italiana, Switzerland*
André Meyer, *University of Zurich, Switzerland*
Delano Oliveira, *UFPE, Brazil*
Jyoti Prakash, *University of Passau, Germany*
Nikitha Rao, *Carnegie Mellon University, United States*
Shawn Rasheed, *Massey University, New Zealand*
Larissa Rocha, *UFBA, Brazil*
Thaís Rocha, *UFPE - Universidade Federal do Agreste de Pernambuco, Brazil*
Hareem Sahar, *University of Alberta, Canada*
Jordan Samhi, *University of Luxembourg, Luxembourg*
Nima Shiri Harzevili, *York University, Canada*
Md Saeed Siddik, *University of Alberta, Canada*
Yongqiang Tian, *University of Waterloo; Hong Kong University of Science and Technology, Canada; Hong Kong*
Bianca Trinkenreich, *Northern Arizona University, United States*
Rosalia Tufano, *Università della Svizzera Italiana, Switzerland*
Asif Kamal Turzo, *Wayne State University, United States*
Anderson Uchôa, *Federal University of Ceará (UFC), Brazil*
pablo valenzuela, *Universidad de Chile, Chile*
Zehao Wang, *Concordia University, Canada*
Supatsara Wattanakriengkrai, *Nara Institute of Science and Technology, Japan*
Anna-Katharina Wickert, *Technische Universität Darmstadt, Germany*
Zhou Yang, *Singapore Management University, Singapore*

FOSS Award

FOSS Award Co-Chairs

Julia Lawall, *Inria, France*

Fabio Palomba, *University of Salerno, Italy*

Stefano Zacchiroli, *Télécom Paris, Polytechnic Institute of Paris, France*

Committee Members

Sebastian Baltes, *SAP SE & University of Adelaide, Germany*

Kelly Blincoe, *University of Auckland, New Zealand*

Jesus M. Gonzalez-Barahona, *Universidad Rey Juan Carlos, Spain*

Letizia Jaccheri, *Norwegian University of Science and Technology, Norway*

Tom Mens, *University of Mons, Belgium*

Lucas Nussbaum, *University of Lorraine, France*

Allison Randal, *University of Cambridge, USA*

Tushar Sharma, *Dalhousie University, Canada*

Bogdan Vasilescu, *Carnegie Mellon University, USA*

Vadim Zaytsev, *University of Twente, Netherlands*

Thomas Zimmermann, *Microsoft Research, USA*

MSR Awards

MSR Award Co-Chairs

Alberto Bacchelli, *University of Zurich, Switzerland*

Miryung Kim, *University of California, Los Angeles, USA*

Committee Members

Bram Adams, *Queen's University, Canada*

Olga Baysal, *Carleton University, Canada*

Kelly Blincoe, *University of Auckland, New Zealand*

Sarah Nadi, *University of Alberta, Canada*

Bogdan Vasilescu, *Carnegie Mellon University, USA*