

PAUL GROTH

pgroth@isi.edu

<http://www.isi.edu/~pgroth>

Objective

Investigate the juncture between knowledge representation and distributed systems to facilitate the production of new knowledge.

Education

Jan 2004 –

Oct 2007

PhD Computer Science, University of Southampton

- Thesis title: “The Origin of Data: Enabling the Determination of Provenance in Multi-institutional Scientific Systems through the Documentation of Processes”
- My research focused on provenance in e-Science applications. It develops novel concepts for creating, recording, and organizing provenance information. I have applied these concepts to a large scale bioinformatics application.

1999 - 2002

B.S. Computer Science, University of West Florida (Magna Cum Laude)

Fall 2001 –

Spring 2002

University of Ulm, Germany (Study Abroad)

Summer 2001

Fukuoka Communication Arts, Japan (Study Abroad)

Experience

Oct 2007 -

present

Postdoctoral Research Associate

Intelligent Systems Division

Information Sciences Institute

University of Southern California

My work at ISI has taken place in the context of two large multi-team projects, Windward and Bootstrapped Learning.

Windward: Scalable Knowledge Discovery Through Grid Workflows

This project aims to develop workflow technologies for large scale data analysis and knowledge discovery tasks through the combination of artificial intelligence and Grid technologies.

Key Contributions:

- Designed the provenance ontology and specification that was adopted by the project. This enabled the end-to-end determination of the provenance of data through the various stages and components of the system including workflow execution, planning, and generation.
- Developed a software library to assist developers in integrating their software with the aforementioned provenance specification.
- Provided support for system integration, testing and documentation tasks.

DARPA Bootstrapped Learning Program

This project aims to develop an electronic student that learns from a teacher using different forms of natural instruction. Within this project, ISI's role is to develop components for learning procedures from tutorial instruction.

Key Contributions:

- Co-designed the architecture for TellMe: a system to learn from human tutorial instruction.

- Developed an implementation of TellMe that was successfully integrated into a larger electronic student that allowed other learning components to leverage the knowledge acquired by TellMe.
- Integrated TellMe into the Wings Workflow System to enable workflows to be input using natural language.
- Cataloging the gaps between natural language descriptions of workflows and their programmatic representation.

In addition to my contributions to these projects, I have also pursued other research avenues. Accomplishments include:

- The development, implementation and evaluation of a distributed provenance query algorithm.
- Investigated the use of workflows in the social networking site, Facebook through the development of prototypes for their use in the domains of privacy and scientific knowledge sharing.
- The integration of the Wings Workflow System with MIT/Harvard Broad Institute's GenePattern workflow system.

2004 -2007

PhD Student

Provenance Aware Service Oriented Architecture (PASOA) project

<http://www.pasoa.org>

This project aims to develop novel computer science techniques and tools for the support of provenance in e-Science.

Key Contributions:

- Collaborated with a bioinformatician to make his software both provenance-aware and grid enabled providing him with larger and better documented data sets.
- Co-developed PReServ, an open source web service that allows for the storage and retrieval of provenance information.
 - <http://twiki.pasoa.ecs.soton.ac.uk/bin/view/PASOA/SoftWare>
 - The software has been used by a number of collaborators including the following institutions and projects:
 - The University of Leeds
 - China Research and Development environment Over Wide-area Network (CROWN Grid)
 - Indiana University
 - The Information Sciences Institute at the University of Southern California
 - The EU Provenance Project

2004 – 2007

PhD Student

EU Provenance Project

<http://www.gridprovenance.org>

This Europe Union funded project aims to develop industrial strength provenance infrastructure demonstrated through applications in aerospace engineering and organ transplant management. The partners involved in the Provenance project are University of Southampton, IBM UK, Cardiff University, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Universitat Politècnica de Catalunya, and Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézet (SZTAKI).

Key Contributions:

- Heavily participated in the design and specification of the architecture used by the project.
- Provided support to project developers through tutorials, presentations, and on and

offline meetings, which allowed them to understand and implement the architecture.

2003

Internship
Fraunhofer-Institut für Produktionstechnik und Automatisierung
<http://www.ipa.fhg.de>

Co-developed a prototype multi-agent system to control semiconductor factories, which successfully integrated with a commercial simulator.

Fall 2001 –
Spring 2002

Internship
University of Ulm, Distributed Systems Department
<http://www-vs.informatik.uni-ulm.de>

Designed and implemented a directory service for the Mobility And Service Adaptation in Heterogeneous Mobile Networks (MASA) project. Project partners were the University of Ulm, NEC Europe Ltd. and Siemens AG.

1998 – 2003

Student Research Assistant
The Institute for Human & Machine Cognition
<http://www.ihmc.us>

Developer on NOMADS (<http://www.ihmc.us/research/projects/Nomads>) a Java-based mobile agent distributed system based on a clean room implementation of the Java virtual machine.

Key Contributions:

- Development of key demonstrators of the system capabilities including its application in both military and Grid contexts.
- Modification of the JVM for controlling CPU resource consumption.
- Development of agent tracking mechanisms.

Developer on two versions of CMapTools (<http://cmap.ihmc.us/>). The software empowers users to easily construct, navigate, share, and criticize knowledge models through concept maps.

Key Contributions:

- Participated in the redesign of the graphical user interface leading to increased usability.
- Designed and implemented three complimentary components for the annotation of concept maps.
- Development of a component for styling concept maps.
- Development of a component that allowed users to manage and attach external resources to concept maps.
- Participated in the design and packaging of a large meteorology knowledge model for the STORM-LK project. (<http://www.ihmc.us/research/projects/StormLK>)
- Supported the development of the NASA Ames CMEX Mars knowledge model for educational outreach. (<http://cmex.ihmc.us/CMEX/index.html>)

1997

Developer
Santa Rosa County School Board Purchasing Department

Designed and implemented a web based automated purchasing system. This site allowed service providers to the county to conduct bid-based purchasing transactions on line.

Program Committees

Program Committee Member for:

- Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)
- International Conference on Intelligent User Interfaces 2009 (IUI-09)
- The First International Workshop on Data and Process Provenance 2009 (WDPP'09)

- 3rd Workshop on Emerging Web Service Technologies (WEWST-2008)
- Second International Provenance and Annotation Workshop (IPAW 2008)

Reviews

Reviewer for journals:

- Interacting with Computers
- Computers & Geosciences
- Computing and Informatics

Reviewer for conferences and workshops where I was not a PC member:

- 9th IEEE/ACM International Conference on Grid Computing (Grid 2008)
- International Supercomputing Conference 2007 (ISC'07)
- Second International Workshop on Provenance and Annotation (IPAW'06)
- Twelfth International European Conference on Parallel and Distributed Computing (Euro-Par'06)
- Web Intelligence 2005
- International Parallel and Distributed Processing Symposium (IPDPS'05)

Panels, Invited Workshops, Invited Talks

- The Role of Biomedical Informatics in Overcoming Barriers in Cancer Research Workshop. May 21-23, 2008. Hosted by the NIH-National Cancer Institute
- Panel discussion on UK-Sino e-Science Collaborations. UK e-Science All Hands Meeting. September 2005. <http://www.allhands.org.uk/2006/programme/workshops/uk-china.html>
- CMapTools: Designing a useable and research oriented learning technology. Learning Societies Lab, School of Electronics and Computer Science, University of Southampton. June 6, 2006.
- What Happened? Using Provenance for Compliance and Verification. Institute for Human and Machine Cognition. Jan. 11, 2006
- The Origin of Data: Determining the Provenance of Data Produced by Multi-institutional Applications. At the AI Seminar Series. Intelligent Systems Division. Information Sciences Institute. Nov. 11, 2008

Teaching

- I am excited about teaching and believe that my passion for computer science and my presentation skills will transfer well into the classroom.
- Invited lecture on Semantic Technologies for Introduction to Grid Computing, Fall 2007 at the University of Southern California
- Presented tutorials about the PASOA project for roughly 15 academic and industrial parties.
- Marking for COMP2002 Computational Systems (e.g. correcting Scheme assignments). Spring 2005. University of Southampton

Outreach and Research Funding

- Over the past one and a half years I have made a concerted effort to learn about research proposal authorship. I have helped my supervisors author three proposals.
- I currently publish a blog about the wider view of provenance as a bridge between my research and a more general audience. The blog can be found at <http://thinklinks.wordpress.com>

Awards & Honors

- Member University of Southampton School of Electronics and Computer Science Graduate School Board
- Global Grid Forum 12 Student Scholarship
- NordiChi 2004 Student Volunteer
- Honorable Mention - Java Programming Language Competition AITP National Collegiate Conference (2001)
- Awarded the Baden-Württemberg-Stipendium for study in Ulm.
- Awarded a scholarship by the Jikei Technical College Group of Japan for study in Fukuoka.

- John C Pace Scholar (University of West Florida's highest scholarship program)
- 8 times on the University of West Florida Dean or Presidents Honor List

Memberships

- Phi Kappa Phi
- Association of Computing Machinery (ACM)
- The Institute of Electrical and Electronics Engineers (IEEE)
- Association for the Advancement of Artificial Intelligence (AAAI)

Courses and Seminars

- 2nd International Summer School on Grid Computing 2004. Vico Equense, Italy
- 1 Day Solution Selling Course, May 30, 2007. London
- 5 Day Seminar in the Restoration of Painting and Canvases. Held at Oro e Colore Workshop, Florence, Italy. January 2002.

Languages

German (Con conversationally fluent)

Publication List

The following publications are listed in reverse chronological order and are organized by type. Significant publications are marked with a *.

Journal

1. *Paul Groth and Luc Moreau. Recording process documentation for provenance. *IEEE Transactions on Parallel and Distributed Systems*, 2008. To be published.
2. *Paul Groth, Simon Miles, and Luc Moreau. A Model of Process Documentation to Determine Provenance in Mash-ups. *Transactions on Internet Technology (TOIT)*, 9(1), 2008
3. Simon Miles, Paul Groth, Ewa Deelman, Karan Vahi, Gaurang Mehta, and Luc Moreau. Provenance: The bridge between experiments and data. *Computing in Science and Engineering*, 10(3):38-46, May/June 2008.
4. Luc Moreau, Paul Groth, Simon Miles, Javier Vazquez-Salceda, John Ibbotson, Sheng Jiang, Steve Munroe, Omer Rana, Andreas Schreiber, Victor Tan, and Laszlo Varga. The provenance of electronic data. *Communications of the ACM*, 51(4):52-58, 2008.
5. Jie Xu, Paul Townend, Nik Looker, and Paul T. Groth. Ft-grid: a system for achieving fault tolerance in grids. *Concurrency and Computation: Practice and Experience*, 20(3):297-309, 2008.
6. Simon Miles, Paul Groth, Miguel Branco, and Luc Moreau. The requirements of using provenance in e-science experiments. *Journal of Grid Computing*, 5(1):1-25, 2007.
7. Simon Miles, Paul Groth, Steve Munroe, Sheng Jiang, Thibaut Assandri, and Luc Moreau. Extracting causal graphs from an open provenance data model. *Concurrency and Computation: Practice and Experience*, 20(5), April 2007.
8. Simon Miles, Sylvia C. Wong, Weijian Fang, Paul Groth, Klaus-Peter Zauner, and Luc Moreau. Provenance-based validation of e-science experiments. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 5:28-38, 2007.
9. David W. Eccles and Paul T. Groth. Wolves, bees, and football: Enhancing coordination in sociotechnological problem solving systems through the study of human and animal groups. *Computers in Human Behavior*, 23(6):2778-2790, 2007.

10. David W. Eccles and Paul T. Groth. Problem solving systems theory: Implications for the design of socio-technological systems. *Technology, Instruction, Cognition and Learning (TICL)*, 3(3-4):323 - 343, 2006.
11. David W. Eccles and Paul T. Groth. Agent coordination and communication in sociotechnological systems: Design and measurement issues. *Interacting with Computers*, 18(6):1170-1185, 2006.

Conference and Workshop

12. *Paul Groth and Yolanda Gil. A scientific workflow construction command line. In *International Conference on Intelligent User Interfaces 2009 (IUI2009)*, 2009.
13. P. Groth. Exposing privacy obligation policies in social networking sites. In *AAAI 2009 Spring Symposium on Social Semantic Web*, 2009. To be published.
14. P. Groth and Y. Gil. Scaffolding instructions to learn procedures from users. In *AAAI 2009 Spring Symposium Agents that Learn from Human Teachers*, 2009. To be published.
15. *Paul T. Groth. A distributed algorithm for determining the provenance of data. In *Proceedings of the fourth IEEE International Conference on e-Science (e-Science'08)*, 2008. To be published.
16. Simon Miles, Paul Groth, and Michael Luck. Handling Mitigating Circumstances for Electronic Contracts. In *Proceedings of the AISB 2008 Symposium on Behaviour Regulation in Multi-agent Systems*, pages 37-42, Aberdeen, UK, April 2008. The Society for the Study of Artificial Intelligence and Simulation of Behaviour.
17. Simon Miles, Ewa Deelman, Paul Groth, Karan Vahi, Gaurang Mehta, and Luc Moreau. Connecting scientific data to scientific experiments with provenance. In *Proceedings of the third IEEE International Conference on e-Science and Grid Computing (e-Science'07)*, Bangalore, India, December 2007.
18. Steve Munroe Michael Luck Luc Moreau Simon Miles, Paul Groth. AgentPrIME: Adapting MAS Designs to Build Confidence. In *Proceedings of the 8th International Workshop on Agent Oriented Software Engineering*, 2007.
19. Paul Groth, Simon Miles, and Steven Munroe. Principles of high quality documentation for provenance: A philosophical discussion. In Luc Moreau and Ian Foster, editors, *Proceedings of Third International Provenance and Annotation Workshop (IPAW'06)*, volume 4145 of *Lecture Notes in Computer Science*, Chicago, IL, 2006. Springer.
20. Victor Tan, Paul Groth, Simon Miles, Sheng Jiang, Steve Munroe, Sofia Tsasakou, and Luc Moreau. Security issues in a soa-based provenance system. In Luc Moreau and Ian Foster, editors, *Proceedings of Third International Provenance and Annotation Workshop (IPAW'06)*, volume 4145 of *Lecture Notes in Computer Science*, Chicago, IL, 2006. Springer.
21. Liming Chen, Victor Tan, Fenglian Xu, Alexis Biller, Paul Groth, Simon Miles, John Ibbotson, Michael Luck, and Luc Moreau. A proof of concept: Provenance in a service oriented architecture. In *Proceedings of the Fourth All Hands Meeting (AHM'05)*, September 2005.
22. Paul Groth, Simon Miles, and Luc Moreau. PReServ: Provenance recording for services. In *Proceedings of the UK OST e-Science Fourth All Hands Meeting (AHM05)*, September 2005.
23. Paul Townsend, Paul Groth, Nik Looker, and Jie Xu. FT-Grid: A fault-tolerance system for e-science. In *Proceedings of the UK OST e-Science Fourth All Hands Meeting (AHM05)*, September 2005.
24. Sylvia C. Wong, Simon Miles, Weijian Fang, Paul Groth, and Luc Moreau. Validation of e-science experiments using a provenance-based approach. In *Proceedings of Fourth All Hands Meeting (AHM'05)*, Nottingham, September 2005.

25. *Paul Groth, Simon Miles, Weijian Fang, Sylvia C. Wong, Klaus-Peter Zauner, and Luc Moreau. Recording and using provenance in a protein compressibility experiment. In Proceedings of the 14th IEEE International Symposium on High Performance Distributed Computing (HPDC'05), July 2005.
26. Paul Townend, Paul Groth, and Jie Xu. A provenance-aware weighted fault tolerance scheme for service-based applications. In Proc. of the 8th IEEE International Symposium on Object-oriented Real-time distributed Computing (ISORC 2005), May 2005.
27. David W. Eccles and Paul T. Groth. Creating expert problem solving systems. In Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05), Jan 2005.
28. Sylvia C. Wong, Simon Miles, Weijian Fang, Paul Groth, and Luc Moreau. Provenance-based validation of e-science experiments. In Proceedings of 4th International Semantic Web Conference (ISWC'05), volume 3729 of Lecture Notes in Computer Science, pages 801-815, Galway, Ireland, Nov 2005. Springer-Verlag.
29. Paul Groth, Michael Luck, and Luc Moreau. A protocol for recording provenance in service-oriented grids. In Proceedings of the 8th International Conference on Principles of Distributed Systems (OPODIS'04), Grenoble, France, December 2004.
30. David W. Eccles and Paul T. Groth. Wolves, football, and ambient computing: facilitating collaboration in problem solving systems through the study of human and animal groups. In Proceedings of the Third Nordic conference on Human-Computer interaction, pages 269-275. ACM Press, October 2004.
31. Paul Groth, Michael Luck, and Luc Moreau. Formalising a protocol for recording provenance in grids. In Proceedings of the UK OST e-Science Second All Hands Meeting 2004 (AHM'04), Nottingham, UK, September 2004.
32. Niranjan Suri, Jeffrey Bradshaw, Andrzej Uszok, Maggie Breedy, Marco Carvalho, Paul Groth, Renia Jeffers, Matt Johnson, Shri Kulkarni, James Lott, Mark Burstein, Brett Benyo, and David Diller. Toward DAML-based policy enforcement for semantic data transformation and filtering in multi-agent systems. In Proceedings of the second international joint conference on Autonomous agents and multiagent systems, pages 1132-1133. ACM Press, 2003.
33. Niranjan Suri, Jeffrey M. Bradshaw, Marco M. Carvalho, Thomas B. Cowin, Maggie R. Breedy, Paul T. Groth, and Raul Saavedra. Agile Computing: Bridging the Gap between Grid Computing and Ad-hoc Peer-to-Peer Resource Sharing. In Proceedings of the 3rd International Symposium on Cluster Computing and the Grid, page 618. IEEE Computer Society, 2003.
34. Niranjan Suri, Marco Carvalho, Jeffrey M. Bradshaw, Maggie R. Breedy, Thomas B. Cowin, Paul T. Groth, Raul Saavedra, and Andrzej Uszok. Enforcement of communications policies in software agent systems through mobile code. In Proceedings of the 4th IEEE International Workshop on Policies for Distributed Systems and Networks, page 247. IEEE Computer Society, 2003.
35. Niranjan Suri, Paul T. Groth, and Jeffrey M. Bradshaw. While You're Away: A System for Load-Balancing and Resource Sharing Based on Mobile Agents. In Proceedings of the 1st International Symposium on Cluster Computing and the Grid, page 470. IEEE Computer Society, 2001.
36. Paul T. Groth and Niranjan Suri. CPU Resource Control and Accounting in the NOMADS Mobile Agent System. In Proceedings of the ACM OOPSLA Workshop on Experiences with Autonomous Mobile Objects and Agent Based Systems, Minneapolis, USA, Oct. 2000., 2000.
37. Niranjan Suri, Jeffrey Bradshaw, Maggie R. Breedy, Paul T. Groth, Gregory A. Hill, and Renia Jeffers. Strong Mobility and Fine-Grained Resource Control in NOMADS. In Friedemann Mattern David Kotz, editor, Proceedings of the Second International Symposium on Agent Systems and Applications and Fourth International Symposium on Mobile Agents, ASA/MA 2000, Zurich, Switzerland, volume 1882 / 2004 of Lecture Notes in Computer Science, pages 2-15. Springer-Verlag, 2000.

38. Niranjan Suri, Jeffrey M. Bradshaw, Maggie R. Breedy, Paul T. Groth, Gregory A. Hill, Renia Jeffers, and Timothy S. Mitrovich. An Overview of the NOMADS Mobile Agent System. In Proceedings of ECOOP'2000, Nice, France, 2000, 2000.
39. Niranjan Suri, Jeffrey M. Bradshaw, Maggie R. Breedy, Paul T. Groth, Gregory A. Hill, Renia Jeffers, Timothy S. Mitrovich, Brian R. Pouliot, and David S. Smith. NOMADS: toward a strong and safe mobile agent system. In Proceedings of the fourth international conference on Autonomous agents, pages 163-164. ACM Press, 2000.

Book Chapter

40. Paul Groth, Steve Munroe, Simon Miles, and Luc Moreau. *In Lucio Grandinetti (ed.), HPC and Grids in Action*, chapter Applying the Provenance Data Model to a Bioinformatics Case. IOS Press, January 2008.

Technical Reports

41. Luc Moreau (Editor), Beth Plale, Simon Miles, Carole Goble, Paolo Missier, Roger Barga, Yogesh Simmhan, Joe Futrelle, Robert McGrath, Jim Myers, Patrick Paulson, Shawn Bowers, Bertram Ludaescher, Natalia Kwasnikowska, Jan Van den Bussche, Tommy Ellkvist, Juliana Freire, Paul Groth. The Open Provenance Model (v1.01). Technical Report, ECS, University of Southampton. Available at <http://eprints.ecs.soton.ac.uk/16148/>
42. *Paul Groth and Sheng Jiang and Simon Miles and Steve Munroe and Victor Tan and Sofia Tsakou and Luc Moreau. An Architecture for Provenance Systems. Technical Report, ECS, University of Southampton. Available at <http://eprints.ecs.soton.ac.uk/13216/>