Latanya Sweeney

2010 WL 5698549 (D.Mass.) (Court Filed Expert Resume)
Filed in Tele-Publishing, Inc. v. Facebook, Inc., et al., (D.Mass., March 29, 2010)

Area of Expertise: Engineering & Science >> Engineers/Engineering
Area of Expertise: Computers & Electronics >> Software

PROFESSION:

Director,

Local contact information for 2010:

Harvard University

Visiting Scholar, Computer Science

School of Engineering and Applied Science

33 Oxford St, MD329
Cambridge, MA 02138

Phone: (617) 496-5661

Fax: (617) 496-6404

Email: latanya@seas.harvard.edu

WWW: *http://dataprivacylab.org/people/sweeney/*
AdvanceHIT Project

AdvanceHIT.org

(*PDF version*)

EDUCATION

• Massachusetts Institute of Technology, Cambridge, MA., Ph.D. in Computer Science 2001. *Computational Disclosure Control: Theory and Practice.*

• Massachusetts Institute of Technology, Cambridge, MA., S.M. 1997 in Electrical Engineering and Computer Science. *Sprees, a Finite-State Orthographic Learning System that Recognizes and Generates Phonologically Similar Spellings.* GPA 4.9/5.0. Finalist in MasterWorks.

• Harvard University, Cambridge, MA., ALB 1995 in Computer Science, Cum Laude. *A Coin Toss: the Dialectical Odds aren't always 50/50.* Honors grades in all courses. Completed graduate courses in computer science, mathematics, physics, educational psychology and philosophy. Delivered graduation speech.

• Massachusetts Institute of Technology, Cambridge, MA. 1979 Undergraduate studies in Electrical Engineering and Computer Science.

• Dana Hall Schools, Wellesley, MA. High school diploma 1977. Among other honors, delivered Valedictorian speech.

ACADEMIC POSITIONS

• 2009 - today. Distinguished Career Professor of Computer Science, Technology and Policy, *Institute for Software Research,* School of Computer Science, Carnegie Mellon University.

• 2008 - today. Visiting Scholar, Computer Science, School of Engineering and Applied Science, Harvard University.

• 2008 - today. MLK Visiting Professor, Computer Science (CSAIL), Massachusetts Institute of Technology.

• 2005 - 2009. Associate Professor of Computer Science, Technology and Policy, *Institute for Software Research,* School of Computer Science, Carnegie Mellon University.

• 2002 - 2004. Assistant Professor of Computer Science, Technology and Policy, *Institute for Software Research,* School of Computer Science, Carnegie Mellon University.

• 2001 - today. Director and Founder, *Laboratory for International Data Privacy,* also known as the "Data Privacy Lab", Institute for Software Research, School of Computer Science, Carnegie Mellon University.

• 2003 - today. Co-Director and co-Founder, *PhD Program in Computation, Organizations and Society,* Institute for Software Research, School of Computer Science, Carnegie Mellon University.

• 2004, 2006 - today. Founder in 2004, Editor-in-Chief from 2006. *Journal of Privacy Technology,* Institute for Software Research, School of Computer Science, Carnegie Mellon University.

• 2003 - today. Director, *Privacy Technology Center,* Institute for Software Research, School of Computer Science, Carnegie Mellon University.

• 1998 - 2002. Assistant Professor of Computer Science and of Public Policy. Dual appointments: H. John Heinz III School of Public Policy and Management and the Center for Automated Learning and Discovery in the School of Computer Science, Carnegie Mellon University.

• 1991-1996, 1999-2000. Teaching Fellow and Instructor, Computer Science Courses, Faculty of Arts and Science, Harvard University. Year 1999-2000 taught distance courses only, using my ActiveTutors and VirtualLectures technologies.

• 1980-1981. Research Assistant, Artificial Intelligence Lab, Massachusetts Institute of Technology.

OTHER ACADEMIC AFFILIATIONS

• 1998 - today. Faculty Advisory Board, *Machine Learning Department,* School of Computer Science, Carnegie Mellon University.

• 2005 - today. Faculty, *Quality of Life Technology Center,* Carnegie Mellon University and the University of Pittsburgh. Co-Leader of the Person and Society Research Thrust, 2005-2006.

• 2002 - today. Faculty, *Aladdin Center,* School of Computer Science, Carnegie Mellon.

• 2002 - today. Faculty, *Information Networking Institute,* Carnegie Mellon University.

• 2004 - today. Faculty, *Cylab,* Carnegie Mellon University

• 2002 - today. Faculty, *H. John Heinz III School of Public Policy and Management.*

CORPORATE AFFILIATIONS

No issued licenses limit my access to these technologies for academic purposes.

• 2003 - today. *Privacert, Inc.* has licenses to some of my technologies. I maintain no ownership interest or hold any official position. I do maintain sufficient involvement to assure my technology is being properly deployed.

• 2002-2003. *Datanon, LLC* had licenses to some of my technologies, which are now available through Privacert, Inc.

• 1998 - today. *CITC* has licenses to some of my technologies. I maintain no ownership interest or hold any official position. I do maintain sufficient involvement to assure my technology is being properly deployed.

• 1981-1991. CESS, Inc., Owner and President. Founded and managed a medium-sized computer company that developed customized AI software and provided supporting retail and repair services. Supervised staffs of programmers, repair technicians, and sales representatives.

NATIONAL AND INTERNATIONAL COMMITTEE APPOINTMENTS (SELECTED)

• Appointed, Federal HIT Policy Committee, Privacy and Security Seat, Obama Administration, 2009-today.

• Appointed, EPIC Advisory Board, 2007-today.

• Appointed, IEEE-USA Medical Technology Policy Committee, 2003-today.

• Appointed member, Biomedical Library Research Committee, National Library of Medicine, National Institutes of Health, 1999-2003.

• Special appointment to the Board of Directors as “Privacy Consultant”, American Psychiatric Association, 2000.

• Health Privacy Project, Working Group, at Georgetown University, 1998.

• Massachusetts State Committee on Healthcare, Medical Record Confidentiality Working Group, 1998.

• National Committee for Quality Assurance, Joint Sessions on Security and Confidentiality of Medical Information, 1998.

• World Wide Web Consortium, Personal Privacy Preferences Working Group, 1996-1997.

GOVERNMENT AND POLICY REFERENCES (SELECTED)

• Federal Register, August 2009, Health Data Breach Regulation. My work and that of my former student is cited explicitly in supporting commentary. See page 42968, Federal Register. Vol. 175, No., Aug 25, 2009. http:// www.ftc.gov/os/2009/08/R911002hbn.pdf

•Privacy Commissioner of Canada, Annual Report, 2000. My work is cited explicitly.

•Federal Register, March and December 2000, Health Insurance Privacy and Portability Act (HIPPA). My work is cited explicitly in supporting commentary on de-identification provisions of the Privacy Rule of **HIPAA**.

PRIMARY RESEARCH

My overall research goal is to create systems that automatically learn strategic and sensitive information from data, and its converse, to create systems that control what can be learned. Most often, the sensitive information I want systems to learn are ways to relate personal identity to seemingly innocents facts (“re-identification” and “identifiability”). And most often

my pursuits in the opposite direction involve creating systems with guarantees that identity cannot be learned ("**anonymity**") while still making sure the **anonymized** results remain useful.

My overall research agenda involves leveraging my work on identifiability and **anonymity** (1) to change the way a computer scientist or engineer thinks about privacy in the technology he creates; (2) to help establish the scientific study of privacy; and, (3) to help decision-makers weave technology and policy together to address today's privacy-technology conflicts.

PUBLICATIONS OF LATANYA SWEENEY, PHD

BOOKS (INCLUDING FORTHCOMING)

1. *Connecting Your Dots: What they know from what you leave behind. Sweeney, L.* (This is not an edited volume.) Forthcoming book publication. 140 pages.

2. *Technology Dialectics: How to construct provably appropriate technology. Sweeney, L.* (This is not an edited volume.) Forthcoming book publication. 257 pages. Readers welcomed. *dataprivacylab.org/dataprivacy/projects/dialectics/index.html*

3. *Computer Guide. Sweeney, L.* and Barrett, S. Cambridge: CESS Publishing. 1983.

CHAPTERS IN BOOKS

4. Face De-identification. Gross, R., *Sweeney, L.,* Cohn, J., de la Torre, F. and Baker, S. *Protecting Privacy in Video Surveillance,* A. Senior, editor. Springer, 2009

5. *Information Explosion. Sweeney, L.*

Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies, L. Zayatz, P. Doyle, J. Theeuwes and J. Lane (eds), Urban Institute, Washington, DC, 2001. *dataprivacylab.org/people/sweeney/explosion.html.*

6. *Datafly: a system for providing **anonymity** in medical data. Sweeney, L.*

Database Security, XI: Status and Prospects, T. Lin and S. Qian (eds), Elsevier Science, Amsterdam, 1998. *dataprivacylab.org/datafly/index.html*

REFEREED JOURNAL PAPERS (INCLUDING TECHNICAL MAGAZINES) - PUBLISHED

7. *Managing End-of-Life Care in Complex Patients Can Reduce Costs Without Shortening Life. Latanya Sweeney, PhD,* Andrew Halpert MD, Joan Waranoff MS MBA American Journal of Managed Care 13 February 2007, pp. 84-92. (impact factor 2.28) (*PDF, HTML*)

Earlier version available as Carnegie Mellon University, Technical Report CMU-ISRI-05-124. *dataprivacylab.org/dataprivacy/projects/complexcare/index.html.*

8. *Protecting Job Seekers from Identity Theft. Sweeney, L.*

IEEE Internet Computing 10 (2) March 2006. (impact factor 1.55) Earlier version available as *A Guardian Angel Protects You From Identity Theft.* Carnegie Mellon University, LIDAP Working Paper 16. Pittsburgh: January 2006. *dataprivacylab.org/dataprivacy/projects/idangel/paper3.html.*

9. *Privacy-Enhanced Linking. Sweeney, L.*

ACM SIGKDD Explorations 7(2) December 2005. Earlier version available as Carnegie Mellon University, School of Computer Science Technical Report CMU-ISRI-05-136. Pittsburgh: November 2005. *(PDF). dataprivacylab.org/dataprivacy/projects/pel/index.html.*

10. *Privacy-Preserving Surveillance using Databases from Daily Life. Sweeney, L.*

IEEE Intelligent Systems, 20 (5), September-October 2005. (impact factor 1.44) Also known as Homeland Security by John Yen , Robert Popp , George Cybenko , K.A. Taipale , Latanya Sweeney, Paul Rosenzweig in same issue. Earlier version available as *Privacy-Preserving Surveillance using Selective Revelation,* Carnegie Mellon University, LIDAP Working Paper 15, February 2005. (*PDF*). *dataprivacylab.org/dataprivacy/projects /selectiverevelation/index.html.*

11. *Preserving Privacy by De-identifying Facial Images. Newton, E., Sweeney, L.,* and *Malin, B.*

IEEE Transactions on Knowledge and Data Engineering, 17 (2) February 2005, pp. 232-243. (impact factor 1.90)

Originally available as Carnegie Mellon University, School of Computer Science, Technical Report, CMU-CS-03-119. Pittsburgh: March 2003. *Abstract,* Paper: 26 pages in *PS,* PDF (*600dpi,* or *300dpi*). *dataprivacylab.org/dataprivacy/projects/video/index.html.*

12. *Navigating Computer Science Research Through Waves of Privacy Concerns: Discussions among Computer Scientists at Carnegie Mellon University. Sweeney, L.*

ACM Computers and Society, 34 (1) April 2004.

Originally available as Carnegie Mellon University, School of Computer Science, Technical Report CMU-CS-03-165, CMU-ISRI-03-102. Pittsburgh: July 2003. *Abstract,* Paper: 18 pages in *PS, PDF. dataprivacylab.org/dataprivacy/projects/csresearch.html*

13. *Finding Lists of People on the Web.*

Sweeney, L.

ACM Computers and Society, 34 (1) April 2004.

Originally available as Carnegie Mellon University, School of Computer Science, Technical Report CMU-CS-03-168, CMU-ISRI-03-104. Pittsburgh: June 2003. *Abstract,* Paper: 12 pages in *PDF. dataprivacy/lab.org/dataprivacy/projects/rosterfinder/index.html*

14. *How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design* ***anonymity*** *protection systems. Malin, B.* and *Sweeney. L.*

Journal of Biomedical Informatics. 2004; 37(3): 179-192. (impact factor 2.0) Also available on *MEDLINE.* **Best of the Year Award** by the International Medical Informatics Association, and as such appears in *The 2005 Year Book of Medical Informatics (more). dataprivacylab.org/dataprivacy/projects/trails/dnaTrails.html*

15. *k-****anonymity****: a model for protecting privacy. Sweeney, L.*

International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570. (impact factor 0.38) Since publication, received **recognition award** as a noteworthy work at the 2004 Workshop on Privacy Enhancing Technologies. The concept "k-**anonymity**," which this paper introduces has been **heavily cited** in other academic publications. *dataprivacylab.org/people/sweeney/kanonymity.html.*

16. *Achieving k-****anonymity*** *privacy protection using generalization and suppression. Sweeney, L.*

International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 571-588. (impact factor 0.38) *dataprivacylab.org/people/sweeney/kanonymity2.html.*

17. *Inferring Genotype from Clinical Phenotype through a Knowledge Based Algorithm. Malin, B.* and *Sweeney, L.*

Pacific Symposium on Biocomputing 2002, R.B. Altman et al. (Eds.) (World Scientific, Singapore, 2002). Also available on *MEDLINE. dataprivacylab.org/dataprivacy/projects /genetic/dna3.html*

18. Commentary: Researchers need not rely on consent or not. *Sweeney, L.*

New England Journal of Medicine, 1998. (impact factor 52.59)

19. *Weaving technology and policy together to maintain confidentiality. Sweeney, L.*

Journal of Law, Medicine and Ethics. 1997, 25:98-110. (impact factor 1.04) Cited and discussed in the commentary of the **HIPAA Privacy Rule.** *dataprivacylab.org/dataprivacy/projects/law/law1.html*

REFEREED JOURNAL PAPERS - SUBMITTED

20. A Multiparty Computation for Learning where People have been Without Knowing Who they Are. *Sweeney, L.*

21. Formal Protection Models: how to prove data are **anonymous**. *Sweeney, L.*

REFEREED CONFERENCE/WORKSHOP PAPER - PUBLISHED

22. Multi-Factor De-Identification of Facial Images. Gross, R., Cohn, F., de la Torre, F., Baker, S., and *Sweeney, L.*

Proceedings of the 2008 American Medical Informatics Association Annual Symposium, 2008. (AMIA acceptance rate 29%)

23. Semi-Supervised Learning of Multi-Factor Models for Face De-Identification. Gross, R. and *Sweeney, L.*

IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, June 2008. (CVPR acceptance rate 28%)

24. Towards Real-World Face De-Identification. Gross, R. and *Sweeney, L.*

IEEE Conference on Biometrics, Washington, DC, September 2007. (BTAS acceptance rate 11%)

25. Robust Hand Geometry Measurements for Person Identification using Active Appearance Models.

Gross, R., Li, Y., and *Sweeney, L.,* Jiang, X., Xu, W., and Yurovsky, D. *IEEE Conference on Biometrics,* Washington, DC, September 2007. (BTAS acceptance rate 11%)

26. New Direction on Contact-free Hand Recognition.

Jiang, X., Xu, W., Li, Y., *Sweeney, L.,* Gross, R., and Yurovsky, D. *IEEE International Conference On Image Processing,* San Antonio, Texas, September 2007. (ICIP acceptance rate 46%)

27. 2D image database indexing: a coefficient-based approach.

Jiang, X., Xu, W., Li, Y., *Sweeney, L.,* Gross, R., and Yurovsky, D. *IEEE International Conference on Multimedia and Expo,* Beijing, China, July 2007. (ICME acceptance rate 22%)

28. Composition and disclosure of unlinkable distributed databases. *Malin, B.* and *Sweeney, L.*

22nd IEEE International Conference on Data Engineering, Atlanta, GA, April 2006. (ICDE acceptance rate 19%)

29. Model-based face de-identification.

R. Gross, *Sweeney, L.,* de la Torre, F., and Baker, S. *IEEE Workshop on Privacy Research in Vision,* 2006.

30. A secure protocol to distribute unlinkable health data. (PDF) *Malin, B.* and *Sweeney, L.*

Proceedings, Journal of the American Medical Informatics Association. Washington, DC. Oct 2005: 485-489. (AMIA acceptance rate 29%)

31. *Email Alias Detection Using Network Analysis.* Holzer, R. Malin, B., and *Sweeney, L.*

Proceedings of the *ACM SIGKDD Workshop on Link Discovery:* Issues, Approaches, and Applications. Chicago, IL. August 2005. (LinkKDD acceptance rate 19%) *dataprivacylab.org/dataprivacy/projects/emailalias/index.html.*

32. *AI Technologies to Defeat Identity Theft Vulnerabilities. Sweeney, L.*

AAAI Spring Symposium, AI Technologies for Homeland Security, 2005. *dataprivacylab.org/dataprivacy/projects/idangel/index.html.*

33. *Mining Images in Publicly-Available Cameras for Homeland Security. Sweeney, L.* and Gross, R.

AAAI Spring Symposium, AI Technologies for Homeland Security, 2005. *dataprivacylab.org/dataprivacy/projects/videocount/index.html.*

34. *Technologies to Defeat Fraudulent Schemes Related to Email Requests.* Airoldi, E. and Malin, B., and *Sweeney, L.*

AAAI Spring Symposium, AI Technologies for Homeland Security, 2005. *dataprivacylab.org/dataprivacy/projects/scamspam/index.html.*

35. *Privacy-Preserving Bio-terrorism Surveillance. Sweeney, L.*

AAAI Spring Symposium, AI Technologies for Homeland Security, 2005. *dataprivacylab.org/dataprivacy/projects/bioterror/index.html.*

36. *Towards a Privacy-Preserving Watchlist Solution. Sweeney, L.*

AAAI Spring Symposium, AI Technologies for Homeland Security, 2005. *dataprivacylab.org/dataprivacy/projects/watchlist/index.html.*

37. *Integrating Utility into Face De-Identification.* Gross, R. Airoldi, E., Malin, B., and *Sweeney, L.*

Workshop on Privacy-Enhanced Technologies, 2005. (PET acceptance rate 26%) *dataprivacylab.org/dataprivacy/projects/videoutility/index.html.*

38. *Guaranteeing **anonymity** when sharing medical data, the datafly system. Sweeney, L.*

Proceedings, Journal of the American Medical Informatics Association. Washington, DC: Hanley & Belfus, Inc., 1997. (AMIA acceptance rate 29%) **Recognition award.** *dataprivacylab.org/datafly/index.html*

39. *Replacing Personally-Identifying Information in Medical Records, the Scrub System. Sweeney, L.*

In: Cimino, JJ, ed. *Proceedings, Journal of the American Medical Informatics Association.* Washington, DC: Hanley & Belfus, Inc., 1996:333-337. (AMIA acceptance rate 29%) **Best paper award.** *dataprivacylab.org/people/sweeney/scrub.html*

40. *Re-Identification of DNA through an Automated Linkage Process.* Malin, B. and *Sweeney, L.*

Proceedings, Journal of the American Medical Informatics Association. Washington, DC: Hanley & Belfus, Inc. Nov 2001; 423-427. (AMIA acceptance rate 29%) *dataprivacylab.org/dataprivacy/projects/genetic/dna2.html*

41. *Electronic Disease Surveillance and Reporting: the eReport System.* Malloy, W. and *Sweeney, L.*

Proceedings, Journal of the American Medical Informatics Association. Washington, DC: Hanley & Belfus, Inc. Nov 2001; 964. (AMIA acceptance rate 29%)

42. *Determining the Identifiability of DNA Database Entries.* Malin, B. and *Sweeney, L.*

Proceedings, Journal of the American Medical Informatics Association. Washington, DC: Hanley & Belfus, Inc. Nov 2000; 537-541. (AMIA acceptance rate 29%) *dataprivacylab.org/dataprivacy/projects/genetic/dnal.html*

43. *Foundations of Privacy Protection from a Computer Science Perspective. Sweeney, L.*

Proceedings, Joint Statistical Meeting, AAAS, Indianapolis, IN. 2000.

44. *An intelligent tutor for teaching basic computer science and Java programming to medical informatics students. Sweeney, L.*

Proceedings, Journal of the American Medical Informatics Association. Washington, DC: Hanley & Belfus, Inc., 1999. (AMIA acceptance rate 29%)

45. *Y2K BEACON: an intelligent web-based crisis center for disseminating Year 2000 biomedical equipment compliance information. Sweeney, L.* and Barrett, S.

Proceedings, Journal of the American Medical Informatics Association. Washington, DC: Hanley & Belfus, Inc., 1999. (AMIA acceptance rate 29%)

46. *Protecting privacy when disclosing information: k-**anonymity** and its enforcement through generalization and suppression.* Samarati, P. and *Sweeney, L.*

Proceedings of the IEEE Symposium on Research in Security and Privacy, May 1998, Oakland, CA. (S&P acceptance rate 10%) Also appears as Protecting respondents identities in microdata release, *IEEE Transactions on Knowledge and Data Engineering,* 2001. (impact factor 1.90)

47. *Generalizing data to provide **anonymity** when disclosing information.* Samarati, P. and *Sweeney, L.*

ACM Principles of Database Systems. Seattle, WA, USA, 1998. (PODS acceptance rate 21%)

48. *Towards the optimal suppression of details when disclosing medical data, the use of sub-combination analysis. Sweeney, L.*

Proceedings, *MEDINFO 98.* International Medical Informatics Association. Seoul, Korea. North-Holland, 1998.

49. *Three computational systems for disclosing medical data in the year 1999. Sweeney, L.*

Proceedings, *MEDINFO 98.* International Medical Informatics Association. Seoul, Korea. North-Holland, 1998.

50. *Computational Disclosure Control for Medical Microdata. Sweeney, L.*

Record Linkage Workshop. Bureau of the Census. Washington, DC: 1997.

TECHNICAL REPORTS AND WHITE PAPERS (EXCLUDING ANY SUBSEQUENTLY PUBLISHED)

51. *The Medical Billing Framework as the Backbone of the National Health Information Infrastructure. Sweeney, L.*

AdvanceHIT Project, White Paper Serices, October 2009. *PDF*

52. *The HandShot Hand Tracking System.* Gross, R. and *Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Technical Report, 2009. In preparation.

53. *Photographing Friction Ridge Information: A Research Notebook. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-43. Pittsburgh: (2007, 2009). In preparation.

54. *Converting a Photograph of Friction Ridge Information to a Corresponding Ink-Like Print: A Research Notebook. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-39.

Pittsburgh: (July 20006, May 2007, 2009). In preparation.

55. *Identifiability of De-identified Clinical Trial Data. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report. Pittsburgh: 2009.

56. *Fine-Grained Segmentation of a Single Foreground Object: A Research Notebook. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-42.

Pittsburgh: 2007.

57. *Risk Assessments of Personal Identification Technologies for Domestic Violence Homeless Shelters. Sweeney, L.*

Pittsburgh: January 2006. Under review for publication.

58. *Inked and Scanned Database over a Common Population.* Weedn, V. and *Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-33.

Pittsburgh: 2006.

59. *Finger Shot Database over a Common Population. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-33.

Pittsburgh: 2006.

60. *Palm Capture Database: a Collection of Naturally Posed Palms. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-32.

Pittsburgh: 2006.

61. *Hand Shake: a Database of Naturally Positioned Hands. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-31.

Pittsburgh: 2006.

62. *Silhouette Database of Naturally Positioned Hands. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-30.

Pittsburgh: 2005. 6

63. *CameraWatch. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report. Pittsburgh: 2005. *dataprivacylab.org/dataprivacy/projects/camwatch/"*

64. *Adding Semantics and Rigor to Association Rule Learning: the GenTree Approach.* Li, Y. and *Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Technical Report, CMU-ISRI-05-101. Pittsburgh: January 2005. *Abstract,* Paper: 25 pages in *PDF. dataprivacylab.org/dataprivacy/projects/gentree/gentree2.html.*

65. *A Fast 3-D Imaging System for Capturing Fingerprints, Palm Prints and Hand Geometry: the HandShot ID System. Sweeney, L.,* Weedn, V., and Gross, R.

Carnegie Mellon University, School of Computer Science, Technical Report, CMU-ISRI-05-105. Pittsburgh: 2004.

66. *Social Security Number Watch: Matching a Person to a Social Security Number. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report. Pittsburgh: 2004. *dataprivacylab.org/dataprivacy/projects/ssnwatch/"*

67. *A Multiparty Computation for Randomly Ordering Players and Making Random Selections. Sweeney, L.* and Shamos, M.

Carnegie Mellon University, School of Computer Science, Technical Report, CMU-ISRI-04-126. Pittsburgh: July 2004. *Abstract,* Paper: 39 pages in *PDF. dataprivacylab.org/dataprivacy/projects/randomorder/index.html*

68. *Detecting Bio-Terrorist Attacks and Naturally Occurring Outbreaks Over a Distributed Network While Protecting Privacy and Confidentiality: the PrivaSum Protocol.* Edo-Eket, S. and *Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Technical Report CMU-ISRI-04-111. Pittsburgh: 2004.

69. *Privacy-Preserving Bio-terrorism Surveillance that is **HIPAA**-Compliant Too: the Privacert System. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Technical Report CMU-ISRI-04-112. Pittsburgh: 2004.

70. *Learning Semantically Robust Rules from Data.* Li, Y. and *Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Tech Report, CMU ISRI 04-107, CMU-CALD-04-100. Pittsburgh: February 2004. *Abstract,* Paper: 21 pages, *PDF. dataprivacylab.org/dataprivacy/projects/gentree/index.html*

71. *Trail Re-identification: Learning Who You are From Where You Have Been.* Malin, B., *Sweeney, L.,* and Newton, E.

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory Technical Report, LIDAP-WP12. Pittsburgh: February 2003. *Abstract,* Paper: 10 pages in *PS* or *PDF. dataprivacylab.org/people/sweeney/trailsl.html.*

72. *Optimal **anonymity** using k-similar, a new clustering algorithm. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report. Pittsburgh: 2003.

73. *Identifiability of De-identified Pharmacy Data. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report. Pittsburgh: 2003.

74. *That's AI?: a history and critique of the field. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Technical Report, CMU-CS-03-106. Pittsburgh: January 2003. *Abstract,* Paper: 28 pages in *PS* or *PDF. dataprivacylab.org/people/sweeney/thatsai.html.*

75. *Uniqueness of Simple Demographics in the U.S. Population. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-WP4. Pittsburgh: 2000.

76. *Re-identification of De-identified Survey Data. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report. Pittsburgh: 2000.

77. *Maintaining Patient Confidentiality When Sharing Medical Data Requires a Symbiotic Relationship Between Technology and Policy. Sweeney, L.*

Artificial Intelligence Laboratory, Massachusetts Institute of Technology, AIWP-WP344, May 1997. (*PDF*) *dataprivacylab.org/dataprivacy/projects/law/law1.html*

78. *Power Learning: using computers as teaching machines. Sweeney, L.*

Massachusetts Institute of Technology, Artificial Intelligence Laboratory: Working Paper. 1998.

79. *Speech perception using real-time phoneme detection, the BeBe System. Sweeney, L.* and Thompson, P.

Massachusetts Institute of Technology, Laboratory for Computer Science: Tech Report MIT-LCS-TR-736. 1998. *dataprivacylab.org/people/sweeney/bebe.html*

80. *Towards the collection of all the data on all the people. Sweeney, L.*

MIT Artificial Intelligence Working Paper, 1998.

81. *Maintaining **anonymity** when sharing medical data, the datafly system. Sweeney, L.*

MIT Artificial Intelligence Laboratory Working Paper. Cambridge: AIWP-WP344 (1997).

82. *Iterative Profiler. Sweeney, L.*

Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report. Pittsburgh: 1997.

83. *Multiple hashed binary storage of words -- tiny, fast and almost perfect. Sweeney, L.*

Massachusetts Institute of Technology, AI Laboratory: Working paper. 1996.

84. *Computational Disclosure Control: Theory and Practice. Sweeney, L.*

Massachusetts Institute of Technology, Ph.D. Thesis in Computer Science, 2001.

85. *Sprees, a Finite-State Orthographic Learning System that Recognizes and Generates Phonologically Similar Spellings. Sweeney, L.*

Massachusetts Institute of Technology, Masters Thesis in Computer Science. Finalist in MasterWorks. 1997.

86. *A Coin Toss: the Dialectical Odds aren't always 50/50.* Sweeney, L.

harvard University, Undergraduate research thesis, 1995.

PATENTS AND INVENTION DISCLOSURES

87. *Systems and Methods for Deidentifying Entries In a Data Source. Sweeney, L.*

Patent 7,269,578, Issued on 9/11/2007.

88. *Method and System for Capturing Fingerprints, Palm Prints and Hand Geometry. Sweeney. L.,* Weedn, V., and Gross, R.

Patent 7660442. Issued on 2/9/2010.

89. Frontal Hand Capture of Fingerprints, Palm Prints amd Hand Geometry Using Contactless Photography. Sweeney, L., Gross, R., and Jiang, X.

Application filed July 2008.

90. *Method for Knowing Where People Have Been Without Knowing Who They Are Sweeney, L.* Application filed May 2009.

91. *Apparatus and Method for Instructing Students.* Intelligent tutoring system. *Sweeney. L.*

92. *Privacert Stream Editing. Sweeney. L.*

SOFTWARE ARTIFACTS

CORE SEMANTIC LEARNING TECHNOLOGIES INVENTED

• *Sprees:* a finite state orthographic learning system that recognizes and generates phonologically similar spellings. This was my Masters thesis. [1995]

• *Scout:* an algorithm and software program that profiles a dataset to learn what fields are present and the kind of information contained within, thereby providing a semantic description of an unknown dataset. [1995]

• Database Profiling Server: an algorithm and software program that relates fields across datasets and tables, thereby identifying fields containing duplicate information across tables and fields in one table that can be reliably linked to those in another table -- learning is based on the actual values that appear in the tables. [1996]

• Identifiability Server: a process and related models for determining how identifiable individuals and other entities within data may be by utilizing summary and aggregate data, e.g., Census data. A finding now cited in hundreds of articles is that "87% of the US population is uniquely identified by |*date of birth, gender, 5-digit ZIPŒ".* This work is the source of that finding when used with Census data. [1997]

• Risk Assessment Server: a system for determining how identifiable individuals within data may be by utilizing an inference engine, a taxonomy and ontology of fields, specific domain knowledge, knowledge about available datasets, and population specifics. This server is licensed commercially to perform **HIPAA** Certifications (see *www.privacert.com*); and, can be used in privacy-preserving bio-terrorism surveillance (see my AAAS Presentation and TAPAC testimony). It has also been the basis, in part, of my numerous expert witness consultations and testimonies. [1998]

• Donor Profiling and Solicitation (DPS): an algorithm and software program that learns a psychological profile of a person based on giving history, and then uses compiled information to determine an optimal personalized solicitation strategy for a

given solicitation attempt. Previously licensed to CESS, Inc. and Share Systems, Inc., who used it with more than 30 leading non-profit organizations including the Democratic National Party, the National Organization for Women, the Harlem Boy's Choir, Bishop Tutu's South African Freedom Campaign, and Greenpeace. [1987]

• DPS Extractor: an algorithm and software program that learns titles, given names, surnames, suffixes, street names, and other information about a household by extracting and interpreting information from mailing labels. This was licensed to CESS, Inc. and to Share Systems, Inc. [1987]

• Biblio: an algorithm and software program that extracts publication references from raw text bibliographies, identifying constituent parts (author names, title, publication, etc.), and then uses learned information to remove duplicates and to construct a searchable database. This was licensed to CESS, Inc. and Chitin, Inc. [1988]

• Scrub Extractor: an algorithm and software program that automatically extracts names, addresses, and other identifying information from letters, notes, articles, and other free text documents. (Basis, in part, for Scrub System, which is a privacy de-identification tool described below.) [1996]

• Iterative Profiler: an algorithm and process by which increasing amounts of information is related to individuals using inferential linkages of data fragments across various kinds of data sources. This has been the basis, in part, of some of my expert witness experiments and testimonies. [1997]

• *RosterFinder:* an algorithm and program that uses the Google API to allow searches for web pages containing rosters (lists of names) of people. A sample use was to locate web pages containing rosters of undergraduate students in computer science; about 18,725 students were found in 39 schools. Results are available in a searchable database *on-line* (thanks in part to Marshall Warfield). [2003]

• *CameraWatch:* a suite of algorithms and programs for locating IP addresses of live webcams and webpages showing images from live webcams. A sample of the results is available in an on-line searchable database (thanks in part to Kishore Madhava). The work has been highlighted on *ABC News, CNN, USA Today,* and *SlashDot,* and has had more than a million hits to the website. [2003]

• *SSNwatch:* a method and process by which public information about Social Security number (SSN) allocations is used to learn past residential information and current age inferences about the person to whom the SSN was assigned. This provides a means to match a person presenting an SSN to the demographics learned about the SSN, which is useful in combating *identity theft.* [2004]

DATA PRIVACY TECHNOLOGIES INVENTED

• *k-**Anonymity**:* any algorithm or process that **anonymizes** data by insuring each entity in the data is indistinguishable from at least a specific number of other such entities in the data. Patent issued. Received a recognition award from the 2004 Workshop on Privacy Enhancing Technologies. Has been cited, discussed and extended in lots of academic work by others across numerous communities and applied to all kinds of data and has inspired other kinds of uses (e.g., k-**anonymous** messaging). [1997]

• Datafly: an algorithm and general-purpose software program that **anonymizes** field-structured data so that the released data adheres to a k-**anonymity** requirement. Specifically, there are k records that are indistinct over the fields sensitive to re-identification. Received a recognition award from the *American Medical Informatics Association.* A license was provided to Datanon, LLC and now subsequently to Privacert, Inc. [1997]

• k-Similar: a general-purpose clustering algorithm that groups the closest information together with the guarantee that there are at least k members to each cluster. The number of clusters is not fixed, making the algorithm the converse of the very well-known k-means cluster algorithm, which guarantees there are at least k clusters having any number of members. Like Datafly, k-Similar can be used to **anonymize** field-structured data so that the released data adheres to a k-**anonymity** requirement.

But unlike Datafly, results from k-Similar maintain the maximum detail possible. A license was provided to Datanon, LLC and now subsequently to Privacert, Inc. [1998]

• *k-Same* (with Elaine Newton and Bradley Malin): an algorithm for de-identifying faces in video surveillance data such that no face recognition software (no matter how good the software may get) can reliably recognize the resulting images even though most facial details are preserved. This is done by averaging image components, which may be the original image pixels (k-Same-Pixel) or eigenvectors (k-Same-Eigen) so that k-**anonymity** is assured. [2003]

• *Scrub:* an algorithm and software program that **anonymizes** unrestricted text such that the identities of individuals and other entities in the data cannot be re-identified. (Converse of Scrub Extractor mentioned earlier.) Based on its use with clinical notes and letters, it received a recognition award from the American Medical Informatics Association in 1996. [1996]

• Policy Explorer: an algorithm and process that characterizes and quantifies data sharing practices by estimating how much information is made available about each entity that is the subject of the data and reporting related measures of risks. Provides a tool for comparing competing policies and performing "what if" analyses. [1997]

• Privacert: a rule-based system with related language ("PrivaCert Editing Language") for expressing and enforcing **anonymity** requirements to render a specific dataset sufficiently de-identified. Results satisfy privacy standards established by the Risk Assessment Server (mentioned above) and not necessarily k-**anonymity**. This technology is licensed to Privacert, Inc. for use in rendering health data sufficently de-identified in accordance to **HIPAA** (U.S. medical privacy regulation). [1997]

• PrivaSum (with Samuel Edo-Eket): a real-world protocol that allows parties to jointly compute an aggregate statistic over a network such that the result is known by all but the contribution by each party remains confidential. Performance is improved over the traditional secret sharing approach, which is deterministic, by providing probabilistic assurances even in the face of widespread collusion by some malicious parties. [2004]

OTHER SEMANTIC LEARNING TECHNOLOGIES INVENTED

• *Collaboration Wheel:* an on-line tool for supporting group joint work on a common document in a networked environment. Java prototype built by Charles Shelton in 1997. [1997]

• *Active Tutor (aka Power Learning):* an on-line teaching-learning environment, where the computer plays the role of an expert teacher. Instructions and practice are seamlessly integrated and adapted to the personal needs of each student. *On-line demonstration available.* [1983-2002]

• Virtual lectures: By taking the expected path of an "average student" through an Active Tutor (described above), the teaching-learning material can be condensed into carefully crafted on-line lectures. Teaching-learning materials for 3 full semesters of the Java Programming language are available on-line *(Javal, Java2, Java3)* and have been used to teach hundreds of students in courses at Carnegie Mellon (e.g., *15-100*) and Harvard University, officially, and many other schools unofficially. Numerous non-traditional students have reported using the on-line materials for self-directed learning. [1998-2002]

• *GridCity:* a visual programming environment in which students write Java programs to control vehicles in a GridWorld. This is an adaptation of *Karel the Robot* by Richard Pattis, but this adaptation exploits the fundamental programming constructs made possible by Java that were not available to Karel in Pascal. On-line *demonstrations* and *software download* available. Used by hundreds of students at Carnegie Mellon and Harvard University. [1999-2002]

• *Bebe:* an algorithm and software program for learning the basics sounds of a human language based on automated detection of phonemes in the analog waveform. No high-level knowledge of the language is used, so performs as well learning the Korean sound system as the American English sound system. Early implementations done jointly with Patrick Thompson. [1997]

• Iris Expert System shell: a method and software program for developing on-line expert systems for diagnostic tasks. A sampler was built for interviewing people about personal behaviors that may place them at risk to AIDS; the program received regional print and radio news coverage in New England. [1989]

• CompuFix: a software program that steps a person through the repair of a personal computer, assuming no prior knowledge of computer repairs. Uses the Iris Expert System shell (mentioned above). Was licensed to American Information Technologies, which sold copies commercially. [1990]

THESES

• *Computational Disclosure Control: Theory and Practice.* . Massachusetts Institute of Technology, Laboratory for Computer Science, Tech Report, **PhD Thesis.** 2001. Committee: Hal Abelson (Chair), Peter Szolovits, and Daniel Weitzner.

• *Sprees, a Finite-State Orthographic Learning System that Recognizes and Generates Phonologically Similar Spellings.* Massachusetts Institute of Technology, Artificial Intelligence Laboratory: Working paper, **Masters Thesis.** 1997. Finalist in MasterWorks. Thesis advisor: Peter Szolovits.

• *A Coin Toss: the Dialectical Odds aren't always 50/50.* Harvard University, Faculty of Arts and Sciences: Working paper, **ALB Thesis.** 1995. Advisors: Henry Leitner (Harvard) and Jon Doyle (MIT).

AWARDS AND HONORS RECEIVED (EXCLUDING KEYNOTE TALKS AND BEST PAPER AWARDS) :

• Elected Fellow, American College of Medical Informatics, 2006. (*Highly distinguished*)

• 2004 Workshop on Privacy Enhancing Technologies, Recognition Award for my work on k-**anonymity**, Toronto, Canada 2004.

• Blue Cross Blue Shield Association of Michigan, Privacy Leadership Award, 2002.

• American Psychiatric Association, Privacy Advocate Award, 2001. *(Highly distinguished)*

• American Medical Informatics Association, Recognition Award, 1997.

• National Library of Medicine, Research Recognition, 1997.

• American Medical Informatics Association, First Place Award, 1996.

• Massachusetts Institute of Technology, Masterworks Finalist, 1996.

• National Library of Medicine, Training Fellow, 1995.

• Harvard University, Academic Achievement Award, 1995.

PUBLIC SERVICE (SELECTED)

EDITORIAL BOARD APPOINTMENTS

o *Journal of Privacy Technology,* Editor-in-Chief, 2006 - today

o Journal of Law and Policy for the Information Society

o Ad hoc reviewer: IEEE Transactions on Knowledge and Data Engineering (*IEEE TKDE*), Journal of the American Medical Association (*JAMIA*), IEEE Security and Privacy Magazine, Science

o Program committee: Knowledge and Discovery in Data (KDD) 2003, Modeling Decisions for Artificial Intelligence (MDAI) *2005,* Workshop on Privacy Enhancing Technologies (PET) *2005.*

AT CARNEGIE MELLON (SELECTED)

o Co-founded a new PhD program in the School of Computer Science at Carnegie Mellon, named the *PhD program is Computation. Organizations and Society.* This was work done with Kathleen Carley, Norman Sadeh, and many in the SCS faculty. August 2003.

o Hosted two brainstorming sessions for faculty at the School of Computer Science at CMU. Topic was computer science research and privacy concerns in emerging technologies. Resulted in a white paper and then to an *ACM publication.* It has also led to the formation of the *Privacy Technology Center* and a related funding proposal. The initial workshops were held in July 2003.

o Helped organize a 2-day workshop for computer scientists on data privacy called *Privacy in D.A.T.A.* Held March 27-28, 2003 at CMU. Sponsored by the NSF Aladdin Center and the Data Privacy Lab at CMU. This was work done with Guy Blelloch, Lenore Blum, and Manuel Blum.

INVITED TALKS OF LATANYA SWEENEY, PH.D.

This list excludes conference and workshop paper presentations.

RECENT TALKS (INVITED, EXCLUDING PAPER PRESENTATIONS)

o "*The Science of Privacy Technology.*"

American Association for the Advancement of Science (AAAS), Topical Lecture at Annual Meeting, St. Louis, MO. February 17, 2006.

o *"De-Identifying Health Data."*

Health Canada (Federal Department of Health), Ottawa, Ontario Canada. December 1, 2005.

o *"Strategies for De-Identifying Patient Data for Research."*

Electronic Health Information and Privacy Conference, Ottawa, Ontario Canada. November 30, 2005.

o *"Risk Assessments of PIN Technologies [identity management] for Domestic Violence Shelters (Updated)."*

Housing and Urban Development. Washington, DC November 18, 2005.

o *"Identity Management: solutions of privacy not security",*

Guest Lecture in Course 17-606, Software Systems Security Engineering, Carnegie Mellon University. Pittsburgh, PA. November 16, 2005.

o *"Identity Management: Dealing with Disclosure."*

6th CACR, Toronto, Canada, November 3, 2005.

o "*Privacy Principles for Ubiquitous Technologies*",

Testimony before the Eurpoean Commission, Brussels Belgium October 25 2005.

o "*Aging with Dignity: Privacy need not be traded for technical assistance.*"

Technology for Life and Living Conference, University of Pittsburgh Medical Center, Pittsburgh, PA, October 21, 2005.

o "*Recommendations to Identify and Combat Privacy Problems in the Commonwealth*",

Testimony before the Pennsylvania House Select Committee on Information Security (HR351), Pittsburgh, PA, October 5, 2005. (*Testimony and Appendices*)

o "*Privacy Technologies for Large Research Databases*"

Spectrum Health and Michigan State University, Grand Rapids, MI, September 23, 2005. (*Slides and Abstract*)

o "*Biometrics Alone Won't Do: Developing Holistic Identity Management Solutions*"

Biometrics Symposium 2005, Arlington, VA, September 19, 2005. (*Slides and Abstract*)

o "*Risk Assessments of PIN Technologies for Domestic Violence Shelters,*"

National HMIS Conference, St. Louis, Missouri, September 13, 2005. (*Slides and Abstract*)

o "*Privacy Technologies for Homeland Security,*"

Testimony before the Privacy and Integrity Advisory Committee of the Department of Homeland Security ("DHS"), Boston, MA, June 15, 2005. (*Testimony and Appendices*)

o "***HIPAA*** *Strategies for De-Identifying Patient Data for Research,*"

American Association of Medical Colleges (AAMC), National Conference, Group on Information Resources, Philadelphia, PA. April 12, 2005. (*Slides and Abstract*)

o "*Privacy Technology in the Face of Information Warfare,*"

Guest Lecture in Course 19-601, Information Warfare, Carnegie Mellon University. Pittsburgh, PA. March 29, 2005. (*Slides and Abstract*)

o "*Privacy Technology: Artificial Intelligence to Save the World,*"

AAAI Spring Symposium (Invited talk). Stanford. Palo Alto, CA. March 23, 2005. (*Slides, References and Abstract*).

o *"Beyond Ickiness is Risk: The Exasperation of Data Privacy Problems by Implanted RFIDs,"*

The Concealed I Conference, University of Ottawa, Ontario Canada. March 4, 2005. (*Slides, References and Abstract*).

o *"Privacy Technology: Computer Scientists Help Save the World,"*

Intel Privacy Forum, Intel Corporation, Hillsboro, Oregon. March 2, 2005. (*Slides, Abstract, References*)

o *"Privacy for Those Who Can Afford It,"*

Guest Lecture in Graduate Course 17-899, Technology and Development for 4 Billion, School of Computer Science, Carnegie Mellon University. Pittsburgh, PA. November 17, 2004. (*Slides and Abstract*)

o "*Tackling Proposals,*"

PhD Program in Computation, Organizations and Society, Graduate Seminar. School of Computer Science, Carnegie Mellon University. Pittsburgh, PA. November 15, 2004. (*Slides and Abstract*)

o *Privacy Technology for the NCPI,*

Special presentation at a joint meeting of Vanderbilt University and the Center for Automated Learning and Discovery, at the School of Computer Science, Carnegie Mellon University. November 4, 2004. (*Slides and References*)

o "*Privacy Technology: The Frontier,*"

13th CACR Information Security Workshop & 5th Annual Privacy and Security Workshop. University of Toronto, Canada. October 29, 2004. (*Slides and References*)

o "*Tracking Terrorism Using Privacy-Preserving Surveillance,*"

Tracking Terrorism in the 21st Century: A national symposium on the role of Science and Law in Detecting, Investigating, and Adjudicating Political Violence. Dusquene University School of Law, The Cyril H. Wecht Institute of Forensic Science and Law. Pittsburgh, PA. October 2004. (*Slides and References*)

o "*Privacy Technology,*"

Access and Privacy Workshop: Connecting citizens and government calls for openness, transparency and strategy.

Office of the Privacy Commissioner. Toronto, Ontario, Canada. October 2004. (*Slides and References*)

o "*Privacy-Preserving Surveillance,*"

United States Department of Homeland Security: Directions for the Data Sciences Research Agenda. Alexandria, VA. September 2004. (*Slides and References*)

o "*Privacy Technology: New Directions for Computer Science,*"

CERIAS Seminar Series, Purdue University. West Lafayette, IN. September 2004. *(Slides and References)*

EMINENT TALKS AND TESTIMONIES (EXCLUDING EXPERT WITNESS TESTIMONIES IN COURTS)

o *"Risk Assessments of PIN Technologies [identity management] for Domestic Violence Shelters (Updated)."* Housing and Urban Development. Washington, DC November 18, 2005.

o *"Privacy Principles for Ubiquitous Technologies",* Testimony before the Eurpoean Commission, Brussels, Belgium, October 25, 2005.

o *"Privacy Technologies for Homeland Security",* Testimony before the Privacy and Integrity Advisory Committee of the Department of Homeland Security ("DHS"), Boston, MA, June 15, 2005. *(Testimony and Appendices)*

o *"Privacy Technology: The Frontier",* 13th CACR Information Security Workshop & 5th Annual Privacy and Security Workshop. Privacy & Security: Seeking the Middle Path. The University of Toronto, Canada. October 29, 2004. *(Slides and References)*

o *"Outsourcing and Off shoring Privacy Risks in light of* ***HIPAA****",* HIMSS Advocacy Day on the Hill. Washington, DC, 2004.

o *"Privacy Technology for Bio-terrorism Surveillance",* Testimony before the Technology and Privacy Advisory Committee (TAPAC), a Federal Advisory Committee. My testimony was on the subject of privacy research and the Total Information Awareness Project at DARPA, Washington, DC 2003.

o *"Privacy-preserving Bio-terrorism Surveillance",* Annual Meeting of the American Association for the Advancement of Science, Seattle, WA 2004.

o *"Vulnerabilities of Genomic Databases: the risks are real",* Annual Meeting of the American Association for the Advancement of Science, Seattle, WA 2004.

o *"The De-identification Provision of* ***HIPAA****",* American Health Lawyers. Atlanta, GA, 2002.

o "*Genetic Privacy*", U.S. Senate Committee Congressional Briefing, Washington, DC, 2000. Hosted by Senator Frist (R) and Senator Rockefeller (D).

o "*Virtual Fingerprints in Medical Data" and Technology for Taming Data,* National Library of Medicine, Board of Regents, Washington, DC, 1999.

o "*Re-Identification of De-Identified Medical Data*", Testimony before the National Center for Vital Health Statistics, on the subject of medical data and privacy (**HIPAA**), Baltimore, Maryland, 1998.

o "***Anonymity****, Technology and Policy*", Massachusetts Institute of Technology, Congressional Staffer Retreat, Cambridge, MA, 1999.

o "*Public-use files: production, linking and* ***anonymity***", United States Health Care and Finance Administration, 1998.

o "*Re-identification of Health Data*", Testimony before the Massachusetts State Healthcare Committee, 1997.

HIGHLY DISTINGUISHED TALKS (SELECTED)

o "*Privacy Technology*", Access and Privacy Workshop: Connecting citizens and government calls for openness, transparency and strategy. Office of the Privacy Commissioner. Toronto, Ontario, Canada. October 2004. (*Slides and References*)

o "*Privacy and the Future of Genomic Research*", Pacific Symposium on Biocomputing, Kauai, HI, 2003. Keynote speaker.

o "*Innovations Possible With Privacy-Enabled Web and Medical Data*", Blue Cross Blue Shield Association of Michigan, Detroit, MI, 2002. Received Privacy Leadership Award.

o "*Medical Privacy in Data and the Role of the Physician*", IPA, Las Vegas, NV 2000.

o "*Rendering Medical Records Sufficiently* ***Anonymous***", California Health Information Association, Sparks, NV, 2000.

o "*Data Privacy Legislation Trends and the Future of Psychiatric Notes*", American Psychiatric Association, Rancho Valencia, CA, 2000.

o "*The Compatibility of Detecting Outbreaks While Protecting Privacy*", State Territorial Eidemioloists New Orleans LA 2000.

o "*No Data Privacy in Canada Either*", Stakeholders Meeting, Canadian Privacy Council, Ottawa, Canada, 2000.

o "*Technology's Challenge to the Privacy of Medical Data*", California Medical Association, Executive Retreat, Palm Springs, CA, 1999.

o *"Data Linkage, Privacy and Fraud Detection"* and *"Federal Data Sharing Concerns",* Defending Cyberspace, Washington, DC, 1999.

o *"Privacy Concerns for Collecting and Sharing Registry Data",* North American Association of Central Cancer Registries, Chicago, IL, 1999.

o *"Technology's Attack on Privacy",* California HealthCare Foundation, Oakland, CA, 1999.

o *"The Identifiability of Data",* Commission on Cancer, American College of Surgeons. Chicago, IL, 1998.

DISTINGUISHED TALKS (SELECTED)

o *"Beyond Ickiness is Risk: The Exasperation of Data Privacy Problems by Implanted RFIDs",* The Concealed I Conference, University of Ottawa, Ontario Canada. March 4, 2005.

o "*Privacy Technology: Computer Scientists Help Save the World*", Intel Privacy Forum, Hillsboro, Oregon. March 2, 2005. (*Slides, Abstract, References*)

o "*Tracking Terrorism Using Privacy-Preserving Surveillance*", Tracking Terrorism in the 21st Century: A national symposium on the role of Science and Law in Detecting, Investigating, and Adjudicating Political Violence. Dusquene University School of Law, The Cyril H. Wecht Institute of Forensic Science and Law. Pittsburgh, PA. October 2004. (*Slides and References*)

o "*Privacy-Preserving Surveillance*", United States Department of Homeland Security: Directions for the Data Sciences Research Agenda. Alexandria, VA. September 2004. (*Slides and References*)

o "*Privacy Technology: New Directions for Computer Science*", CERIAS Seminar Series, Purdue University. West Lafayette, IN. September 2004. (*Slides and References*)

o "*Using Non-Traditional Data Sources to Detect Outbreaks Early*", Centers for Disease Control and Prevention. Atlanta, GA, 2000.

o "*Using Publicly and Semi-Publicly Available Data to Detect Bioterrorist Attacks*", The Pentagon, Washington, DC, 2000.

o "*Towards a Bioterrorist Surveillance System*", The U.S. Department of Defense, DARPA, Washington, DC, 2000.

o "*A Trilogy of Talks: Exploiting Existing Data Collections to Gain Strategic Information*", Centers for Disease Control and Prevention. Atlanta, GA, 2000.

o "*A Trilogy of Talks: Protecting Privacy when Sharing Data*", Centers for Disease Control and Prevention. Atlanta, GA, 2000.

o "*A Trilogy of Talks: Bioterrorism Surveillance*", Centers for Disease Control and Prevention. Atlanta, GA, 2000.

o *"Public-use files: production, linking and* ***anonymity****",* University of Michigan, Ann Arbor, MI, 2000.

o *"Computer Systems that Can Assist in Detecting Outbreaks and in Establishing and Coordinating Information During a Crisis",* Centers for Disease Control and Prevention, Philadelphia, PA, 2000.

o *"Informed Consent Provides No Privacy Protection in Data",* American Psychiatric Association, Washington, DC, 2000.

o *"Privacy? Going, going...",* Smith College, Northampton, MA, 2000.

o *"Privacy Challenges to the Digital Government Initiative",* National Institute for Statistical Sciences, Durham, NC, 1999.

o *"Data Linkage Problems" and "Disclosure Control",* American Statistical Association, Joint Statistical Meeting, Baltimore, MD, 1999.

o *"****Anonymity*** *Concerns are not a Matter of Security",* Association of Computing Machinery, Computers, Freedom and Privacy Conference, Washington, DC, 1999.

o *"Computational Disclosure Control: Problems & Solutions",* National Research Council, Washington, DC, 1999.

o *"Medical Data and Privacy Protections in Today's Technologically-Empowered Society",* Maryland Medical Society, Baltimore, Maryland, 1998.

OTHER NON-CONFERENCE TALKS (SELECTED)

o *"Privacy for Those Who Can Afford It,* Guest Lecture in Graduate Course 17-899, Technology and Development for 4 Billion, School of Computer Science, Carnegie Mellon University. Pittsburgh, PA. November 17, 2004. (*Slides and Abstract*)

o "*Tackling Proposals,* PhD Program in Computation, Organizations and Society, Graduate Seminar. School of Computer Science, Carnegie Mellon University. Pittsburgh, PA. November 15, 2004. (Slides and Abstract)

o *Privacy Technology for the NCPI,* Special presentation at a joint meeting of Vanderbilt University and the Center for Automated Learning and Discovery, at the School of Computer Science, Carnegie Mellon University. November 4, 2004. (*Slides and References*)

o "*Algorithms for Distributed Privacy-Preserving Surveillance*", Carnegie Museum of Science, Pittsburgh, PA, 2004.

o "*Data Detectives and Privacy Protectors*", Berkman Center for Internet and Society, Harvard Law School, Cambridge, MA, 1999.

o "*Data Mining, Data Linkage, Privacy Protection and Instant Epidemiology*" and "*Learning about Entities*", Center for Disease Control and Prevention (CDC), Atlanta, GA, 1999.

o "*Problems and Solutions in Releasing Medical Data*", New York University Medical School, New York, New York, 1998.

o "*Computational Solutions to Releasing Personal Information in the Age of Data Mining*", Carnegie Mellon University, Center for Automated Learning and Discovery, Pittsburgh, Pennsylvania, 1997.

o "*The Difficulties of Generating and Using **Anonymous** Data*", University of Pittsburgh, Medical Informatics Program, Pittsburgh, Pennsylvania, 1997.

o "*The Information Challenge to Privacy, Confidentiality and De-classification*", Carnegie Mellon University, The Heinz School for Public Policy, Pittsburgh, Pennsylvania, 1997.

o "*The Risks of De-Identified and **Anonymous** Data*", SRI International, Database Security, Menlo Park, California, 1997.

o *"Disclosing **Anonymous** Data in a Globally-Networked Society",* Stanford University, Database Group, Palo Alto, California, 1997.

o *"Yesterday Transistor Logic, Today Digital Computers and Tomorrow Atomic Automata",* Harvard Summer School, Computer Science, Cambridge, Massachusetts, 1997.

o *"Problems and Solutions in Releasing De-identified Patient Records",* National Library of Medicine, Annual Trainee Meeting, 1997.

o *"How to Disclose **Anonymous** Medical Data",* Massachusetts General Hospital, Laboratory of Computer Science, Boston, Massachusetts, 1997.

o *"Who Gets Access to What Kind of Patient Data",* Beth Israel Hospital, BIRT Lecture, Boston, Massachusetts, 1997.

o "*Generating **Anonymous** Data on-the-Fly*", Carnegie Mellon University, The Heinz School of Public Policy, Pittsburgh, Pennsylvania, 1997.

o "*How to De-Identify Medical Text*", Massachusetts General Hospital, Laboratory of Computer Science, Boston, Massachusetts, 1996.

o "*Replacing Personally-Identifying Information in Medical Records, the Scrub System*", Massachusetts Institute of Technology, MasterWorks, Cambridge, Massachusetts, 1996.

o "*Team Wheel: a new Collaboration Tool*", Massachusetts Institute of Technology, Clinical Decision-Making Group, Cambridge, Massachusetts, 1995.

NON-ACADEMIC WORKS BY OTHERS THAT REFERENCE LATANYA SWEENEY'S WORK (SELECTED)

POPULAR PRESS ARTICLES THAT REFERENCE MY WORK (SELECTED FROM MORE THAN 125)

o *Scientific American,* July 2007, "A Little Privacy, Please" by Chip Walter (*text*)

o *Pittsburgh Post-Gazette,* December 26, 2005, "The Thinkers: Data privacy drives CMU expert's work" *(text)*

o *CBS News, Denver,* "Angel Protects Those Who Might be Targets for ID Theft," October 20, 2005 *(text, video)*

o *CBS News, New York,* "Why A Resume Could Bring A Job, But Also ID Theft: Identity Theft From Online Resumes On The Rise," Septmber 28, 2005 *(text, video)*

o *Boston Phoenix* and various affiliates around the USA, February 15, 2005, "Spying Eyes: America's Most Popular Search Engine Is Keeping Tabs On Us" *(text)*

o ABC News, June 13, 2004, "University Web Site Watches Public Spies". *(text)*

o *CBS News,* Associated Press, March 15, 2004, "Privacy Safeguards Quietly Killed". (*text*)

o *CNN,* October 11, 2003, "Privacy researcher: Public Web cams troublesome". (*text*)

o *USA Today,* August 25, 2003, "CameraWatch, Who watches the watchmen?". (*text*)

o *CBS News,* Associated Press, November 4, 2002, "Germ Patrol: Like Never Before". (*text*)

o *Computer World,* October 14, 2002, "Privacy Algorithms." (*text*)

o *Pittsburgh Post-Gazette,* August 15, 2002, "Seeing profits in privacy, recent transplant launches firm to protect individual identities." (*more*)

o Wall Street Journal, February 2001. Author Glen Simpson. Privacy of Census Data.

o Consumer Reports, August 2000, Section: Consumer Interest: Health and Medical Records Privacy. "Who knows your medical secrets?" pages 22-26. http:// www.consumerreports.org

o Newsweek, October 16, 2000. "It Doesn't Take Much to Make You Stand Out".

o Newsday, November 21, 2000 Section: Health & Discovery, "When Medical Data Goes Public", pages C8-C10, Author Earl Lane (Washington Bureau), http:// library.newsday.com/

o *National Journal's Technology Daily,* July 14, 2000, "Senator predicts three-pronged medical privacy plan." (*text*)

o Wall Street Journal, February 2001. Author Glen Simpson. Privacy of Census Data.

o 20/20 Television News Magazine [20 second clip regarding pharmacy data]

o *New York Times,* February 16, 2000.

o *ZDNet,* April 7, 1999, "Free speech and privacy forever." (*text*)

o *Los Angeles Times,* February 8, 1999, "A new push is on for Patients Privacy Law." (*text*)

MEMBERSHIPS

• American Association for the Advancement of Science

• American Association for Artificial Intelligence

• American Medical Informatics Association

• Association for Computing Machinery

• Institute of Electrical and Electronic Engineers

• International Federation for Information Processing

• Sigma Xi

TEACHING EXPERIENCE

• *Technology Dialectics* (graduate), *Constructing Appropriate Technology* (undergraduate). This course teaches students how to infuse cross-disciplinary knowledge into research or technology development to make sure created technology is provably appropriate for a given personal, societal, organizational, and/or legal context.

2007-present *Carnegie Mellon University*

• *Privacy Technology.* This course prepares students to conduct research in privacy technology and to develop privacy-enhancing technology in industry.

2006-present *Carnegie Mellon University*

• *Computation, Organizations and Society Lab.* Intensive introduction of methods and traditions necessary for doing graduate research in computer science and writing publishable results such that the resulting technology that is the subject of the research is provably appropriate to the setting in which it will be deployed.

2004-present *Carnegie Mellon University, School of Computer Science.*

• *Privacy and **Anonymity** in Data.* Intensive undergraduate and graduate computer science courses providing research experience in data re-identification, privacy technology, and integrated legal and policy issues. Taught as Associate Professor, having about 300 students total.

2000-present *Carnegie Mellon University*

• Java Programming Courses: *I, II, III,* Intensive 3-semester courses that provide mastery of programming in the Java Programming language. Taught as Associate Professor, having about 500 students total. Course materials available on-line and used by students at 10 other schools. 1999-present *Carnegie Mellon University* Computer Science Department and MISM and MSIT Program.

1999-2000 *Harvard University* Distance course. Graduate option.

• *How to Really get **Anonymity** and Privacy in Data.* Half-day tutorial for professionals.

1999 *American Medial Informatics Association,* Annual meeting. Received best ratings possible in student evaluations.

• Other computer science courses.

1991-1996 *Harvard University,* Faculty of Arts and Sciences, as Instructor or Head Teaching Fellow. Received 7, on a maximum scale of 7, in student evaluations in each course taught. Taught most of these courses for several semesters, having about 900 students total.

o *Data Structures and Algorithms in C* (graduate and undergraduate)

o *Data Structures and Algorithms in Pascal* (graduate and undergraduate)

o *Great Ideas in Computer Science* (undergraduate)

o *Computer science and programming in C* (graduate and undergraduate)

o *Computer Architecture* (graduate)

CURRENT AND RECENT STUDENTS

• *Bradley Malin.*

Research work on genetic privacy and on learning algorithms. PhD Thesis, 2006, *Trail Re-identification and Unlinkability in Distributed Databases.* Now faculty member at Vanderbilt University.

• *Samuel Edoho-Eket.*

Masters student in the Information Networking Institute, Carnegie Mellon University. Graduated 2004. Thesis: *Simple Data Collection Protocol: A Privacy-Preserving Approach to Distributed Internet Surveillance.* This work relates to PrivaSum, described above. Now working in national security.

• William Gronim. Undergraduate student in computer science. Carnegie Mellon University. Thesis, 2006, *Name extraction from rosters.* Now working at Amazon.com.

• *Ralf Holzer.*

Masters student in the Information Networking Institute, Carnegie Mellon University. Graduated 2005. Thesis: *Email Alias Detection Using Social Network Analysis.* Now working at a major consulting corporation.

• *Ralph Gross.*

PhD student in Computation, Organizations and Society, School of Computer Science, Carnegie Mellon University. Research work on entity detection in and de-identification of video data. A recent accomplishment: paper accepted in *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2004, which is a prestigious computer science journal.

• Xiaoqian Jiang.

PhD student in Computation, Organizations and Society, School of Computer Science, Carnegie Mellon University. Research on privacy algorithms.

• *Laurie Jones.*

PhD student in Computation, Organizations and Society, School of Computer Science, Carnegie Mellon University. Research on privacy technology.

• Yiheng Li.

PhD student in Computation, Organizations and Society, School of Computer Science, Carnegie Mellon University. Has earned a Masters in CALD. Research on **anonymizing** and learning from **anonymized** data.

• *Brooke Singer.*

Masters thesis on exposing personal privacy problems. School of Art, Carnegie Mellon University. Graduated 2002. Now accomplished artist, and faculty member at SUNY.

• Wanhong Xu.

PhD student in Computation, Organizations and Society, School of Computer Science, Carnegie Mellon University. Research on policy specification and enforcement.

• Daniel Yurovsky.

Undergraduate student in computer science, Carnegie Mellon University. Research on identity management. Graduated 2007. Now graduate student in psychology working on thinking machines.

Last modified 3/2009 by *latanya@dataprivacylab.org*

End of Document