

---

## Curriculum Vitae

### **Boris Glavic**

Department of Computer Science  
Stuart Building, Room 206 B  
10 W 31st Street  
Chicago, IL 60616, USA

*Email:* bglavic@iit.edu  
*Homepage:* <http://www.cs.iit.edu/~glavic/>  
*Group Webpage:* <http://www.cs.iit.edu/~dbgroup>  
*Phone:* +1 (312) 567 5205 (work)

---

## Research Interests

Database systems, in particular, data provenance, scalable distributed query processing, and data cleaning and data integration.

## Education

- 2010 **Doktor (Ph.D.)**, summa cum laude, in Computer Science  
University of Zurich, Switzerland  
*Thesis Advisors:* Michael H. Böhlen, Gustavo Alonso  
*Thesis:* Perm: Efficient Provenance Support for Relational Databases
- 2005 **Diploma (M.Sc.)** in Computer Science  
RWTH Aachen, Germany  
*Thesis Advisors:* Ralph Krieger, Ira Assent  
*Thesis:* Subspace Clustering for Sequences of Ordered Categorical Data

## Employment and Appointments

- 2016– **Adjunct Assistant Professor (Status only)**  
University of Toronto, Toronto, Canada
- 2012– **Assistant Professor (Tenure-track)**  
IIT, Chicago, USA
- 2010–2012 **Postdoctoral Fellow**  
University of Toronto, Canada  
*Supervisor:* Renée J. Miller
- 2005–2010 **Ph.D. Student and Research Assistant**  
University of Zurich, Switzerland  
*Supervisors:* Klaus R. Dittrich, Carl-Christian Kanne, Michael H. Böhlen

## Publications

### Proceedings Editor

- [1] Marta Mattoso, **Boris Glavic**. *Provenance and Annotation of Data and Processes - 6th International Provenance and Annotation Workshop, IPAW 2016, McLean, VA, USA, June 7-8, 2016, Proceedings*, Lecture Notes in Computer Science 9672, Springer 2016, ISBN 978-3-319-40592-6

### Guest Editorials

- [1] Mohamed Eltabakh, **Boris Glavic**. *Guest editorial: large-scale data curation and metadata management*, Distributed and Parallel Databases (**DaPD**), Springer 2018, ISBN 0926-8782

### Peer-Reviewed Journals

- [1] Xing Niu, Raghav Kapoor, **Boris Glavic**, Dieter Gawlick, Zhen Hua Liu, Vasudha Krishnaswamy, Venkatesh Radhakrishnan. *Heuristic and Cost-based Optimization for Diverse Provenance Tasks* IEEE Transactions on Knowledge and Data Engineering (**TKDE**), 2018 (to appear).
- [2] Bahareh Arab, Su Feng, **Boris Glavic**, Seokki Lee, Xing Niu, Qitian Zeng. *GProM-A Swiss Army Knife for Your Provenance Needs* In **IEEE Data Engineering Bulletin**, vol. 41, no. 1, pp. 51–62, 2018.
- [3] Bahareh S. Arab, Dieter Gawlick, Vasudha Krishnaswamy, Venkatesh Radhakrishnan, **Boris Glavic**. *Using Reenactment to Retroactively Capture Provenance for Transactions* IEEE Transactions on Knowledge and Data Engineering (**TKDE**), vol. 30, no. 3, pp. 599–612, 2018.
- [4] Xing Niu, Bahareh S. Arab, Seokki Lee, Su Feng, Xun Zou, Dieter Gawlick, Vasudha Krishnaswamy, Zhen Hua Liu, **Boris Glavic**. *Debugging Transactions and Tracking their Provenance with Reenactment*. Proceedings of the VLDB Endowment (**PVLDB**), vol. 10, no. 12, pp. 1857–1860, 2017.
- [5] Patricia C. Arocena, **Boris Glavic**, Giansalvatore Mecca, Renée J. Miller, Paolo Papotti, and Donatello Santoro, *Benchmarking Data Curation Systems*, In **IEEE Data Engineering Bulletin**, vol. 39, no.2, pp. 47–62, 2016.
- [6] Patricia C. Arocena, **Boris Glavic**, Radu Ciucanu, and Renée J. Miller. *The iBench Integration Metadata Generator*. In Proceedings of the VLDB Endowment (**PVLDB**), vol. 9, no. 3, pp. 108–119, 2015.
- [7] Patricia C. Arocena, **Boris Glavic**, Giansalvatore Mecca, and Renée J. Miller, Paolo Papotti, Donatello Santoro. *Messing Up with BART: Error Generation For Evaluating Data-Cleaning Algorithms*. In Proceedings of the VLDB Endowment (**PVLDB**), vol. 9, no. 2, pp. 36–47, 2015.
- [8] Patricia C. Arocena, Radu Ciucanu, **Boris Glavic**, and Renée J. Miller. *Gain Control over your Integration Evaluations*. In Proceedings of the VLDB Endowment (**PVLDB**) (Demonstration Track), vol. 8, no. 12, pp. 1960 - 1971, 2015.
- [9] Quan Pham, Richard Whaling, **Boris Glavic**, and Tanu Malik. *Sharing and reproducing database applications*. Proceedings of the VLDB Endowment (**PVLDB**) (Demonstration Track), vol. 8, no. 12, pp. 1988 - 1999, 2015.
- [10] **Boris Glavic**, Kyumars Sheykh Esmaili, Peter M. Fischer, and Nesime Tatbul. *Efficient Stream Provenance via Operator Instrumentation*. In Transactions on Internet Technology (**TOIT**), vol. 14, no. 1, pp. 7:1-7:26, 2014.
- [11] **Boris Glavic**, Jiang Du, Renée J. Miller, Gustavo Alonso, and Laura M. Haas. *Debugging Data Exchange with Vagabond*. In Proceedings of the Very Large Data Bases Endowment (**PVLDB**) (Demonstration Track), vol. 4, no. 12, pp. 1383–1386, 2011.
- [12] **Boris Glavic**, Gustavo Alonso, Renée J. Miller, and Laura M. Haas. *TRAMP: Understanding the Behavior of Schema Mappings through Provenance*. In Proceedings of the Very Large Data Bases Endowment (**PVLDB**), vol. 3, no. 1, pp. 1314–1325, 2010.

- [13] Ira Assent, Ralph Krieger, **Boris Glavic**, and Thomas Seidl. *Clustering Multidimensional Sequences in Spatial and Temporal Databases*. In International Journal on Knowledge and Information Systems (**KAIS**), vol. 16, no. 1, pp. 29-51, 2008.

## Peer-Reviewed Conferences

- [1] Seokki Lee, Sven Köhler, Bertram Ludäscher, **Boris Glavic**, *A SQL-Middleware Unifying Why and Why-Not Provenance for First-Order Queries*, In Proceedings of the 33rd IEEE International Conference on Data Engineering (**ICDE**), 2017
- [2] Xing Niu, Raghav Kapoor, **Boris Glavic**, Dieter Gawlick, Zhen Hua Liu, Vasudha Krishnaswamy, Venkatesh Radhakrishnan, *Provenance-aware Query Optimization*, In Proceedings of the 33rd IEEE International Conference on Data Engineering (**ICDE**), 2017
- [3] Jiang Du, **Boris Glavic**, Wei Tan, Renée J. Miller, *DeepSea: Adaptive Workload-Aware Partitioning of Materialized Views in Scalable Data Analytics*, In Proceedings of the 20th International Conference on Extending Database Technology (**EDBT**), pp. 198-209, 2017
- [4] William Spoth, Bahareh Arab, Eric S. Chan, Dieter Gawlick, Adel Ghoneimy, **Boris Glavic**, Beda Hammerschmidt, Oliver Kennedy, Seokki Lee, Zhen Hua Liu, Xing Niu, Ying Yang, *Adaptive Schema Databases*, In Proceedings of the 8th Biennial Conference on Innovative Data Systems (**CIDR**), 2017
- [5] James Wagner, Alexander Rasin, **Boris Glavic**, Karen Heart, Jacob Furst, Lucas Bressan and Jonathan Grier. *Carving Database Storage to Detect and Trace Security Breaches*, Digital Forensics Research Conference (**DFRWS**), 2017
- [6] Bahareh Arab, Dieter Gawlick, Vasudha Krishnaswamy, Venkatesh Radhakrishnan, and **Boris Glavic**, *Reenactment for Read-Committed Snapshot Isolation*, In Proceedings of the 25th International Conference on Information and Knowledge Management (**CIKM**), pp. 841-850, 2016
- [7] Donatello Santoro, Patricia C. Arocena, **Boris Glavic**, Giansalvatore Mecca, Renée J. Miller, and Paolo Papotti, *BART in Action: Error Generation and Empirical Evaluations of Data-Cleaning Systems*, In Proceedings of the 42nd International Conference on Management of Data (**SIGMOD**) (Demonstration Track), 2016
- [8] Quan Pham, Tanu Malik, **Boris Glavic**, and Ian Foster. *LDV: Light-weight Database Virtualization*, In Proceedings of the 31st IEEE International Conference on Data Engineering (**ICDE**), pp. 1179-1190, 2015
- [9] Patricia C. Arocena, **Boris Glavic**, and Renée J. Miller. *Value Invention for Data Exchange*. In Proceedings of the 39th International Conference on Management of Data (**SIGMOD**), pp. 157-168, 2013
- [10] **Boris Glavic**, Kyumars Sheykh Esmaili, Peter M. Fischer, and Nesime Tatbul. *Ariadne: Managing Fine-Grained Provenance on Data Streams*. In Proceedings of the 7th ACM International Conference on Distributed Event-Based Systems (**DEBS**), pp. 39-50, 2013
- [11] Christian Tilgner, **Boris Glavic**, Michael H. Böhlen, and Carl-Christian Kanne. *Declarative Serializable Snapshot Isolation*. In Proceedings of the 15th International Conference on Advances in Database and Information Systems (**ADBIS**), pp. 170-184, 2011
- [12] Christian Tilgner, **Boris Glavic**, Michael H. Böhlen, and Carl-Christian Kanne. *Smile: Enabling Easy and Fast Development of Domain-Specific Scheduling Protocols*. In Proceedings of the 28th British National Conference on Databases (**BNCOD**), pp. 128-131, 2011
- [13] **Boris Glavic** and Gustavo Alonso. *The Perm Provenance Management System in Action*. In Proceedings of the 35th ACM SIGMOD International Conference on Management of Data (**SIGMOD**) (Demonstration Track), pp. 1055-1058, 2009
- [14] **Boris Glavic** and Gustavo Alonso. *Perm: Processing Provenance and Data on the same Data Model through Query Rewriting*. In Proceedings of the 25th IEEE International Conference on Data Engineering (**ICDE**), pp. 174-185, 2009
- [15] **Boris Glavic** and Gustavo Alonso. *Provenance for Nested Subqueries*. In Proceedings of the 12th International Conference on Extending Database Technology (**EDBT**), pp. 982-993, 2009

- [16] **Boris Glavic** and Klaus R. Dittrich. *Data Provenance: A Categorization of Existing Approaches*. In Proceedings of the 12th GI Conference on Database Systems for Business, Technology, and Web (**BTW**), pp. 227-241, 2007
- [17] **Boris Glavic**. *Subspace Sequence Clustering - Datamining zur Entscheidungsunterstützung in der Hydrologie*. In Proceedings of the 11th GI Conference on Database Systems for Business, Technology, and Web (**BTW**) (Student Track), pp. 15-17, 2005

## Book Chapters

- [1] **Boris Glavic**, Gustavo Alonso, and Renée J. Miller. *Using SQL for Efficient Generation and Querying of Provenance Information*. In search of elegance in the theory and practice of computation: a Festschrift in honour of Peter Buneman, Lecture Notes in Computer Science, pp. 291-320, 2013.

## Workshops

- [1] Bahareh S. Arab, **Boris Glavic**. *Answering Historical What-if Queries with Provenance, Reenactment, and Symbolic Execution* In Proceedings of the 8th USENIX Workshop on the Theory and Practice of Provenance (**TaPP**), 2017
- [2] Seokki Lee, Xing Niu, Bertram Ludäscher, **Boris Glavic**. *Integrating Approximate Summarization with Provenance Capture*. In Proceedings of the 8th USENIX Workshop on the Theory and Practice of Provenance (**TaPP**), 2017
- [3] Juliana Freire, **Boris Glavic**, Oliver Kennedy, and Heiko Müller. *The Exception that Improves the Rule* In SIGMOD Workshop on Human-In-the-Loop Data Analytics (**HILDA**), 2016
- [4] Xing Niu, Bahareh Arab, Dieter Gawlick, Zhen Hua Liu, Vasudha Krishnaswamy, Oliver Kennedy, and **Boris Glavic**, *Provenance-aware Versioned Dataworkspaces* In Proceedings of the 8th USENIX Workshop on the Theory and Practice of Provenance (**TaPP**), 2016
- [5] Xing Niu, Raghav Kapoor, and **Boris Glavic**. *Heuristic and Cost-based Optimization for Provenance Computation*. In Proceedings of the 7th USENIX Workshop on the Theory and Practice of Provenance (**TaPP**) (Poster), 2015.
- [6] Xing Niu, Raghav Kapoor, Dieter Gawlick, Zhen Hua Liu, Vasudha Krishnaswamy, Venkatesh Radhakrishnan, and **Boris Glavic**. *Interoperability for provenance-aware databases using prov and json*. In Proceedings of the 7th USENIX Workshop on the Theory and Practice of Provenance (**TaPP**), 2015.
- [7] **Boris Glavic**, Tanu Malik, and Quan Pham. *Making database applications shareable*. In Proceedings of the 7th USENIX Workshop on the Theory and Practice of Provenance (**TaPP**) (Poster), 2015.
- [8] **Boris Glavic**, Sven Köhler, Sean Riddle, and Bertram Ludäscher. *Towards constraint-based explanations for answers and non-answers*. In Proceedings of the 7th USENIX Workshop on the Theory and Practice of Provenance (**TaPP**), 2015.
- [9] Bahareh Arab, Dieter Gawlick, Venkatesh Radhakrishnan, Hao Guo, **Boris Glavic**. *A Generic Provenance Middleware for Queries, Updates, and Transactions*. In Proceedings of the 6th USENIX Workshop on the Theory and Practice of Provenance (**TaPP**), 2014.
- [10] **Boris Glavic**. Invited talk. *Big Data Provenance: Challenges and Potential Implications for Benchmarking*. In Proceedings of the 2nd Workshop on Big Data Benchmarking (**WBDB**), 2013.
- [11] **Boris Glavic**. *Provenance for Data Mining*. In Proceedings of the 5th USENIX Workshop on the Theory and Practice of Provenance (**TaPP**), 2013.
- [12] **Boris Glavic** and Renée J. Miller. *Reexamining Some Holy Grails of Data Provenance*. In Proceedings of the 3rd USENIX Workshop on the Theory and Practice of Provenance (**TaPP**), 2011.
- [13] **Boris Glavic**, Kyumars Sheykh Esmaili, Peter M. Fischer, and Nesime Tatbul. *The Case for Fine-Grained Stream Provenance*. In Proceedings of the 1st Workshop on Data Streams and Event Processing (**DSEP**) collocated with BTW, pp. 58-61, 2011
- [14] Ira Assent, Ralph Krieger, **Boris Glavic**, and Thomas Seidl. *Spatial Multidimensional Sequence Clustering*. In Proceedings of the 1st International Workshop on Spatial and Spatio-temporal Data Mining (**SSTDM**) collocated with ICDM, pp. 343-348, 2006.

- [15] **Boris Glavic** and Klaus Dittrich. *Sesam: Ensuring Privacy for an Interdisciplinary Longitudinal Study*. In Proceedings of the 1st Workshop Elektronische Datentreuhänderschaft - Anwendungen, Verfahren, Grundlagen collocated with GI Jahrestagung, pp. 736-743, 2006.

## Invited Talks

- 2017 *Reenacting Transactions for Provenance, Debugging, and World Peace*  
DB Seminar, **Brown University**
- 2016 *HRDBMS: How to scale like an elephant and perform like a fruit*  
**Midwest Big Data Opportunities and Challenges (MBDOC) Workshop**, sponsored by NSF Midwest Big Data Hub  
*Towards Rigorous Evaluation of Data Integration Systems - It's All About the Tools*  
**QDB workshop (co-located with VLDB 2016)**  
*Reenacting Transactions to Compute their Provenance*  
UBDB Seminar, **SUNY Buffalo**  
*Answering Why- and Why-not Provenance Questions Efficiently and Comprehensibly.*  
CMD Research Colloquium, **DePaul University**  
*Reenacting Transactions to Compute their Provenance.*  
The Yahoo!-DAIS Seminar, **University of Illinois at Urbana Champaign**
- 2015 *Reenacting Transactional Histories to Compute Their Provenance.*  
DIS Seminar Series, **Free University of Bozen, Italy**  
*Big Data Benchmarking needs Big Metadata Generation.*  
Workshop on Big Data Benchmarking (**WBDB**)  
*Reenacting Transactions to Compute their Provenance.*  
Database Seminar @ NCSA, **University of Illinois at Urbana Champaign**  
*Vagabond: Automatic Generation and Ranking of Explanations for Data Exchange Errors.*  
CMD Research Colloquium, **DePaul University**
- 2014 *Vagabond: Automatic Generation and Ranking of Explanations for Data Exchange Errors.*  
Workshop on Best Practices in Empirical Evaluation of Big Data Integration and Cleaning  
*Computing Provenance for Database Updates and Transactions.*  
CMD Research Colloquium, **DePaul University**  
*Computing Provenance for Database Updates and Transactions.*  
**IIT** Data Science Seminar
- 2012 *Big Data Provenance: Challenges and Implications for Benchmarking.*  
Workshop on Big Data Benchmarking (**WBDB**)
- 2010 *Integrating Data Provenance Support in Database Systems.*  
DIS Seminar Series, **Free University of Bozen, Italy**

## Technical Reports

- [1] Su Feng, Aaron Huber, **Boris Glavic**, Oliver Kennedy. *Uncertainty annotated databases - a lightweight approach for dealing with uncertainty*. Technical report IIT/CS-DB-2018-01, Illinois Institute of Technology, 2018.
- [2] Seokki Lee, Sven Köhler, Bertram Ludäscher, **Boris Glavic**. *Efficiently computing provenance graphs for queries with negation*. Technical report IIT/CS-DB-2016-03, Illinois Institute of Technology, 2016.
- [3] Xing Niu **Boris Glavic**. *Optimizing provenance capture and queries - algebraic transformations and cost-based optimization*. Technical report IIT/CS-DB-2016-02, Illinois Institute of Technology, 2016.
- [4] Bahareh Arab, Dieter Gawlick, Vasudha Krishnaswamy, Venkatesh Radhakrishnan, **Boris Glavic**. *Reenactment for read-committed snapshot isolation (long version)*. Technical report, Illinois Institute of Technology, 2016.
- [5] Arindam Nandi, Ying Yang, Oliver Kennedy, **Boris Glavic**, Ronny Fehling, Zhen Hua Liu, and Dieter Gawlick. *Mimir: Bringing CTables into Practice*, Technical report, CoRR, 2016.
- [6] Bahareh Arab, Dieter Gawlick, Vasudha Krishnaswamy, Venkatesh Radhakrishnan, and **Boris Glavic**. *Formal Foundations of Reenactment and Transaction Provenance*, Technical report, IIT, 2016.
- [7] Seokki Lee, Yuchen Tang, Sven Köhler, Bertram Ludäscher, and **Boris Glavic**. *An Efficient Implementation Of Game Provenance In DBMS*, Technical Report, IIT, 2015.
- [8] Patricia C. Arocena, **Boris Glavic**, Gianni Mecca, Renée Miller, Paolo Papotti, and Donatello Santoro. *Error Generation for Evaluating Data Cleaning Algorithms*. Technical report TR-01-2015, Università della Basilicata, 2015.
- [9] Patricia C. Arocena, **Boris Glavic**, Radu Ciucanu, and Renée J. Miller. *The iBench Integration Metadata Generator*. Technical report, University of Toronto, 2015.
- [10] Seokki Lee, Zhen Wang, **Boris Glavic**, and Renée J. Miller. *Automatic Generation and Ranking of Explanations for Mapping Errors*. Technical report, IIT, 2015.
- [11] Quan Pham, Tanu Malik, **Boris Glavic**, Ian Foster. *LDV: Light-weight Database Virtualization*. Technical Report, IIT, 2014.
- [12] Bahareh Arab, Dieter Gawlick, Vasudha Krishnaswamy, Venkatesh Radhakrishnan, **Boris Glavic**. *Reenacting Transactions to Compute their Provenance*. Technical Report, IIT, 2014.
- [13] **Boris Glavic**. *A Primer on Database Provenance*. Technical Report, IIT, 2014.
- [14] Patricia C. Arocena, Mariana D'Angelo, **Boris Glavic**, Renée J. Miller. *iBench First Cut*. Technical Report, University of Toronto, 2013.
- [15] Javed Siddique, **Boris Glavic**, Renée J. Miller. *Provenance Management for Frequent Itemsets*. Technical report, University of Toronto, 2013.
- [16] **Boris Glavic**, Kyumars Sheykh Esmaili, Peter M. Fischer, Nesime Tatbul. *Ariadne: Managing Fine-Grained Provenance on Data Streams*. Technical report, ETH Zurich, 2012.
- [17] **Boris Glavic**. *Formal Foundation of Contribution Semantics and Provenance Computation through Query Rewrite in TRAMP*. Technical report, University of Zurich, 2010.
- [18] Christian Tilgner, **Boris Glavic**, Michael H. Böhlen, and Carl-Christian Kanne. *Correctness Proof of the Declarative SS2PL Protocol Implementation*. Technical report, University of Zurich, 2010.

## Extended Abstracts and Posters

- [1] Xing Niu, and **Boris Glavic**, *Debugging Transactions and Tracking their Provenance with Reenactment*, 6th Greater Chicago Area System Research Workshop (**GCASR**) - Student Poster, 2017
- [2] Seokki Lee, Sven Köhler, Bertram Ludäscher, and **Boris Glavic**, *A SQL-Middleware Unifying Why and Why-Not Provenance for First-Order Queries*, 6th Greater Chicago Area System Research Workshop (**GCASR**) - Student Poster, 2017
- [3] Xing Niu, Raghav Kapoor and **Boris Glavic**, *Heuristic and Cost-based Optimization for Provenance Computation*, 5th Greater Chicago Area System Research Workshop (**GCASR**) - Student Poster, 2016

- [4] Seokki Lee, Sven Köhler, Bertram Ludäscher, and **Boris Glavic**, *Implementing Unified Why- and Why-Not Provenance Through Games*, In Proceedings of the 8th USENIX Workshop on the Theory and Practice of Provenance (**TaPP**) (Poster), 2016
- [5] Jason Arnold, **Boris Glavic**, Ioan Raicu. *HRDBMS: A NewSQL Database for Analytics*, In Proceedings of the IEEE International Conference on Cluster Computing (**Cluster**) (Poster), 2015
- [6] Richard Whaling. *LDV – Lightweight Database Virtualization for Scientific Reproducibility*. 4th Greater Chicago Area System Research Workshop (**GCASR**) - Student Poster, 2015
- [7] Seokki Lee. *Automatic Generation and Ranking of Explanations for Mapping Errors*. 4th Greater Chicago Area System Research Workshop (**GCASR**) - Student Poster, 2015
- [8] Xing Niu, Raghav Kapoor. *Heuristic and Cost-based Optimization for Provenance Computation*. 4th Greater Chicago Area System Research Workshop (**GCASR**) - Student Poster, 2015
- [9] Bahareh Arab, Hao Guo. *Provenance for Database Transactions*. 3rd Greater Chicago Area System Research Workshop (**GCASR**) - Student Poster, 2014
- [10] **Boris Glavic**. *The (Potential) Perks of Integrating Provenance Support into Database Engines*, 2nd Greater Chicago Area System Research Workshop (**GCASR**), 2013

<b>Funding</b>
----------------

2018-2019	\$95,248	PI	Unrestricted Gift from the <b>Oracle</b> Cooperation <i>Exploiting Provenance to Enhance Data Management</i>
2017-2018	\$24,000 (\$8,000)	PI at IIT (collaborative)	<b>NSF</b> REU Supplement for <i>CIF21 DIBBs: EI: Vizier, Streamlined Data Curation</i>
2017-2020	\$2,725,699 (\$404,979)	PI at IIT	<b>NSF</b> <i>CIF21 DIBBs: EI: Vizier, Streamlined Data Curation</i>
2016-2017	\$90,871	PI	Unrestricted Gift from the <b>Oracle</b> Cooperation <i>Provenance for Big Data</i>
2014-2015	\$95,829	PI	Unrestricted Gift from the <b>Oracle</b> Cooperation <i>Provenance using temporal databases (extension)</i>
2013-2014	\$85,000	PI	Unrestricted Gift from the <b>Oracle</b> Cooperation <i>Provenance using temporal databases</i>
2013-2013	\$25,000	PI	<b>IIT</b> Educational and Research Initiative Fund (ERIF) Grant <i>An interdisciplinary approach for assessing treatment motivation among patients undergoing antiretroviral therapy: Integrating multidimensional scaling with data provenance techniques</i>



## Professional Service

### Program Chair

2016 International Provenance and Annotation Workshop (**IPAW**)

### Associate Editor

2018 - current IEEE Transactions on Knowledge and Data Engineering (**TKDE**)

### Guest Editor

2016 Distributed and Parallel Databases (**DAPD**) - Special issue on Large-Scale Data Curation and Meta-data Management

### Program Committee Member

- 2019 International Conference on Extending Database Technology (**EDBT**)  
ACM SIGMOD International Conference on Management of Data (**SIGMOD**)  
International Conference on Very Large Databases (**VLDB**)  
IEEE International Conference on Data Engineering (**ICDE**)
- 2018 ACM International Conference on Information and Knowledge Management (**CIKM**)  
International Workshop on Bidirectional Transformations (**Bx**)  
USENIX Workshop on the Theory and Practice of Provenance (**TaPP**)  
IEEE International Conference on Data Engineering (**ICDE**)  
International Conference on Extending Database Technology (**EDBT**)  
International Conference on Very Large Databases, demonstrations (**VLDB**)  
International Workshop on Big Data Stream Processing (**BigStream**)
- 2017 International Workshop on Data Management on Virtualized Active Systems (**Active**)  
ACM International Conference on Information and Knowledge Management (**CIKM**)  
International Conference on Data Management Technologies and Applications (**DATA**)  
IEEE International Conference on Data Engineering (**ICDE**), Ph.D. symposium  
ACM SIGMOD International Conference on Management of Data (**SIGMOD**)  
International Conference on Very Large Databases (**VLDB**)
- 2016 IEEE International Conference on Data Science and Advanced Analytics (**DSSA**)  
IEEE International Conference on Data Engineering (**ICDE**), demonstrations  
ACM SIGMOD International Conference on Management of Data (**SIGMOD**)
- 2015 Workshop on Big Data Benchmarking Canada (**WBDB Canada**)  
International Workshop on the Web and Databases (**WebDB**)  
USENIX Workshop on the Theory and Practice of Provenance (**TaPP**)  
International Conference on Data Management Technologies and Applications (**DATA**)  
ACM SIGMOD International Conference on Management of Data (**SIGMOD**)  
IEEE International Conference on Data Engineering (**ICDE**)  
International Conference on Extending Database Technology (**EDBT**)  
ACM SIGMOD International Conference on Management of Data, demonstrations (**SIGMOD**)
- 2014 ACM SIGMOD International Conference on Management of Data, demonstrations (**SIGMOD**)  
USENIX Workshop on the Theory and Practice of Provenance (**TaPP**)  
International Conference on Data Management Technologies and Applications (**DATA**)
- 2013 Workshop on Big Data Benchmarking China (**WBDB China**)  
Workshop on Big Data Benchmarking US (**WBDB.us**)  
USENIX Workshop on the Theory and Practice of Provenance (**TaPP**)

International Conference on Scientific and Statistical Database Management, demonstrations  
(**SSDBM**)

2012 Alberto Mendelzon Workshop on Foundations of Data Management (**AMW**)

2011 ACM SIGMOD International Conference on Management of Data (**SIGMOD**)

## Journal Reviews

2018 Elsevier Information Systems

ACM Transactions on Database Systems (**TODS**)

IEEE Transactions on Knowledge and Data Engineering (**TKDE**)

2017 International Journal on Semantic Web and Information Systems (**IJSWIS**)

ScienceDirect Science of Computer Programming Journal (**SCICO**)

Springer Information Systems Frontiers (**ISF**)

The VLDB Journal

2016 IEEE Transactions on Software Engineering (**TSE**)

ACM Transactions on Internet Technology (**TOIT**)

The VLDB Journal

IEEE Transactions on Knowledge and Data Engineering (**TKDE**)

2015 Elsevier Journal of Biomedical Informatics (**JBIM**)

Springer Artificial Intelligence Review (**ARTIF INTELL REV**)

Elsevier Information Systems

The VLDB Journal

2014 The VLDB Journal

ACM Journal of Data and Information Quality (**JDIQ**)

IEEE Computer (**Computer**)

2013 IEEE Transactions on Knowledge and Data Engineering (**TKDE**)

ACM Transactions on Internet Technology (**TOIT**)

2012 The VLDB Journal

ACM Transactions on Database Systems (**TODS**)

2011 ACM Transactions on Database Systems (**TODS**)

IEEE Transactions on Knowledge and Data Engineering (**TKDE**)

## NSF Panels

2013 Information and Intelligent Systems (CISE-IIS)

## Other Proposal Reviews

2018 European Research Council (ERC) - Proposal Review

2017 Canadian Natural Sciences and Engineering Research Council (NSERC) - Proposal Review

2016 Israel Science Foundation (ISF) - Proposal Review

2015 Fields Institute, Toronto - Event Proposal Review

2014 Marsden Fund (The Royal Society of New Zealand) - Proposal Review

## External Reviewer

2019 International Conference on Database Theory (**ICDT**)

2015 ACM/IEEE Symposium on Logic In Computer Science (**LICS**)

2010 IEEE International Conference on Data Engineering (**ICDE**)

2009 GI Conference on Database Systems for Business, Technology, and Web (**BTW**)

2008 IEEE International Conference on Data Engineering (**ICDE**)

- 2007 International Conference on Very Large Databases (**VLDB**)
- ACM SIGMOD International Conference on Management of Data (**SIGMOD**)
- International Conference on Objects, Models, Components, Patterns (**TOOLS**)
- International Conference on Business Information Systems (**BIS**)
- 2006 International Conference on Very Large Databases (**VLDB**)

## Teaching and Advisor Experience

### Courses as Instructor

- 2018 CS520 - Data Integration, Warehousing, and Provenance  
IIT  
CS425 - Advanced Database Organization  
IIT
- 2017 CS425 - Advanced Database Organization  
IIT  
CS525 - Advanced Database Organization  
IIT
- 2016 CS425 - Database Organization  
IIT  
CS520 - Data Integration, Warehousing, and Provenance  
IIT
- 2015 CS525 - Advanced Database Organization  
IIT  
CS520 - Data Integration, Warehousing, and Provenance  
IIT
- 2014 CS425 - Advanced Database Organization  
IIT  
CS525 - Advanced Database Organization  
IIT  
CS695 - Doctoral Seminar  
IIT
- 2013 CS425 - Database Organization  
IIT  
CS525 - Advanced Database Organization  
IIT
- 2012 CS595 - Hot topics in database systems: Data Provenance  
IIT
- 2009 327 - Informatik für Ökonomen II (Teil 1) - Einführung in Datenbanken (Computer science for economists II (part 1) - introduction to databases)  
University of Zurich

### Supervised Ph.D. Students

- |                   |   |
|-------------------|---|
| current B. Arab   | Provenance using Temporal Databases   |
| current J. Arnold | HRDMS - a Distributed Relational Database System<br>(co-supervised with Ioan Raicu) |
| current S. Feng   | Vizier  |
| current S. Lee    | Vagabond, Game Provenance   |
| current X. Niu    | Optimizing of Provenance Computations   |
| current Q. Zeng   | TBA   |

## Ph.D. Student Thesis Committee Member

2018 (expected)	University of Toronto CS	J. Du
	University of Toronto CS	J. Siddique
	University of Tübingen, Germany	T. Müller
2017	IIT CS	X. Zhang
	IIT CS	T. Jung
	University of Basilicata CS, Italy	E. Veltri
2016	IIT CS	J. Kroclic
	IIT CS	I. Sadooghi
	IIT CS	M. Sharma
2015	IIT Psychology	K. Niel
	IIT ECE	Q. Tong
	IIT CS	X. Zhang
2013	IIT CS	T. Li
	IIT CS	U. Das
	IIT CS	R. Summerscales

## Supervised Master and Diploma Theses

2018	S. Samson	Optimization Communication in Distributed Query Execution
2015	A. Cornudella	Computing Candidate Keys of Relational Operators for Optimizing Rewrite-Based Provenance Computations
2014	Z. Wang	Efficient Ranking of Explanations for Data Exchange Errors in Vagabond
2010	B. Steiger	Data lineage/provenance in X-query (co-supervised with P.M. Fischer and K.S. Esmaili)
	A. Vagliardo	MidPerm - Realisierung einer Provenance SQL-Erweiterung als Middleware Lösung
2009	O. Wirz	Entwurf und Implementierung eines SQL-DDL-Präprozessors zur Unterstützung von Datenbankentwurfsmustern
2008	S. Gerster	Entwurf und Umsetzung eines Prototyps zur effizienten Speicherung von Hoch-Volumen Prozessdaten
	S. Blatti	Entwurf und Implementierung eines Provenance Browsers für die Visualisierung von Data Provenance
	S. Nedkoff	DBDoc Entwurf und Implementierung einer Anwendung zur partiellen Automation des Dokumentations-Prozesses für Datenbanken
2007	P. Hochstrasser	Entwurf und Implementierung einer Anwendung zur computergestützten Durchführung von klinischen Interviews
	C. Humard	Entwurf und Implementierung einer modular erweiterbaren Anwendung für Eingabe und Import von heterogenen Daten
	T. Schlaginhaufen	Design and Implementation of a Database Client Application for Inserting, Modifying, Presentation and Export of Bitemporal Personal Data

## Supervised Bachelor Theses

2008	M. Keller	Visualisierung von datenbankunterstützten Prozessen
	S. Schurgast	Export von Datenbankinhalten in Datenformate von Statistikprogrammen

## Supervised Student Projects

2018	P. Rotti	CS497 - GProM - Creating uncertainty labelings from incomplete databases
------	----------	--

	J. Sun	CS497 - GProM - Creating uncertainty labelings from probabilistic databases
	V. Polavarapu	CS597 - HRDBMS - Non-blocking queues
2017	T. Nguyen	HRDBMS - External Table types
	A. Alsabbagh	NSF REU - Vizier
	F. Morgan	CS597 - HRDBMS - Fast external load
	M. Kuzmyn	CS597 - HRDBMS - External table support
	F. Morgan	CS597 - HRDBMS - External table support
	N. Brahmankar	CS597 - HRDBMS - expression compilation
	K. Juneja	CS597 - HRDBMS - expression compilation
	J. Chen	CS597 - HRDBMS - expression compilation
	M. Mistry	CS597 - GProM - SparkSQL backend
2016	S. Kunali	CS597 - HRDBMS - External table support
	M. Kuzmyn	CS597 - An effective encoding of temporal data and its implementation in an ORM tool
	P. Li	CS597 - Developing a debugger GUI for transactional workloads
	C. Liu	CS597 - Extending reenactment with temporal aspects
	D. Patel	CS597 - User-defined domain specification for Datalog provenance
	L. Patil	CS597 - Adding support for JSON querying to HRDBMS
	A. Zhen	Summer Internship - Datageneration in iBench
	R. Alves Da Silva	BSMP Undergraduate Summer Intern - iBench
	D. Campos Faial	BSMP Undergraduate Summer Intern - GProM
	R. Leitao Oliveira	BSMP Undergraduate Summer Intern - HRDBMS
	P. Miranda Van-develd	BSMP Undergraduate Summer Intern - HRDBMS
	J. Monte Carvalho	BSMP Undergraduate Summer Intern - HRDBMS
	K. Queiroz Giulianetti C Moura	BSMP Undergraduate Summer Intern - HRDBMS
	F. Soares Di Biase Moura	BSMP Undergraduate Summer Intern - iBench
2015	R. Sreenidhi	CS597 - Vagabond - Interactive confirmation of ranked explanations
	S. Dasadawala	CS597 - GProM - LDV JDBC instrumentation
	J. Zhu	CS597 - GProM - Enhancing Algebra Graph Support
	R. Kapoor	CS597 - GProM - JSON support and PROV import/export
	C. Abreu	BSMP Undergraduate Summer Intern - GProM
	R. Barsanulfo Fernandes	BSMP Undergraduate Summer Intern - iBench
	C. Borges Da Silva	BSMP Undergraduate Summer Intern - iBench
	W. Castro	BSMP Undergraduate Summer Intern - GProM
	C. Dos Santos Sousa	BSMP Undergraduate Summer Intern - Vagabond
	R. Franco	BSMP Undergraduate Summer Intern - iBench
	A. Hirai	BSMP Undergraduate Summer Intern - iBench
	A. Paiva	BSMP Undergraduate Summer Intern - Vagabond
	M. Rocha	BSMP Undergraduate Summer Intern - iBench

	J. Vieira	BSMP Undergraduate Summer Intern - Vagabond
2014	R. Kapoor	CS597 - GProM - Heuristic and Cost-based Optimization
	R. Li	CS597 - Vagabond GUI - Fixes and Support for Query-based Error Marking
	J. Zhu	CS597 - GProM - Add Support for Algebra Graphs to Provenance Rewriter
	M. Lucera	BSMP Undergraduate Summer Intern - A GUI for Transaction Provenance
	A. Iype	CS597 - Functional Dependency Compliant Data Generator
	H. Guo	CS597 - Implementing a Query-Rewrite Based Provenance Database Middle- ware - Provenance Support for Update and Transactions
2013	B. Dan	CS597 - Implementing a Query-Rewrite Based Provenance Database Middle- ware - Temporal Features
	H. Guo	CS597 - Implementing a Query-Rewrite Based Provenance Database Middle- ware - Provenance Support for Update and Transactions
	Z. Lin	CS597 - Implementing a Query-Rewrite Based Provenance Database Middle- ware - PI-CS Provenance
	Y. Ni	CS597 - Implementing a Query-Rewrite Based Provenance Database Middle- ware - Nested Subqueries
	P. Purandare	CS597 - Implementing a Query-Rewrite Based Provenance Database Middle- ware - PI-CS Provenance
	S. Xie	CS597 - Implementing a Query-Rewrite Based Provenance Database Middle- ware - Collections and internal model
	A. Iype	CS597 - Database-side Batch Handling of Explanations and Error Sets
	S. Xie	CS597 - MDS Provenance
	G. Shankar	CS597 - Implementation and Integration of a Provenance-aware Aggregation Operator
2011	V. Suen	Undergraduate Student Research Awards Summer Project (NSERC USRA): Automatic Generation of Explanations for Data Exchange Errors
	G. Pourshahid	Project for Research Topics in Data Management (CSC 2525): A User Interface for TRAMP
2010	K. Bühler	Performance Analysis of Provenance Computation using Query Rewrite Tech- niques on a Commercial Database System