Renée J. Miller

Degrees

- Ph.D. in Computer Science, **University of Wisconsin-Madison** 1991-1994
  PhD Thesis: *Managing Structural Heterogeneity.* **Advisor:** Yannis Ioannidis

- M.S. in Computer Science, **University of Wisconsin-Madison** 1987-1989

- B.S. in Mathematics, **Massachusetts Institute of Technology** 1983-1987

- B.S. in Cognitive Science, **Massachusetts Institute of Technology** 1983-1987

Employment

UNIVERSITY WATERLOO Waterloo, ON

- Professor & Canada Excellence Research Chair (CERC) in Data Intelligence 2024

NORTHEASTERN UNIVERSITY Boston, MA

- University Distinguished Professor 2018

UNIVERSITY of TORONTO Toronto, ON

- Professor of Computer Science 2004
  Bell Canada Chair of Information Systems 2006-2016

- Associate Professor of Computer Science with tenure 2001-2004

- Assistant Professor of Computer Science 1998-2001

STANFORD UNIVERSITY Palo Alto, CA

- Visiting Professor of Computer Science 2001-2002

IBM ALMADEN RESEARCH CENTER San Jose, CA

- Visiting scientist 2001-2002

OHIO STATE UNIVERSITY Columbus, OH

- Assistant Professor of Computer Science 1994-1998

UNIVERSITY of WISCONSIN Madison, WI

- Research & teaching assistant 1991-1994

AT&T-BELL LABORATORIES Holmdel, NJ

- Distributed information system architectures 1989-1991

Distinctions, Awards, Honors

- Fellow of AAAS (American Association for the Advancement of Science) 2024

- Alonzo Church Award for Outstanding Contributions to Logic 2021

  Citation: *"Awarded to Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, Lucian Popa, and Wang Chiew Tan for their ground-breaking work on laying the logical foundations for data exchange."*

- EDBT Best Paper Award 2021

- CS Canada Lifetime Achievement Award in Computer Science 2020

  Citation: *"Professor Renée Miller is world-renowned as a pioneer in data management, especially for the creation of tools to integrate, transform, query and analyze information. Her contributions to data integration, schema mapping, and data exchange have established her as one of the seminal thinkers in database systems. Her work stands out for its originality and elegance, while being at the same time one of those rare and much-needed instances of academic research that bridges in a principled way theory and practice."*

- VLDB Women in Database Research Award 2018

- PVLDB Best Demonstration Award "Interactive Navigation of Open Data Linkages" 2017

- Theoretical Computer Science (Journal) 2nd Most Cited Paper 2005-2014 2015

- ICDT Test-of-Time Award 2013

  This award recognizes a paper presented at the ICDT conference from 10 years prior that has best met the test of time and was awarded to "Data Exchange: Semantics and Query Answering", by Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. This work was praised for *"initiating a new line of research, and providing a beautiful conceptual contribution, rich of challenging questions, which has constantly attracted the community in the last ten years, and still attracts many valuable research efforts."* ICDT and PODS are the two top international research forums in database theory.

- NSERC Accelerator Award for Exceptional Opportunity 2013

- Fellow of the Royal Society of Canada (FRSC) 2011

  The Royal Society of Canada is the Canadian National Academy, counterpart of the US National Academy of Science. Citation: *Professor Renée Miller is world-renowned as a pioneer in the field of Database Systems, a core topic for Computer Science. Her work has focused on the long-standing open problem of data integration and has achieved the goal of building practical data integration systems. At the same time, Professor Miller has been a leader among her peers in Canada and abroad, spearheading research and research activity. Her profile is unique in that it combines gracefully theoretical elegance, reflected by citations and scholarly awards, with industrial impact, reflected by patents and successful industrial products.*

- ACM Fellow   2009

  Citation: *For innovations in metadata management, especially the creation of tools to integrate, transform, query and analyze information.*

- NSERC Accelerator Award for Exceptional Opportunity   2008

- Bell Canada Chair of Information Systems (Endowed Chair)   2006-2016

- Premier's Research Excellence Award (PREA)   1999

  PREA recognizes the most innovative Ontario researchers in any academic discipline who are beginning their careers.

- IBM University Partnership Faculty Award   1998-2001

  This competitive award recognizes the quality of faculty research programs and their importance to industry.

- Ohio State University (OSU) Distinguished Diversity Enhancement Award   1998

  This University award recognizes leadership in the creation and enrichment of educational opportunities for underrepresented minorities.

- OSU Engineering Research Accomplishment Award   1998

  This College of Engineering award recognizes an exceptional research accomplishment.

- OSU Dept. of Computer & Info. Science Award for Outstanding Service   1998

  This Departmental award recognizes exceptional contributions to the department.

- US Presidential Early Career Award for Scientists and Engineers (**PECASE**)   1997

  The PECASE program replaced the Presidential Faculty Fellow program in 1996 and is the highest honor bestowed by the US government on outstanding scientists and engineers beginning their careers. Only three computer scientists received this award in 1997.

- NSF Faculty Early Career Award (**CAREER**)   1997

  The CAREER program replaced the Presidential Young Investigator (PYI) program in 1995. Awardees are selected from among the most meritorious junior investigators funded by the National Science Foundation (NSF).

- Ameritech Faculty Fellow   1996

  The Ameritech Fellows Program promotes and showcases innovation in teaching and research. Ameritech is one of the seven regional telecommunications companies formed by the breakup of AT&T.

- Member of Phi Beta Kappa and Sigma Xi, an honorary scientific research society

Professional Activities

- **Associate Editor, Communications of the ACM (CACM)** — 2024-2027

- **Editor-in-Chief, VLDB Journal** — 2015-2024

- **Board of Trustees/Advisory Committees**

  - President VLDB Endowment — 2010-2014
  - Trustee VLDB Endowment — 2004-2010
    The Board oversees the activities of the Endowment, including the VLDB Conference (and its proceedings PVLDB), the VLDB Journal, and the VLDB tutorial series held in developing countries in order to assist with the education of their young researchers. New members are elected by the board of 21 distinguished database researchers.
  - PVLDB Advisory Board — 2014-2018
  - ACM SIGMOD Executive Committee — 2014-2017
  - Scientific Advisory Board, Max Planck Institute for Informatics — 2015-2021
  - Executive Committee of IEEE TCDE — 2016-2019
    Technical Committee on Data Engineering (oversees IEEE ICDE)
  - Steering Committee VADA (Value Added Data Systems) Consortium UK — 2016-2021

- **Editorial Boards** VLDB Journal, TKDE and ACM TODS are the three best journals in data management.

  - VLDB Journal — 2009-2015
  - IEEE Transactions on Knowledge and Data Engineering (TKDE) — 2009-2013
  - Journal of Information and Data Management — 2009-2014
    JIDM is a new journal started in 2009 with a goal of increasing the visibility and excellence of the South American DB Research Community.

- **Program Committee Chair**

  - PC Chair ACM **SIGMOD** (Special Interest Group on Modeling of Data) — 2011
  - PC Co-Chair VLDB Business Intelligence Workshop — 2009
    Co-Chaired with Umesh Dayal and Malu Castellanos
  - PC Vice-Chair IEEE **ICDE** (International Conference on Data Engineering) — 2007
    Vice-Chair of Metadata and Integration Track.
  - PC Co-Chair **VLDB** (International Conference on Very Large Data Bases) — 2004
    Co-Chaired with Donald Kossmann
  - PC Chair CASCON Conference — 1999
    CASCON is IBM's Academic-Industry Joint Conference.

- **Associate Editor**

  - SIGMOD — 2020, 2021, 2025
  - PVLDB 2018 — 2018

- Encylopedia of Database Systems 2007-2009
  - ACM SIGMOD DiSC (Digital Symposium Collection) 1998-2006
  - IEEE Data Engineering Bulletin 2002-2003

- **Conference Officer**

  - **ACM SIGMOD Awards** Committee 2019-2022
  - **IEEE Data Engineering** Awards Chair 2016-2019
  - **IEEE Data Engineering** Awards Committee 2014-2016
  - **ACM SIGMOD** Keynote Selection 2015
  - **ACM SIGMOD** Dissertation Award Selection Committee 2014-2016
  - **IEEE Data Engineering** Workshop Chair 2010
  - **ACM SIGMOD** Industrial Sponsorship Chair 2007-2009
  - **VLDB** Board of Trustee Conference Liason 2008
  - **ACM SIGKDD** Workshop Chair 2002

- **Royal Society of Canada** Fellow selection committee 2018–2022

- **NSERC Service**

  - Computer Science Grant Selection Committee 2002-2005
  - Selected as **GSC Chair** 2004-2005

- **NSF Service**

  - Steering Commitee: Workshop on Future Trends in Info. Integration & Informatics 2010
  - NSF IIS/III Review Panels (most years 1996-present, most recent is listed) 2023

- **ACM Service**

  - **ACM Grace Murray Hopper Award Selection Committee** 2016–2020
  - **ACM Fellow Selection Committee** 2010-2015
    * Selected as **Chair** 2013-2014
  - **ACM Paris Kanellakis Awards Committee** 2003-2008
    * **Chair** 2005-2006

- **CRA-W** 2002-2009
  The goal of the CRA (Computing Research Association) Committee on the Status of Women in Computing Research (CRA-W) is to take positive action to increase the number of women participating in Computer Science and Engineering research and education at all levels. The importance and impact of this committee is perhaps best understood by two awards it earned over this period.

  - US Presidential Award for Excellence in Science, Mathematics & Engineering Mentoring 2004
  - National Science Board's Public Service Award 2005

Research Interests

Big data. Metadata management. Data and knowledge curation. Data provenance. Data quality and data cleaning. Heterogeneous databases. Data integration. Data exchange.

Research Awards

- *Canada Excellence Research Chair in Data Intelligence*   2024
  From NSERC   Can $8,000,000
  From Waterloo & Partners   Can $15,800,000

- *NSF III: Small: Semantic Version Management in Data Lakes*   2023
  PI   US$600,000

- *NSF Collaborative Research: III: Medium: From Open Data to Open Data Science*   2021
  PI (co-PI Glavic and Nargesian)   US$1,000,000

- *NSF III: Medium: Table-as-Query: Unifying Data Discovery and Alignment*   2020
  PI (co-PI Gatterbauer and Reidewald)   US $1,000,000

- *NSERC Discovery and Accelerator*
  Enabling Data Science over Open Data and Massive Data Lakes   2018
  PI $88,000

- *NSERC Strategic Project*
  Data-Driven Knowledge Mobilization, Translation & Innovation   2015-2018
  co-PI with K. Lyons, E. Stroulia, K. Booth $249,000

- *MITACS*
  MITACS Accelerate Post-Doc Supplement   2014
  PI $20,000

- *NSERC & Industry*
  Strategic Network, Business Intelligence Network   2009-2015
  PI $6,000,000

  I was the Network Leader and PI for this partnership of researchers from seven universities and several industry labs and start-ups across Canada. IBM and SAP each gave $500,000 in cash to the network and provided more than triple this amount in in-kind contributions (mostly in time of their researchers). This grant funded over 210 graduate student years and a dozen post-docs.

- *NSERC*
  Strategic Network Enhancement Initiative   2010-2015
  PI $137,050

- *NSERC*
  Discovery Accelerator Supplement   2013-2015
  Sole PI $120,000

  Discovery Grant, Big Data Curation   2013-2017
  Sole PI $220,000

  Equipment Grant, Web-Scale Data Integration   2012-2014
  Sole PI $137,500

- *MITACS and IBM*
  Elevate Industrial Post-doctoral Fellowship   2010
  Application with Dr. Richard Hull of IBM $75,000

- *NSERC*
  Discovery Accelerator Supplement   2008-2011
  Sole PI $120,000

  Strategic Grant, Data Sharing in Peer Networks   2007-2009
  (co-PIs J. Mylopoulos & I. Kiringa) $408,000

  Discovery Grant, Data and Metadata Alignment   2008-2012
  Sole PI $240,000
  Discovery Grant, Integration Management   2003-2007
  Sole PI $215,000
  Discovery Grant, Managing Schematic Heterogeneity in Database Management Systems   1999-2003
  Sole PI $168,000

  CRD, Supporting Data Privacy   2008-2009
  Sole PI $58,000
  CRD, Implementation of Peer-to-Peer Multidatabases   2003-2005
  (co-PIs J. Mylopoulos & I. Kiringa) $93,000
  Equipment Grant, Databases and Modern Hardware   2005
  (co-PIs N. Koudas, J. Mylopoulos, A. Mendelzon, K. Sevcik) $110,265
  Equipment Grant, Storage and Server Pool for Internet-Scale Data Management Research   2004
  (co-PIs M. Consens, N. Koudas, J. Mylopoulos, A. Mendelzon, K. Sevcik) $60,000
  Equipment Grant   1999
  (co-PIs J. Mylopoulos, A. Mendelzon, K. Sevcik) $144,601

- *Bell Canada University Labs*
  Addressing Data Quality Issues,   2005
  (co-PI N. Koudas and J. Mylopoulos) $28,750
  Data Mining, Seed Grant   1999-2000
  Sole PI $25,000

- *Communications & Information Technology Ontario (CITO),*
  Deploying Networks of Peer Data Managment Systems   2005-2007

(co-PIs J. Mylopoulos & I. Kiringa)	$264,600
Implementation of Peer-to-Peer Multidatabases	2003-2005
(co-PIs J. Mylopoulos & I. Kiringa)	$226,000
Structure Discovery in Large Data Sets	2000-2002
(co-PIs K. Sevcik & A. Mendelzon)	$300,000

- *MITACS*, Towards Interactive Data Mining   2002-2004
  (co-PIs R. Ng, L. Lakshmanan, K. Sevcik) $260,000

- *Connaught Foundation*
  Matching Start-up Grant   1998-1999
  Sole PI $40,000

- *Premier's Research Excellence Award*   1999-2003
  Sole PI $100,000

- *IBM*, University Partnership Award   1998-2001
  Sole PI $174,000

- *NSF*
  Science of Design, Data and Meta-Data Integration   2004-2008
  Senior Personnel with L. Getoor (PI)   US$510,200
  CAREER Award   1997-2002
  Sole PI   US$542,269
  Experimental Software Systems   1997-2001
  (co-PIs M. J. Harrold, A. Porter, G. Rothermel)   US$1,553,149

- *GE Foundation*, Faculty for the Future   1997-2000
  Sole PI   US$581,502

Patents

- System and Method for Schema Mappings. Inventors: L. M. Haas, M. A. Hernández-Sherrington, R. J. Miller. Patent application filed 2000. U.S. Patent No. 7010539 issued 2006.

- Efficient Determination of Homogeneous Rectangles in a Binary Matrix. Inventors: J. Edmonds, J. Gryz, D. Liang and R. J. Miller. Patent application filed 2000. U.S. Patent No. 7113955 issued 2006.

- Method for schema mapping and data transformation. Inventors: R. Fagin, L. M. Haas, M. A. Hernández-Sherrington, R. J. Miller, F. Naumann, L. Popa, L. Yan. Patent application filed 2002. U.S. Patent No. 7149746 issued 2006.

- System and Method for Translating Data from a Source Schema to a Target Schema. Inventors: R. Fagin, M. A. Hernández-Sherrington, R. J. Miller, L. Popa and Y. Velegrakis. Patent application filed 2003. U.S. Patent No. 20040199905 issued 2004.

- Method for Generating Nested Mapping Specifications in a Schema Mapping Formalism. Inventors: A. Fuxman, M. A. Hernández-Sherrington, C.-T. Ho, R. J. Miller, P. Papotti and L. Popa. Patent application filed 2007. U.S. Patent No. 20080243765 issued 2008..

Selected Scholarly and Professional Work

The four most prestigious and competitive publishing forums within Database Systems are VLDB (the International Conference on Very Large Databases), SIGMOD (ACM's SIGMOD International Conference on Modeling of Data), ICDE (IEEE's International Conference on Data Engineering), and EDBT. Since 2008, VLDB has published its papers in the journal PVLDB (the Proceedings of the VLDB Endowment). These conferences have rigorous reviewing standards (with 15-17% acceptance rates).

Post-doctoral and student co-authors (including graduate and undergraduate students) are indicated with a * after their names.

References

[LCL+24]   Aristotelis Leventidis*, Martin Pekár Christensen, Matteo Lissandrini, Laura Di Rocco, Katja Hose, and Renée J. Miller. A large scale test corpus for semantic table search. In *SIGIR*, pages 1142–1151. ACM, 2024.

[PKSM24]   Koyena Pal*, Aamod Khatiwada*, Roee Shraga, and Renée J. Miller. ALT-GEN: Benchmarking Table Union Search using Large Language Models. In *VLDB Workshop on Tabluar Data Analysis (TaDA)*, 2024. **Best Paper Award.** To appear.

[FSM24a]   Grace Fan*, Roee Shraga, and Renée J. Miller. Finding Support for Tabular LLM Outputs. In *VLDB Workshop on Tabluar Data Analysis (TaDA)*, 2024. To appear.

[FSM24b]   Grace Fan*, Roee Shraga, and Renée J. Miller. Gen-T: Table reclamation in data lakes. In *IEEE ICDE*, pages 3532–3545. IEEE, 2024.

[Gla+24]   Boris Glavic, Giansalvatore Mecca, Rene J. Miller, Paolo Papotti, Donatello Santoro, Enzo Veltri: Similarity Measures For Incomplete Database Instances. EDBT, 461-473, 2024.

[LRG+23]   Aristotelis Leventidis*, Laura Di Rocco, Wolfgang Gatterbauer, Renée J. Miller, and Mirek Riedewald. Domainnet: Homograph detection and understanding in data lake disambiguation. *ACM Trans. Database Syst. (TODS)*, 48(3):9:1–9:40, 2023.

[FWL+23]   Grace Fan*, Jin Wang, Yuliang Li, Dan Zhang, and Renée J. Miller. Semantics-aware dataset discovery from data lakes with contextualized column-based representation learning. *Proc. VLDB Endow.*, 16(5):1726-1739, 2023.

[KFS+23]   Aamod Khatiwada*, Grace Fan*, Roee Shraga*, Zixuan Chen*, Wolfgang Gatterbauer, Renée J. Miller, and Mirek Riedewald. SANTOS: relationship-based semantic table union search. *Proc. ACM Manag. Data (PACMMOD), previously named SIGMOD*, 9:1-9:25, 2023.

[NPB+23]   Fatemeh Nargesian, Ken Q. Pu, Bahar Ghadiri Bashardoost*, Erkang Zhu, and Renée J. Miller. Data lake organization. *IEEE Trans. Knowl. Data Eng.*, 35(1):237–250, 2023.

[SM23]   Roee Shraga* and Renée J. Miller. Explaining dataset changes for semantic data versioning with explain-da-v. *Proc. VLDB Endow.*, 16(6):1587–1600, 2023.

[KSGM22]   Aamod Khatiwada*, Roee Shraga*, Wolfgang Gatterbauer, and Renée J. Miller. Integrating data lake tables. *Proc. VLDB Endow.*, 16(4):932–945, 2022.

[LRG+21] Aristotelis Leventidis*, Laura Di Rocco*, Wolfgang Gatterbauer, Renée J. Miller, and Mirek Riedewald. DomainNet: Homograph detection for data lake disambiguation. In EDBT, pages 13–24, 2021. **Best Paper Award.**

[OSN+21] Paul Ouellette*, Aidan Sciortino*, Fatemeh Nargesian, Bahar Ghadiri Bashardoost*, Erkang Zhu, Ken Pu, and Renée J. Miller. RONIN: data lake exploration. *Proceedings of the VLDB Endowment (PVLDB)*, 14(12):2863–2866, 2021.

[BLM21] Bahar Ghadiri Bashardoost, Kelly A. Lyons, and Renée J. Miller. Towards knowledge exchange: State-of-the-art and open problems. In Tomás Bures, Riccardo Dondi, Johann Gamper, Giovanna Guerrini, Tomasz Jurdzinski, Claus Pahl, Florian Sikora, and Prudence W. H. Wong, editors, *SOFSEM - Theory and Practice of Computer Science*, volume 12607 of *Lecture Notes in Computer Science*, pages 13–27. Springer, 2021.

[NPZ+20] Fatemeh Nargesian, Ken Q. Pu, Erkang Zhu, Bahar Ghadiri Bashardoost*, and Renée J. Miller. Organizing data lakes for navigation. In SIGMOD, pages 1939–1950. ACM, 2020.

[BMLN20] Bahar Ghadiri Bashardoost*, Renée J. Miller, Kelly A. Lyons, and Fatemeh Nargesian. Knowledge translation. *Proc. VLDB Endow.*, 13(11):2018–2032, 2020.

[CMG+20] Christina Christodoulakis*, Eric Munson, Moshe Gabel, Angela Demke Brown, and Renée J. Miller. Pytheas: Pattern-based table discovery in CSV files. *Proc. VLDB Endow.*, 13(11):2075–2089, 2020.

[NZMPA19] Fatemeh Nargesian*, Erkang Zhu*, Rene J. Miller, Ken Q. Pu, Patricia C. Arocena: Data Lake Management: Challenges and Opportunities. PVLDB 12(12): 1986-1989 (2019)

[ZDNM19] Erkang Zhu*, Dong Deng, Fatemeh Nargesian*, and Renée J. Miller. JOSIE: Overlap set similarity search for finding joinable tables in data lakes. In *ACM Int'l Conf. on the Management of Data (SIGMOD)*, 847-864, 2019.

[KMMG19] Angelika Kimmig, Alex Memory, Renée J. Miller, and Lise Getoor. A collective, probabilistic approach to schema mapping using diverse noisy evidence. *IEEE Transactions on Knowledge and Data Engineering*, 2019. 31(8): 1426-1439 (2019).

[Mil18] Renée J. Miller. Open data integration. *PVLDB*, 11(12):2130–2139, 2018.

[NZPM18] Fatemeh Nargesian*, Erkang Zhu*, Ken Q. Pu, and Renée J. Miller. Table Union Search on Open Data. *Proceedings of the VLDB Endowment (PVLDB)*, 11(7):813–825, 2018.

[MZP+18] Renée J. Miller, Erkang Zhu*, Ken Q. Pu, Fatemeh Nargesian*, and Christina Christodoulakis*, Periklis Andritsos. Making Open Data Transparent. *IEEE Data Engineering Bulletin*, 41(2):59-70 , 2018.

[KMMG17] A. Kimmig, A. Memory*, R. .J. Miller, and L. Getoor. A Collective, Probabilistic Approach to Schema Mapping In *IEEE Proc. of the Int'l Conf. on Data Eng. (ICDE)*, pages 921–932, 2017.

[DGTM17] J. Du*, B. Glavic, W . Tan, and R. J. Miller. Deepsea: Progressive Workload-aware Partitioning of Materialized Views in Scalable Data Analytics. In *Advances in Database Technology - Int'l Conf. on Extending Database Technology (EDBT)*, pages 198–209, 2017.

[CKTM17] C . Christodoulakis*, E. Kandogan, I. G. Terrizzano and R. J. Miller. VIQS: Visual Interactive Exploration of Query Semantics. In *IUI Workshop: Exploratory Search and Interactive Data Analytics*. ACM, pages 25-32, 2017.

[SDD+17] Shazia Wasim Sadiq, Tamraparni Dasu, Xin Luna Dong, Juliana Freire, Ihab F. Ilyas, Sebastian Link, Renée J. Miller, Felix Naumann, Xiaofang Zhou, and Divesh Srivastava. Data quality: The role of empiricism. *SIGMOD Record*, 46(4):35–43, 2017.

[ZNPM16] E. Zhu*, F. Nargesian*, K. Q. Pu, and R. J. Miller. LSH Ensemble: Internet Scale Domain Search. *Proceedings of the VLDB Endowment (PVLDB)*, 9(12): 1185-1196 (2016)

[AGM+16] P. C. Arocena*, B. Glavic, G. Mecca, R. J. Miller, P. Papotti, and D. Santoro*. Benchmarking Data Curation Systems. *IEEE Data Eng. Bull.*, 39(2):47–62, 2016.

[AGCM15] Patricia C. Arocena*, Boris Glavic, Radu Ciucanu*, and R. J. Miller. The iBench Integration Metadata Generator. *Proceedings of the VLDB Endowment (PVLDB)*, 9(3):108–119, 2015.

[AGM+15] P. C. Arocena*, B. Glavic, G. Mecca, R. J. Miller, P. Papotti, and D. Santoro*. Messing up with BART: Error Generation for Evaluating Data-Cleaning Algorithms. *Proceedings of the VLDB Endowment (PVLDB)*, 9(2):36–47, 2015.

[PSC+15] N. Prokoshyna*, J. Szlichta*, Fei Chiang, R. J. Miller, and Divesh Srivastava. Combining Quantitative and Logical Data Cleaning. *Proceedings of the VLDB Endowment (PVLDB)*, 9(4):300–311, 2015.

[HM15] O. Hassanzadeh and R. J. Miller. Automatic Curation of Clinical Trials Data in LinkedCT. In *Int'l Conf. on the Semantic Web (ISWC)*, pages 270–278, 2015.

[KRS+15] E. Kandogan, M. Roth, P. M. Schwarz, J. Hui, I. Terrizzano, Christina Christodoulakis*, and R. J. Miller. Labbook: Metadata-driven Social Collaborative Data Analysis. In *IEEE Int'l Conf. on Big Data*, pages 431–440, 2015.

[VCSM14] M. Volkovs*, F. Chiang*, J. Szlichta*, and R. J. Miller. Continuous Data Cleaning. In *IEEE Proc. of the Int'l Conf. on Data Eng. (ICDE)*, pages 244-255, April 2014.

[HPY+13] O. Hassanzadeh, K. Q. Pu, S. Hassas Yeganeh*, R. J. Miller, L. Popa, M. A. Hernández, and H. Ho. Discovering Linkage Points over Web Data. *Proceedings of the VLDB Endowment (PVLDB)*, 6(6): 444-456, 2013.

[AGM13] P. C. Arocena*, B. Glavic*, and R. J. Miller. Value Invention in Data Exchange. In *ACM SIGMOD Int'l Conf. on the Mgmt. of Data (SIGMOD)*, pages 157-168, 2013.

[DHMT13] N. Dindar*, N. Tatbul, R. J. Miller, L. Haas and I. Botan*. Modeling the Execution Semantics of Stream Processing Engines with SECRET. *The Int'l Journal on Very Large Data Bases (VLDB Journal)*, 22(4):421-446, 2013.

[GSAM13] B. Glavic*, J. Siddique*, P. Andristos, and R. J. Miller. Provenance for Data Mining. In *USENIX Workshop on the Theory and Practice of Provenance*, 5 pages, 2013.

[XHF+13] R. S. Xin*, O. Hassanzadeh*, C. Fritz, S. Sohrabi*, and R. J. Miller. Publishing Bibliographic Data on the Semantic Web using BibBase. *Semantic Web*, 4(1):15–22, 2013.

[GMA13] B. Glavic*, R. J. Miller, and G. Alonso. Using SQL for efficient generation and querying of provenance information. In Search of Elegance in the Theory and Practice of Computation. Festschrift for Peter Buneman. Lecture Notes in Computer Science Volume 8000, pp. 291-320, 2013.

[AMM13a] P. C. Arocena*, R. J. Miller, and J. Mylopoulos. The Vivification Problem in Real-Time Business Intelligence: A Vision. In M. Castelanos, U. Dayal, and E. A. Rundensteiner, editors, *Enabling Real-Time Business Intelligence*, volume 154 of *Lecture Notes in Business Information Processing*, pages 37–49. Springer, 2013.

[HKL⁺12] O. Hassanzadeh*, A. Kementsietsidis, L. Lim, R. J. Miller, and M. Wang. Semantic link discovery over relational data. In Y. Velegrakis R. De Virgilio, F. Guerra, editor, *Semantic Search over the Web*, pages 193–224. Springer, 2012.

[SMKV11] T. K. Sellis, R. J. Miller, A. Kementsietsidis, and Y. Velegrakis, editors. *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2011, Athens, Greece, June 12-16, 2011*. ACM, 2011. I was PC Chair and co-editor of the proceedings.

[CM11] F. Chiang* and R. J. Miller. A unified model for data and constraint repair. In *IEEE Proc. of the Int'l Conf. on Data Eng. (ICDE)*, pages 446–457, 2011.

[YHM11] S. Hassas Yeganeh*, O. Hassanzadeh*, and R. J. Miller. Linking semistructured data on the web. In *ACM SIGMOD Workshop on the Web and Databases (WebDB 2011)*, 6 pages, 2011.

[GM11] B. Glavic* and R. J. Miller. Reexamining Some Holy Grails of Data Provenance. In *USENIX Workshop on the Theory and Practice of Provenance*, 6 pages, 2011.

[GAMH10] B. Glavic*, G. Alonso, R. J. Miller, and L. M. Haas. TRAMP: Understanding the Behavior of Schema Mappings through Provenance. *Proceedings of the VLDB Endowment (PVLDB)*, 3(1):1314–1325, 2010.

[BDD⁺10] I. Botan*, R. Derakhshan*, N. Dindar*, L. Haas, R. J. Miller, and N. Tatbul. SECRET: A Model for Analysis of the Execution Semantics of Stream Processing Systems. *Proceedings of the VLDB Endowment (PVLDB)*, 3(1):232–243, 2010.

[AFM10] P. C. Arocena*, A. Fuxman, and R. J. Miller. Composing Local-As-View Mappings: Closure & Applications. In *Proc. of the Int'l Conf. on Database Theory (ICDT)*, pp. 209-218, 2010.

[FEM10] P. M. Fischer*, K. Sheykh Esmaili*, and R. J. Miller. Stream schema: providing and exploiting static metadata for data stream processing. In *Advances in Database Technology - Int'l Conf. on Extending Database Technology (EDBT)*, pages 207–218, 2010.

[PHDM10] K. Q. Pu, O. Hassanzadeh*, R. Drake*, and R. J. Miller. Online annotation of text streams with structured entities. In *Proc. of the Conf. on Information and Knowledge Management (CIKM)*, pages 29–38, 2010.

[CMRV10] M. Consens, R. J. Miller, F. Rizzolo*, and A. A. Vaisman. Exploring XML web collections with DescribeX. *ACM Transactions on the Web (TWEB)*, 4(3), 2010. (46 pages)

[CDM10] M. Castellanos, U. Dayal, and R. J. Miller, editors. *Enabling Real-Time Business Intelligence - Third International Workshop, BIRTE 2009, Held at the 35th International Conference on*

Very Large Databases, VLDB 2009, Lyon, France, August 24, 2009, Revised Selected Papers, volume 41 of *Lecture Notes in Business Information Processing*. Springer, 2010. I was PC co-Chair and co-editor of the workshop proceedings.

[HM09]   O. Hassanzadeh* and R. J. Miller. Creating probabilistic databases from duplicated data. *The Int'l Journal on Very Large Data Bases*, 18(5):1141–1166, October 2009.

[HCLM09] O. Hassanzadeh*, F. Chiang*, H. C. Lee*, and R. J. Miller. Framework for evaluating clustering algorithms in duplicate detection. *Proceedings of the VLDB Endowment (PVLDB)*, 2(1):1282–1293, 2009.

[HHKM09] L. M. Haas, M. Hentschel*, D. Kossmann, and R. J. Miller. Schema AND Data: A Holistic Approach to Mapping, Resolution and Fusion in Information Integration. In *Int'l Conf. on Conceptual Modeling (ER)*, pages 27–40, 2009. Invited keynote paper.

[FHH$^+$09] R. Fagin, L. M. Haas, M. A. Hernández, R. J. Miller, L. Popa, and Y. Velegrakis. Clio: Schema mapping creation and data exchange. In *Conceptual Modeling: Foundations and Applications. Festschrift for John Mylopoulos. Springer*, pages 198–236, 2009.

[HKL$^+$09a] O. Hassanzadeh*, A. Kementsietsidis, L. Lim, R. J. Miller, and M. Wang. A framework for semantic link discovery over relational data. In *Proc. of the Conf. on Information and Knowledge Management (CIKM)*, pages 1027–1036, 2009.

[CM08b]   F. Chiang* and R. J. Miller. Discovering Data Quality Rules. *Proceedings of the VLDB Endowment (PVLDB)*, 1(1):1166–1177, 2008.

[ACMT08] B. Alexe*, L. Chiticariu*, R. J. Miller, and W.-C. Tan. Muse: Mapping understanding and design by example. In *IEEE Proc. of the Int'l Conf. on Data Eng. (ICDE)*, pages 10-19, 2008.

[UGM07]   O. Udrea*, L. Getoor, and R. J. Miller. Leveraging data and structure in ontology integration. In *ACM SIGMOD Int'l Conf. on the Management of Data (SIGMOD)*, pages 449–460, 2007.

[ABMM07] Y. An*, A. Borgida, R. J. Miller, and J. Mylopoulos. A Semantic Approach to Discovering Schema Mapping Expressions. In *IEEE Proc. of the Int'l Conf. on Data Engineering (ICDE)*, pages 206–215, 2007.

[HSM07]   O. Hassanzadeh*, M. Sadoghi*, and R. J. Miller. Accuracy of Approximate String Joins Using Grams. In *5th International Workshop on Quality in Databases at VLDB (QDB)*, pages 11 – 18, 2007.

[FM07]   A. Fuxman* and R. J. Miller. First-Order Query Rewriting for Inconsistent Databases. *Journal of Computer and System Sciences*, 73(4):610–635, 2007.

[LLM07]   H. C. Lee*, H. Liu*, R. J. Miller. Geographically-Sensitive Link Analysis. IEEE/ACM Conference on Web Intelligence, 628–634, 2007. **Best Paper Award.**

[FKMT06] A. Fuxman*, P. G. Kolaitis, R.J. Miller, and W.-C. Tan. Peer Data Exchange. *ACM Trans. on Database Sys. (TODS)*, 31(4):1454–1498, 2006.

[FHH$^+$06] A. Fuxman*, M. A. Hernandez, H. Ho, R. J. Miller, P. Papotti*, and L. Popa. Nested mappings: Schema Mapping Reloaded. In *Proc. of the Int'l Conf. on Very Large Data Bases (VLDB)*, pages 67–78, 2006.

[AFM06]   P. Andritsos*, A. Fuxman*, and R.J. Miller. Clean Answers over Dirty Databases: A Proba-
           bilistic Approach. In *IEEE Proc. of the Int'l Conf. on Data Engineering (ICDE)*, pages 30–42,
           2006.

[KMMZ06]   Y. Kanza*, A. O. Mendelzon, R. J. Miller, and Z. Zhang*. Authorization-Transparent Access
           Control for XML Under the Non-Truman Model. In *Advances in Database Technology - Int'l
           Conf. on Extending Database Technology (EDBT)*, pages 222–239, 2006.

[FKMT05]   A. Fuxman*, P. G. Kolaitis, R.J. Miller, and W.-C. Tan. Peer Data Exchange. In *Proc. of the
           ACM Symp. on Principles of Database Systems (PODS)*, pages 160–171, 2005. **Selected as
           one of the five best papers of PODS05.**

[FFM05]   A. Fuxman*, E. Fazli*, and R.J. Miller. Efficient Management of Inconsistent Databases. *ACM
           SIGMOD Int'l Conf. on the Management of Data (SIGMOD)*, 34(2):155-166, 2005.

[FKMP05]   R. Fagin, P. G. Kolaitis, R. J. Miller, and L. Popa. Data Exchange: Semantics and Query
           Answering. *Theoretical Computer Science*, 336(1):89-124, 2005.

[VMM05]   Y. Velegrakis*, R. J. Miller, and J. Mylopoulos. Representing and Querying Data Transforma-
           tions. In *IEEE Proc. of the Int'l Conf. on Data Engineering (ICDE)*, pages 81–92, 2005.

[FM05]   A. Fuxman* and R. J. Miller. First-Order Query Rewriting for Inconsistent Databases. *Proc.
           of the Int'l Conf. on Database Theory (ICDT)*, pages 337–351, 2005. LNCS Volume 3363/2004.

[NOK+04]   M. A. Nascimento, M. T. Özsu, D. Kossmann, R. J. Miller, J. A. Blakeley, and K. B. Schiefer,
           editors. *Proceedings of the Thirtieth International Conference on Very Large Data Bases
           (VLDB)*, Toronto, Canada, 2004. Morgan Kaufmann.

[VMP04]   Y. Velegrakis*, R. J. Miller, and L. Popa. On Preserving Mapping Consistency under Schema
           Changes. *The Int'l Journal on Very Large Data Bases (VLDB Journal)*, 13(3):274–293, 2004.

[AMT04]   P. Andritsos*, R. J. Miller, and P. Tsaparas*. Information-Theoretic Tools for Mining Database
           Structure from Large Data Sets. *ACM SIGMOD Int'l Conf. on the Management of Data
           (SIGMOD)*, 33(2):731–742, 2004.

[ATMS04]   P. Andritsos*, P. Tsaparas*, R. J. Miller, and K. Sevcik. LIMBO: Scalable Clustering of
           Categorical Data. In *Int' Conf. on Extending Database Technology (EDBT)*, pp. 123-146,
           2004.

[AFK+04]   P. Andritsos, A. Fuxman, A. Kementsietsidis, R. J. Miller, and Y. Velegrakis. Kanata: Adap-
           tation and Evolution in Data Sharing Systems. *SIGMOD Record*, 33(4):32–37, 2004.

[VMP03]   Y. Velegrakis*, R. J. Miller, and L. Popa. Mapping Adaptation under Evolving Schemas. In
           Proc. of Int'l Conf. on Very Large Data Bases (VLDB), pp. 584-595, 2003. **Selected as one
           of the five best papers of VLDB03.**

[MA03]   R. J. Miller and P. Andritsos*. On Schema Discovery. *IEEE Data Engineering Bulletin*,
           26(3):39-44, 2003.

[AKK+03]   M. Arenas*, V. Kantere*, A. Kementsietsidis*, I. Kiringa, R. J. Miller, and J. Mylopoulos. The
           Hyperion Project: From Data Integration to Data Coordination. *SIGMOD Record*, 32(3):53–
           58, 2003.

[KAM03]   A. Kementsietsidis*, M. Arenas*, and R. J. Miller. Mapping Data in Peer-to-Peer Systems: Semantics and Algorithmic Issues. In *ACM SIGMOD Int'l Conf. on the Management of Data (SIGMOD)*, pp. 325-336, 2003.

[KAM03a]   A. Kementsietsidis*, M. Arenas*, and R. J. Miller. Managing Data Mappings in The Hyperion Project. Short Paper. In *IEEE Int'l Conf. on Data Engineering (ICDE)*, pp. 732–734, 2003.

[EGLM03]   J. Edmonds, J. Gryz, D. Liang*, and R. J. Miller. Mining for Empty Spaces in Large Data Sets. *Theoretical Computer Science*, 3(296):435–452, 2003.

[FKMP03]   R. Fagin, P. G. Kolaitis, R. J. Miller, and L. Popa. Data Exchange: Semantics and Query Answering. In *Int'l Conf. on Database Theory (ICDT)*, pp. 207–224, 2003. **Selected as one of the five best papers of ICDT03. In 2013 received the ICDT Test-of-Time Award.**

[AFF⁺02]   P. Andritsos*, R. Fagin, A. Fuxman*, L. Haas, M. Hernández, H. Ho, A. Kementsietsidis*, R. J. Miller, F. Nauman, L. Popa, Y. Velegrakis*, C. Vilarem*, and L.-L. Yan. Schema Management. *IEEE Data Engineering Bulletin*, 25(3):32–38, 2002.

[PVM⁺02a]   L. Popa, Y. Velegrakis*, R. J. Miller, M. A. Hernandez, and R. Fagin. Mapping Generation and Data Translation of Heterogeneous Web Data. In *Second International Workshop on Data Integration over the Web (DIWeb)*, Toronto, Canada, May 2002.

[PVM⁺02b]   L. Popa, Y. Velegrakis*, R. J. Miller, M. A. Hernandez, and R. Fagin. Translating Web Data. In *Int'l Conf. on Very Large Data Bases (VLDB)*, pp. 598-609, August 2002.

[PM02]   I. Popivanov* and R. J. Miller. Similarity Search Over Time-Series Data Using Wavelets. In *IEEE Int'l Conf. on Data Engineering (ICDE)*, San Jose, CA, pp. 212-224, February 2002.

[YMHF01]   L. Yan, R. J. Miller, L. Haas, and R. Fagin. Data-Driven Understanding and Refinement of Schema Mappings. *ACM SIGMOD Int'l Conf. on the Management of Data (SIGMOD)*, 30(2):485–496, May 2001.

[MHH⁺01]   R. J. Miller, M. A. Hernández, L. M. Haas, L. Yan, C. T. H. Ho, R. Fagin, and L. Popa. The Clio Project: Managing Heterogeneity. *SIGMOD Record*, 30(1):78-83, March 2001.

[EGLM01]   J. Edmonds, J. Gryz, D. Liang*, and R. J. Miller. Mining for Empty Rectangles in Large Data Sets. In *International Conference on Database Theory (ICDT)*, pages 174-188, London, UK, January 2001. **Selected as one of the five best papers of ICDT01.**

[AM01]   P. Andritsos* and R. J. Miller. Reverse Engineering Meets Data Analysis. In *Ninth International Workshop on Program Comprehension*, Toronto, Canada, pp. 157-166 May 2001.

[MHH00]   R. J. Miller, L. M. Haas, and M. Hernández. Schema Mapping as Query Discovery. In *Int'l Conf. on Very Large Data Bases (VLDB)*, pages 77–88, Cairo, Egypt, September 2000.

[JM00]   C. Jermaine* and R. J. Miller. Approximate Query Answering in High-Dimensional Data Cubes. In *2000 ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD 2000)*, pages 31–36, Dallas, TX, May 2000.

[GM99]   M. N. Garofalakis and R. J. Miller. Very Large Databases. In J.G. Webster, editor, *Wiley Encyclopedia of Electrical and Electronics Engineering*, volume 23, pages 133–139. John Wiley and Sons, 1999.

[MG99]   R. J. Miller and A. Gujarathi*. Mining for Program Structure. *International Journal on Software Engineering and Knowledge Engineering*, 9(5):499–517, 1999.

[HMN⁺99] L. M. Haas, R. J. Miller, B. Niswonger, M. Tork Roth, P. M. Schwarz, and E. L. Wimmers. Transforming Heterogeneous Data with Database Middleware: Beyond Integration. *IEEE Data Engineering Bulletin*, 22(1):31–36, 1999.

[Mil98]   R. J. Miller. Using Schematically Heterogeneous Structures. *ACM SIGMOD Int'l Conf. on the Management of Data (SIGMOD)*, 27(2):189–200, June 1998.

[WML98]   C.-H. Wu*, R. J. Miller, and M. Liu. Querying Multimedia Presentations. *Computer Communications Journal*, 21(14):1212–1225, September 1998.

[MTW97]   R. J. Miller, O. G. Tsatalos, and J. H. Williams. DataWeb: Customizable Database Publishing for the Web. *IEEE Multimedia*, 4(4):14–21, Oct-Dec 1997.

[MY97]   R. J. Miller and Y. Yang*. Association Rules over Interval Data. *ACM SIGMOD Int'l Conf. on the Management of Data (SIGMOD)*, 26(2):452–461, May 1997.

[WML97]   C.-H. Wu*, R. J. Miller, and M. Liu. An Integrated Composition and Query Model for Multimedia Presentation Systems. In *IEEE Conf. on Protocols for Multimedia Systems – Multimedia Networking (PROMS-MmNet'97)*, pages 64–73, Santiago, Chile, November 1997. **Selected as one of the six best papers from the conference.**

[HMPR97] M. J. Harrold, R. J. Miller, A. Porter, and G. Rothermel. A Collaborative Investigation of Program-Analysis-Based Testing and Maintenance. In *International Workshop on Experimental Studies of Software Maintenance*, pages 51–56, Bari, Italy, October 1997.

[HMV96]   S. Heiler, R. J. Miller, and V. Ventrone. Using Metadata to Address Problems of Semantic Interoperability in Large Object Systems. In *IEEE Metadata*, Washington, D.C., April 1996. 14 pages.

[MTW95]   R. J. Miller, O. G. Tsatalos, and J. H. Williams. Integrating Hierarchical Navigation and Querying: A User Customizable Solution. In *ACM Multimedia Workshop on Effective Abstractions in Multimedia Layout, Presentation, and Interaction*, San Francisco, CA, November 1995. 10 pages.

[MIR94]   R. J. Miller, Y. E. Ioannidis, and R. Ramakrishnan. Schema Equivalence in Heterogeneous Systems: Bridging Theory and Practice. *Information Systems*, 19(1):3–31, 1994.

[MIR94]   R. J. Miller, Y. E. Ioannidis, and R. Ramakrishnan. Schema Equivalence in Heterogeneous Systems: Bridging Theory and Practice. In *Proc. of the International Conference on Extending Database Technology (EDBT)*, pages 73–80, Cambridge, UK, March 1994. **Best Paper Award.**

[MIR93]   R. J. Miller, Y. E. Ioannidis, and R. Ramakrishnan. Understanding Schemas. In *Third Int'l Workshop on Research Issues in Data Engineering (RIDE): Interoperability in Multidatabase Systems*, pages 170–173, Vienna, Austria, April 1993.

[ILH⁺93]   Y. Ioannidis, M. Livny, E. Haber, R. J. Miller, O. Tsatalos, and J. L. Wiener. Desktop Experiment Management. *IEEE Data Engineering Bulletin: Special Issue on Scientific Databases*, 16(1):19–23, March 1993.

Refereed System Demonstrations

[KSM23]   A. Khatiwada*, R. Shraga*, R. J. Miller: DIALITE: Discover, Align and Integrate Open Data Tables. In *ACM SIGMOD*, pages 187-190, 2023.

[CZBM19] H. Choi*, E. Zhu*, A. Bangash*, R. J. Miller: VISE: Vehicle Image Search Engine with Traffic Camera. *PVLDB* 12(12): 1842-1845 (2019). System Demonstration.

[ZNPM17] E. Zhu*, K. Q. Pu, F. Nargesian*, and R. J. Miller. Interactive Navigation of Open Data Linkages *Proceedings of the VLDB Endowment (PVLDB)*, 10(12):1837-1840, 2017. System Demonstration. **Received Best Demo Award.**

[SAG+16] D. Santoro*, P. C. Arocena*, B. Glavic, G. Mecca, R. J. Miller, and P. Papotti. BART in Action: Error Generation and Empirical Evaluations of Data-Cleaning Systems. In *ACM SIGMOD* , pages 2161-2164, 2016. System Demonstration.

[ACGM15] P. C. Arocena*, R. Ciucanu*, B. Glavic, and R. J. Miller. Gain Control Over Your Integration Evaluations. *Proceedings of the VLDB Endowment (PVLDB)*, 8(12):1960–1971, 2015. System Demonstration.

[HMNZ15] O. Hassanzadeh, R. J Miller, F. Nargesian*, and E. Zhu*. LinkedCT Live: Platform for Online Curation of Clinical Trials Data. *International Semantic Web Conference (ISWC)*. 4 pages, 2015. System Demonstration.

[BCY+15] B. Ghadiri Bashardoost*, C. Christodoulakis*, S.-H. Yeganeh*, R. J. Miller, K. Lyons, and O. Hassanzadeh. VizCurator: A visual tool for curating open data. In *Int'l Conf. on the World Wide Web (WWW)*, pages 195–198, 2015. System Demonstration.

[CFM14]   C. Christodoulakis∗, C. Faloutsos, and R. J. Miller. VoidWiz: Resolving Incompleteness Using Network Effects. In *IEEE Proc. of the Int'l Conf. on Data Eng. (ICDE)*, pages 1230–1233, 2014. System Demonstration.

[CAZM12] F. Chiang*, P. Andritsos, E. Zhu*, and R. J. Miller. Autodict: Automated Dictionary Discovery. In *IEEE Int'l Conf. on Data Engineering (ICDE)*, pages 1277–1280, 2012. System Demonstration.

[GDM+11] B. Glavic*, J. Du*, R. J. Miller, G. Alonso, and L. M. Haas. Debugging Data Exchange with Vagabond. In *Proceedings of the VLDB Endowment (PVLDB)*, pages 1383–1386, 2011. System Demonstration.

[HHM10]   M. Hentschel*, L. Haas, and R. J. Miller. Just-in-time Data Integration in Action. *Proceedings of the VLDB Endowment (PVLDB)*, 3(2):1621–1624, 2010. System Demonstration.

[HXF+10] O. Hassanzadeh*, R. Xin*, C. Fritz, Y. Yang*, J. Du*, M. Zhao*, and R. J. Miller. Bibbase Triplified. In *I-SEMANTICS*, 2010. System Demonstration for Triplification Challenge.

[HXK+09] O. Hassanzadeh*, R. Xin*, R. J. Miller, A. Kementsietsidis, L. Lim, and M. Wang. Linkage Query Qriter. *Proceedings of the VLDB Endowment (PVLDB)*, 2(2):1590–1593, 2009. System Demonstration.

[DCBM09b] F. Duchateau*, R. Coletta, Z. Bellahsene, and R. J. Miller. Yam: a Schema Matcher Factory. In *Proc. of the Conf. on Information and Knowledge Management (CIKM)*, pages 2079–2080, 2009. System Demonstration.

[ACMPT08] B. Alexe*, L. Chiticariu*, R. J. Miller, D. Pepper*, and W.-C. Tan. Muse: A System for Understanding and Designing Mappings. In *ACM SIGMOD Int'l Conf. on the Management of Data (SIGMOD)*, pages 1281–1284, 2008.

[UGM07a] O. Udrea*, L. Getoor, and R. J. Miller. Homer: Ontology Alignment Visualization and Analysis. In *International Semantic Web Conference (ISWC)*, 2007.

[HHP+07] M. Hernandez, H. Ho, L. Popa, T. Fukuda, A. Fuxman*, R. J. Miller, and P. Papotti*. Creating nested mappings with Clio: System demonstration. In *IEEE Proc. of the Int'l Conf. on Data Engineering (ICDE)*, pages 1487–1488, 2007.

[FFM05b] A. Fuxman*, D. Fuxman*, and R. J. Miller. ConQuer: A System for Efficient Querying Over Inconsistent Databases, *System Demonstration*. In *Proc. of the Int'l Conf. on Very Large Data Bases (VLDB)*, pages 1354–1357, 2005.

[RGJ+05] P. Rodríguez-Gianolli*, M. Garzetti, L. Jiang*, A.Kementsietsidis*, I. Kiringa, M. Masud*, R. J. Miller, and J. Mylopoulos. Data Sharing in the Hyperion Peer Database Systens, *System Demonstration*. In *Proc. of the Int'l Conf. on Very Large Data Bases (VLDB)*, pages 1291–1294, 2005.

[YMP04] Y. Velegrakis*, R. J. Miller, L. Popa, and J. Mylopoulos. ToMAS: A System for Adapting Mappings as Schemas Evolve, *System Demonstration*. In *IEEE Int'l Conf. on Data Engineering (ICDE)*, pp. 862, 2004.

[PHVM02] L. Popa, M. A. Hernández, Y. Velegrakis*, R. J. Miller, F. Naumann*, and C.-T. Ho. Mapping XML and Relational Schemas with Clio, *System Demonstration*. In *IEEE Int'l Conf. on Data Engineering (ICDE)*, pp. 498-499 2002.

[HMH+01] M. A. Hernández, R. J. Miller, and L. Haas. Clio: A Semi-Automatic Tool for Schema Mapping, *System Demonstration. ACM SIGMOD Int'l Conf. on the Management of Data (SIGMOD)*, 30(2):607, 2001.

[MS96] R. J. Miller and S. Subramanian*. INQuire: Integrating Navigation & Querying. Student Exhibit & System Demonstration. *Advances in Digital Libraries*, 1996.

Education Articles

[Mil97] R. J. Miller. Establishing a Diversity Program at Ohio State University. *Computing Research News*, 9(5), November 1997. Computing Research Association (CRA).

Some Recent Invited Lectures and Keynotes

- **ACM PODS**, Invited Talk, Bridging Theory and Practice   2024

- **ACM SIGMOD HILDA Workshop**, Semantic Benchmark Generation: Can LLMs Generate Better Benchmarks than Humans?   2024

- **VLDB TaDa Workshop**, From Discovery to Integration of Data Lake Tables   2023

- **VLDB Quality in Databases (QDB) Workshop**, Semantic Version Management   2023

- **ACM CODS COMAD** (flagship conference in Data Science & Databases in India), Keynote Speaker   2021

- **47th International Conference on Current Trends in Theory and Practice of Computer Science**, Keynote Speaker   2021

- **ACM SIGMOD Panelist** Grand Challenges in Exploring, Understanding, and Searching a Billion Datasets   2020

- **VLDB Workshop on Challenges and Experiences from Data Integration to Knowledge Graphs, Keynote Speaker**   2020

- **Rome Chamber of Commerce**, Data Lakes, Open Data, and Democratizing Data, Speaker 2020

- **Very Large Data Base (VLDB) Conference, Keynote Speaker**   2018

- **ACM KDD (Knowledge Discovery & Data Mining), Keynote Speaker**   2017

- **Oxford University CS Dept. Distinguished Speaker**, Data Curation at Scale   2016

- **University of Lausanne, Winter School**, Taming Big Data   2016

- **BICOD (British International Conf. on Databases) Keynote**, Big Data Curation   2015

- **Big Data Summit of the German Federal Ministry of Education & Research**   2015

- **COMAD (Indian National Database Conference) Keynote**, Big Data Curation   2014

- **School of Computing, National University of Singapore**, Big Data Curation   2014

- **Inria, Lille, France**, Value Invention in Data Exchange   2014

- **Dept. de Math. et Informatique, Université Paris Descartes**, Big Data Curation   2014

- **Department of Computer Science, USC**, Big Data Curation   2014

- **Department of Computer Science, UC San Diego**, Big Data Curation   2014

- **ISI (Information Sciences Institute), Los Angeles**, Big Data Curation   2014

- **IBM Almaden Research Labs**, Value Invention in Data Exchange   2013

- **Department of Computer Science, Memorial University**, Big Data Curation   2013

- **Data Science Institute, UC Santa Cruz**, Big Data Curation   2013

- **School of Informatics, Univ. of Edinburgh**, Scalable Data Provenance   2013

- **DIMACS, Rutgers University**, Big Data Curation   2013

- Invited Talk **DIMACS Workshop Big Data Integration, Rutgers**   2013

- Distingished Lecture **School of Computer Science, University of Waterloo** 2013

- Cyber Center Seminar Series, **Purdue University**, 2012

- Distinguished Lecture **School of Computing, Queen's University** 2012

- Keynote: **IEEE International Conference on Data Mining** 2011

- Distinguished Lecture **Dept. Computer Science, Cornell University** 2011

- Distinguished Lecture **Dept. Computer Science, Brown University** 2010

Students

- Graduated PhD   University of Toronto

  - Periklis Andritsos   PhD September, 2004
    Scalable Clustering of Categorical Data and Applications
    Current employment: Associate Professor, Univ. of Toronto, School of Information

  - Yannis Velegrakis (co-advisor with John Mylopoulos)   PhD December, 2004
    Managing Schemas and Mappings
    Current employment: Full Professor, University Utrecht, Netherlands

  - Tasos Kementsietsidis (co-advisor with John Mylopoulos) PhD January, 2005
    Data Exchange and Querying in Peer-to-Peer Data Mgmt Systems
    Current employment: Google AI

  - Ken Pu (co-advisor with Alberto Mendelzon)   PhD July, 2006
    On Formal Methods of Multidimensional Databases
    Current employment: Professor (former Chair), Ontario Institute of Technology

  - Ariel Fuxman   PhD December, 2006
    Efficient Query Processing Over Inconsistent Databases
    Received **ACM SIGMOD Jim Grey Dissertation Award**
    Current employment: Google Research

  - Flavio Rizzolo (co-advisor with Alberto Mendelzon)   PhD June, 2008
    DescribeX: A Framework for Exploring and Querying XML Web Collections
    Current employment: Researcher, Statistics Canada

  - Oktie Hassanzadeh   PhD April, 2012
    Record Linkage over Web Data
    Current employment: Researcher, IBM Watson Research Center

  - Fei Chiang   PhD September, 2012
    Data Quality Through Active Constraint Discovery and Maintenance
    Current employment: Associate Professor, McMaster University, Canada

  - Patricia Rodríguez-Gianolli Arocena (co-advisor with John Mylopoulos)   PhD March, 2013
    Embracing Incompleteness in Schema Mappings
    Current employment: VP Data Science, RBC (Royal Bank of Canada)

  - Jiang Du   PhD February, 2019
    Online Physical Design and Materialization in Scalable Data Management
    Current employment: Facebook

- Erkang Zhu   PhD April, 2019
  Search and Join Algorithms for Tables in Data Lakes
  Current employment: Microsoft Research

- Fatemeh Nargesian   PhD June, 2019
  Relational Data Enrichment by Discovery and Transformation
  Current employment: Assistant Professor University of Rochester

- Bahar Ghadiri Bashardoost (co-supervised by Kelly Lyons)   PhD August 2022
  Knowledge Base Exchange
  Current employment: SchemaApp (Toronto Start-up)

- Current PhD Students   Northeastern University

  - Aristotelis Sigiouan Leventidis   PhD expected Fall 2024
  - Grace Fan   PhD expected Fall 2024
  - Aamod Khatiwada   PhD expected Fall 2024
  - Koyena Pal   PhD expected 2026

- Post-doctoral Fellows

  - Periklis Andritsos (co-advisor with Ken Sevcik)   2004-2005
    Associate Professor, ISchool, University of Toronto
  - Yaron Kanza (co-advisor with Alberto Mendelzon)   2004-2006
    Researcher, AT&T Research Labs
  - Lee Hyun Chul   2007-2008
    Samsung VP
  - Hojjat Ghaderi   2010-2011
    Two Sigma Investments
  - Boris Glavic   2010-2012
    Associate Professor, University of Illinois, Chicago
  - Patricia (Rodríguez-Gianolli) Arocena   2013-2016
    VP Data Science, RBC (Royal Bank of Canada)
  - Laura Di Rocco   2018-2020
    Knowledge Engineer, Bayer AG
  - Roee Shraga   2021-2023
    Assistant Professor, Worcester Polytechnic Institute

- External PhD Dissertation Advisory or Committee Member

  - Lan Jiang   HPI, University of Potsdam, Germay, 2022
    Discovering Metadata in Data Files
    (Advisor Felix Naumann)
  - Masayo Ota   NYU, 2020
    Towards Data Integration for Open Structured Data
    (Advisor Juliana Friere)

- Rihan Hai   Aachen University, Germany, 2020
  Data Integration and Metadata Management in Data Lakes
  (Advisor Matthias Jarke)

- Yodsawalai Chodpathumwan   Univ. Illinois, Urbana-Champaign, 2018
  Cost-Effective Data Structual Preparation
  (advisor Marianne Winslett)

- Alex Memory   University of Maryland, 2018
  Collective Relational Data Integration with Diverse and Noisy Evidence
  (advisor Lise Getoor)

- Ahmed El-Roby   University of Waterloo, 2018
  Web Data Integration for Non-Expert Users
  (advisor Ashraf Alboulnaga)

- Lucas Braun   ETH, Zurich 2016
  Analytics in Motion
  (advisor Donald Kossmann)

- Babak Salimi   Carleton University, 2015
  Query-Answer Causality in Databases and Connections with Reverse Reasoning Tasks in Data and Knowledge Management
  (advisor Leo Bertossi)

- Gunes Aluc   University of Waterloo, 2015
  Workload Matters: A Robust Approach to Physical RDF Database Design
  (advisor Tamer Ozsu), Software Engineer at SAP

- Donatello Santoro   University of Rome (Roma Tre), 2014
  Advanced Techniques for Mapping and Cleaning
  (advisor Gianni Mecca), Postdoc University of Bascillicata

- Yoones Asgharzadeh Sekhavat   Memorial University, 2013
  A Framework for Information Integration using Ontological Foundations
  (advisor Jeffrey Parsons), Postdoc University of Alberta

- Thanh Nguyen   NYU-Poly & Univ. Utah, 2012
  Integrating Structured Data on the Web
  (advisor Juliana Freire)

- George Beskales   University of Waterloo, 2012
  Modelling and Querying Uncertainty in Data Cleaning
  (advisor Ihad Ilyas), Co-founder Tamr

- Kyumars Sheykh Esmaili   ETH Zurich, 2011
  Data Stream Processing in Complex Applications
  (advisor Donald Kossmann)

- Salvatore Raunich   University of Rome (Roma Tre), 2010
  Quality of Mappings for Data Exchange Applications
  (advisor Paolo Atzeni and Giansalvatore Mecca)

- Laura Chiticariu   University of California, Santa Cruz, 2008
  Enabling Data Interoperability with Provenance and Schema Integration
  (advisor Wang-Chiew Tan)

- Hanady Abdulsalam   Queens University, 2008
  Streaming Random Forests
  (advisor Pat Martin)

- Alessandro Raffio   University of Milan, 2008
  Schema Mapping for Semi-Structured Data
  (advisor Stefano Ceri)

- Zhen Zhang   U. Illinois, Urbana-Champagne, 2007
  Large Scale Information Integration on the Web: Finding, Understanding and Querying Web Databases
  (advisor Kevin Chang)

- Loreto Bravo   Carlton University, Ottawa, 2007
  Handling Inconsistency in Databases and Data Integration Systems
  (advisor Leo Bertossi)

- Paolo Papotti   University of Rome, 2007
  Methods and Techniques for Data Model Translation
  (advisor Riccardo Torlone)

- Xuerong Tang   University of Waterloo, 2003
  A High-level Specification Language for Structured Document Transformation
  (advisor Frank Tompa)

- Chris Jermaine   Georgia Institute of Technology, 2002
  Approximation for Decision Support
  (advisor Ed Omiecinski)

- Wang-Chiew Tan   University of Pennsylvania, 2002
  Data Annotations, Provenance, and Archiving
  (advisor Peter Buneman)

- Ling Ling Yan   University of Alberta, 2000
  Building Scalable and Flexible Mediation: the AURORA Approach
  (advisor Tamer Özsu)

- Chao-Hui Wu   Ohio State University, 1998
  Querying and Synchronizing Multimedia Presentations
  (advisor Ming T. Liu)

- Wang-Chien Lee   Ohio State University, 1996
  Indexing Techniques for Non-traditional Database Systems
  (advisors Dik Lun Lee and Douglas S. Kerr)

- Budi Yuwono   Ohio State University, 1995
  An Architecture for Cooperative Distributed Internet Resource Discovery Systems
  (advisor Ten-Hwang Lai)

- Advisor for Masters (with Thesis)   University of Toronto

  - Erkang (Eric) Zhu   January 2016
    Searching Open Data

  - Christina Christodoulakis   January 2015
    New Applications of Recommendations to Data Science