

## Juliana Freire de Lima e Silva

<http://www.cs.utah.edu/juliana>

School of Computing  
50 S. Central Campus Dr. Rm. 3190  
Salt Lake City, UT 84112

Phone: (801) 585-5047  
Fax: (801) 581-5843  
[juliana@cs.utah.edu](mailto:juliana@cs.utah.edu)

---

### 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

### Research Interests

- Scientific data management: scientific workflows; infrastructure to support the scientific discovery process; provenance management and analytics;
- Data integration: integration of a large number of autonomous, heterogeneous information sources; query processing and optimization; hidden-Web data retrieval and integration;
- Internet applications and Web information systems: focused Web crawling, document clustering and classification, personalization, information delivery to diverse environments;
- XML/semi-structured data management: storage, indexing, query processing and optimization;
- Databases and knowledge-base systems: query evaluation strategies, semantic Web.

### Professional Experience

- *Professor* July 2011 –  
Polytechnic Institute of New York University – New York University, New York, New York.
- *Faculty Member* July 2009 – June 2011  
Scientific Computing and Imaging (SCI) Institute – University of Utah, Salt Lake City, Utah
- *Associate Professor* July 2008 – June 2011  
School of Computing – University of Utah, Salt Lake City, Utah
- *Director* 2007 – 2008  
Center for Management of Provenance and Exploratory Workflows – University of Utah, Salt Lake City, Utah

- *Co-founder and CTO* 2007 – present  
VisTrails Inc.
- *Assistant Professor* July 2005 – June 2008  
School of Computing – University of Utah, Salt Lake City, Utah
- *Visiting Professor* August 2006 – present  
Department of Computer Science – Linköping University, Sweden
- *Visiting Assistant Professor* August 2004 – June 2005  
School of Computing – University of Utah, Salt Lake City, Utah
- *Assistant Professor* September 2002 – April 2006  
Department of Computer Science and Engineering – OGI School of Science & Engineering, Oregon Health & Science University, Beaverton, Oregon
- *Member of Technical Staff* December 1997 – September 2002  
Database Systems Research Department – Bell Laboratories, Murray Hill, New Jersey
- *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, NY
- *Programmer* July – August 1992  
Periphonics Corporation, Bohemia, NY

## Awards/Honors

- *Best paper award.* Educator program, Eurographics 2010.
- *Best poster award.* Brazilian Symposium on Databases (SBBDD), 2009.
- *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.* IBM, 2008.
- *Best paper award.* IEEE Visualization Conference, 2007.
- *Best paper runner-up.* WWW10 commendation, 2001.
- *First prize (for Syntactica).* AT&T Student Software Awards, 1997.
- *Award to attend the 1997 Grace Hopper Celebration, San Jose.* USENIX, 1997.
- *Research Travel Award.* Sigma XI, The Scientific Research Society, 1997.
- *Doctoral Fellowship.* CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, 1992.

## 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) totaling over 8 million dollars. 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*. Supplement for CAREER: Storing, Querying and Re-Using Provenance of Computational Tasks. **PI: Juliana Freire**. US\$ 60,000 (2010–2013).
2. *National Science Foundation, IIS 1050422*. III:EAGER:Collaborative Research: A Community Experiment Platform for Reproducibility and Generalizability. **PI: Juliana Freire**. US\$ 190,000 (2010–2012). Collaborative grant with Dennis Shasha, NYU Courant.
3. *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).
4. *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).
5. *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).
6. *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).
7. *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 (2009–2012).
8. *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.)
9. *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).
10. *National Science Foundation, IIS 0746500*. CAREER: Storing, Querying and Re-Using Provenance of Computational Tasks. **PI: Juliana Freire**. US\$ 499,999 (2008–2013)
11. *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 (2007–2010)
12. *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)
13. *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)

14. *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)

### Past Grants

1. *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)
2. *National Science Foundation, IIS-DMS 0534628*. XML Data Management: Taking Order and Updates into Account. **PI: Juliana Freire**. US\$ 270,000. (2006–2009)
3. *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).
4. *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).
5. *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).
6. *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).
7. *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).
8. *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

#### Books Co-Authored

- [1] *Semantica*. R. Larson, D. Warren, J. Freire, K. Sagonas and P. Gomez. MIT Press, 1997. ISBN: 0-262-62117-7
- [2] *Syntactica*. R. Larson, D. Warren, J. Freire and K. Sagonas. MIT Press, 1996. ISBN: 0-262-62106-1

#### Journals

- [3] *Exploring the Coming Repositories of Repeatable Experiments: Challenges and Opportunities*, J. Freire, P. Bonnet, and D. Shasha. In Proceedings of the VLDB Endowment. To appear, 2011. *Papers from the International Conference on Very Large Databases (VLDB)*.
- [4] *Synthesizing Products for Online Catalogs*, H. Nguyen, A. Fuxman, S. Paparizos, J. Freire, R. Agrawal. In Proceedings of the VLDB Endowment, vol. 4, no. 7, pp. 409-418, 2011. *Papers from the International Conference on Very Large Databases (VLDB)*.

- [5] 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. Schollwck, C. Silva, B. Surer, S. Todo, S. Trebst, M. Troyer, M.L. Wall, P. Werner, S. Wessel. JSTAT. To appear, 2011.
- [6] *XML Management for Bioinformatics Applications*. L. Strömbäck and J. Freire. IEEE Computing in Science & Engineering, 2011. To appear.
- [7] *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. (13 citations.)
- [8] *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.
- [9] *Indexing Web Form Constraints*, R. Mello, R. Pinnamaneni and J. Freire. Journal of Information and Data Management (JIDM), 1(2), pp. 343-358, 2010.
- [10] *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.
- [11] *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.
- [12] *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.
- [13] *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.
- [14] *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.
- [15] *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)*. (8 citations.)
- [16] *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*.
- [17] *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. (54 citations.)
- [18] *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.

- [19] *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.
- [20] *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. (45 citations.)
- [21] *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. (95 citations.)
- [22] *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. (60 citations.)
- [23] *Provenance for Visualizations: Reproducibility and Beyond*, C. Silva, J. Freire, S. P. Callahan. In *IEEE Computing in Science & Engineering*, 9(5): 82-90 (2007). (29 citations.)
- [24] *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*. (46 citations.)
- [25] *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). (9 citations.)
- [26] *Integrating Network Devices in a Meta-Directory: The MetaComm Experience*, J. Freire, D. Lieuwen and Joann Ordille. *Information Systems*, 27(3): 193-217 (2002).
- [27] *Automating Web Navigation with the WebVCR*, V. Anupam, J. Freire, B. Kumar and D. Lieuwen. *Computer Networks* 33(1-6): 503-517 (2000).
- [28] *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. (82 citations.)

### Conferences and Workshops

- [29] *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)*, 2011. To appear.
- [30] *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*, 2011. To appear.
- [31] *PruSM: A Prudent Schema Matching Strategy for Web Forms*. T. Nguyen, H. Nguyen and J. Freire. In *Proceedings of CIKM*, pp. 1385-1388, 2010.

- [32] *Using Latent-Structure to Detect Objects on the Web*. L. Barbosa and J. Freire. In Proceedings of WebDB, 2010.
- [33] *Querying Wikipedia Documents and Relationships*. H. Nguyen, T. Nguyen, H. Nguyen and J. Freire. In Proceedings of WebDB, 2010.
- [34] *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.
- [35] *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.
- [36] *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.
- [37] *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*.
- [38] *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.
- [39] *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.
- [40] *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.
- [41] *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.
- [42] *Desenvolvimento de Estruturas de Controle Explícito para o SGWfC VisTrails*, Fernando Seabra Chirigati, Rafael Dahis, Sergio Manuel Serra da Cruz, Juliana Freire, Claudio Silva, Marta Mattoso. Proceedings of the Brazilian Symposium on Databases (SBBD), 2009. *Best Poster Award*.
- [43] *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.
- [44] *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.
- [45] *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.
- [46] *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.

- [47] *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.
- [48] *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. .
- [49] *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.
- [50] *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.
- [51] *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. (9 citations.)
- [52] *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.
- [53] *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. (8 citations.)
- [54] *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. (18 citations.)
- [55] *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. (78 citations.)
- [56] *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. (9 citations.)
- [57] *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.
- [58] *Towards Enabling Social Analysis of Scientific Data*, J. Freire and Cláudio Silva. In CHI Social Data Analysis Workshop, 2008. (7 citations.)
- [59] *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. (26 citations.)
- [60] *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. (32 citations.)
- [61] *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. (23 citations.)

- [62] *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.
- [63] *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.*
- [64] *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. *(20 citations.)*
- [65] *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.
- [66] *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. *(14 citations.)*
- [67] *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. *(93 citations.)*
- [68] *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.
- [69] *Visualizing Uncertainty with Uncertainty Multiples*, R. B. Gilbert, F. Tonon, J. Freire, C. Silva, and D. R. Maidment. In ASCE 2006 GeoCongress.
- [70] *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. *(4 citations.)*
- [71] *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. *(82 citations.)*
- [72] *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. *(36 citations.)*
- [73] *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. *(52 citations.)*
- [74] *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.
- [75] *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. *(50 citations.)*

- [76] *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. (71 citations.)
- [77] *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. (21 citations.)
- [78] *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. (35 citations.)
- [79] *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. (7 citations.)
- [80] *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.
- [81] *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. (59 citations.)
- [82] *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. (24 citations.)
- [83] *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. (12 citations.)
- [84] *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. (6 citations.)
- [85] *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. (32 citations.)
- [86] *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. (133 citations.)
- [87] *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. (311 citations.)
- [88] *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. (102 citations.)
- [89] *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.* (89 citations.)

- [90] *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. (9 citations.)
- [91] *Web Services and Information Delivery for Diverse Environments*, J. Freire and B. Kumar. In Proceedings of VLDB Workshop on Technologies for E-Services, 2000.
- [92] *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. (85 citations.)
- [93] *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.
- [94] *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. (18 citations.)
- [95] *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. (9 citations.)
- [96] *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. (84 citations.)
- [97] *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. (9 citations.)
- [98] *Scheduling Strategies in SLG Revisited - Abstract*, J. Freire, T. Swift and D. Warren. Workshop on Tabulation in Parsing and Deduction (TAPD), 1998.
- [99] *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. (31 citations.)
- [100] *Using Logic Programming to Efficiently Evaluate Complex Queries*, J. Freire. In Proceedings of the Grace Hopper Technical Track, pp. 25-29, 1997.
- [101] *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. (125 citations.)
- [102] *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. (63 citations.)
- [103] *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.

- [104] *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. (22 citations.)
- [105] *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. (12 citations.)

### Invited Papers

- [106] 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.
- [107] *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.
- [108] *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.
- [109] *XML Selectivity Estimation*, M. Ramanath, J. Freire and A. Polyzotis. In Encyclopedia of Database Systems, L. Liu and T. Özsu (Eds.), 2009.
- [110] *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.
- [111] *Introducing the VisTrails Provenance Explorer Plugin for ParaView*, C. Silva, J. Freire, and J. Schreiner. Kitware Source, July 2009.
- [112] *Provenance Management: Challenges and Opportunities*, J. Freire. In Datenbanksysteme in Business, Technologie und Web (BTW), pp. 4, 2009. Abstract for keynote talk.
- [113] *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. (35 citations.)
- [114] *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.* (105 citations.)
- [115] *Accessing and Personalizing Web Content and Services*, J. Freire. Invited to WWW2002 Workshop on Mobile Search, May 2002. *Paper corresponding to invited talk.* (18 citations.)

### Book Chapters

- [116] *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.
- [117] *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.

- [118] *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

- [119] Provenance in Web Applications. IEEE Internet Computing, vol. 15, no 1, Jan/Feb 2011. Eds. Geetika T. Lakshmanan, Francisco Curbera, Juliana Freire, Amit Sheth.
- [120] 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.
- [121] 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.
- [122] *Proceedings of International Provenance and Annotation Workshop (IPAW), 2008*, Eds. J. Freire, D. Koop and L. Moreau.
- [123] *Proceedings of the ACM SIGMOD International Workshop on Web and Databases (WebDB), 2003*, Eds. V. Christophides, J. Freire.
- [124] *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.
- [125] *The XSB Programmer's Manual*. K. Sagonas, T. Swift, D. Warren, J. Freire, P. Rao. Available at <http://xsb.sourceforge.net>.

### Patents Issued

- [126] *Applications of executable shopping lists* U.S. patent number 7,103,566 (September 5, 2006). Assigned to Lucent Technologies.
- [127] *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.
- [128] *Active Database Trigger Processing Using a Trigger Gateway*. U.S. patent number 6,594,656 (July 15, 2003). Assigned to Lucent Technologies.
- [129] *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.
- [130] *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.
- [131] *Method for providing fast access to dynamic content on the world wide web*. European patent number EP1081607 (March 7, 2001). Assigned to Lucent Technologies.
- [132] *Database Trigger Gateway and its Operating Method*. Japanese patent number JP2000215092 (August 4, 2000). Assigned to Lucent Technologies.

### Patents Pending

- [133] *Method and System for Adaptive Discovery of Content on a Network*, patent filed by the University of Utah, February 2008.
- [134] *Method and System for Clustering Identified Forms*, patent filed by the University of Utah, January 2008.
- [135] *Managing Provenance of the Evolutionary Development of Workflows*, patent filed by the University of Utah, April 2007.
- [136] *Analogy-Based Updates for Rapid Development of Data Processing Results*, patent filed by the University of Utah, April 2007.
- [137] *Managing Provenance for an Evolutionary Workflow Process in a Collaborative Environment*, patent filed by the University of Utah, April 2007.

### Freely-Available Software Systems and Web Sites

- **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. Since its beta release in 2007, VisTrails has been downloaded over 15,000 times.  
<http://www.vistrails.org>
- **DEFOG:** DEFOG is a new information visualization system. It is based on a model for visualization 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. It is currently being used in a number of projects, including the analysis of metabolic pathway (collaboration with researchers at the University of Washington) and large-scale genotype/phenotype data analysis (collaboration with the School of Medicine at the University of Utah).
- **DeepPeep.org:** DeepPeep is 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.
- **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.

- **ACHE:** ACHE is a focused crawler specialized in locating forms, which are sparsely distributed on the Web. ACHE is able to efficiently locate forms by using the contents of pages to focus the crawl on a topic, by prioritizing promising links within the topic; and by also following links that may not lead to immediate benefit. In addition, ACHE uses active learning: it automatically learns patterns of promising links and adapts its focus as the crawl progresses, thus greatly reducing the amount of required manual setup and tuning.
- **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

- *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

- *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

- *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.

### Selected Conference Talks

- *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

- 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**

### **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.

### **External Service**

- 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.

### Minority Involvement

- *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

- **Editorial Boards:**
  - *IEEE Transactions on Knowledge and Data Engineering.*
  - *IEEE Internet Computing.*
  - *Journal of Information and Data Management.*
  - *Associate Editor of the Journal of Web Engineering.*
- **International liaison**, Brazilian Computer Society.
- **Conference Organization:**
  1. *Program Committee Group Leader*: ACM International Conference on Management of Data (SIGMOD), 2012.
  2. *Program Chair*: Experiments and Analysis Track of the International Conference on Very Large Data Bases–Istanbul, September 2012.
  3. *Program Chair*: Workshop on the Theory and Practice of Provenance (TAPP)– Crete, Greece, June 2011.
  4. *Program Chair*: International World Wide Web Conference (WWW), 2010.
  5. *Chair*: The Second International Workshop on Role of Semantic Web in Provenance Management (SWPM 2010) – Shanghai, China, November 2010.
  6. *Co-Chair*: First International Workshop on Traceability and Compliance of Semi-Structured Processes (TC4SP10) – Hoboken, New Jersey, USA, September 2010.
  7. *Co-Chair*: International Workshop on Role of Semantic Web in Provenance Management (SWPM) – Washington D.C., USA, October 2009.
  8. *Workshops Co-Chair*: International Conference on Very Large Databases (VLDB) – Lyon, France, August 2009.
  9. *Track Chair*: ACM Conference on Information and Knowledge Management (CIKM) – Napa Valley, California, October 2008.

10. *Co-organizer*: Symposium on Provenance in Scientific Workflows, Salt Lake City, Utah, October 2008.
11. *Co-Chair*: International Provenance and Annotation Workshop (IPAW)– Salt Lake City, UT, June 2008.
12. *Vice-Chair*: International World Wide Web Conference (WWW), XML and Web Data track – Beijing, China, April 2008.
13. *Co-Chair*: HPDC Provenance Challenge Workshop – Monterey Bay, CA, June 2007.
14. *Vice-Chair*: IEEE International Conference on Data Engineering (ICDE) 2007 – Istanbul, Turkey, April 2007.
15. *Deputy Vice-Chair*: International World Wide Web Conference (WWW) 2007 XML and Web Data track – Banff, Canada, May 2007.
16. *Co-Chair*: IEEE Visualization Workshop on Grid-Based Visualization – Baltimore, October 2006.
17. *Vice-Chair*: International World Wide Web Conference (WWW) 2005 Browsers and User Interfaces track – Edinburgh, Scotland, May 2005.
18. *Deputy Vice-Chair*: International World Wide Web Conference (WWW) 2003 Browsers and User Interfaces track – Budapest, Hungary, May 2003.
19. *Chair*: SIGMOD 2003 Demonstrations Committee – San Diego, June 2003.
20. *Co-Chair*: SIGMOD International Workshop on the Web and Databases (WebDB 2003) – San Diego, CA, June 2003.
21. *Co-Chair*: First International Workshop on Tabling in Logic Programming – Leuven, Belgium, July 1997.

● **Program committee member of:**

1. International Conference on Scientific and Statistical Database Management (SSDBM), 2011.
2. IEEE International Conference on e-Science 2010.
3. International Conference on Scientific and Statistical Database Management (SSDBM), 2010.
4. International World Wide Web Conference (WWW), 2010.
5. International Conference on Data Engineering (ICDE), 2010.
6. International Workshop on the Web and Databases (WebDB), 2009.
7. Brazilian Symposium on Databases (SBBD), 2009.
8. International Workshop on Role of Semantic Web in Provenance Management (SWPM), 2009.
9. First International Workshop on Data and Process Provenance (WDPP), 2009.
10. ACM International Conference on Management of Data (SIGMOD), 2009 (tutorial program).
11. International Conference on Very Large Databases (VLDB) Ph.D. Workshop, 2009.
12. IEEE International Workshop on Scientific Workflows (SWF), 2009.
13. International Conference on Very Large Databases (VLDB), 2008.
14. ACM Conference on Information and Knowledge Management (CIKM), 2008.
15. Brazilian Symposium on Databases (SBBD), 2008.
16. International Provenance and Annotation Workshop (IPAW), 2008.

17. Data Integration in the Life Sciences (DILS), 2008.
18. First International Workshop on REsource Discovery (RED), 2008.
19. IEEE International Workshop on Scientific Workflows (SWF), 2008.
20. International World Wide Web Conference (WWW), 2008.
21. International Conference on Data Engineering (ICDE), 2008.
22. IEEE Workshop on Scientific Workflows, 2007.
23. Brazilian Symposium on Databases (SBBD), 2007.
24. International Workshop on the Web and Databases (WebDB), 2007.
25. International Conference on Very Large Databases (VLDB), 2007.
26. ACM International Conference on Management of Data (SIGMOD), 2007 (industrial program).
27. International Conference on Data Engineering (ICDE), 2007.
28. International World Wide Web Conference (WWW), 2007.
29. International Workshop on Challenges in Web Information Retrieval and Integration (WIRI), 2006.
30. International World Wide Web Conference (WWW), 2006.
31. Brazilian Symposium on Databases (SBBD), 2006.
32. ACM Conference on Information and Knowledge Management (CIKM), 2005.
33. International Workshop on the Web and Databases (WebDB), 2005.
34. Latin American Web Congress (LAWeb), 2005.
35. Brazilian Symposium on Databases (SBBD), 2005.
36. International Conference on Data Engineering (ICDE), 2005.
37. International World Wide Web Conference (WWW), 2005.
38. International XML Database Symposium (XSym), 2004.
39. International Conference on Web Engineering (ICWE), 2004.
40. Brazilian Symposium on Databases (SBBD), 2004.
41. International Workshop on the Web and Databases (WebDB), 2004.
42. ACM International Conference on Management of Data (SIGMOD), 2004.
43. International Conference on Very Large Databases (VLDB), 2003.
44. International Workshop on the Web and Databases (WebDB), 2003.
45. International World Wide Web Conference (WWW), 2003.
46. IJCAI Workshop on Information Integration on the Web, 2003.
47. Symposium on String Processing and Information Retrieval (SPIRE), 2003.
48. International Symposium on Practical Aspects of Declarative Languages (PADL), 2003.
49. Workshop on Foundations of Models and Languages for Data and Objects (FMLDO), 2002.
50. Workshop on Efficient Web-based Information Systems (EWIS), 2002.
51. Symposium on String Processing and Information Retrieval (SPIRE), 2002.
52. Brazilian Symposium on Multimedia and Hypermedia Systems (SBMIDIA), 2002.
53. International Workshop on the Web and Databases (WebDB), 2002.

54. SIGMOD Demonstrations, 2002.
  55. International World Wide Web Conference (WWW), 2002.
  56. Web Engineering track - International World Wide Web Conference (WWW), 2002.
  57. International Conference on Data Engineering (ICDE), 2002.
  58. International World Wide Web Conference (WWW10), 2001.
  59. International Conference on Data Engineering (ICDE), 2001.
  60. International World Wide Web Conference (WWW9), 2000.
  61. International Conference on Computational Logic (CL), 2000.
  62. International Workshop on the Web and Databases (WebDB), 2000.
  63. International Conference on E-Commerce and Web (EC&Web), 2000.
  64. First International Workshop on Tabling in Logic Programming, 1997.
- *Reviewer* for Information Systems Journal–2010; Frontiers in Neuroinformatics–2009; Communications of the ACM (CACM)–2008, 2009; ACM Computing Surveys–2008; ACM International Conference on Management of Data (SIGMOD)–2008; IEEE Information Visualization Conference (INFOVIS)–2007; Journal of Data and Knowledge Engineering (DKE)–2005; ACM Transactions on Databases (TODS)–2005; ACM Transactions on Internet Technology (TOIT)–2005; VLDB Journal–2000, 2005, 2008; IEEE Transactions on Knowledge Bases and Engineering (TKDE)–2002, 2003,2004, 2005, 2006, 2009; Journal of Web Engineering –2004; IEEE Software–2003; Very Large Databases Conference (VLDB)–1998, 1999, 2000, 2002; International Conference on Management of Data (SIGMOD)–1998, 2001; World Wide Web Journal–1999; Journal of Functional and Logic Programming–1998.
  - *Publicity Chair*: International Symposium on Logic Programming, ILPS’97 – Port Jefferson, October 1997.
  - *Member* of IEEE, ACM and ACM SIGMOD.

### Review of Grant Proposals

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

### Teaching

- *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
  - Lauro Lins (2007-present) *Co-mentored by Cláudio Silva.*
- Advisor (students graduated)
  - Emanuele Santos (Ph.D. 2010, University of Utah. Co-advised with Cláudio Silva. )
  - Luciano Barbosa (Ph.D. 2009, 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)

- David Koop (Ph.D. expected 2011, University of Utah)
- Hoa Nguyen (Ph.D. expected 2011, University of Utah)
- Thanh Nguyen (Ph.D. expected 2012, University of Utah)
- Huong Nguyen (Ph.D. expected 2013, University of Utah)
- Sergio Mergen (co-advised with Carlos Heuser. Ph.D. expected 2011, Federal University of Rio Grande do Sul—Brazil)
- Karane Vieira (co-advised with Altigran Silva. Ph.D. expected 2011, Federal University of Amazonas—Brazil)
- Ph.D. Committee Member
  - 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 Heuse 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)
- 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
  - 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 Assistant Professor at University of Alberta, Canada)
  
- 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 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)