

## Juliana Freire

<https://vgc.engineering.nyu.edu/juliana>

Department of Computer Science and Engineering  
New York University  
370 Jay Street, 11th floor  
Brooklyn, NY 11201

Phone: (646) 997-4128  
[juliana.freire@nyu.edu](mailto:juliana.freire@nyu.edu)

---

### Research Interests

Data management, data science, big data, data analysis and visualization, provenance management and analytics, computational reproducibility, scientific data management, urban data analytics, large-scale information curation and integration, web information discovery

### Professional Experience

- *Elected Chair* July 2017 – June 2021  
ACM Special Interest Group on Management of Data (SIGMOD).
- *Council Member* July 2017 – June 2020  
Computing Research Association's Computing Community Consortium (CCC).
- *PI and Executive Director* September 2015 – November 2019  
Moore Sloan Data Science Environment, New York University.
- *Director of Graduate Studies* July 2014 – 2017  
Center for Data Science, New York University.
- *Professor* January 2014 –  
Department of Computer Science and Engineering, New York University.
- *Professor* July 2011 – December 2013  
Polytechnic Institute of New York University.
- *Associated Faculty* September 2013 –  
Center for Data Science (CDS), New York University.
- *Associated Faculty* September 2013 –  
Center for Urban Science & Progress (CUSP), New York University.
- *Affiliated Faculty* December 2011 –  
Courant Institute of Mathematical Sciences, New York University.
- *Faculty Member* July 2009 – June 2011  
Scientific Computing and Imaging (SCI) Institute, University of Utah.
- *Associate Professor* July 2008 – June 2011  
School of Computing, University of Utah.
- *Director* 2007 – 2008  
Center for Management of Provenance and Exploratory Workflows, University of Utah.

- *Co-founder* 2007  
VisTrails Inc.
- *Assistant Professor* July 2005 – June 2008  
School of Computing, University of Utah.
- *Visiting Professor* August 2006 – 2011  
Department of Computer Science, Linköping University, Sweden.
- *Visiting Assistant Professor* August 2004 – June 2005  
School of Computing, University of Utah.
- *Assistant Professor* September 2002 – April 2006  
Department of Computer Science and Engineering, OGI School of Science & Engineering, Oregon Health & Science University.
- *Member of Technical Staff* December 1997 – September 2002  
Database Systems Research Department, Bell Laboratories, Lucent Technologies.
- *Research Assistant* September 1992 – November 1997  
XSB Group–Department of Computer Science – SUNY at Stony Brook, NY
- *Research Assistant* January 1993 – December 1995  
Department of Computer Science & Linguistics, SUNY at Stony Brook.

## Education

- Ph.D. in Computer Science** December 1997  
State University of New York at Stony Brook  
*Advisor:* David S. Warren  
*Dissertation:* Scheduling Strategies for Evaluation of Recursive Queries over Memory and Disk-Resident Data
- M.S. in Computer Science** December 1992  
State University of New York at Stony Brook
- B.S. in Computer Science** July 1991  
Universidade Federal do Ceará, Brazil

## Awards/Honors

- *AAAS Fellow*, 2021.
- *ACM SIGMOD Contributions Award*, 2020.
- *ACM Fellow*, 2014.
- *IBM Faculty Award*, 2014.
- *Google Faculty Research Award*, 2013.
- *Dean's Teaching Commendation*. University of Utah, Spring 2009.

- *Utah Innovation Award for the VisTrails Plug-In for Autodesk's Maya.* Stoel Rives and the Utah Technology Council, 2009.
- *Utah's Forty under 40 business leaders.* Utah Business Magazine, 2009.
- *CAREER award.* National Science Foundation, 2008.
- *IBM Faculty Award,* 2008.
- *First prize (for Syntactica).* AT&T Student Software Awards, 1997.
- *Doctoral Fellowship.* Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES Brazil), 1992.

## Paper Awards

- *Best paper award.* Seventh Workshop on Data Management for End-to-End Machine Learning, ACM SIGMOD, 2023.
- *Best paper award.* International Workshop on Traffic Measurements for Cybersecurity (WTMC), IEEE Symposium on Security and Privacy, 2021.
- *Best paper award.* Workshop on Computational Methods in Online Misbehavior (CyberSafety), The Web Conference, 2020.
- *FOSS Impact Paper Award.* International Conference on Mining Software Repositories (MSR), 2019.
- *Distinguished Paper Award.* ACM SIGSOFT International Conference on Mining Software Repositories (MSR), 2019.
- *Best demo award.* ACM Special Interest Group on Management of Data (SIGMOD), 2018.
- *Most reproduciblle paper award.* ACM Special Interest Group on Management of Data (SIGMOD), 2017.
- *Best demo honorable mention award.* ACM Special Interest Group on Management of Data (SIGMOD), 2017.
- *Best paper honorable mention award.* IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2016.
- *Best paper award.* Educator program, Eurographics 2010.
- *Best poster award.* Brazilian Symposium on Databases (SBBD), 2009.
- *Best paper award.* IEEE Visualization Conference, 2007.
- *Best paper runner-up.* WWW10 commendation, 2001.

## Research Funding

From December 1997 through August 2002, I worked at Bell Labs (Lucent Technologies) where all funding was internal. I started to actively pursue grants after I moved to academia, in late 2002. Since then, I have received grants, contracts and gifts (funded by NSF, DOE, NIH, State of Utah, IBM, Microsoft, Google, Moore Foundation, Sloan Foundation, Keck Foundation, Yahoo!, AT&T, DARPA). Besides academic research funding, I have helped VisTrails Inc. obtain funding from the University of Utah, State of Utah, NSF, and DOE for business and product development.

## Active Grants

1. *National Science Foundation*. D-ISN/Collaborative Research: An Interdisciplinary Approach to the Discovery, Analysis, and Disruption of Wildlife Trafficking Networks. **PI: Juliana Freire**; Co-PI: Jennifer Jacquet. US\$655,765.00. August 2022-July 2024.
2. *National Science Foundation*. III: Medium: Dataset Search and Ranking for Data Augmentation and Explanation. **PI: Juliana Freire**; Co-PI: Chris Musco. US\$1,093,195.00. September 2021-August 2024.
3. DARPA. Streamlining Model Design, Comparison and Curation (technology transition). **PI: Juliana Freire**. US\$500,000. August 2021-July 2022.
4. DARPA. Streamlining Model Design, Comparison and Curation. **PI: Juliana Freire**. US\$3,800,000. March 2017-June 2021.
5. *National Science Foundation*. CIF21 DIBBs: EI: Vizier, Streamlined Data Curation. **PI: Oliver Kennedy**; Co-PIs: Boris Glavic and **Juliana Freire**. US\$2,725,699 (NYU portion: US\$1,236,355). January 2017.

## Past Grants

1. DARPA. Towards Locating and Exploring Hard-to-Find Information on the Web. **PI: Juliana Freire**. US\$3,600,000. September 2014–May 2018.
2. *Moore-Sloan Data Science Initiative*. Joint effort by New York University, the University of California, Berkeley and the University of Washington. US\$37.8 million. (NYU portion: \$12.6 million) **NYU PIs: Juliana Freire and Yann LeCun**. November 2013-2019.
3. *National Science Foundation*. CI-EN: Enhancing and Supporting a Community-Based Data Analysis, Visualization, and Provenance Platform. **PI: Juliana Freire**; Co-PI: David Koop. US\$ 499,962. September 2014.
4. *National Science Foundation*. MRI: Acquisition of an infrastructure for prototyping next-generation algorithms for large-scale visualization, data processing and analysis. **PI: Cláudio Silva**; Co-PIs: Huy Vo, **Juliana Freire**, John Iacono, Torsten Suel. US\$ 799,999 (2012–2016).
5. *W. M. Keck Foundation*. Development of a Laboratory X-ray Tomography System with Stationary Sample and Phase Contrast Imaging. Collaborative effort with Louisiana State University. US\$500,000. (NYU portion US\$50,000)
6. *Amazon*. AWS in Education Coursework grant. **PI: Juliana Freire**. US\$12,000. January 2016.
7. *Amazon*. AWS in Education Coursework grant. **PI: Juliana Freire**. US\$7,000. January 2015.
8. *Amazon*. AWS in Education Coursework grant. **PI: Juliana Freire**. US\$6,000. Aug 2014.
9. *Google Faculty Research Award*. **PI: Juliana Freire**. US\$58,436. August 2013.
10. *NYU Poly Faculty Seed Grant*. **PIs: Juliana Freire, Vicki Been, and Ingrid Gould Ellen**; US\$100,000. July 2013.
11. *Amazon*. AWS in Education Coursework grant. **PI: Juliana Freire**. US\$4,100. April 2013.

12. *Sloan Foundation*. Computational Reproducibility: Understanding the Requirements and Building the Necessary Infrastructure. **PI: Juliana Freire**; Co-PIs: Cláudio Silva and Dennis Shasha. US\$74,398. October 2012–2014.
13. *Amazon*. AWS Education Research grant. **PI: Juliana Freire**. US\$5,000. October 2012.
14. *Amazon*. AWS in Education Coursework grant. **PI: Juliana Freire**. US\$3,000. September 2012.
15. *National Science Foundation*. Supplement for CAREER: Storing, Querying and Re-Using Provenance of Computational Tasks. **PI: Juliana Freire**. US\$ 60,000 (2010–2013).
16. *National Science Foundation, IIS 1050422*. III:EAGER:Collaborative Research: A Community Experiment Platform for Reproducibility and Generalizability. **PI: Juliana Freire**. US\$ 190,000 (2010–2013).
17. *National Science Foundation, IIS 0905385*. III: Medium: Provenance Analytics: Exploring Computational Tasks and their History. **PI: Juliana Freire**; Co-PI: Cláudio Silva. US\$ 957,467 (2009–2013).
18. *National Science Foundation, IIS 0746500*. CAREER: Storing, Querying and Re-Using Provenance of Computational Tasks. **PI: Juliana Freire**. US\$ 499,999 (2008–2013)
19. *National Science Foundation, CNS 0751152*. CRI: IAD: A Service-Oriented Architecture for The Computation, Visualization, and Management of Scientific Data. **PI: Cláudio Silva**; Co-PIs: **Juliana Freire**, Mike Kirby and Sarang Joshi. US\$ 500,000 (2007–2012)
20. *National Science Foundation, IIS 0713637*. III-COR: Discovering and Organizing Hidden-Web Sources. **PI: Juliana Freire**. US\$ 336,000. International Travel Supplement US\$27,212; REU Supplement US\$15,000. (2007–2011)
21. *Department of Energy*. SBIR Phase I and Phase II: Provenance-Enabling DOE Visualization Applications. **PI: David Koop**; Co-PIs: *Juliana Freire* and Cláudio Silva. US\$ 850,000 (2008-2011).
22. *Sandia National Labs – Department of Energy*. Supporting Pipelines of Retrieval, Analysis and Visualization of Web Data. **PI: Juliana Freire**; Co-PI: Cláudio Silva. US\$ 103,430 (2009-2010).
23. *Sandia National Labs – Department of Energy*. Provenance Analytics Tools to Improve the Measurement of Usability and Insight in Visualization Applications. **PI: Cláudio Silva**; **Co-PI: Juliana Freire**. US\$ 100,000 (2009-2010).
24. *National Institutes of Health*. NCCR ARRA Administrative Supplement - Translational. **PI: Donald A. McClain**; Co-PIs: Lisa Cannon-Albright, Perry Renshaw, Cláudio Silva, **Juliana Freire**, Deborah A. Yurgelun-Todd. US\$ 998,137 (2009–2011).
25. *National Science Foundation, IIS 0844546*. Where the Ocean Meets the Cloud: Ad Hoc Longitudinal Analysis and Collaboration Over Massive Mesh Data. **PI: Cláudio Silva**; Co-PI: **Juliana Freire**. US\$ 190,001 (2009–2011). (This is a collaborative proposal with B. Howe, University of Washington.)
26. *National Science Foundation, OIA - Science & Technology Centers*. Science and Technology Center for Coastal Margin Observation and Prediction. **PI: António Baptista, OHSU**; Co-PIs for Utah sub-contract: **Juliana Freire** and Cláudio Silva. Total: US\$ 20,000,000; Utah portion: US\$ 478,563. (2006–11)
27. *National Science Foundation, CNS-0524096*. CT-T: A Laboratory Workbench for Security Research. **PI: Juliana Freire**; Co-PI: Eric Eide. US\$ 1,466,000. (2005–2010)

28. *State of Utah, Centers of Excellence*. Center for Management of Provenance and Exploratory Workflows. Greg Jones, Juliana Freire and Cláudio Silva. US\$200,000 (2008–2010).
29. *National Science Foundation, IIS-SEIII-0513692*. Managing Complex Visualizations. **PI: Juliana Freire**; Co-PI: Cláudio Silva. US\$ 518,252. REU Supplement US\$12,000. (2005–2009)
30. *National Science Foundation, IIS-DMS 0534628*. XML Data Management: Taking Order and Updates into Account. **PI: Juliana Freire**. US\$ 270,000. (2006–2009)
31. *National Science Foundation, IIP 0712592*. SBIR Phase I and IB: A Collaborative Architecture to Support Large-Scale Exploratory Workflows. *PI: Steven P. Callahan*; Co-PIs: **Juliana Freire** and Cláudio Silva US\$ 150,000 (2007–2008).
32. *Department of Energy, Los Alamos National Laboratory*. Obtaining Ray Tracing for Unstructured Grids Using Advanced Hardware and Target Improvements to the VisTrails Software. *PI: Cláudio Silva*; Co-PIs: **Juliana Freire**, Charles Hansen. US\$ 200,000 (2007–2008).
33. *State of Utah, Centers of Excellence*. Center for Management of Exploratory Workflows–Business Team. **PI: Juliana Freire**; Co-PI: Cláudio Silva. US\$50,000 (2007–2008).
34. *University of Utah Research Foundation Funding Incentive Seed Grant*. Infrastructure for the Automatic Construction of High-Quality Hidden-Web Directories. **PI: Juliana Freire**. US\$27,000 (2006–2007).
35. *National Science Foundation, CNS-0541560*. A Cluster Infrastructure to Support Retrieval, Management and Visualization of Massive Amounts of Data. **PI: Juliana Freire**; Co-PI: Cláudio Silva. US\$ 165,000 (2003–2006).
36. *Department of Defense (Army STTR), W911INF-05-C-0107*. A Scalable System for Enormous Dataset Volume Visualization on Commodity Hardware (Phase I). *Visual Influence PI: D. Weinstein; U. Utah PI: Cláudio Silva; Co-PI: Juliana Freire*. US\$ 100,000. (2005)

## Publications

*According to Google Scholar, my h-number is 61 and my work has received over 16,300 citations as of 3/8/2022.*

## Books Co-Authored

- [1] *Reproducibility and Replicability in Science*. National Academies of Sciences, Engineering, and Medicine. 2019. Washington, DC. The National Academies Press.  
<https://doi.org/10.17226/25303>
- [2] *Semantica*. R. Larson, D. Warren, J. Freire, K. Sagonas and P. Gomez. MIT Press, 1997. ISBN: 0-262-62117-7
- [3] *Syntactica*. R. Larson, D. Warren, J. Freire and K. Sagonas. MIT Press, 1996. ISBN: 0-262-62106-1

## Journals

- [4] *The Seattle report on database research*. Daniel Abadi, Anastasia Ailamaki, David G. Andersen, Peter Bailis, Magdalena Balazinska, Philip A. Bernstein, Peter A. Boncz, Surajit Chaudhuri, Alvin Cheung, AnHai Doan, Luna Dong, Michael J. Franklin, Juliana Freire, Alon Y. Halevy, Joseph M. Hellerstein, Stratos Idreos, Donald Kossmann, Tim Kraska, Sailesh Krishnamurthy, Volker Markl,

- Sergey Melnik, Tova Milo, C. Mohan, Thomas Neumann, Beng Chin Ooi, Fatma Ozcan, Jignesh M. Patel, Andrew Pavlo, Raluca A. Popa, Raghu Ramakrishnan, Christopher Ré, Michael Stonebraker, Dan Suciu. *Commun. ACM* 65(8): 72-79 (2022).
- [5] *The Right Tool for the Job: Data-Centric Workflows in Vizier*. Oliver Kennedy, Boris Glavic, Juliana Freire, Mike Brachmann: *IEEE Data Eng. Bull.* 45(3): 129-144 (2022).
- [6] *Diversity and Inclusion Activities in Database Conferences: A 2021 Report*. Sihem Amer-Yahia, Yael Amsterdamer, Sourav S. Bhowmick, Angela Bonifati, Philippe Bonnet, Renata Borovica-Gajic, Barbara Catania, Tania Cerquitelli, Silvia Chiusano, Panos K. Chrysanthis, Carlo Curino, Jérôme Darmont, Amr El Abbadi, Avrielia Floratou, Juliana Freire, Alekh Jindal, Vana Kalogeraki, Georgia Koutrika, Arun Kumar, Sujaya Maiyya, Alexandra Meliou, Madhulika Mohanty, Felix Naumann, Nele Sina Noack, Fatma Özcan, Liat Peterfreund, Wenny Rahayu, Wang-Chiew Tan, Yuanyuan Tian, Pinar Tözün, Genoveva Vargas-Solar, Neeraja J. Yadwadkar, Meihui Zhang. *SIGMOD Rec.* 51(2): 69-73 (2022).
- [7] *The Right Tool for the Job: Data-Centric Workflows in Vizier*. Oliver Kennedy, Boris Glavic, Juliana Freire, and Michael Brachmann. *IEEE Data Engineering Bulletin*, vol 5, no. 3, 2022.
- [8] *BugDoc: Iterative debugging and explanation of pipelines*. Raoni Lourenço, Juliana Freire, Eric Simon, Gabriel Weber, and Dennis Shasha. *The VLDB Journal*, 2022.
- [9] *Understanding and improving the quality and reproducibility of Jupyter notebooks*. João Felipe Pimentel, Leonardo Murta, Vanessa Braganholo, and Juliana Freire. *Empirical Software Engineering* volume 26, Article number: 65 (2021).
- [10] *Interactive Data Visualization in Jupyter Notebooks*. Jorge Piazzentin Ono, Juliana Freire, and Claudio Silva. In *IEEE Computing in Science & Engineering*, vol. 23, no. 2, pp. 99-106, 2021.
- [11] *PipelineProfiler : A Visual Analytics Tool for the Exploration of AutoML Pipelines*. Jorge Piazzentin Ono, Sonia Castelo, Roque Lopez, Enrico Bertini, Juliana Freire, and Claudio Silva. In *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 27, no. 2, pp. 390-400, 2021.
- [12] *Winds from Seattle: Database Research Directions*. Peter Bailis, Magda Balazinska, Xin Luna Dong, Juliana Freire, Raghu Ramakrishnan, Michael Stonebraker, Joseph M. Hellerstein. In *Proceedings of the VLDB Endowment (PVLDB)*, vol. 13, no. 12, 2020.
- [13] *Exploring Reproducibility in Visualization*. Jean-Daniel Fekete and Juliana Freire. *IEEE Computer Graphics and Applications*, vol. 40, no. 5, pp. 108-119, 2020.
- [14] *Data-Driven Domain Discovery for Structured Datasets*. Masayo Ota, Heiko Muller, Juliana Freire, and Divesh Srivastava. *Proceedings of the VLDB Endowment*, 13(7), pp. 953-967, 2020.
- [15] *Efficient Discovery of Meaningful Outlier Relationships*. Aline Bessa, Juliana Freire, Divesh Srivastava, and Tamraparni Dasu *ACM Transactions on Data Science*, vol. 1, no. 2, article no. 12, 2020.
- [16] *The Seattle Report on Database Research*. Daniel Abadi, Anastasia Ailamaki, David Andersen, Peter Bailis, Magdalena Balazinska, Philip Bernstein, Peter Boncz, Surajit Chaudhuri, Alvin Cheung, Anh-Hai Doan, Luna Dong, Michael J. Franklin, Juliana Freire, Alon Halevy, Joseph M. Hellerstein, Stratos Idreos, Donald Kossmann, Tim Kraska, Sailesh Krishnamurthy, Volker Markl, Sergey Melnik, Tova Milo, C. Mohan, Thomas Neumann, Beng Chin Ooi, Fatma Ozcan, Jignesh Patel, Andrew Pavlo, Raluca Popa, Raghu Ramakrishnan, Christopher Ré, Michael Stonebraker and Dan Suciu. *SIGMOD Record* vol. 48, No. 4, pg. 44-53 2019.

- [17] *A Survey on Collecting, Managing, and Analyzing Provenance from Scripts*. João Felipe Pimentel, Juliana Freire, Leonardo Murta, and Vanessa Braganholo. *ACM Surveys*, 52(3), 2019.
- [18] *Spatio-Temporal Urban Data Analysis: A Visual Analytics Perspective*. Harish Doraiswamy, Juliana Freire, Marcos Lage, Fabio Miranda, and Claudio Silva. *IEEE Computer Graphics and Applications*, 38(5):26-35, 2018.
- [19] *Time Lattice: A Data Structure for the Interactive Visual Analysis of Large Time Series*. Fábio Miranda, Marcos Lage, Harish Doraiswamy, Charlie Mydlarz, Justin Salamon, Yitzchak Lockerman, Juliana Freire, and Cláudio Silva. *Computer Graphics Forum*, 37: 23-35, 2018. doi:10.1111/cgf.13398 (Presented at *Eurographics Conference on Visualization (EuroVis)*, 2018.)
- [20] *Provenance and the Different Flavors of Reproducibility*. J. Freire and F. Chirigati. *IEEE Data Engineering Bulletin*, 41(1):15-26, 2018.
- [21] *ARIES: Enabling Visual Exploration and Organization of Art Image Collections*. Lhaylla Crissaff, Louisa Ruby, Samantha Deutch, Luke DuBois, Jean-Daniel Fekete, Juliana Freire, and Cláudio T. Silva. *IEEE Computer Graphics & Applications (CG&A)*, 38(1): 91-108, 2018.
- [22] *Connecting Visualization and Data Management Research (Dagstuhl Seminar 17461)*. Remko Chang, Jean-Daniel Fekete, Juliana Freire and Carlos Scheidegger. *Dagstuhl Reports* 7(11), 2018.
- [23] *Data Quality: The Role of Empiricism*. Shazia Sadiq, Tamraparni Dasu, Xin Luna Dong, Juliana Freire, Ihab F. Ilyas, Sebastian Link, Renee J. Miller, Felix Naumann, Xiaofang Zhou, and Divesh Srivastava. *ACM SIGMOD Record Volume 46, Number 4, December 2017*.
- [24] *GPU Rasterization for Real-Time Spatial Aggregation over Arbitrary Polygons*. Eleni Tzirita Zacharotou, Harish Doraiswamy, Anastasia Ailamaki, Cláudio Silva, and Juliana Freire. In *PVLDB vol. 11(2)*, pp. 352-365, 2017.
- [25] *Real-time understanding of humanitarian crises via targeted information retrieval*. Kien Pham, Prasanna Sattigeri, Amit Dhurandhar, Arpith Jacob, Maja Vukovic, Patrice Chataigner, Juliana Freire, Aleksandra Mojsilović, and Kush Varshney. *IBM Journal of Research and Development*, 61(6), pp 7:1–7:12, 2017.
- [26] *noWorkflow: a Tool for Collecting, Analyzing, and Managing Provenance from Python Scripts*. João F. N. Pimentel, Juliana Freire, Vanessa Braganholo, and Leonardo Murta. In *PVLDB vol 10(12)*, pp. 1841-1844, 2017.
- [27] *ReproZip: The Reproducibility Packer*. Rémi Rampin, Fernando Chirigati, Dennis Shasha, Juliana Freire and Vicky Steeves. *The Journal of Open Source Software*. <http://dx.doi.org/10.21105/joss.00107>
- [28] *STaRS: Simulating Taxi Ride Sharing at Scale*. Masayo Ota, Huy Vo, Cláudio T. Silva, and Juliana Freire. *IEEE Transactions on Big Data*, 3(3) 349–361, 2017.
- [29] *Reproducibility of Data-Oriented Experiments in e-Science (Dagstuhl Seminar 16041)*. Juliana Freire, Norbert Fuhr and Andreas Rauber. *Dagstuhl Reports* 6(1):108–159, 2016.
- [30] *Exploring What not to Clean in Urban Data: A Study Using New York City Taxi Trips*. J. Freire, A. Bessa, F. Chirigati, H. Vo, and K. Zhao. *IEEE Data Engineering Bulletin*, 39(2):63-77, 2016.



- [31] *A collaborative approach to computational reproducibility*. Fernando Chirigati, Rebecca Capone, Rémi Rampin, Juliana Freire, Dennis Shasha Information Systems, Volume 59 (C): 95-97, 2016.
- [32] *Finding seeds to bootstrap focused crawlers*. Karane Vieira, Luciano Barbosa, Altigran Soares da Silva, Juliana Freire, Edleno Moura. World Wide Web 19(3): 449-474, 2016.
- [33] *Visually Exploring Transportation Schedules*. Cesar Palomo, Zhan Guo, Cláudio T. Silva, and Juliana Freire. IEEE Transactions on Visualization and Computer Graphics (TVCG) 22(1): 170-179, 2016.
- [34] *YesWorkflow: A User-Oriented, Language-Independent Tool for Recovering Workflow Information from Scripts*. Timothy McPhillips, Tianhong Song, Tyler Kolisnik, Steve Aulenbach, Khalid Belhajjame, R. Kyle Bocinsky, Yang Cao, James Cheney, Fernando Chirigati, Saumen Dey, Juliana Freire, Christopher Jones, James Hanken, Keith W. Kintigh, Timothy A. Kohler, David Koop, James A. Macklin, Paolo Missier, Mark Schildhauer, Christopher Schwalm, Yaxing Wei, Mark Bieda, Bertram Ludäscher. International Journal of Digital Curation, 10(1):298-313, 2015.
- [35] *Exploring Traffic Dynamics in Urban Environments Using Vector-Valued Functions*. Jorge Poco, Harish Doraiswamy, Huy. T. Vo, João L. D. Comba, Juliana Freire and Cláudio T. Silva. Computer Graphics Forum, 34 (3): 161-170, 2015.
- [36] *Riding from Urban Data to Insight Using New York City Taxis*. J. Freire, C. Silva, H. Vo, H. Doraiswamy, N. Ferreira, and J. Poco. IEEE Data Engineering Bulletin, 37(4):43-55, 2014.
- [37] *Structured Open Urban Data: Understanding the Landscape*. L. Barbosa, K. Pham, C. Silva, M. Vieira, and J. Freire. Big Data Journal, 2(3): 144-154, 2014.
- [38] *Using Topological Analysis to Support Event-Guided Exploration in Urban Data*. H. Doraiswamy, N. Ferreira, T. Damoulas, J. Freire, and C. Silva. IEEE Transactions on Visualization and Computer Graphics (TVCG), 20(12): 2634-2643, 2014.
- [39] *A computational reproducibility benchmark*. F. Chirigati, M. Troyer, D. Shasha, and J. Freire. IEEE Data Engineering Bulletin, 36(4):54-59, 2013.
- [40] *Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips*. N. Ferreira, J. Poco, H. Vo, J. Freire, and C. Silva. IEEE Transactions on Visualization and Computer Graphics (TVCG) 19(12): 2149-58, 2013.
- [41] *Making Computations and Publications Reproducible with VisTrails*. J. Freire and Cláudio Silva. Computing in Science and Engineering 14(4): 18-25, 2012.
- [42] *BirdVis: Visualizing and Understanding Bird Populations*. N. Ferreira, L. Lins, D. Fink, S. Kelling, C. Wood, J. Freire, and C. Silva. IEEE Transactions on Visualization and Computer Graphics (TVCG) 17(12): 2374-2383, 2011.
- [43] *Multilingual Schema Matching for Wikipedia Infoboxes*. T. Nguyen, V. Moreira, H. Nguyen, H. Nguyen, and J. Freire. In Proceedings of the VLDB Endowment (PVLDB), 5(2): 133-144, 2011.
- [44] *XML Management for Bioinformatics Applications*. Lena Strömbäck and J. Freire. Computing in Science and Engineering 13(5): 12-23, 2011.
- [45] *Guest Editors' Introduction: Provenance in Web Applications*. Geetika T. Lakshmanan, Francisco Curbera, J. Freire, and Amit P. Sheth. IEEE Internet Computing 15(1): 17-21, 2011.

- [46] *Repeatability and Workability Evaluation of SIGMOD 2011*, Philippe Bonnet, Stefan Manegold, Matias Bjorling, Wei Cao, Javier Gonzales, Joel Granados, Nancy Hall, Stratos Idreos, Milena Ivanova, Ryan Johnson, David Koop, Tim Kraska, René Müller, Dan Olteanu, Paolo Papotti, Christine Reilly, Dimitris Tsirogiannis, Cong Yu, Juliana Freire and Dennis Shasha. In SIGMOD Record, vol. 40, no. 2, pp. 45-48, 2011.
- [47] *Exploring the Coming Repositories of Repeatable Experiments: Challenges and Opportunities*, J. Freire, P. Bonnet, and D. Shasha. In Proceedings of the VLDB Endowment (PVLDB), vol. 4, no. 12, pp. 1494-1497, 2011.
- [48] *Synthesizing Products for Online Catalogs*, H. Nguyen, A. Fuxman, S. Pappas, J. Freire, R. Agrawal. In Proceedings of the VLDB Endowment (PVLDB), vol. 4, no. 7, pp. 409-418, 2011.
- [49] *The ALPS project release 2.0: Open source software for strongly correlated systems*. B. Bauer, L. D. Carr, H.G. Evertz, A. Feiguin, J. Freire, S. Fuchs, L. Gamper, J. Gukelberger, E. Gull, S. Guertler, A. Hehn, R. Igarashi, S.V. Isakov, D. Koop, P.N. Ma, P. Mates, H. Matsuo, O. Parcollet, G. Pawłowski, J.D. Picon, L. Pollet, E. Santos, V.W. Scarola, U. Schollwöck, C. Silva, B. Surer, S. Todo, S. Trebst, M. Troyer, M.L. Wall, P. Werner, S. Wessel. Journal of Statistical Mechanics: Theory and Experiment (JSTAT), 2011(05), 2011.
- [50] *The Open Provenance Model - Core Specification (v1.1)*, L. Moreau, B. Clifford, J. Freire, Y. Gil, P. Groth, J. Futrelle, N. Kwasnikowska, S. Miles, P. Missier, J. Myers, Y. Simmhan, E. Stephan, and J. Van den Bussche. Future Generation Computer Systems, 27(6), 743-756, 2011.
- [51] *Using VisTrails and Provenance for Teaching Scientific Visualization*. C. Silva, E. Anderson, E. Santos, and J. Freire, Computer Graphics Forum, 30(1), pp. 75-84, 2011.
- [52] *Indexing Web Form Constraints*, R. Mello, R. Pinnamaneni and J. Freire. Journal of Information and Data Management (JIDM), 1(2), pp. 343-358, 2010.
- [ [53]] *Querying Structured Information Sources on the Web*, S. Mergen, J. Freire, and C. Heuser. In International Journal of Metadata Semantics and Ontologies, 5(3), pp. 208-221, 2010.
- [54] *Siphoning Hidden-Web Data through Keyword-Based Interfaces*, L. Barbosa and J. Freire. Journal of Information and Data Management (JIDM) special issue including the most cited papers of the Brazilian Database Symposium, 1(1), pp. 133-144, 2010.
- [55] *Siphoning Hidden-Web Data through Keyword-Based Interfaces: Retrospective*, L. Barbosa and J. Freire. Journal of Information and Data Management (JIDM) special issue including the most cited papers of the Brazilian Database Symposium, 1(1), pp. 145-146, 2010.
- [56] *VisMashup: Streamlining the Creation of Custom Visualization Applications*, E. Santos, L. Lins, J. Ahrens, J. Freire, and C. Silva. IEEE Transactions on Visualization and Computer Graphics, 15(6), pp. 1539-1546, 2009.
- [57] *On Finding Templates on Web Collections*, K. Vieira, A. L. da C. Carvalho, K. Berlt, E. S. de Moura, A. S. da Silva and J. Freire. In World Wide Web Journal, 12(2), pp. 171-211, 2009.
- [58] *Learning to Extract Form Labels*, H. Nguyen, T. Nguyen, and J. Freire. In Proceedings of the VLDB Endowment, 1(1), pp. 684-694, 2008. *Papers from the International Conference on Very Large Databases (VLDB)*.

- [59] *VisComplete: Automating Suggestions for Visualization Pipeline*, D. Koop, C. E. Scheidegger, S. P. Callahan, H. T. Vo, J. Freire, C. Silva. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), pp. 1691-1698, 2008. *Papers from the IEEE Visualization Conference*.
- [60] *Provenance for Computational Tasks: A Survey*, J. Freire, D. Koop, E. Santos, C. Silva. In *IEEE Computing in Science & Engineering*, 10(3), pp. 11-21, 2008.
- [61] *Software Infrastructure for Exploratory Visualization and Data Analysis: Past, Present, and Future*, C. Silva and J. Freire. In *Journal of Physics: Conference Series*, vol. 125, SciDAC 2008 Conference, 2008.
- [62] *Scientific Exploration in the Era of Ocean Observatories*, A. Baptista, B. Howe, J. Freire, D. Maier, and C. Silva. In *IEEE Computing in Science & Engineering*, 10(3), pp. 53-58, 2008.
- [63] *Tackling the Provenance Challenge One Layer at a Time*, C. E. Scheidegger, D. Koop, E. Santos, H. T. Vo, S. P. Callahan, J. Freire, C. Silva. *Concurrency and Computation: Practice and Experience*, 20(5), pp. 473-483, 2008.
- [64] *The First Provenance Challenge*, L. Moreau, B. Ludäscher, I. Altintas, R. Barga, S. Bowers, S. P. Callahan, G. Chin Jr., B. Clifford, S. Cohen, S. Cohen-Boulakia, S. Davidson, E. Deelman, L. Di-giampietri, I. Foster, J. Freire, J. Frew, J. Futrelle, T. Gibson, Y. Gil, C. Goble, J. Golbeck, P. Groth, D. A. Holland, S. Jiang, J. Kim, D. Koop, A. Krenek, T. McPhillips, G. Mehta, S. Miles, D. Metzger, S. Munroe, J. Myers, B. Plale, N. Podhorszki, V. Ratnakar, E. Santos, C. E. Scheidegger, K. Schuchardt, M. Seltzer, Y. L. Simmhan, C. Silva, P. Slaughter, E. Stephan, R. Stevens, D. Turi, H. T. Vo, M. Wilde, J. Zhao, and Y. Zhao. In *Concurrency and Computation: Practice and Experience*, 20(5), pp. 409-418, 2008.
- [65] *Provenance in Scientific Workflow Systems*, S. Davidson, S. Cohen-Boulakia, A. Eyal, B. Ludascher, T. McPhillips, S. Bowers, M. K. Anand, and J. Freire. *IEEE Data Engineering Bulletin*, 32(4), pp. 44-50, 2007.
- [66] *Provenance for Visualizations: Reproducibility and Beyond*, C. Silva, J. Freire, S. P. Callahan. In *IEEE Computing in Science & Engineering*, 9(5): 82-90 (2007).
- [67] *Querying and Creating Visualizations by Analogy*, C. E. Scheidegger, H. T. Vo, D. Koop, J. Freire, C. Silva. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), pp. 1560-1567, 2007. *Best paper award at the IEEE Visualization Conference*.
- [68] *Managing XML Data: An Abridged Overview*, J. Freire and M. Benedikt. In *IEEE Computing in Science & Engineering*, pp. 12-19, July/August 2004 (Vol. 6, No. 4).
- [69] *Integrating Network Devices in a Meta-Directory: The MetaComm Experience*, J. Freire, D. Lieuwen and Joann Ordille. *Information Systems*, 27(3): 193-217 (2002).
- [70] *Automating Web Navigation with the WebVCR*, V. Anupam, J. Freire, B. Kumar and D. Lieuwen. *Computer Networks* 33(1-6): 503-517 (2000).
- [71] *Beyond Depth-First: Improving Tabled Logic Programs through Alternative Scheduling Strategies*, J. Freire, T. Swift and D. Warren. In *Journal of Functional and Logic Programming*, vol 1998, number 3 MIT Press, April 1998.

**Conferences and Workshops**

- [72] *AlphaD3M: An Open-Source AutoML Library for Multiple ML Tasks*. Roque López, Raoni Lourenço, Sonia Castelo, Aécio Santos, Jorge Piazzentin Ono, Cláudio Silva, Juliana Freire. In ACM SIGMOD Workshop on Data Management for End-to-End Machine Learning, pp. 1-11, 2023.
- [73] *Weighted Minwise Hashing Beats Linear Sketching for Inner Product Estimation*. Aline Bessa, Majid Daliri, Juliana Freire, Cameron Musco, Christopher Musco, Aécio Santos, Haoxiang Zhang. Proceedings of the ACM Symposium on Principles of Database Systems (PODS), pp. 169-181, 2023.
- [74] *Using Pipeline Performance Prediction to Accelerate AutoML Systems*. Haoxiang Zhang, López, Aécio Santos, Jorge Piazzentin Ono, Aline Bessa, Juliana Freire. In AutoML Conference, 2023. *To appear*.
- [75] *DataPrism: Exposing Disconnect between Data and Systems*. Sainyam Galhotra, Anna Fariha, Raoni Lourenço, Juliana Freire, Alexandra Meliou, Divesh Srivastava. In ACM SIGMOD, 2022.
- [76] *What we learned about The Gateway Pundit from its own web traffic data*. Zhouhan Chen, Haohan Chen, Joshua A. Tucker, Juliana Freire, Jonathan Nagler. In ICWSM Workshops, 2022.
- [77] *A Sketch-based Index for Correlated Dataset Search*. Aécio S. R. Santos, Aline Bessa, Christopher Musco, Juliana Freire. ICDE 2022: 2928-2941.
- [78] *SPADE: GPU-Powered Spatial Database Engine for Commodity Hardware*. Harish Doraiswamy, Juliana Freire. IEEE ICDE 2022: 2669-2681.
- [79] *DSDD: Domain-Specific Dataset Discovery on the Web*. Haoxiang Zhang, Aécio Santos and Juliana Freire. In ACM CIKM, pp. 2527-2536, 2021.
- [80] *Interactive Audience Expansion On Large Scale Online Visitor Data*. Gromit Yeuk-Yin Chan, Tung Mai, Anup Rao, Ryan A. Rossi, Fan Du, Claudio Silva, Juliana Freire. In ACM KDD, pp. 2621-2631, 2021.
- [81] *Auctus: A Dataset Search Engine for Data Discovery and Augmentation*. Sonia Castelo, Rémi Rampin, Aécio Santos, Aline Bessa, Fernando Chirigati and Juliana Freire. In VLDB, 2021. *To appear*.
- [82] *From Papers to Practice: The openclean Open-Source Data Cleaning Library*. Heiko Müller, Sonia Castelo, Munaf Qazi, and Juliana Freire. In VLDB, 2021. *To appear*.
- [83] *Correlation Sketches for Approximate Join-Correlation Queries*. Aécio Santos, Aline Bessa, Fernando Chirigati, Christopher Musco, and Juliana Freire. In ACM SIGMOD, pp. 2384-2388, 2021.
- [84] *An Ecosystem of Tools for Modeling Political Violence*. Aline Bessa, Sonia Castelo, Rémi Rampin, Aécio Santos, Mike Shoemate, Vito D’Orazio, and Juliana Freire. In ACM SIGMOD, pp. 1531-1544, 2021.
- [85] *An automatic framework to continuously monitor multi-platform information spread*. Zhouhan Chen, Kevin Aslett, Jen Rosiere Reynolds, Juliana Freire, Jonathan Nagler, Joshua A. Tucker and Richard Bonneau. In Workshop on Misinformation Integrity in Social Networks (Misinfo), 2021. *To appear*.
- [86] *Discovering and Measuring Malicious URL Redirection Campaigns from Fake News Domains*. Zhouhan Chen and Juliana Freire. International Workshop on Traffic Measurements for Cybersecurity (WTMC), 2021. *To appear*.

- [87] *Next 5 years: What opportunities should the database community seize to maximize their impact?* Magda Balazinska, Surajit Chaudhuri, Anastasia Ailamaki, Juliana Freire, Sailesh Krishnamurthy, and Michael Stonebraker. ACM SIGMOD, 2020.
- [88] *A GPU-friendly Geometric Data Model and Algebra for Spatial Queries.* Harish Doraiswamy and Juliana Freire. In ACM SIGMOD, pp. 1875-1885, 2020.
- [89] *BugDoc: Algorithms to Debug Computational Processes.* Raoni Lourenco, Juliana Freire, and Dennis Shasha. In ACM SIGMOD, pp. 463-478, 2020.
- [90] *BugDoc: A System for Debugging Computational Pipelines.* Raoni Lourenco, Juliana Freire, and Dennis Shasha. In ACM SIGMOD, pp. 2733-2736, 2020.
- [91] *Real-Time Clustering for Large Sparse Online Visitor Data.* Gromit Yeuk-Yin Chan, Fan Du, Ryan Rossi, Anup Rao, Eunye Koh, Claudio Silva and Juliana Freire The Web Conference (WWW), pp. 1049-1059, 2020.
- [92] *Detecting, Combating, and Identifying Dis and Mis-information.* Nadya Bliss and Juliana Freire AAAS Annual Meeting, 2020.
- [93] *Proactive Discovery of Fake News Domains from Real-Time Social Media Feeds.* Zhouhan Chen and Juliana Freire WWW Workshop on CyberSafety – Proceedings of the Web Conference 2020, pp. 584-592, 2020. *Best paper award.*
- [94] *Your notebook is not crumby enough, REPLace it.* Mike Brachmann, William Spoth, Oliver A Kennedy, Boris Glavic, Heiko Mueller, Sonia Castelo Quispe, Carlos Bautista, Juliana Freire. In Conference on Innovative Data Systems Research (CIDR), 2020.
- [95] *An Information Theory Approach to Detect Media Bias in News Websites.* Victoria Patricia Aires, Juliana Freire, Fabiola G. Nakamura, Altigran Silva, and Eduardo Nakamura In ACM KDD Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM), 2020.
- [96] *Understanding Spatio-Temporal Urban Processes.* Lais M. A. Rocha, Aline Bessa, Fernando Chirigati, Eugene OFriel, Mirella Moro, and Juliana Freire. IEEE International Conference on Big Data, pp. 563-572, 2019.
- [97] *Automatic Machine Learning by Pipeline Synthesis using Model-Based Reinforcement Learning and a Grammar.* Iddo Drori, Yamuna Krishnamurthy, Remi Rampin, Raoni de Paula Lourencco, Jorge Piazentin Ono, Kyunghyun Cho, Claudio Silva, Juliana Freire. In ICML Workshop on Automated Machine Learning, 2019.
- [98] *Debugging Machine Learning Pipelines.* Raoni Lourenco, Juliana Freire, and Dennis Shasha. In ACM SIGMOD Workshop on Data Management for End-to-End Machine Learning (DEEM), 2019.
- [99] *Visus: An Interactive System for Automatic Machine Learning Model Building and Curation.* Aécio Santos, Sonia Castelo, Cristian Felix, Jorge Ono, Bowen Yu, Ray Hong, Cláudio Silva, Enrico Bertini, Juliana Freire. In ACM SIGMOD Workshop on Human-In-the-Loop Data Analytics (HILDA), 2019.
- [100] *Bootstrapping Domain-Specific Content Discovery on the Web.* Kien Pham, Aécio Santos, Juliana Freire. In The Web Conference (WWW), pp. 1476–1486, 2019.
- [101] *A Topic-Agnostic Approach to Identify Fake News Pages.* Sonia Castelo, Thais Almeida, Anas Elghafari, Aecio Santos, Kien Pham, Eduardo Nakamura, and Juliana Freire. In The Web Conference (WWW) MisinfoWorkshop, pp. 975–980, 2019.

- [102] *A Large-scale Study about Quality and Reproducibility of Jupyter Notebooks*. João Felipe Nicolaci Pimentel, Vanessa Braganholo, Leonardo Murta, and Juliana Freire. In Proceedings of Mining Software Repositories (MSR), 2019. *Distinguished Paper Award and FOSS Impact Paper Award*.
- [103] *Data Debugging and Exploration with Vizier*. Mike Brachmann, Carlos Bautista, Sonia Castelo, Su Feng, Juliana Freire, Boris Glavic, Oliver Kennedy, Heiko Müller, Rémi Rampin, William Spoth, Ying Yang. In ACM SIGMOD, pp. 1877–1880, 2019.
- [104] *AlphaD3M: Machine learning pipeline synthesis*. Iddo Drori, Yamuna Krishnamurthy, Remi Rampin, Raoni de Paula Lourenco, Jorge Piazentin Ono, Kyunghyun Cho, Claudio Silva, Juliana Freire. In ICML Workshop on AutoML, 2018. *Invited for oral presentation*.
- [105] *Interactive Visual Exploration of Spatio-Temporal Urban Data Sets using Urbane*. Harish Doraiswamy, Eleni Tzirita Zacharitou, Fabio Miranda, Marcos Lage, Anastasia Ailamaki, Cláudio T. Silva, and Juliana Freire. In ACM SIGMOD, pp. 1693-1696, 2018. *ACM SIGMOD Best Demo Award*.
- [106] *Learning to Discover Domain-Specific Web Content*. Kien Pham, Aécio Santos, and Juliana Freire. In ACM International Conference on Web Search and Data Mining (WSDM), pp. 432–440, 2018.
- [107] *Querying and Exploring Polygamous Relationships in Urban Spatio-Temporal Data Sets*. Yeuk-Yin Chan, F. Chirigati, H. Doraiswamy, C. Silva, and J Freire. In ACM SIGMOD, pp. 1643-1646, 2017.
- [108] *Predicting Taxi Demand at High Spatial Resolution: Approaching the Limit of Predictability*. Kai Zhao, Denis Khryashchev, Juliana Freire, Claudio Silva, and Huy Vo. IEEE International Conference on Big Data, 2016.
- [109] *Anonymizing NYC Taxi Data: Does It Matter?* Marie Douriez, Harish Doraiswamy, Claudio Silva, and Juliana Freire. In IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2016. *Best paper honorable mention*.
- [110] *A Unified Index for Spatio-Temporal Keyword Queries*. T. Hoang-Vu, H. Vo, and J. Freire. In ACM CIKM, pp. 135-144, 2016.
- [111] *Interactive Exploration for Domain Discovery on the Web*. Y. Krishnamurthy, K. Pham, A. Santos, and J. Freire. In ACM KDD Workshop on Interactive Data Exploration and Analytics (IDEA), pp. 64-71, 2016.
- [112] *Understanding Website Behavior based on User Agent*. Kien Pham, Aécio S. R. Santos, and Juliana Freire. In ACM SIGIR, pp. 1053-1056, 2016.
- [113] *Data Polygamy: The Many-Many Relationships among Urban Spatio-Temporal Data Sets*. F. Chirigati, H. Doraiswamy, T. Damoulas, and J Freire. In ACM SIGMOD, pp. 1011-1025, 2016. *Received “Most reproducible paper award in 2017”*.
- [114] *ReproZip: Computational Reproducibility With Ease*. F. Chirigati, R. Rampin, D. Shasha, and J. Freire. In ACM SIGMOD, pp. 2085-2088, 2016.
- [115] *The Exception that Improves the Rule*. J. Freire, B. Glavic, O. Kennedy, and H Mueller. In ACM SIGMOD Workshop on Human-In-the-Loop Data Analytics (HILDA), 2016.
- [116] *A GPU-Based Index to Support Interactive Spatio-Temporal Queries over Historical Data*. Harish Doraiswamy, Huy T. Vo, Cláudio Silva, and Juliana Freire. In Proceedings of IEEE International Conference on Data Engineering (ICDE), pp. 1086-1097, 2016.

- [117] *Virtual Lightweight Snapshots for Consistent Analytics in NoSQL Stores*, Fernando Chirigati, Jerome Simeon, Martin Hirzel, and Juliana Freire. In Proceedings of IEEE International Conference on Data Engineering (ICDE), pp. 1310-1321, 2016.
- [118] *A First Study on Temporal Dynamics of Topics on the Web*, Aécio Santos, Bruno Pasini and Juliana Freire. In Proceedings of WWW Workshop on Temporal Web Analytics, pp. 849-854, 2016.
- [119] *Prov Viewer: a graph-based visualization tool for interactive exploration of provenance data*. Troy Kohwalter, Thiago Oliveira, Juliana Freire, Esteban Clua, and Leonardo Murta. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pp. 71-82, 2016.
- [120] *Tracking and Analyzing the Evolution of Provenance from Scripts*. João F. N. Pimentel, Juliana Freire, Vanessa Braganholo, and Leonardo Murta. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pp. 16-28, 2016.
- [121] *Fine-Grained Provenance Collection over Scripts Through Program Slicing*. João F. N. Pimentel, Juliana Freire, Vanessa Braganholo, and Leonardo Murta. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pp. 199-203, 2016.
- [122] *A Scalable Approach for Data-Driven Taxi Ride-Sharing Simulation*. Masayo Ota, Huy Vo, Cláudio T. Silva, and Juliana Freire. In Proceedings of IEEE International Conference on Big Data, pp. 888-897, 2015.
- [123] *RioBusData: Outlier Detection in Bus Routes of Rio de Janeiro*. Aline Bessa, Fernando de Mesentier Silva, Rodrigo Frassetto Nogueira, Enrico Bertini, and Juliana Freire. Proceedings of Symposium on Visualization in Data Science (VDS), 2015.
- [124] *Visualizing the Evolution of Module Workflows*. Marcel Hlawatsch, Michael Burch, Fabian Beck, Juliana Freire, Cláudio T. Silva, and Daniel Weiskopf. In Proceedings of the IEEE International Conference on Information Visualisation, pp. 40-49, 2015.
- [125] *Collecting and Analyzing Provenance on Interactive Notebooks: when IPython meets noWorkflow*. João Felipe Nicolaci Pimentel, Vanessa Braganholo, Leonardo Murta, Juliana Freire. In Proceedings of the USENIX Workshop on the Theory and Practice of Provenance (TaPP), 2015.
- [126] *Reproducibility Made Easy*. J. Freire. In Proceedings of EuroRV3: EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization, 2015.
- [127] *An Urban Data Profiler*. D. Ribeiro, H. Vo, J. Freire and C. Silva. In Proceedings of the 24th International Conference on World Wide Web Companion, pp. 1389-1394, 2015. *Presented at the WWW Workshop on Data Science and Smart Cities*.
- [128] *Managing and Reusing Provenance as a Critical Capability for (Data) Scientists*. J. Freire. In Proceedings of the AAAS Annual Meeting, 2015.
- [129] *Should we all be teaching “intro to data science” instead of “intro to databases”?* B. Howe, M. J. Franklin, J. Freire, J. Frew, T. Kraska, R. Ramakrishnan. In Proceedings of ACM SIGMOD, pp. 917-918, 2014.
- [130] *A model project for reproducible papers: critical temperature for the Ising model on a square lattice*. M. Dolfi, J. Gukelberger, A. Hehn, J. Imriska, K. Pakroutski, T. F. Rønnow, M. Troyer, I. Zintchenko, F. Chirigati, J. Freire, and D. Shasha. CoRR abs/1401.2000, 2014.

- [131] *Bridging Vocabularies to Link Tweets and News*. T. Hoang-Vu, A. Bessa, L. Barbosa, and J. Freire. In Proceedings of WebDB, 2014.
- [132] *Towards Understanding Real-Estate Ownership in New York City: Opportunities and Challenges*. T. Hoang-Vu, V. Been, I. Gould Ellen, M. Weselcouch, and J. Freire. In Proceedings of Data Science for Macro-Modeling with Financial and Economic Datasets (DSMM), 2014.
- [133] *noWorkflow: Capturing and Analyzing Provenance of Scripts*. L. Murta, V. Braganholo, F. Chirigati, D. Koop, and J. Freire. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pp. 71-83, 2014.
- [134] *Reorganizing workflow evolution provenance*. D. Koop and J. Freire. In USENIX Workshop on the Theory and Practice of Provenance (TaPP), 2014.
- [135] *ReproZip: Using Provenance to Support Computational Reproducibility*. F. Chirigati, D. Shasha, and J. Freire. In USENIX Workshop on the Theory and Practice of Provenance (TaPP), 2013.
- [136] *Packing Experiments for Sharing and Publication* F. Chirigati, D. Shasha, and J. Freire. In ACM SIGMOD, 2013.
- [137] *Visual Summaries for Graph Collections*. D. Koop, J. Freire, C. Silva. In Proceedings of the IEEE Pacific Visualization Symposium (PacificVis), pp. 57-64, 2013.
- [138] VisTrails provenance traces for benchmarking. Fernando Seabra Chirigati, Juliana Freire, David Koop, Cláudio T. Silva. EDBT/ICDT Workshops 2013: 323-324.
- [139] *Clustering Wikipedia Infoboxes to Discover their Types*. T. Nguyen, H. Nguyen , V. Moreira and J. Freire. In Proceedings of CIKM, pp. 2134-2138, 2012.
- [140] *Computational reproducibility: state-of-the-art, challenges, and database research opportunities*. J. Freire, P. Bonnet, D. Shasha. In Proceedings of SIGMOD, pp. 593-596, 2012.
- [141] *Towards Integrating Workflow and Database Provenance: A Practical Approach*. F. Chirigati and J. Freire. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pp. 11-23, 2012.
- [142] *Designing a Provenance-Based Climate Data Analysis Application*. E. Santos, D. Koop, T. Maxwell, C. Doutriaux, T. Ellqvist, G. Potter, J. Freire, D. Williams and C. Silva. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pp. 214-219, 2012.
- [143] *Parallel Visualization on Large Clusters Using MapReduce*. H. Vo, J. Bronson, B. Summa, J. Comba, J. Freire, B. Howe, V. Pascucci and C. Silva In Large Scale Data Analysis and Visualization Symposium (LDAV), pp. 81-88, 2011.
- [144] *VisCareTrails: Visualizing Trails in the Electronic Health Record with Timed Word Trees, a Pancreas Cancer Use Case*. L. Lins, M. Heilbrun, J. Freire, and C. Silva. Workshop on Visual Analytics in Healthcare (VAHC), 2011.
- [145] *crowdLabs: Social Analysis and Visualization for the Sciences*. P. Mates, E. Santos, J. Freire, and C. Silva. In Proceedings of the 23rd International Conference on Scientific and Statistical Database Management (SSDBM), pp. 555-564, 2011.



- [146] *A Provenance-Based Infrastructure to Support the Life Cycle of Executable Papers*. D. Koop, E. Santos, P. Mates, H. Vo, P. Bonnet, B. Bauer, B. Surer, M. Troyer, D. Williams, J. Tohline, J. Freire and C. Silva. In Proceedings of the International Conference on Computational Science. *Procedia Computer Science*, vol. 4, pp. 648-657, 2011.
- [147] *Provenance-Enabled Data Exploration and Visualization with VisTrails*. C. Silva, J. Freire, E. Santos and E. Anderson. In Proceedings of the 23rd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), pp. 1-9, 2010.
- [148] *PruSM: A Prudent Schema Matching Strategy for Web Forms*. T. Nguyen, H. Nguyen and J. Freire. In Proceedings of CIKM, pp. 1385-1388, 2010.
- [149] *Using Latent-Structure to Detect Objects on the Web*. L. Barbosa and J. Freire. In Proceedings of WebDB, 2010.
- [150] *Querying Wikipedia Documents and Relationships*. H. Nguyen, T. Nguyen, H. Nguyen and J. Freire. In Proceedings of WebDB, 2010.
- [151] *The Provenance of Workflow Upgrades*. D. Koop, C. E. Scheidegger, J. Freire, and C. Silva. In Proceedings of the International Provenance and Annotation Workshop (IPAW), pp. 2-16, 2010.
- [152] *Bridging Workflow and Data Provenance using Strong Links*. D. Koop, E. Santos, B. Bauer, M. Troyer, J. Freire, and C. Silva. In Proceedings of the 22nd International Conference on Scientific and Statistical Database Management (SSDBM), pp. 397-415, 2010.
- [153] *Creating and Exploring Web Form Repositories*, L. Barbosa, H. Nguyen, T. Nguyen, R. Pinnamaneni and J. Freire. In Proceedings of SIGMOD, pp. 1175-1178, 2010.
- [154] *Using VisTrails and Provenance for Teaching Scientific Visualization*, C. Silva, E. Anderson, E. Santos and J. Freire. In Proceedings of the Eurographics Education Program, 2010. *Best Paper Award*.
- [155] *Indexing Relations on the Web*, S. Mergen, J. Freire, and C. Heuser. In Proceedings of the 13th International Conference on Extending Database Technology (EDBT), pp. 430-440, 2010.
- [156] *Towards Supporting Collaborative Data Analysis and Visualization in a Coastal Margin Observatory*, E. Santos, P. Mates, E. Anderson, B. Grimm, J. Freire, C. Silva. In Proceedings of the ACM CSCW Workshop on The Changing Dynamics of Scientific Collaborations, 2010.
- [157] *Enabling Advanced Visualization Tools in a Simulation Monitoring System*, E. Santos, J. Tierny, A. Khan, B. Grimm, L. Lins, J. Freire, V. Pascucci and C. Silva. In Proceedings of the IEEE International Conference on e-Science, pp. 358-365, 2009.
- [158] *Using Mediation to Achieve Provenance Interoperability*, T. Ellkvist, D. Koop, E. Santos, J. Freire, C. Silva, and L. Stromback. In Proceedings of the IEEE 2009 Third International Workshop on Scientific Workflows (SWF 2009). Congress on Services-I, pp.291-298, 2009.
- [159] *Desenvolvimento de Estruturas de Controle Explícito para o SGWfC VisTrails*, Fernando Seabra Chirigati, Rafael Dahis, Sergio Manuel Serra da Cruz, Juliana Freire, Cláudio Silva, Marta Mattoso. Proceedings of the Brazilian Symposium on Databases (SBBD), 2009. *Best Poster Award*.
- [160] *A First Study on Strategies for Generating Workflow Snippets*, T. Ellkvist, L. Stromback, L. Lins, and J. Freire. In Proceedings of the ACM SIGMOD International Workshop on Keyword Search on Structured Data (KEYS), 2009.

- [161] *Using Workflow Medleys to Streamline Exploratory Tasks*, E. Santos, D. Koop, H. Vo, E. Anderson, J. Freire, and C. Silva. In Proceedings of the 21st International Conference on Scientific and Statistical Database Management (SSDBM), pp. 292-301, 2009.
- [162] *Simplifying the Design of Workflows for Large-Scale Data Exploration and Visualization*. J. Freire and C. Silva. In Proceedings of the Microsoft eScience Workshop, 2008.
- [163] *End-to-End eScience: Integrating Workflow, Query, Visualization, and Provenance at an Ocean Observatory*. B. Howe, P. Lawson, R. Bellinger, E. Anderson, E. Santos, J. Freire, C. E. Scheidegger, A. Baptista and C. Silva. In Proceedings of the IEEE International Conference on e-Science, pp. 127-134, 2008.
- [164] *Using Mediation to Achieve Provenance Interoperability (Extended Abstract)*, T. Ellkvist, D. Koop, E. Santos, J. Freire, C. Silva, and L. Stromback. In Proceedings of the IEEE International Conference on e-Science, pp. 398-399, 2008.
- [165] *Siphon++: A Hidden-Web Crawler for Keyword-Based Interfaces (Poster)*, K. Vieira, L. Barbosa, J. Freire and A. Silva. In Proceedings of the ACM 17th Conference on Information and Knowledge Management (CIKM), 2008.
- [166] *Querying Structured Information Sources on the Web*, S. Mergen, C. Heuser and J. Freire. In Proceedings of the International Workshop on Resource Discovery (RED), 2008.
- [167] *A Search Engine for Querying Structured Information Sources on the Web*, S. Mergen, C. Heuser and J. Freire. In Proceedings of the Brazilian Symposium on Databases (SBBD), 2008.
- [168] *A First Study on Clustering Collections of Workflow Graphs*, E. Santos, L. Lins, J. P. Ahrens, J. Freire, C. Silva. In Proceedings of International Provenance and Annotation Workshop (IPAW), pp. 160-173, 2008.
- [169] *Towards Provenance-Enabling ParaView*, S. P. Callahan, J. Freire, C. E. Scheidegger, C. Silva, H. T. Vo. In Proceedings of International Provenance and Annotation Workshop (IPAW), pp. 120-127, 2008.
- [170] *Using Provenance to Support Real-Time Collaborative Design of Workflows*, T. Ellkvist, D. Koop, E. Anderson, J. Freire, C. Silva. In Proceedings of International Provenance and Annotation Workshop (IPAW), pp. 266-279, 2008.
- [171] *Querying and Re-Using Workflows with VisTrails*, C. E. Scheidegger, H. T. Vo, D. Koop, J. Freire, C. Silva. In Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 1251-1254, 2008.
- [172] *Provenance and Scientific Workflows: Challenges and Opportunities*, S. Davidson and J. Freire. In Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 1345-1350, 2008.
- [173] *Examining Statistics of Workflow Evolution Provenance: A First Study*, L. Lins, D. Koop, E. Anderson, S. P. Callahan, E. Santos, C. E. Scheidegger, J. Freire, and C. Silva. In Proceedings of International Conference on Scientific and Statistical Database Management (SSDBM), pp. 573-579, 2008.
- [174] *Automatically Extracting Form Labels*, H. Nguyen, E. Y. Kang, and J. Freire. In Proceedings of IEEE International Conference on Data Engineering (ICDE), pp. 1498-1500, 2008.

- [175] *Towards Enabling Social Analysis of Scientific Data*, J. Freire and Cláudio Silva. In CHI Social Data Analysis Workshop, 2008.
- [176] *Combining Classifiers to Identify Online Databases*, L. Barbosa and J. Freire. In Proceedings of the 16th International World Wide Web Conference (WWW), pp. 431-440, 2007.
- [177] *An Adaptive Crawler for Locating Hidden-Web Entry Points*, L. Barbosa and J. Freire. In Proceedings of the 16th International World Wide Web Conference (WWW), pp. 441-450, 2007.
- [178] *Organizing Hidden-Web Databases by Clustering Visible Web Documents*, L. Barbosa, J. Freire, and A. Silva. In Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 326-335, 2007.
- [179] *Automatically Constructing a Directory of Molecular Biology Databases*, L. Barbosa, S. Tandon, and J. Freire. In Proceedings of the International Workshop on Data Integration in the Life Sciences (DILS), pp. 6-16, 2007.
- [180] *VisTrails: Using Provenance to Streamline Data Exploration*, E. Anderson, S. P. Callahan, D. A. Koop, E. Santos, C. E. Scheidegger, H. T. Vo, J. Freire and C. T. Silva. In Poster Proceedings of the International Workshop on Data Integration in the Life Sciences (DILS), pp. 8, 2007. *Invited for oral presentation.*
- [181] *A Fast and Robust Method for Web Page Template Detection and Removal*, K. Vieira, A. Silva, N. Pinto, E.S. Moura, J. Cavalcanti and J. Freire. In Proceedings of the ACM 15th Conference on Information and Knowledge Management (CIKM), pp. 258-267, 2006.
- [182] *Automatically Constructing Collections of Online Databases (Poster)*, L. Barbosa and J. Freire. In Proceedings of the ACM 15th Conference on Information and Knowledge Management (CIKM), pp. 796-797, 2006.
- [183] *Integrated scientific workflow management for the Emulab network testbed*, E. Eide, T. Stack, L. Stoller, J. Freire and J. Lepreau. In Proceedings of USENIX, pp. 363-368, 2006.
- [184] *VisTrails: Visualization Meets Data Management*, S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C.T. Silva and H. T. Vo. In Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 745-747, 2006.
- [185] *Managing the Evolution of Dataflows with VisTrails*, S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C.T. Silva and H. T. Vo. In Proceedings of IEEE Workshop on Workflow and Data Flow for Scientific Applications (SciFlow), 2006.
- [186] *Visualizing Uncertainty with Uncertainty Multiples*, R. B. Gilbert, F. Tonon, J. Freire, C. Silva, and D. R. Maidment. In ASCE 2006 GeoCongress.
- [187] *Looking at both the Present and the Past to Efficiently Update Replicas of Web Content*, L. Barbosa, A. C. Salgado, F. Carvalho, J. Robin and J. Freire. In Proceedings of the ACM International Workshop on Web Information and Data Management (WIDM), 2005.
- [188] *VisTrails: Enabling Interactive Multiple-View Visualizations*, L. Bavoil, S. P. Callahan, P.J. Crossno, J. Freire, C. E. Scheidegger, C. Silva and H. T. Vo. In IEEE Visualization, pp. 135-142, 2005.
- [189] *Designing Information-Preserving Mapping Schemes for XML*, D. Barbosa, J. Freire and A. Mendelson. In Proceedings of the International Conference on Very Large Databases (VLDB), pp. 109-120, 2005.

- [190] *Searching for Hidden-Web Databases*, L. Barbosa and J. Freire. In Proceedings of the International Workshop on the Web & Databases (WebDB), pp. 1-6, 2005.
- [191] *IMAX: Incremental Maintenance of Schema-based XML Statistics*, M. Ramanath, L. Zhang, J. Freire and J. Haritsa. In Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 273-284, 2005.
- [192] *A Comprehensive Solution to the XML-to-Relational Mapping Problem*, F. Du, S. Amer-Yahia and J. Freire. In Proceedings of the ACM International Workshop on Web Information and Data Management (WIDM), pp. 31-38, 2004.
- [193] *Siphoning Hidden-Web Data through Keyword-Based Interfaces*, L. Barbosa and J. Freire. In Proceedings of Brazilian Symposium on Databases (SBBD), pp. 309-321, 2004.
- [194] *Information Preservation in XML-to-Relational Mappings*, D. Barbosa, J. Freire and A. Mendelzon. In Proceedings of the International XML Database Symposium (XSym), pp. 66-81, 2004.
- [195] *ShreX: Managing XML Documents in Relational Databases*, F. Du, S. Amer-Yahia and J. Freire. In Proceedings of International Conference on Very Large Databases (VLDB), pp. 1297-1300, 2004.
- [196] *Supporting Exploratory Queries in Database-Centric Web Applications*, A. Kadlag, A. V. Wanjari, J. Freire, J. Haritsa. In International Conference on Database Systems for Advanced Applications (DASFAA), pp. 594-605, 2004.
- [197] *A Flexible Infrastructure for Gathering XML Statistics and Estimating Query Cardinality*, J. Freire, M. Ramanath, L. Zhang. In Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 857, 2004.
- [198] *Capturing both Types and Constraints in Data Integration*, M. Benedikt, C.Y. Chan, W. Fan, J. Freire, R. Rastogi. In Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 277-288, 2003.
- [199] *Searching for Efficient XML-to-Relational Configurations*, M. Ramanath, J. Freire, J. Haritsa, P. Roy. In Proceedings of the VLDB XML Database Symposium (XSym), pp. 19-36, 2003.
- [200] *Bridging the XML-Relational Divide with LegoDB: A Demonstration*, P. Bohannon, J. Freire, J. Haritsa, M. Ramanath, P. Roy and J. Siméon. In Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 759-760, 2003.
- [201] *Adaptive XML Shredding: Architecture, Implementation, and Challenges*, J. Freire and J. Siméon. In Proceedings of the VLDB Workshop on Efficiency and Effectiveness of XML Tools and Techniques (EEXTT), pp. 104-116, 2002. Also appears in Efficiency and Effectiveness of XML Tools and Techniques and Data Integration over the Web, LNCS vol 2590, 2003.
- [202] *LegoDB: Customizing Relational Storage for XML Documents*, P. Bohannon, J. Freire, J. Haritsa, M. Ramanath, P. Roy and J. Siméon. In Proceedings of International Conference on Very Large Databases (VLDB), pp. 1091-1094, 2002.
- [203] *StatiX: Making XML Count*, J. Freire, J. Haritsa, M. Ramanath, P. Roy and J. Siméon. In Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 209-218, 2002.
- [204] *From XML Schema to Relations: A Cost-Based Approach to XML Storage*, P. Bohannon, J. Freire, P. Roy and J. Siméon. In Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 64-75, 2002.

- [205] *VeriWeb: A Platform for Automating Web Site Testing*, M. Benedikt, J. Freire and P. Godefroid. In Proceedings of the World Wide Web Conference (WWW) – Web Engineering track, 2002.
- [206] *WebViews: Accessing Personalized Web Content and Services*, J. Freire, B. Kumar and D. Lieuwen. In Proceedings of the World Wide Web Conference (WWW10), pp. 576-586, 2001. *This paper was the best paper runner-up at WWW10 and was also the Selected Area Highlight for the Browsers and Tools area of the Refereed Paper Track.*
- [207] *Efficient Acquisition of Web Data through Restricted Query Interfaces*, S. Byers, J. Freire and C. Silva. In Poster Proceedings of the WWW10 World Wide Web Conference, pp 184-185, 2001.
- [208] *Web Services and Information Delivery for Diverse Environments*, J. Freire and B. Kumar. In Proceedings of VLDB Workshop on Technologies for E-Services, 2000.
- [209] *Automating Web Navigation with the WebVCR*, V. Anupam, J. Freire, B. Kumar and D. Lieuwen. In Proceedings of the WWW9 World Wide Web Conference, pp. 503-517, 2000.
- [210] *MetaComm: a Meta-Directory for Telecommunications*, J. Freire, D. Lieuwen, J. Ordille, L. Garg, M. Holder, H. Urroz, G. Michael, J. Orbach, L. Tucker, Q. Ye and R. Arlein. In Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 211-219, 2000.
- [211] *Personalizing the Web Using Site Descriptions*, V. Anupam, Y. Breitbart, J. Freire and B. Kumar. In Proceedings of DEXA International Workshop on Database & Expert Systems Applications, pp. 732-738, 1999.
- [212] *Making LDAP Active with the LTAP Gateway: Case Study in Providing Telecom Integration and Enhanced Services*, R. Arlein, J. Freire, N. Gehani, D. Lieuwen, and J. Ordille. In Proceedings of the Very Large Databases Conference (VLDB) Workshop on Databases in Telecommunication, pp. 54-73, 1999.
- [213] *A Layered Architecture for Querying Dynamic Web Content*, H. Davulcu, J. Freire, M. Kifer, I.V. Ramakrishnan. In Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 491-502, 1999.
- [214] *Practical Problems in Coupling Deductive Engines with Relational Databases*, J. Freire. In Proceedings of the 5th Workshop on Knowledge Representation meets Databases (KRDB), pp. 11.1-11.7, 1998.
- [215] *Scheduling Strategies in SLG Revisited - Abstract*, J. Freire, T. Swift and D. Warren. Workshop on Tabulation in Parsing and Deduction (TAPD), 1998.
- [216] *Taking I/O Seriously: Resolution Reconsidered for Disk*, J. Freire, T. Swift and D. Warren. In Proceedings of the International Conference on Logic Programming (ICLP), pp. 198-212, 1997.
- [217] *Using Logic Programming to Efficiently Evaluate Complex Queries*, J. Freire. In Proceedings of the Grace Hopper Technical Track, pp. 25-29, 1997.
- [218] *XSB - A System for Efficiently Computing Well-Founded Semantics*. P. Rao, K. Sagonas, T. Swift, D.S. Warren and J. Freire. In Proceedings of the 4th International Conference on Logic Programming and Non-Monotonic Reasoning (LP & NMR), pp. 431-441, 1997.
- [219] *Beyond Depth-First: Improving Tabled Logic Programs through Alternative Scheduling Strategies* J. Freire, T. Swift and D. Warren. In Proceedings of the Eighth International Symposium on Programming Languages, Implementations, Logics and Programs (PLILP), pp. 243-258, 1996.

- [220] *Logic Programming and Databases Integrated at Last? (Poster Abstract)* J. Freire, T. Swift and D. Warren. In Proceedings of the Joint International Conference and Symposium on Logic Programming (JICSLP), pp. 538, 1996.
- [221] *Exploiting Parallelism in Tabled Evaluations* J. Freire, R. Hu, T. Swift and D. Warren. In Proceedings of the Seventh International Symposium on Programming Languages, Implementations, Logics and Programs (PLILP), pp. 115-132, 1995.
- [222] *Parallelizing Tabled Evaluations* J. Freire, R. Hu, T. Swift and D. Warren. In Proceedings of the International Logic Programming Symposium (ILPS) Workshop on Parallel Logic Programming Systems, pp. 18-31, 1994.

### Invited Papers

- [223] *Provenance and Reproducibility*. Fernando Chirigati and Juliana Freire. In Encyclopedia of Database Systems, L. Liu and T. Özsu (Eds.), 2018.
- [224] *Provenance in Workflows*. David Koop, Marta Mattoso, and Juliana Freire. In Encyclopedia of Database Systems, L. Liu and T. Özsu (Eds.), 2018.
- [225] *XML Storage*. Denilson Barbosa, Philip Bohannon, Juliana Freire, Carl-Christian Kanne, Ioana Manolescu, Vasilis Vassalos, Masatoshi Yoshikawa. In Encyclopedia of Database Systems, L. Liu and T. Özsu (Eds.), 2018.
- [226] *Biological Resource Discovery*. Zoé Lacroix, Cartik R. Kothari, Peter Mork, Rami Rifaieh, Mark D. Wilkinson, Juliana Freire, Sarah Cohen Boulakia. In Encyclopedia of Database Systems, L. Liu and T. Özsu (Eds.), 2018.
- [227] *XML Selectivity Estimation*. Maya Ramanath, Juliana Freire, Neoklis Polyzotis. In Encyclopedia of Database Systems, L. Liu and T. Özsu (Eds.), 2018.
- [228] Editor's Letter: Scientific Data Management. IEEE Data Engineering Bulletin, vol. 36, no 4, Dec 2013, pp. 2. Ed. Juliana Freire.
- [229] Editor's Letter: Data Management Beyond Database Systems. IEEE Data Engineering Bulletin, vol. 35, no 2, Sept 2012, pp. 6. Ed. Juliana Freire.
- [230] Letter from the the Associate Editors. Eds. Gustavo Alonso, Juliana Freire. PVLDB 5(7): i-vii (2012).
- [231] Guest Editors' Introduction: Provenance in Web Applications. IEEE Internet Computing, vol. 15, no 1, Jan/Feb 2011, pp. 17-21. Eds. Geetika T. Lakshmanan, Francisco Curbera, Juliana Freire, Amit Sheth.
- [232] *Provenance Management for Data Exploration*, J. Freire. In International Conference on Data Integration in the Life Sciences (DILS), pp. 1-2, 2010. Abstract for keynote talk.
- [233] *Biological Resource Discovery*, Z. Lacroix, C. Kothari, P. Mork, R. Rifaieh, M. Wilkinson, J. Freire and S. Cohen-Boulakia. In Encyclopedia of Database Systems, L. Liu and T. Özsu (Eds.) , 2009.
- [234] *XML Selectivity Estimation*, M. Ramanath, J. Freire and A. Polyzotis. In Encyclopedia of Database Systems, L. Liu and T. Özsu (Eds.), 2009.
- [235] *XML Storage*, D. Barbosa, P. Bohannon, J. Freire, C. Kanne, Y. Manolescu, V. Vassalos and M. Yoshikawa. In Encyclopedia of Database Systems, L. Liu and T. Özsu (Eds.), 2009.

- [236] *Introducing the VisTrails Provenance Explorer Plugin for ParaView*, C. Silva, J. Freire, and J. Schreiner. Kitware Source, July 2009.
- [237] *Provenance Management: Challenges and Opportunities*, J. Freire. In *Datenbanksysteme in Business, Technologie und Web (BTW)*, pp. 4, 2009. Abstract for keynote talk.
- [238] *The Open Provenance Model: An Overview*, L. Moreau, J. Freire, J. Futrelle, R. McGrath, J. Myers and P. Paulson. In *Proceedings of International Provenance and Annotation Workshop (IPAW)*, pp. 323-326, 2008.
- [239] *Managing Rapidly-Evolving Scientific Workflows*, J. Freire, C.T. Silva, S. P. Callahan, E. Santos, C. E. Scheidegger and H. T. Vo. In *Proceedings of International Provenance and Annotation Workshop (IPAW)*, pp. 10-18, 2006. *Invited paper corresponding to keynote talk.*
- [240] Environmental observation and forecasting systems: Vision, challenges and successes of a prototype. A. Baptista, T. Leen, Y. Zhang, A. Chawla, D. Maier, W.C. Feng, W.C. Feng, J. Walpole, C. Silva, and J. Freire, 2003. *Encyclopedia of Physical Science and Technology (RA Meyers, Ed.)*, Academic Press., 5, pp.565-581.
- [241] *Accessing and Personalizing Web Content and Services*, J. Freire. Invited to WWW2002 Workshop on Mobile Search, May 2002. *Paper corresponding to invited talk.*

### Book Chapters

- [242] *Reproducibility using VisTrails*. J. Freire, D. Koop, F. Chirigati, and C. Silva. In *Stodden et al., Implementing Reproducible Research*, 2014.
- [243] *VisTrails*, J. Freire, D. Koop, E. Santos, C. Scheidegger, C. Silva, and H. T. Vo. In *A. Brown and G. Wilson, The Architecture of Open Source Applications*, 2011.
- [244] *Scientific Process Automation and Workflow Management*, B. Ludäscher, I. Altintas, S. Bowers, J. Cummings, T. Critchlow, E. Deelman, D. D. Roure, J. Freire, C. Goble, M. Jones, S. Klasky, T. McPhillips, N. Podhorszki, C. Silva, I. Taylor, and M. Vouk. In *A. Shoshani and D. Rotem, editors, Scientific Data Management: Challenges, Existing Technology, and Deployment, Computational Science Series*, chapter 13. Chapman & Hall/CRC, 2009.
- [245] *Combining Scheduling Strategies in an SLG Evaluation*, J. Freire and D. Warren. In *Parallelism and Implementation of Logic and Constraint Logic Programming*, Inês de Castro Dutra et al (editors). Nova Science Publishers, Inc, 1999.

### Manual, Thesis, Special Issues and Proceedings

- [246] *Proceedings of the Workshop on Human-In-the-Loop Data Analytics, HILDA@SIGMOD 2018*. June, 2018. Eds. Carsten Binnig, Juliana Freire, Eugene Wu.
- [247] *IEEE Data Engineering Bulletin. Special Issue on Data Management for Scientific Applications*. December, 2013. Ed. Juliana Freire.
- [248] *IEEE Data Engineering Bulletin. Special Issue on Data Management beyond Database Systems*. September, 2012. Ed. Juliana Freire.
- [249] *Proceedings of the 6th Alberto Mendelzon International Workshop on Foundations of Data Management*, Ouro Preto, Brazil, June 27-30, 2012. Eds. Juliana Freire and Dan Suciu.

- [250] Provenance in Web Applications. IEEE Internet Computing, vol. 15, no 1, Jan/Feb 2011. Eds. Geetika T. Lakshmanan, Francisco Curbera, Juliana Freire, and Amit Sheth.
- [251] Proceedings of the 19th International Conference on World Wide Web (WWW 2010), Raleigh, North Carolina, USA, April 26-30, 2010. Eds. Michael Rappa, Paul Jones, Juliana Freire, and Soumen Chakrabarti.
- [252] Proceedings of the First International Workshop on the role of Semantic Web in Provenance Management (SWPM 2009), collocated with the 8th International Semantic Web Conference (ISWC-2009), Washington DC, USA, October 25, 2009 Eds. Juliana Freire, Paolo Missier, Satya Sanket Sahoo.
- [253] *Proceedings of International Provenance and Annotation Workshop (IPAW), 2008*, Eds. J. Freire, D. Koop and L. Moreau.
- [254] *Proceedings of the ACM SIGMOD International Workshop on Web and Databases (WebDB), 2003*, Eds. V. Christophides, J. Freire.
- [255] *Scheduling Strategies for Evaluation of Recursive Queries over Memory and Disk-Resident Data*. J. Freire. Ph.D. thesis, Department of Computer Science, State University of New York at Stony Brook, 1997.
- [256] *The XSB Programmer's Manual*. K. Sagonas, T. Swift, D. Warren, J. Freire, P. Rao. Available at <http://xsb.sourceforge.net>.

### Technical Reports

- [257] *Using Provenance to Streamline Data Exploration through Visualization*, S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. Silva and H. T. Vo. SCI Institute Technical Report, No. UUSCI-2006-016, University of Utah, 2006.
- [258] *Visualization in Radiation Oncology: Towards Replacing the Laboratory Notebook*, E. W. Anderson, S. P. Callahan, G. T. Y. Chen, J. Freire, E. Santos, C. E. Scheidegger, C. Silva and H. T. Vo. SCI Institute Technical Report, No. UUSCI-2006-17, University of Utah, 2006.

### Patents Issued

- [259] *Product synthesis from multiple sources*, U.S. patent number 9,384,233 (July 5, 2016). Assigned to Microsoft.
- [260] *Method and System for Adaptive Discovery of Content on a Network*, U.S. patent number 8,965,865 (February 24, 2015). Assigned to University of Utah.
- [261] *Analogy based workflow identification*, U.S. patent number 8,762,186 (June 24, 2014). Assigned to University of Utah.
- [262] *Product synthesis from multiple sources*, U.S. patent number 8,352,473 (January 8, 2013). Assigned to Microsoft Corporation.
- [263] *Enabling Provenance Management for Pre-Existing Applications*, U.S. patent number 8,190,633 (May 29, 2012). Assigned to University of Utah.
- [264] *Analogy based workflow identification*, U.S. patent number 8,060,39 (November 15, 2011). Assigned to University of Utah.



- [265] *Method and System for Clustering Identified Forms*, U.S. patent number 7,996,390 (August 9, 2011). Assigned to University of Utah.
- [266] *Automatic exploration and testing of dynamic Web sites*, U.S. patent number 7,716,322 (May 11, 2010). Assigned Alcatel-Lucent USA Inc.
- [267] *Applications of executable shopping lists* U.S. patent number 7,103,566 (September 5, 2006). Assigned to Lucent Technologies.
- [268] *Method and apparatus for web-site-independent personalization from multiple sites having user-determined extraction functionality* U.S. patent number 6,976,210 (December 13, 2005). Assigned to Lucent Technologies.
- [269] *Active Database Trigger Processing Using a Trigger Gateway*. U.S. patent number 6,594,656 (July 15, 2003). Assigned to Lucent Technologies.
- [270] *Method for creating and playing back a smart bookmark that automatically retrieves a requested Web page through a plurality of intermediate Web pages*. U.S. patent number 6,535,912 (March 18, 2003). Assigned to Lucent Technologies.
- [271] *Method And Apparatus For Web-Site-Independent Personalization From Multiple Sites Having User-Determined Individual Refresh Rates*. European patent number EP1389319 (February 18, 2004) and International patent WO0116802 (March 8, 2001). Assigned to Lucent Technologies.
- [272] *Method for providing fast access to dynamic content on the world wide web*. European patent number EP1081607 (March 7, 2001). Assigned to Lucent Technologies.
- [273] *Database Trigger Gateway and its Operating Method*. Japanese patent number JP2000215092 (August 4, 2000). Assigned to Lucent Technologies.

### Freely-Available Software Systems and Web Sites

- **OpenClean:** openclean is an extensible Python library that brings together open source data profiling and data cleaning tools into an easy-to-use environment. Through the integration with the rich Python ecosystem, openclean supports a wide range of data cleaning requirements. <https://github.com/VIDA-NYU/openclean>
- **Vizier:** Vizier is an open-source system that combines the flexibility of notebooks with the easy-to-use data manipulation interface of spreadsheets. The system automatically tracks provenance for both data and computational steps, and uses this provenance to support reproducibility, reflective reasoning and streamline data curation. <https://vizierdb.info>
- **AlphaD3M:** AutoML system that automatically searches for models and synthesizes end-to-end pipelines that read, pre-process the data, and train a model. AlphaD3M uses deep learning to learn how to incrementally construct these pipelines. The process progresses by self play with iterative self improvement. <https://gitlab.com/VIDA-NYU/d3m/alphad3m>
- **Visus:** system designed to support the model building process and curation of machine learning and data processing pipelines generated by AutoML systems. Visus integrates visual analytics techniques and allows users to perform interactive data augmentation and visual model selection. <http://gitlab.com/VIDA-NYU/d3m/ta3>

- **Auctus:** search engine for structured data that supports data augmentation, both for machine learning and analytics tasks. Auctus discovers datasets on the Web, and profiles them to automatically infer metadata. It supports a wide range of queries including search and integration queries over spatial, temporal, categorical and numerical data. <https://github.com/VIDA-NYU/auctus>
- **Aries:** web-based platform that enables art historians to explore and manipulate digital images. Aries was featured as one of the most notable artifacts in the Frick Collection. <https://artimageexplorationspace.com>
- **Data Polygamy:** scalable topology-based framework that automatically identifies relationships between salient features across large collections of spatio-temporal data sets. <https://github.com/VIDA-NYU/data-polygamy>
- **STIG/MongoDB:** GPU-based spatio-temporal index integrated into MongoDB. This index provides sub-second response times to queries over large spatio-temporal data sets and is orders of magnitude faster than commercial and open-source databases. <https://github.com/harishd10/mongodb>
- **ReproZip:** ReproZip is a tool that automatically captures the provenance of experiments and packs all the necessary files, library dependencies and variables to reproduce the results. Reviewers can then unpack and run the experiments without having to install any additional software. <https://github.com/VIDA-NYU/reprozip>
- **TaxiVis:** TaxiVis is a visual analytics systems that supports the interactive analysis of spatio-temporal data. <https://github.com/VIDA-NYU/taxivis>, <http://taxivis.org>
- **ReproMatch:** ReproMatch was designed to help users find the tools that best match their reproducibility needs. The tools in the ReproMatch catalog are classified according to different reproducibility tasks, which we organized in a taxonomy. <http://repromatch.poly.edu>
- **Domain Discovery Tool (DDT):** DDT is an open-source, visual analytics framework for interactive domain discovery that augments the functionality provided by search engines. The system supports the exploratory analysis of web pages, and translates the expert's interactions with this data into a computational model of the domain of interest. [https://github.com/VIDA-NYU/domain\\_discovery\\_tool](https://github.com/VIDA-NYU/domain_discovery_tool)
- **ACHE:** ACHE is an open-source, focused Web crawler. <https://github.com/VIDA-NYU/ache>
- **VisTrails:** VisTrails, an open-source, provenance-aware scientific workflow management system that provides support for exploratory computational tasks, such as simulations, data analysis and visualization. <http://www.vistrails.org>
- **noWorkflow:** noWorkflow transparently captures provenance of scripts and enables reproducibility. The system is non-intrusive relies on techniques from Software Engineering, including abstract syntax tree analysis, reflection, and profiling, to collect different types of provenance without requiring a version control system or an instrumented environment. noWorkflow is available as open source software at <https://github.com/gems-uff/noworkflow>
- **BirdVis:** BirdVis is a tool that supports interactive analysis and visualization of spatio-temporal predictive models for bird distribution. The system was developed as part of a collaboration among computer scientists, statisticians, biologists and ornithologists. <http://www.birdvis.org>

- **DEFOG:** DEFOG is an information visualization system based on a model where each element is backed by all of its data regardless of how it is represented in the visualization. Users are free to add, copy, select, reorder, and group objects before and after a visualization has been applied. Objects from different visualizations can be easily combined or a subset of objects in a visualization can be selected and visualized in a different manner.
- **DeepPeep:** DeepPeep was a search engine specialized in Web forms. It helps users discover the entry points to content in Deep Web (aka Hidden Web) sites, including online databases and Web services. The system was featured in the New York Times.
- **crowdLabs.org:** crowdLabs adopts the model used by social Web sites and integrates a set of usable tools and a scalable infrastructure to provide an environment for scientists to collaboratively analyze and visualize data. The system combines benefits of social Web sites and science portals while at the same time addressing their limitations. Similar to social Web sites, crowdLabs aims to foster collaboration, but unlike these sites, it was specifically designed to support the needs of computational scientists, including the ability to access high-performance computers and manipulate large volumes of data. By providing mechanisms that simplify the publishing and use of analysis pipelines, it allows users to collaboratively construct and refine portals. This lowers the barriers for the use of scientific analyses and enables broader audiences to contribute insights to the scientific exploration process, without the high costs incurred by traditional portals.
- **Siphon++:** Siphon++ is a hidden-Web crawler that automatically retrieves contents hidden behind keyword-based query interfaces.
- **ShreX:** ShreX is a freely-available system for shredding, loading and querying XML documents in relational databases. It supports a wide range of XML-to-relational mappings; provides generic document shredding and query translation capabilities; and works with virtually any RDBMS.
- **StatiX++:** StatiX++ is a freely-available result estimator for XML queries. It uses histograms to uniformly capture both the structural and value skew present in documents. It leverages schema information to produce high-quality and concise statistical summaries.
- **XSB:** XSB is a deductive database/Prolog system that runs on over a dozen platforms (e.g., Linux, Windows 95/98 and NT, Solaris, SGI, etc), and has thousands registered users world wide. I was part of the XSB development team from 1993 to 1997.
- **Syntactica and Semantica:** Syntactica and Semantica are applications (workbooks) used to teach introductory courses of syntax and semantics. The applications and their manuals are available at bookstores or directly from MIT Press at <http://mitpress.mit.edu>.

### Keynote Presentations and Distinguished Lectures

- *Embracing Computational Reproducibility: Challenges, Solutions, and Cultivating Trust in Data-Driven Science.* Keynote at 2023 ACM Conference on Reproducibility and Replicability, UC Santa Cruz, Santa Cruz, California, June 27, 2023.
- *Dataset Search for Data Discovery, Augmentation, and Explanation.* Keynote at 2023 EDBT/ICDT Joint Conference, Ioannina, Greece, March 29, 2023.
- *Towards Usability and Trust for Data-Driven Models.* Keynote at International Conference on AI-ML Systems. Bangalore, India, Oct 21, 2021.

- *Towards Usability, Interactivity, and Trust for Data-Intensive Computations*. *Keynote* at the ACM Symposium on Cloud Computing (SoCC). October 21, 2020.
- *Towards Usability, Transparency, and Trust for Data-Intensive Computations*. *Keynote* at the joint Eurographics and Eurovis Conference, Norrköping, Sweden, May 29, 2020.
- *Democratizing Urban Data Exploration*. *Keynote* at the National Database Conference of China (NDBC). October 11, 2019.
- *Towards Accessibility, Transparency, and Trust in Data-Intensive Computations*. *Aula Inaugural*, Instituto de Computação, Universidade Federal Fluminense, October 1, 2020.
- *Towards Accessibility, Transparency, and Trust in Data-Intensive Computations*. NSF CISE Distinguished Lecture, January 28, 2020.
- *Democratizing Urban Data Exploration*. Center for Data and Computing (CDAC) Distinguished Speaker Series, University of Chicago, December 6, 2019.
- *Towards Accessibility, Transparency, and Trust in Data-Driven Exploration*. *Keynote* at the ACM SIGMOD Workshop on Human-In-the-Loop Data Analytics (HILDA), Amsterdam, Netherlands, July 5, 2019.
- *Democratizing Urban Data Exploration*, Laboratoire d'Informatique Paris Descartes (LIPADE), *Keynote* at the Open Day, L'Université de Paris Descartes, Paris, France, June 17, 2019.
- *Towards Accessibility, Transparency, and Trust in Data-Driven Exploration*. *Keynote* at the IEEE International Conference on Data Engineering, Macau, April 2019.
- *Towards Democratizing Urban Data Exploration*, *Keynote* at the ACM SIGMOD ExploreDB Workshop, Chicago, May 19th, 2017.
- *Democratizing Urban Data Analysis*, *Keynote* at IEEE Computing in the 21st Century Conference, Alberta, Canada, October 18, 2016.
- *Democratizing Urban Data Analysis*, *Keynote* at Computing in the 21st Century Conference, Seoul, Korea, November 3, 2016.
- *Provenance for Computational Reproducibility and Beyond*, *Keynote* at International Symposium on Experimental Algorithms (SEA), St. Petersburg, Russia, June 5, 2016.
- *Exploring Big and Polygamous Urban Data*, *Keynote* at the Alberto Mendelzon International Workshop on Foundations of Data Management, Panama City, Panama, June 10, 2016.
- *Exploring Big Urban Data*, Data Science Initiative Distinguished Lecture, Brown University, Providence, April 28, 2016.
- *Exploring Big Urban Data*, *Keynote* at the WWW BigData Innovators Gathering, Montreal, Canada, April 12, 2016.
- *Exploring Big Urban Data*, LIG Distinguished Seminar, CNRS, Grenoble, France, March 5, 2015.
- *Exploring Big Urban Data*, *Keynote* at the AAAI Workshop on AI for Cities, Austin, San Francisco, January 15, 2015.

- *Provenance and Data Exploration*. *Keynote* at the IEEE VIS 2014 Workshop on Provenance for Sensemaking, Paris, France, November 10th, 2014.
- *Exploring Big Urban Data*. *Keynote* at the Brazilian Symposium on Databases (SBBD), Curitiba, Brazil, October 9th, 2014.
- *Exploring Big Urban Data*. *Keynote* at the IKDD Conference on Data Sciences (CoDS 2014), Delhi, India, March 21-23, 2014.
- *Exploring Big Urban Data*. *Keynote* at the DIMACS/CCICADA Workshop on Big Data Integration, Rutgers University, New Jersey, June 20 - 21, 2013.
- *Exploring Big and not so Big Data*. Distinguished Lecture Series, Department of Computer Science, University of Toronto, Toronto, December 4, 2012.
- *Exploring Big and not so Big Data: Opportunities and Challenges*. Distinguished Seminar Series, Mechanical & Industrial Engineering, University of Toronto, Toronto, October 26, 2012.
- *Provenance-Rich Science*. *Keynote* at the DB/IR Day, AT&T Shannon Labs, Florham Park, NJ, October 22, 2010.
- *Provenance Management for Data Exploration*. *Keynote* at the International Conference on Data Integration in the Life Sciences (DILS), Sweden, August 2010.
- *Provenance Management: Challenges and Opportunities*. *Keynote* at the GI-Fachtagung Datenbanksysteme für Business, Technologie und Web (BTW), Münster, Germany, March 4, 2009.
- *VisTrails: Using Provenance to Streamline Data Exploration*. *Keynote* at the e-Science WorkShop, João Pessoa, Brazil, October 18, 2007
- *Managing Rapidly Evolving Workflows*. *Keynote* at the International Provenance and Annotation Workshop (IPAW), Chicago, IL, May 3–5, 2006

## Selected Tutorials

- *Computational reproducibility: state-of-the-art, challenges, and database research opportunities*. J. Freire, Philippe Bonnet, Dennis Shasha. Tutorial at the ACM SIGMOD Conference, Scottsdale, Arizona, May 2012.
- *Provenance-Enabled Data Exploration and Visualization with VisTrails, with Cláudio Silva, Emanuele Santos and Erik Anderson*. Tutorial at the SIBGRAPI Conference on Graphics, Patterns and Images, Gramado, Brazil, August 2010.
- *Provenance-Enabled Data Exploration and Visualization (with Cláudio Silva, Emanuele Santos, David Koop, Erik Anderson)*. Tutorial at the IEEE Visualization Conference, Atlantic City, October 2009.
- *Provenance-Enabled Data Exploration and Visualization with VisTrails (with Emanuele Santos, and Cláudio Silva)*. Tutorial at the Department of Energy SciDAC Conference, San Diego, June 2009.
- *Streamlining Data Exploration and Visualization through Scientific Workflows and Provenance (with Emanuele Santos, and Cláudio Silva)*. Tutorial at the IEEE International Conference on e-Science, Indianapolis, December 2008.

- *Provenance and Scientific Workflows: Challenges and Opportunities (with S. Davidson)*. Tutorial at the ACM International Conference on Management of Data (SIGMOD), Vancouver, Canada, June 2008.
- *Provenance Management: Challenges and Opportunities (with Cláudio Silva)*. Invited tutorial at Brazilian Symposium on Databases. João Pessoa, Brazil, October 2007.
- *XML and Data Management (with M. Benedikt, M. Fernandez and A. Sahuguet)*. Tutorial presented at: World Wide Web Conference (WWW2002) - Hawaii, May 2002; Extreme Markup Languages - Montreal, Canada, August 2002.
- *Querying the Web (with D. Florescu)*. Tutorial presented at the Brazilian Conference on Databases, September 2000

### Selected Invited Talks

- *Provenance and Trust for Computational Workflows*. FONDA Lecture Series, Humboldt-Universität zu Berlin, Germany, June 8, 2021.
- *Towards Usability, Interactivity, and Trust for Data-Intensive Computations*. Women in Science Seminar Series for Department of Informatics, King's College, London, November 25, 2020.
- *Towards Usability, Interactivity, and Trust for Data-Intensive Computations*. Data Systems and Foundations seminar, University of California, Berkeley, November 2, 2020.
- *Democratizing Urban Data Exploration*. CS Colloquium Series, Harvard University, September 26, 2019.
- *Reproducibility in Data Science*. Data Council Conference, New York City, November 2019.
- *Democratizing Urban Data Exploration*, Women in Data Science Conference, IT University of Copenhagen, Copenhagen, Denmark, April 16, 2019.
- *Urban Computing*. INRIA, Saclay, France, March 2019.
- *Towards Accessibility, Transparency, and Trust in Data-Driven Exploration*. JASON Fall Meeting, Mitre, McLean, Virginia, November, 2018.
- *Building a Data Science Environment: A View from the Trenches*. Canadian Data Science Workshop, Toronto, Canada, May 2018.
- *Computational Reproducibility*, Data Abuse Prevention (DAP) Workshop, DARPA Information Science and Technology (ISAT), Arlington, August 2, 2018.
- *ReproZip and ReproServer*, ACM Workshop on Reproducibility in Publications, New York City, December 7-8, 2017.
- *Visualization and Reproducibility*, Scientific Visualization Conference, Flatiron Institute, New York, October 27, 2017.
- *Democratizing Urban Data Exploration*, Computer Science Colloquium, Cornell University, Ithaca, September 28, 2017.
- *Big (Urban) Data Analytics*, 1st ACM Europe Summer School, Athens, Greece, July 13-19, 2017.

- *Democratizing Reproducibility*, Moore-Sloan Data Science Summit, New York University, October 26, 2016.
- *Provenance for Computational Reproducibility and Beyond*, RDI2 Workshop: Reproducibility in Experimental and Computational Science Research - A Rutgers 250 Event, Rutgers University, October 10, 2016.
- *Large-Scale Information Integration*, Urban Analytics Workshop, The Alan Turing Institute, London, England, July 7-8, 2016.
- *Reproducibility Made Easy*, EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization, Cagliari, Italy, May 25-26, 2015.
- *Exploring Big Urban Data*, Campus Visualization Partnership (CVP), University of Maryland, May 1, 2015.
- *Reproducibility Made Easy*, Workshop on Algorithmic Transparency in the Media, New York, March 27, 2015.
- *Managing and Re-Using Provenance as a Critical Capability for Data Scientists*, AAAS Annual Meeting, San Jose, February 15, 2015.
- *Principles of Working with Big Data*, National Academies, Washington D.C., April 11, 2014.
- *Exploring Big and not so Big Data: Opportunities and Challenges*, Los Alamos National Laboratory, June 10, 2013.
- *VisTrails: Using Provenance to Streamline Data Exploration and Visualization*, Brookhaven National Laboratory, March 11, 2013.
- *Making Computations and Publications Reproducible with VisTrails*, ICERM Reproducibility in Computational and Experimental Mathematics, Brown University, Providence, December 13, 2012.
- *Synthesizing Products for Online Catalogs*, Database seminar, Department of Computer Science, University of Toronto, Toronto, December 5, 2012.
- *Managing Provenance for Reproducibility and Beyond*. Google, Mountain View, November 16, 2012.
- *Integrating Structured Data on the Web*. Federal University of Pernambuco, Recife, Brazil, July 31st, 2012.
- *A Provenance-Based Infrastructure for Creating Reproducible Papers*. AMP 2011 Workshop on Reproducible Research. Vancouver, Canada. July 14, 2011.
- *Provenance-Rich Science*. Juliana Freire. FORTH. Crete, Greece. June 22, 2011.
- *Publishing Reproducible Results with VisTrails*. Juliana Freire and Cláudio Silva. SIAM Mini-Symposium on Reproducible Research. Las Vegas. March 4, 2011.
- *Provenance for Reproducibility and Beyond*. IBM T.J. Watson. Hawthorne, NY. February 15th, 2011.
- *Towards and Infrastructure to Create Reproducible Papers*. Beyond the PDF Workshop. San Diego. January 19th, 2011.

- *Infrastructure for Understanding Human Knowledge*. ICiS Workshop on Integrating, Representing, and Reasoning over Human Knowledge: A Computational Grand Challenge for the 21st Century. Snowbird, August 8, 2010.
- *Supporting Provenance-Rich Science with VisTrails*. CScADS Scientific Data and Analytics for Petascale Computing Workshop. Snowbird, July 26, 2010.
- *The WebDB Group: Research Overview*. Federal University of Amazonas, Manaus, Brazil, June 29th, 2010.
- *Provenance in Science: Challenges and Opportunities*. International Workshop on The Next Generation of Quantum Simulations, Moorea, French Polynesia, May 6, 2009.
- *Provenance in Science: Challenges and Opportunities*. University of Utah Campus CyberInfrastructure Day, Salt Lake City, UT, March 13, 2009.
- *Querying and Creating Visualizations by Analogy*. University of Southampton, Southampton, UK, August 14, 2007
- *Discovering and Organizing Hidden-Web Sources*. Microsoft Search Labs, Mountain View, CA, August 6, 2007
- *Provenance Analytics: Exploring Science Trails*. SciFoo, The Googleplex, Mountain View, CA, August 5, 2007
- *VisTrails: Using Provenance to Support Data Exploration through Workflows*. IBM Almaden, San Jose, CA, August 3, 2007
- *Managing Provenance for Exploratory Tasks*. Institute for Systems Biology, Seattle, WA, March 6, 2007
- *Managing Exploratory Workflows*. Microsoft eScience Workshop at The Johns Hopkins University, Baltimore, MD, October 13–15, 2006
- *Managing Exploratory Workflows*. Southern California Earthquake Center (SCEC) Workflow Workshop, Los Angeles, CA, October 25, 2006
- *Managing Rapidly-Evolving Scientific Workflows with VisTrails*. Universidade Federal do Rio Grande do Sul, Porto Alegre (Brazil), August 3, 2006.
- *The Future of Databases*, Engineering National Advisory Council (ENAC), University of Utah, October 2005.
- *Uncovering the Hidden-Web*, Yahoo Research, Sunnyvale, CA, September 2005.
- *The Future of Information Integration*, invited panel presentation at the WebDB workshop, co-located with SIGMOD – Baltimore, MD, June 2005.
- *Managing XML Data*, Medical Informatics Department, University of Utah, Salt Lake City, UT, March 2005.
- *A Brief Introduction to Database Systems*, Medical Informatics Department, University of Utah, Salt Lake City, UT, March 2005.



- *Using Wrappers for Device Independent Web Access: Opportunities, Challenges and Limitations*, Portland State University, Portland, OR, February 2003.
- *Adaptive XML Shredding: Architecture, Implementation, and Challenges*, University of Toronto, Toronto, CA, January 2003.
- *Accessing and Personalizing Web Content and Services*, WWW2002 Workshop on Mobile Search, Honolulu, HI, May 2002.
- *Navigating the Hidden-Web with the WebVCR*, WhizBang! Labs, Pittsburgh, PA, April 2002.
- *From XML Schema to Relations: A Cost-Based Approach to XML Storage*
  - Seminar on Foundations for Semi-Structured Data. Dagstuhl (Germany), September 2001.
  - AT&T Research, Florham Park, NJ, March 2002.
  - University of Massachusetts, Amherst, MA, February 2002.
  - University of California, Santa Barbara, CA, March 2002.
  - Rutgers University, New Brunswick, NJ, March 2002.
  - University of Pittsburgh, Pittsburgh, PA, April 2002.
  - Michigan State University, East Lansing, MI, December 2001.
- *WebViews: Accessing Personalized Web Content and Services*, AT&T Research, Florham Park, May 2001.
- *A Career in a Research Lab*, CRAW Distinguished Speakers Series. University of California at Irvine, January 2001.
- *Integrating and Querying Dynamic Web Content*
  - Seminar on Foundations for Data Integration. Dagstuhl (Germany), June 1999.
  - University of Maryland, College Park, MD, April 2000.
- *Taking I/O Seriously: Resolution Reconsidered for Disk*
  - IBM Almaden – San Jose, CA – August 1997.
  - HP Labs – Palo Alto, CA – August 1997.
  - IBM T.J. Watson – Hawthorne, NY – July 1997.
  - Universidade do Porto – Porto, Portugal – July 1997.

## Panels

- Winds from Seattle: Database Research Directions. VLDB Conference, Tokyo, Japan, 2020.
- The Next 5 Years: What Opportunities Should the Database Community Seize to Maximize its Impact?. ACM SIGMOD, Portland, 2020.
- Detecting, Combating, and Identifying Dis and Mis-information. AAAS, Seattle, 2020.
- ML for CI and CI for ML. NSF CSSI PI Meeting, 2020.

- Exploring Reproducibility in Vis: Expanding on the National Academies' Report on Reproducibility and Replicability in Science. IEEE VIS 2019.
- Is There a Data Science and Engineering Brain Drain? If So, How Can We Rebalance Them? IEEE International Conference on Data Engineering (ICDE) 2019.
- Advice from Early PhD to Early Career. New Researcher Symposium, ACM SIGMOD 2018.
- Future Talent 2040. Moderator: Tim Pan. Panelists: Juliana Freire, Sue Moon, Fred Shneider, Xiaofan Wang. Microsoft Research Asia Faculty Summit, 2016.
- Big Data Quality — Whose problem is it? Moderators: Paolo Papotti and Shazia Sadiq. Panelists: Feliz Naumann, Tamraparni Dasu, Juliana Freire, Ihab Ilyas, and Eric Simon. IEEE ICDE, 2016.
- Should we all be teaching “intro to data science” instead of “intro to databases”? Panelists: Bill Howe, Michael J. Franklin, Juliana Freire, James Frew, Tim Kraska, Raghu Ramakrishnan. ACM SIGMOD, 2014.
- German Science Foundation Panel on Big Data. Moderator: Cláudio Silva (NYU CUSP). Panelists: Juliana Freire (NYU Poly), Raghu Ramakrishnan (Microsoft), Gerhard Weikum (MPI, Germany). German Center for Research and Innovation, United Nations, New York. June 27, 2013.
- SIGMOD New Researcher Symposium. ACM SIGMOD 2013.
- Reproducible Visualization Research: How Do We Get There? Moderator: Enrico Bertini. Panelists: Juliana Freire (NYU Poly), Gordon Kindlmann (University of Chicago), Tamara Munzner (University of British Columbia), Tim Dwyer (Microsoft Research). IEEE VisWeek 2012.
- Maximizing Impact. Moderator: Ed Lazowska (University of Washington). Panelists: David DeWitt (Microsoft), Juliana Freire (NYU Polytechnic), Ed Lazowska (University of Washington), Sam Madden (MIT), Jennifer Widom (Stanford). VLDB 2011.

### Selected Conference Talks

- *Multilingual Schema Matching for Wikipedia Infoboxes*. Very Large Databases Conference, Istanbul, Turkey, August 2012.
- *Simplifying the Design of Workflows for Large-Scale Data Exploration and Visualization*. Microsoft eScience Workshop, Indianapolis, December 2008.
- *Automatically Constructing a Directory of Molecular Biology Databases*. Workshop on Data Integration in the Life Sciences – Philadelphia, June 2007.
- *VisTrails: Using Provenance to Streamline Data Exploration*. Workshop on Data Integration in the Life Sciences – Philadelphia, June 2007.
- *Searching for Hidden-Web Databases*. ACM SIGMOD WebDB workshop – Baltimore, June 2005.
- *StatiX: Making XML Count*. SIGMOD – Madison, June 2002
- *VeriWeb: A Platform for Automating Web Site Testing*. WWW2002 – Hawaii, May 2002
- *From XML Schema to Relations: A Cost-Based Approach to XML Storage*. ICDE – San Jose, February 2002.

- *WebViews: Accessing Personalized Web Content and Services*. WWW10 – Hong Kong, May 2001
- *Automating Web Navigation with the WebVCR*. WWW9 – Amsterdam, May 2000.
- *Practical Problems in Coupling Deductive Engines with Relational Databases*. SIGMOD Workshop on Knowledge Representation meets Databases (KRDB) – Seattle, WA, June 1998.
- *Taking I/O Seriously: Resolution Reconsidered for Disk*. ICLP – Leuven, Belgium, July 1997.
- *Beyond Depth-First: Improving Tabled Logic Programs through Alternative Scheduling Strategies*. PLILP – Aachen, Germany, September 1996.
- *Exploiting Parallelism in Tabled Evaluations*. PLILP – Utrecht, Netherlands, September 1995.
- *Parallelizing Tabled Evaluations*. ILPS Workshop on Parallel Logic Programming Systems – Ithaca, November 1994.

### Media Coverage

- BBC Brasil, July 19th, 2014. Internet oculta: os segredos de um universo paralelo.
- Elsevier Connect, August 19th, 2013. Harnessing the power of big data is no small feat.
- The Economist, August 12th, 2011. Difference Engine: Happy anniversary? <https://www.economist.com/babbage/2011/08/12/difference-engine-happy-anniversary>
- NPR Weekend Edition, Dec 27, 2009. The Web That's Hidden From You.
- LiveScience, Dec 2, 2009. Conquering the Digital Data Overload: Juliana Freire Profiled on Live Science.
- The Guardian, Nov 26, 2009. Over to the dark side of the Web.
- New York Times, Feb 22, 2009. Exploring a 'Deep Web' That Google Can't Grasp.
- Herald Tribune, Feb 23, 2009. Emerging search technologies aim for Web's hidden depths.
- Diário do Comercio, Brazil, March 2, 2009. Superando o Google.
- NSF Discoveries, March 3, 2009. A New Vision for Scientific Visualizations.

### Professional Activities

#### Funding Agencies

- P41 Reviewer for NIH, 2014.
- Panelist for NSF Division of Information and Intelligent Systems, 2012.
- Panelist for NSF Office of Cyberinfrastructure, 2012.
- Panelist for NIH, 2010.
- Panelist for NSF Division of Information and Intelligent Systems, 2008.
- Panelist for NSF Division of Computer and Network Systems, 2008.

- Participant of NSF/Mellon Foundation Workshop on Workflow Interoperability, 2007.
- Participant of NSF Workshop on the Challenges of Scientific Workflows, 2007.
- Panelist for NSF Division of Information and Intelligent Systems, 2007.
- Panelist for NSF Division of Information and Intelligent Systems, 2003.
- Panelist for NSF Division of Information and Intelligent Systems, 2001.
- Panelist for NSF Division of Experimental & Integrative Activities, 2003.

### Special Committees

- Member, Harvard Data Science Initiative Trust in Science External Advisory Board, 2021-.
- Member, NSF Advisory Committee for Cyberinfrastructure, 2020-.
- Member, Search committee for the Editor-in-Chief ACM Transactions on Database Systems, 2020.
- Member, National Academies Committee on Reproducibility and Replicability in Science, November 2017-2019.
- Advisory Board Member, Center for Reproducible Biomedical Modeling, 2018-.
- Council Member, Computing Community Consortium, 2017-2020.
- Member, PVLDB Executive Committee.
- International liaison, Brazilian Computer Society.
- ACM Task Force on Data, Software, and Reproducibility in Publication, 2015-.
- Reviewer, ACM Future of Computing, 2017.
- Member, ACM Grace Murray Hopper Award Committee, 2017.
- Member, ACM Grace Murray Hopper Award Committee, 2016.
- Chair, ACM SIGMOD Jim Gray Doctoral Dissertation Award Committee, 2016.
- Member, PVLDB advisory committee, 2015-.
- Deputy chair, ACM SIGMOD Jim Gray Doctoral Dissertation Award Committee, 2015.
- Member, NYC Taxi & Limousine Commission Data/Technology Advisory Committee, 2015-.
- Member, ACM Digital Library Committee: Project on Data, Software, Replicability, 2015-.
- Chair, ACM Grace Murray Hopper Award Committee, 2015.
- Deputy Chair, ACM Grace Murray Hopper Award Committee, 2014.
- Advisory Board, SOCIAM Project, United Kingdom.
- Expert Advisor, Open Data 500, 2014-.

- Member, ACM Grace Murray Hopper Award Committee, 2013.
- Member, New York City Council High Tech-Academia Working Group, 2013-.
- Board of Trustees, Very Large Data Bases Endowment Inc., 2012-.
- Member, ACM SIGMOD Jim Gray Doctoral Dissertation Award Committee, 2013-2014.

### Minority Involvement

- *Member* of the Diversity and Inclusion Task Force, ACM SIGMOD, 2020–.
- *Member* of the DBCares committee, ACM SIGMOD and VLDB, 2019–.
- *Founding Chair* of the Diversity and Outreach committee, School of Computing, University of Utah, 2007–2009.
- *Host* for CRA-W Distinguished Lecture Series, University of Utah, September 2007.
- *Member* of the Academic Alliance for the National Center for Women in Technology (NCWIT).
- *Panelist* for the CRAW program to recruit undergrads, especially women, to consider going to graduate school in computer science and engineering rather than straight into the industry with their Bachelors.
  - UC Irvine, January 2001
  - Princeton, November 2001
- *Mentor*: MentorNet program, 1997 and 1998 (Mentornet is a program that pairs women who are studying engineering or science with professional scientists and engineers working in industry, and help them form e-mail based mentoring relationships).

### Review of Scholarly Work

#### Conference Organization:

1. *Program Committee Co-Chair*: International Conference on Very Large Databases (VLDB), Sydney, Australia, 2022.
2. *Area Chair*: Proceedings of the VLDB Endowment (PVLDB), 2021.
3. *Co-Chair*: CCC Roundtable on Misinformation, Washington, D.C., March 2019.
4. *Area Chair*: Proceedings of the VLDB Endowment (PVLDB), 2018.
5. *Co-chair*: SciPy Mini Symposium on Open Data and Reproducibility, Austin, Texas, July 10-16, 2017.
6. *Co-organizer*: Dagstuhl Workshop on Connecting Visualization and Data Management Research, Dagstuhl, Germany, November 13 – 17, 2017.
7. *Program Committee Group Leader*: ACM International Conference on Management of Data (SIGMOD) 2017.
8. *Co-organizer*: Dagstuhl Workshop on Reproducibility of Data-Oriented Experiments in e-Science, Dagstuhl, Germany, January 24 – 29 , 2016.

9. *Chair of Demo Program*: ACM International Conference on Management of Data (SIGMOD), 2015.
10. *Chair and organizer*: Workshop on Software Infrastructure for Reproducibility in Science–New York, May 30th-31st, 2013.
11. *Chair and organizer*: Workshops on Reproducibility in Science–New York, May 1st and May 29th, 2013.
12. *Chair*: ACM International Conference on Management of Data (SIGMOD) Repeatability Committee—New York, May 2013.
13. *Program Committee Group Leader*: ACM International Conference on Management of Data (SIGMOD), Scottsdale, Arizona, May 2012.
14. *Program Chair*: Experiments and Analysis Track of the International Conference on Very Large Data Bases–Istanbul, September 2012.
15. *Program Chair*: Alberto Mendelzon International Workshop on Foundations of Data Management–Ouro Preto, Brazil, June 2012.
16. *Organizer*: Synergy Workshop–Dresden, Germany, March 2012.
17. *Program Chair*: Usenix Workshop on the Theory and Practice of Provenance (TAPP)– Crete, Greece, June 2011.
18. *Program Chair*: International World Wide Web Conference (WWW), 2010.
19. *Chair*: The Second International Workshop on Role of Semantic Web in Provenance Management (SWPM 2010) – Shanghai, China, November 2010.
20. *Co-Chair*: First International Workshop on Traceability and Compliance of Semi-Structured Processes (TC4SP’10) – Hoboken, New Jersey, USA, September 2010.
21. *Co-Chair*: International Workshop on Role of Semantic Web in Provenance Management (SWPM) – Washington D.C., USA, October 2009.
22. *Workshops Co-Chair*: International Conference on Very Large Databases (VLDB) – Lyon, France, August 2009.
23. *Track Chair*: ACM Conference on Information and Knowledge Management (CIKM) – Napa Valley, California, October 2008.
24. *Co-organizer*: Symposium on Provenance in Scientific Workflows, Salt Lake City, Utah, October 2008.
25. *Co-Chair*: International Provenance and Annotation Workshop (IPAW)– Salt Lake City, UT, June 2008.
26. *Vice-Chair*: International World Wide Web Conference (WWW), XML and Web Data track – Beijing, China, April 2008.
27. *Co-Chair*: HPDC Provenance Challenge Workshop – Monterey Bay, CA, June 2007.
28. *Vice-Chair*: IEEE International Conference on Data Engineering (ICDE) 2007 – Istanbul, Turkey, April 2007.

29. *Deputy Vice-Chair*: International World Wide Web Conference (WWW) 2007 XML and Web Data track – Banff, Canada, May 2007.
30. *Co-Chair*: IEEE Visualization Workshop on Grid-Based Visualization – Baltimore, October 2006.
31. *Vice-Chair*: International World Wide Web Conference (WWW) 2005 Browsers and User Interfaces track – Edinburgh, Scotland, May 2005.
32. *Deputy Vice-Chair*: International World Wide Web Conference (WWW) 2003 Browsers and User Interfaces track – Budapest, Hungary, May 2003.
33. *Chair*: SIGMOD 2003 Demonstrations Committee – San Diego, June 2003.
34. *Co-Chair*: SIGMOD International Workshop on the Web and Databases (WebDB 2003) – San Diego, CA, June 2003.
35. *Co-Chair*: First International Workshop on Tabling in Logic Programming – Leuven, Belgium, July 1997.

**Program committee member of:**

1. International Conference on Automated Machine Learning (AutoML), 2022.
2. International Workshop on Computational Fact-Checking (The Web Conference), 2019
3. International Conference on Data Integration in the Life Sciences (DILS), Lisbon, Portugal, July 2014.
4. Workshop on Data-Centric Programming (DCP), San Diego, January 2014.
5. EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV<sup>3</sup>).
6. Alberto Mendelzon International Workshop on Foundations of Data Management–Puebla/Cholula, Mexico, May 2013.
7. Usenix Workshop on the Theory and Practice of Provenance (TAPP), 2013.
8. International Conference on Very Large Databases (VLDB), 2013.
9. ACM International Conference on Management of Data (SIGMOD) Best Paper Award, 2012.
10. International Conference on Scientific and Statistical Database Management (SSDBM), 2011.
11. IEEE International Conference on e-Science 2010.
12. International Conference on Scientific and Statistical Database Management (SSDBM), 2010.
13. International World Wide Web Conference (WWW), 2010.
14. International Conference on Data Engineering (ICDE), 2010.
15. International Workshop on the Web and Databases (WebDB), 2009.
16. Brazilian Symposium on Databases (SBBD), 2009.
17. International Workshop on Role of Semantic Web in Provenance Management (SWPM), 2009.

18. First International Workshop on Data and Process Provenance (WDPP), 2009.
19. ACM International Conference on Management of Data (SIGMOD), 2009 (tutorial program).
20. International Conference on Very Large Databases (VLDB) Ph.D. Workshop, 2009.
21. IEEE International Workshop on Scientific Workflows (SWF), 2009.
22. International Conference on Very Large Databases (VLDB), 2008.
23. ACM Conference on Information and Knowledge Management (CIKM), 2008.
24. Brazilian Symposium on Databases (SBBD), 2008.
25. International Provenance and Annotation Workshop (IPAW), 2008.
26. Data Integration in the Life Sciences (DILS), 2008.
27. First International Workshop on REsource Discovery (RED), 2008.
28. IEEE International Workshop on Scientific Workflows (SWF), 2008.
29. International World Wide Web Conference (WWW), 2008.
30. International Conference on Data Engineering (ICDE), 2008.
31. IEEE Workshop on Scientific Workflows, 2007.
32. Brazilian Symposium on Databases (SBBD), 2007.
33. International Workshop on the Web and Databases (WebDB), 2007.
34. International Conference on Very Large Databases (VLDB), 2007.
35. ACM International Conference on Management of Data (SIGMOD), 2007 (industrial program).
36. International Conference on Data Engineering (ICDE), 2007.
37. International World Wide Web Conference (WWW), 2007.
38. International Workshop on Challenges in Web Information Retrieval and Integration (WIRI), 2006.
39. International World Wide Web Conference (WWW), 2006.
40. Brazilian Symposium on Databases (SBBD), 2006.
41. ACM Conference on Information and Knowledge Management (CIKM), 2005.
42. International Workshop on the Web and Databases (WebDB), 2005.
43. Latin American Web Congress (LAWeb), 2005.
44. Brazilian Symposium on Databases (SBBD), 2005.
45. International Conference on Data Engineering (ICDE), 2005.
46. International World Wide Web Conference (WWW), 2005.



47. International XML Database Symposium (XSym), 2004.
48. International Conference on Web Engineering (ICWE), 2004.
49. Brazilian Symposium on Databases (SBBD), 2004.
50. International Workshop on the Web and Databases (WebDB), 2004.
51. ACM International Conference on Management of Data (SIGMOD), 2004.
52. International Conference on Very Large Databases (VLDB), 2003.
53. International Workshop on the Web and Databases (WebDB), 2003.
54. International World Wide Web Conference (WWW), 2003.
55. IJCAI Workshop on Information Integration on the Web, 2003.
56. Symposium on String Processing and Information Retrieval (SPIRE), 2003.
57. International Symposium on Practical Aspects of Declarative Languages (PADL), 2003.
58. Workshop on Foundations of Models and Languages for Data and Objects (FMLDO), 2002.
59. Workshop on Efficient Web-based Information Systems (EWIS), 2002.
60. Symposium on String Processing and Information Retrieval (SPIRE), 2002.
61. Brazilian Symposium on Multimedia and Hypermedia Systems (SBMIDIA), 2002.
62. International Workshop on the Web and Databases (WebDB), 2002.
63. SIGMOD Demonstrations, 2002.
64. International World Wide Web Conference (WWW), 2002.
65. Web Engineering track - International World Wide Web Conference (WWW), 2002.
66. International Conference on Data Engineering (ICDE), 2002.
67. International World Wide Web Conference (WWW10), 2001.
68. International Conference on Data Engineering (ICDE), 2001.
69. International World Wide Web Conference (WWW9), 2000.
70. International Conference on Computational Logic (CL), 2000.
71. International Workshop on the Web and Databases (WebDB), 2000.
72. International Conference on E-Commerce and Web (EC&Web), 2000.
73. First International Workshop on Tabling in Logic Programming, 1997.

**Editorial Boards:**

- *Editor in Chief for PVLDB, 2022-.*
- *Associate Editor for PVLDB, 2020-2021.*
- *IEEE Data Engineering Bulletin.*
- *VLDB Journal.*
- *IEEE Transactions on Knowledge and Data Engineering.*
- *IEEE Internet Computing.*
- *Journal of Information and Data Management.*
- *Journal of Web Engineering.*

**Steering Committes:**

- ACM SIGMOD Executive Committee.
- pVLDB Advisory Board.
- ACM SIGMOD Workshop on Human-In-the-Loop Data Analytics (HILDA).
- International Provenance and Annotation Workshop (IPAW).
- Workshop on the Theory and Practice of Provenance (TaPP).
- ACM SIGMOD Repeatability Evaluation.
- PVLDB Repeatability Evaluation.

**Review of Grant Proposals**

- Sloan Foundation.
- National Science Foundation (NSF)
- Engineering and Physical Sciences Research Council (EPSRC), UK
- Austrian Research Foundation
- National Research Council Canada
- European Science Foundation
- Swedish Research Council
- Research Council of Norway
- Research Grants Council–Hong Kong

**Internal Service (NYU)**

- Faculty Executive Committee (FEC), 2020-.
- Financial Policies Committee (FPC), 2020-.
- Director of Graduate Studies, Center for Data Science, 2014-2017.
- Member of the School of Engineering Tenure and Promotion committee, 2017-.
- Chair of faculty search committee – joint Center for Data Science and Computer Science and Engineering, 2017.
- Chair of faculty search committee – joint Center for Data Science and Computer Science and Engineering, 2018.

**Internal Service (University of Utah)**

- Women in Technology Distinguished Lecture Series, 2007– (creator and organizer).
- M.S. and Ph.D. in Computing, Information Track, 2007– (director).
- M.S. in Computing, Masters in Information Technology Track (joint with Business School), 2007– (director).
- Diversity and Outreach committee, 2007–2009 (founding chair).
- Curriculum committee, 2006– (member).
- Development Committee/Industrial Liaison, 2005– (member).
- Graduate Admissions Committee, 2006 (member).
- Graduate Studies Committee, 2005–2006 (member).
- Committee for creating MS degree in Computing in eBusiness, 2005–2007 (member).
- Faculty search committee, Medical Informatics Department, 2005–2006 (member).
- Research talk at the Engineering National Advisory Council (ENAC) meeting, University of Utah, October 2005.
- Gave two invited lectures in the Bioinformatics course, Medical Informatics, March 2005.

**Teaching**

- *Big Data*: graduate, NYU Tandon School of Engineering – CS-GY 6513 (Spring 2018, Fall 2019, Spring 2020, Fall 2020, Spring 2021, Fall 2021).
- *Big Data*: graduate, NYU Center for Data Science and Courant Institute – DS-GA 1004 (Spring 2015, 2016, 2017).
- *Massive Data Analysis*: graduate, NYU Poly– cs9223 (Fall 2014).
- *Big Data*: graduate, NYU Center for Data Science and Courant Institute – DS-GA 1004 (Spring 2014).

- *Massive Data Analysis*: graduate, NYU Poly– cs9223 (Fall 2013).
- *Advanced Visualization*: graduate, NYU Poly– cs9223 (Fall 2012).
- *Massive Data Analysis*: graduate, NYU Poly– cs9223 (Fall 2012).
- *Advanced Databases*: graduate, NYU Poly– cs6093 (Spring 2012).
- *Research Topics in Web and Databases*: graduate, University of Utah – cs7965 (Spring 2009). Enrollment: 8 graduate students.
- *Research Topics in Database Systems*: graduate, University of Utah – cs7962 (Spring 2009). Enrollment: 10 graduate students.
- *Database Systems*: graduate and undergraduate, University of Utah – cs5530/cs6530 (Fall 2008). Enrollment: 54 undergraduate and 9 graduate students.
- *Web Mining*: graduate, University of Utah – cs7960 (Spring 2008). Enrollment: 9 graduate students.
- *Database Systems*: graduate and undergraduate, University of Utah – cs5530/cs6530 (Fall 2007). Enrollment: 37 undergraduate and 10 graduate students.
- *Advanced Database Systems*: graduate, University of Utah – cs7965 (Spring 2007). Enrollment: 14 graduate students.
- *Database Systems*: graduate and undergraduate, University of Utah – cs5530/cs6530 (Fall 2006). Enrollment: 55 undergraduate and 19 graduate students.
- *Advanced Database Systems*: graduate, University of Utah – cs7965 (Spring 2006). Enrollment: 5 graduate students.
- *Database Systems*: graduate and undergraduate, University of Utah – cs5530/cs6530 (Fall 2005). Enrollment: 42 undergraduate and 7 graduate students.
- *Advanced Database Systems*: graduate, University of Utah – cs7965 (Spring 2005). Enrollment: 5 graduate students.
- *Database Systems*: graduate and undergraduate, University of Utah – cs5530/cs6530 (Fall 2004). Enrollment: 52 undergraduate and 3 graduate students.
- *XML Data Management*: graduate, OGI/OHSU – cse506 (Spring 2004).
- *Web Data Management*: graduate, OGI/OHSU – cse582 (Spring 2003).

### **Advising/Mentoring**

- Post-doctoral mentor
  - Aline Bessa (2020–2021).
  - Raoni Lourenço (2020–2021).
  - Sheng Wang (2019–2021) *Co-mentored with Daniel Neill*.
  - Ray Hong (2019-2020) *Co-mentored with Enrico Bertini*.
  - Fernando Chirigati (2018-2020).
  - Viviane Moreira (2011).
  - Ronaldo Mello (2010).

- Lauro Lins (2007-2012) *Co-mentored with Cláudio Silva.*
- Advisor (students graduated)
  - Zhouhan Chen (Ph.D. 2022, NYU CDS).
  - Gromit Chan (Ph.D. 2021, NYU CSE). *Co-advised with Cláudio Silva.*
  - Raoni Lourenço (Ph.D. 2020, NYU CSE).
  - Aline Bessa (Ph.D. 2020, NYU CSE).
  - Masayo Ota (Ph.D. 2020, NYU CSE).
  - Fernando Chirigati (Ph.D. 2018, New York University).
  - Kien T. Pham (Ph.D. 2018, New York University).
  - Tuan-Ahn Huang-Vu (Ph.D. 2017, New York University).
  - Thanh Nguyen (Ph.D. 2013, University of Utah).
  - Karane Vieira (co-advised with Altigran Silva. Ph.D. 2012, Federal University of Amazonas—Brazil).
  - Sergio Mergen (co-advised with Carlos Heuser. Ph.D. 2011, Federal University of Rio Grande do Sul—Brazil).
  - Hoa Nguyen (Ph.D. 2011, University of Utah).
  - David Koop (Ph.D. 2011, University of Utah).
  - Emanuele Santos (Ph.D. 2010, University of Utah. Co-advised with Cláudio Silva. )
  - Luciano Barbosa (Ph.D. 2009, University of Utah).
  - Huong Nguyen (M.S. 2011, University of Utah).
  - Tommy Ellkvist (Licentiate 2010, Linkoping University).
  - Ramesh Pinnamaneni (M.S. 2010, University of Utah).
  - Sumit Tandon (M.S. 2008, University of Utah).
  - Lorena Carlo (M.S. 2008, University of Utah).
  - Thanh Nguyen (M.S. 2008, University of Utah).
  - Eun Yong Kang (M.S. 2007, University of Utah).
  - Radhika Bheemidi (M.S. 2007, University of Utah).
  - Fang Du (M.S. 2004, OGI/OHSU).
  - Lingzhi Zhang (M.S. 2004, OGI/OHSU).
  - Leo Wang (M.S. 2003, OGI/OHSU).
- Advisor (in progress)
  - Aécio Santos (Ph.D. expected 2023, NYU CSE).
  - Haoxiang Zhang (Ph.D. expected 2023, NYU CSE).
- Ph.D. Committee Member
  - Nabeel Abdur Rehman (Advisor: Rumi Chunara), NYU Tandon. Defended in 2019.
  - Debjyoti Paul (Advisor: Feifei Li), University of Utah. Defended in 2020.
  - Eleni Tzirita Zacharatou (Advisor: Anastasia Ailamaki), EPFL Switzerland. Defended in 2019.

- Cristian Felix (Advisor: Enrico Bertini), NYU Tandon. Defended in 2019.
- Nivan Ferreira (Advisor: Cláudio Silva), NYU Poly. Defended in 2015.
- Jorge Poco (Advisor: Cláudio Silva), NYU Poly. Defended in 2015.
- Maria Christoforaki (Advisor: Torsten Suel), NYU Poly. Defended in 2014.
- Avishek Saha (Advisor: Suresh Venkatasubramanian), University of Utah, defended in August 2012.
- Dominic Jones (Advisor: Lee Hollaar), University of Utah, defended in April 2007.
- Paul Groth (Advisor: Luc Moreau), University of Southampton, England, defended in August 2007.
- Carlos Scheidegger (Advisor: Cláudio Silva), University of Utah, defended October 2009.
- Erik Anderson (Advisor: Cláudio Silva), University of Utah.
- Huy T. Vo (Advisor: Cláudio Silva), University of Utah, defended March 2011.
- Maurício Moraes (Advisors: Carlos Heuser and Viviane Moreira), Federal University of Rio Grande do Sul, Brazil.
- Graduate mentoring
  - Clifton Brooks (M.S. 2009-2010, University of Utah)
  - Pravin Chandru (M.S. 2008-2009, University of Utah)
  - Anusua Trivedi (M.S. 2009-2010, University of Utah)
  - Ravindra Lanka (M.S. 2007, University of Utah)
  - Harsh Doshi, Independent Study (University of Utah, 2007)
  - Kumar Chheda, Independent Study (University of Utah, 2007)
  - Mayank Maheshwari, Independent Study (University of Utah, 2005-2006)
  - Aaron Demille, Independent Study (University of Utah, 2006)
  - Marie Douriez, Internship at NYU, 2015 (B.S., Ecole Polytechnique, 2015)
  - Marcio Bastos, Internship at NYU, 2015 (B.S., Instituto Militar de Engenharia, Brazil, 2015)
  - Felipe Moraes, Internship at NYU, 2015 (M.S., Universidade Federal de Minas Gerais, Brazil, 2016)
  - Ferdinand Legros, Internship at NYU, 2016 (B.S., Ecole Polytechnique, 2016)
  - Clement Goubet, Internship at NYU, 2016 (B.S., Ecole Polytechnique, 2016)
  - João Felipe Pimentel, Internship at NYU, 2017 (PhD, Universidade Federal Fluminense, TBD)
  - Troy Kohwater, Internship at NYU, 2017 (PhD, Universidade Federal Fluminense, TBD)
  - Eleni Tzirita Zacharitou, Internship at NYU, 2017 (PhD, EPFL, Switzerland, 2019)
  - Lais Mota, Internship at NYU, 2018 (MS, Universidade Federal de Minas Gerais, Brazil, 2019)
  - Thais Gomes Almeida, Internship at NYU, 2018 (MS, Universidade Federal do Amazonas, Brazil, 2019)
- Undergraduate mentoring
  - Clifton Brooks, NSF-REU (University of Utah, 2008–2009) *Co-supervised with Cláudio Silva.*
  - Nathan Smith, NSF-REU (University of Utah, 2006–2007) *Co-supervised with Cláudio Silva.*

- Qing Zhang, Undergraduate Research Opportunity (UROP) (University of Utah, 2005) (currently a Ph.D. student at UCSD)
- Eun-Yong Kang, NSF-REU (University of Utah, 2005–2006) (currently a Ph.D. student at UCLA)
- Troy Wardrop, Independent Study (University of Utah, 2005)
- Internship mentor
  - Aline Bessa (Federal University of Minas Gerais—Brazil), NYU Poly, 2013
  - Fernando Seabra (Federal University of Rio de Janeiro—Brazil), University of Utah, 2009
  - Rafael Dahis (Federal University of Rio de Janeiro—Brazil), University of Utah, 2009
  - Karane Vieira (Federal University of Amazonas—Brazil), University of Utah, 2008
  - Marcelo Nery dos Santos (Federal University of Pernambuco—Brazil), University of Utah, 2006
  - Luciano Barbosa (Federal University of Pernambuco—Brazil), OGI/OHSU 2003 & 2004
  - Lourena Rocha (IMPA—Brazil), OGI/OHSU 2003
  - Denilson Barbosa (University of Toronto), OGI/OHSU 2003 (currently an Associate Professor at University of Alberta, Canada)
- Mentor, NYU Tandon School of Engineering’s Applied Research Innovations in Science and Engineering (ARISE) for High-School Students
  - Lihan Wang, Summer 2021
  - Becca Foulén, Summer 2021
- Mentor, Bell Labs Global Scholar
  - Leo Wolpert (University of Michigan), Summer 2002
  - Tao Yue (MIT), Summer 2001
- Mentor, Bell Labs Summer Internship Program
  - Michalis Petropoulos (UCSD), Summer 2002 (currently a faculty member at University at Buffalo, SUNY)
  - Maya Ramanath (Indian Institute of Science), Summer 2001 (a post-doc at Max-Planck-Institut, Germany; currently an Assistant Professor at IIT-Delhi, India)
  - Ioana Manolescu (INRIA), Summer 1999 (currently a senior researcher at INRIA)
  - Hasan Davulcu (University at Stony Brook), Summer 1998 (currently an Associate Professor at Arizona State University)
- Graduate Student Committees
  - Mark Valentine, M.S. Spring 2010 (Advisor: Olivia Sheng)

## Personal

Citizenship: U.S.A.

Fluent in English and Portuguese.