Latanya Arvette Sweeney, Ph.D.

Professor of Government and Technology in Residence
Department of Government

Editor-in-Chief
Technology Science
techscience.org

Director, Data Privacy Lab
dataprivacylab.org

Harvard University 1737 Cambridge Street, CGIS K310
Cambridge, MA 02138

Curriculum Vitae

(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 and Government Positions

- 2012 - Professor of Government and Technology in Residence, Department of Government, Harvard University.
- 2014 Chief Technology Officer, U.S. Federal Trade Commission.
- 2011 - today. Director and Founder, Data Privacy Lab, Harvard University.
- 2008 - 2012. Visiting Professor, Visiting Scholar, Computer Science, School of Engineering and Applied Science, Harvard University.
- 2009 - 2011. Distinguished Career Professor of Computer Science, Technology and Policy, School of Computer Science, Carnegie Mellon University.
- 2008 - 2011. MLK Visiting Professor, Computer Science (CSAIL), Massachusetts Institute of Technology.
- 2005 - 2009. Associate Professor of Computer Science, Technology and Policy, School of Computer Science, Carnegie Mellon University.

- 2002 - 2004. Assistant Professor of Computer Science, Technology and Policy, School of Computer Science, Carnegie Mellon University.
- 2001 - 2011. Director and Founder, Data Privacy Lab, Carnegie Mellon University.
- 2003 - 2011. Co-Director and co-Founder, PhD Program in Computation, Organizations and Society, School of Computer Science, Carnegie Mellon University.
- 2004, 2006 - 2009. Founder in 2004, Editor-in-Chief from 2006. Journal of Privacy Technology, Institute for Software Research, School of Computer Science, Carnegie Mellon University.
- 2003 - 2006. 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

- 2015 - Editor-in-Chief, Technology Science.
- 2005 - 2011. 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 - 2011. Faculty, Aladdin Center, School of Computer Science, Carnegie Mellon.
- 2002 - 2011. Faculty, Information Networking Institute, Carnegie Mellon University.
- 2004 - 2011. Faculty, Cylab, Carnegie Mellon University
- 2002 - 2011. Faculty, H. John Heinz III School of Public Policy and Management.
- 1998 - 2011. Faculty Advisory Board, Machine Learning Department, School of Computer Science, Carnegie Mellon University.

Corporate Affiliations

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

- 2013 - 2013. Wugster, LLC. has licenses to some of my technologies and is using these licenses to make the methods and results openly accessible.
- 2003 - 2013. Privacert has licenses to some of my technologies and is using these licenses to make the methods and results openly accessible.
- 2009 - 2013 Hand ID owns patents to some technologies created by 2013 colleagues and me. I maintain some ownership interest and hold an official position.
- 2002-2003. Datanon, LLC had licenses to some of my technologies, which are now available through Privacert, Inc.
- 1998 - 2013. 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 CEO. 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, GAO Appointment, 2009-2013.
- Appointed, EPIC Advisory Board, 2007-2013.
- Appointed, IEEE-USA Medical Technology Policy Committee, 2003-2009.

- Appointed member, Biomedical Library Research Committee, National Library of Medicine, National Institutes of Health, 1999-2003.
- Patient privacy Rights, Board of Directors, 2011-2013.
- Patient privacy Rights, Board of Directors, 2011-2013.
- Dana Hall Schools, Board of Directors, 2011-2013.
- Health Record Banking Alliance, Advisory Board, 2011-2013.
- 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.

Publications of Latanya Sweeney, PhD

By Type | Chronological

Books (including Forthcoming)

1. *The Sweet Spot: Harmonizing Technology and Society*.
   Sweeney, L. (This is not an edited volume.)
   Forthcoming book publication. 220 pages.

2. *Computer Guide*.
   Sweeney, L. and Barrett, S.
   Cambridge: CESS Publishing. 1983. (1425 pages) dataprivacylab.org/computerguide1983/index.html

Chapters in Books

3. 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.
   dataprivacylab.org/projects/facedeid/index.html.

4. 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/dataprivacy/projects/explosion/index.html.

5. 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 - Published

6. Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps.
   Zang J, Dummit K, Graves J, Lisker P, and Sweeney L.
   Technology Science, 2015103001. October 30, 2015.

7. Sharing Sensitive Data with Confidence: The Datatags System.
   Sweeney L Crosas M, and Bar-Sinai M.
   Technology Science, 2015101601. October 16, 2015.

8. De-anonymizing South Korean Resident Registration Numbers Shared in Prescription Data.
   Sweeney L and Yoo J.
   Technology Science, 2015092901. September 29, 2015.

9. Only You, Your Doctor, and Many Others May Know.
   Sweeney L.
   Technology Science, 2015092903. September 29, 2015.
   Earlier technical report: Sweeney L. Matching Known Patients to Health Records in Washington State Data. Harvard University. Data Privacy Lab. 1089-1. June 2013.

10. Automating Open Science for Big Data.
    Crosas M, King G, Honaker J and Sweeney L.
    The ANNALS of the American Academy of Political and Social Science May 2015 vol. 659 no. 1 260-273.

11. Discrimination in Online Ad Delivery.
    Sweeney L.
    Communications of the Association of Computing Machinery (CACM), Vol. 56 No. 5, Pages 44-54 DOI 10.1145/2447976.2447990 http://cacm.acm.org/magazines/2013/5/163753-discrimination-in-online-ad-delivery/abstract
    Earlier technical report: Data Privacy Lab White Paper 1071-1. Harvard University. Cambridge. January 2013. SSRN: http://ssrn.com/abstract=2208240 and arXiv: http://arxiv.org/abs/1301.6822.

12. Putting Health IT on the Path to Success
    William Yasnoff, MD, Latanya Sweeney, PhD, Edward Shortliffe, MD
    Journal of the American Medical Association (JAMA) 2013;309(10):989-990.
    doi:10.1001/jama.2013.1474

13. 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)
    Earlier version available as Carnegie Mellon University, Technical Report CMU-ISRI-05-124. dataprivacylab.org/dataprivacy/projects/complexcare/index.html.

14. 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.

15. 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. dataprivacylab.org/dataprivacy
    /projects/pel/index.html.

16. 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. dataprivacylab.org/dataprivacy/projects/selectiverevelation
    /index.html.

17. 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. dataprivacylab.org/dataprivacy/projects
    /video/index.html.

18. 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.
    dataprivacylab.org/dataprivacy/projects/csresearch.html

19. 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.
    dataprivacylab.org/dataprivacy/projects/rosterfinder/index.html

20. 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

21. *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/dataprivacy/projects/kanonymity/kanonymity.html.

22. 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/dataprivacy/projects/kanonymity
    /kanonymity2.html.

23. 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

24. Commentary: Researchers need not rely on consent or not.
    Sweeney, L.
    New England Journal of Medicine, 1998. (impact factor 52.59) dataprivacylab.org/projects
    /research/index.html

25. 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

26. A study to assess the impact of re-identification on the willingness of patients to contribute
    personal information to a genomics project.
    Sweeney, L., Abu A., Winn J. and Zang J.

27. Identifying Participants in the Personal Genome Project by Name.
    Sweeney, L., Abu A. and Winn J.

28. Publicly Available State Health Databases are not HIPAA Compliant.
    Hooley S. and Sweeney, L.

29. Privacy-Preserving Marketplaces.
    Sweeney, L., Pentland, S., Chong, S., and Parkes, D.

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

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

Refereed Conference/Workshop Paper - Published

32. Adaptive Gaussian Process for Wind Speed Forecast
    Jiang, X., Dong, B., Xie, L., and Sweeney, L.
    The 19th European Conference on Artificial Intelligence (ECAI 2010).
    dataprivacylab.org/projects/sensor/index.html

33. From Sensor Networks to Social Networks - A Time Series M3N Approach

Jiang, X., Bing, D., and Sweeney, L.
NIPS Workshop Analyzing Networks and Learning with Graphs, 2009.
dataprivacylab.org/projects/sensor/index2.html

34. Multi-Factor De-Identification of Facial Images.
Gross, R., and Sweeney, L., Cohn, F., de la Torre, F., and Baker, S.
Proceedings of the 2008 American Medical Informatics Association Annual Symposium, 2008.
(AMIA acceptance rate 29%) dataprivacylab.org/projects/facedeid/index2.html

35. 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%) dataprivacylab.org/projects/facedeid/index3.html

36. Towards Real-World Face De-Identification.
Gross, R. and Sweeney, L.
IEEE Conference on Biometrics, Washington, DC, September 2007. (BTAS acceptance rate
11%) dataprivacylab.org/projects/facedeid/index4.html

37. 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%) dataprivacylab.org/projects/hand/index.html

38. 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%) dataprivacylab.org/projects/hand/index2.html

39. 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%) dataprivacylab.org/projects/hand/index3.html

40. 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%) dataprivacylab.org/dataprivacy/projects/trails/index4.html

41. Model-based face de-identification.
R. Gross, Sweeney, L., de la Torre, F., and Baker, S.
IEEE Workshop on Privacy Research in Vision, 2006. dataprivacylab.org/projects/facedeid
/index5.html

42. A secure protocol to distribute unlinkable health data.
Malin, B. and Sweeney, L..
Proceedings, Journal of the American Medical Informatics Association. Washington, DC. Oct
2005: 485-489. (AMIA acceptance rate 29%) dataprivacylab.org/dataprivacy/projects/trails
/index5.html

43. 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.

44. 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.

45. 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.

46. 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.

47. Privacy-Preserving Bio-terrorism Surveillance.
    Sweeney, L.
    AAAI Spring Symposium, AI Technologies for Homeland Security, 2005.
    dataprivacylab.org/dataprivacy/projects/bioterror/index.html.

48. Towards a Privacy-Preserving Watchlist Solution.
    Sweeney, L.
    AAAI Spring Symposium, AI Technologies for Homeland Security, 2005.
    dataprivacylab.org/dataprivacy/projects/watchlist/index.html.

49. 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.

50. 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/index4.html

51. 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/projects/scrub/index.html

52. 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

53. 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%)
    dataprivacylab.org/projects/ereport/index.html

54. 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/dna1.html

55. Foundations of Privacy Protection from a Computer Science Perspective.
Sweeney, L.
Proceedings, Joint Statistical Meeting, AAAS, Indianapolis, IN. 2000.
dataprivacylab.org/projects/disclosurecontrol/index.html

56. 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%) dataprivacylab.org/projects/tutor1/index.html

57. 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%) dataprivacylab.org/projects/crisis/index.html

58. 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)
dataprivacylab.org/dataprivacy/projects/kanonymity/index3.html

59. 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%)
dataprivacylab.org/dataprivacy/projects/kanonymity/index4.html

60. 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. dataprivacylab.org/dataprivacy/projects/kanonymity/index5.html

61. 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. dataprivacylab.org/dataprivacy/projects/kanonymity/index6.html

62. Computational Disclosure Control for Medical Microdata.
Sweeney, L.
Record Linkage Workshop. Bureau of the Census. Washington, DC: 1997.
dataprivacylab.org/dataprivacy/projects/kanonymity/index7.html

Commentaries and News Articles

63. Privacy Issues Overlooked
    Sweeney, L.
    Modern Healthcare. September 9, 2010. dataprivacylab.org/projects/nhin3/

Technical Reports and White Papers (excluding any subsequently published or submitted)

64. Patient Privacy Risks in U.S. Supreme Court Case Sorrell v. IMS Health Inc.: Response to Amici Brief of El Emam and Yakowitz
    Sweeney, L.
    Data Privacy Lab Working Paper 1027. Cambridge 2011. dataprivacylab.org/projects/identifiability/pharma2.html

65. Patient Identifiability in Pharmaceutical Marketing Data
    Sweeney, L.
    Data Privacy Lab Working Paper 1015. Cambridge 2011. Earlier version available as *Identifiability of De-identified Pharmacy Data*. Pittsburgh 2003. dataprivacylab.org/projects/identifiability/pharma1.html

66. The Medical Billing Framework as the Backbone of the National Health Information Infrastructure.
    Sweeney, L.
    Data Privacy Lab. AdvanceHIT Series. Working Paper 1001. October 2009.
    dataprivacylab.org/projects/nhin1/

67. Demonstration of a Privacy-Preserving System that Performs an Unduplicated Accounting of Services across Homeless Programs.
    Sweeney, L.
    Data Privacy Lab Working Paper 902. Pittsburgh 2007, October 2008.
    dataprivacylab.org/projects/homeless/index2.html

68. *The HandShot Hand Tracking System*.
    Gross, R. and Sweeney, L.
    Carnegie Mellon University, School of Computer Science, Technical Report, 2009. In preparation.

69. *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.

70. *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.

71. *Identifiability of De-identified Clinical Trial Data*.
    Sweeney, L.
    Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report. Pittsburgh: 2009.

72. *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.

73. Risk Assessments of Personal Identification Technologies for Domestic Violence Homeless Shelters.
Sweeney, L.
Carnegie Mellon University, School of Computer Science. Technical Report CMU-ISRI-05-133. Pittsburgh: November 2005. Data Privacy Lab Working Paper 901. Pittsburgh 2005.
dataprivacylab.org/projects/homeless/index.html

74. *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.

75. *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.

76. *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.

77. *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.

78. *Silhouette Database of Naturally Positioned Hands*.
Sweeney, L..
Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-30.
Pittsburgh: 2005.

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

80. *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. dataprivacylab.org/dataprivacy/projects/gentree/gentree2.html

81. 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. dataprivacylab.org/dataprivacy/projects/handshot/index.html

82. *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/index.html

83. 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. dataprivacylab.org/dataprivacy/projects/randomorder/index.html

84. *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.

85. *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.

86. *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. dataprivacylab.org/dataprivacy/projects/gentree/index.html

87. *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. dataprivacylab.org/dataprivacy/projects/trails/index3.html

88. *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.

89. *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. dataprivacylab.org/dataprivacy/projects/thatsai/index.html

90. 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.
Shorter version: Simple Demographics Often Identify People Uniquely.
Sweeney, L.
Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

dataprivacylab.org/projects/identifiability/index.html.

91. *Re-identification of De-identified Survey Data*.
    Sweeney, L.
    Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical
    Report. Pittsburgh: 2000.

92. 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. dataprivacylab.org/dataprivacy/projects/law/law1.html

93. Power Learning: using computers as teaching machines.
    Sweeney, L.
    Massachusetts Institute of Technology, Artificial Intelligence Laboratory: Working Paper. 1998.
    dataprivacylab.org/projects/plcp/index.html

94. *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/projects/bebe/index.html

95. *Towards the collection of all the data on all the people*.
    Sweeney, L.
    MIT Artificial Intelligence Working Paper, 1998.

96. *Maintaining anonymity when sharing medical data, the datafly system*.
    Sweeney, L.
    MIT Artificial Intelligence Laboratory Working Paper. Cambridge: AIWP-WP344 (1997).

97. *Iterative Profiler*.
    Sweeney, L.
    Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical
    Report. Pittsburgh: 1997.
    The Iterative Profiler sifts through publicly available data, using inferential linkages of data
    fragments across data sources to construct profiles of people whose information appears in the
    data. After it re-identified the names of children who appeared in a cancer registry, an Illinois
    court ordered the approach sealed. For the appellate decision, which upheld the sealing of the
    methodology while highly praising my approach, see http://www.state.il.us/court/Opinions
    /AppellateCourt/2004/5thDistrict/June/Html/5020836.htm.

98. *Multiple hashed binary storage of words -- tiny, fast and almost perfect*.
    Sweeney, L.
    Massachusetts Institute of Technology, AI Laboratory: Working paper. 1996.

99. *Computational Disclosure Control: Theory and Practice*.
    Sweeney, L.
    Massachusetts Institute of Technology, Ph.D. Thesis in Computer Science, 2001.

100. *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.

101. *A Coin Toss: the Dialectical Odds aren't always 50/50.*
   Sweeney, L.
   Harvard University, Undergraduate research thesis, 1995.

Patents and Invention Disclosures

102. *Systems and methods for deidentifying entries in a data source.*
   Sweeney, L.
   Patent 7,269,578. Issued on 9/11/2007.

103. *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.

104. *Frontal Hand Capture of Fingerprints, Palm Prints amd Hand Geometry Using Contactless Photography.*
   Sweeney, L., Gross, R., and Jiang, X.
   Application filed July 2008.

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

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

107. *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 ZIP*}". 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 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, 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 (Java1, 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

- Journal of Privacy Technology, Editor-in-Chief, 2006 - today
- Journal of Law and Policy for the Information Society
- 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
- 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)

- 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.
- 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.
- 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)

- *"Gov 2.0"*
  Kennedy School of Government, Harvard University,
  Cambridge, MA January 16, 2013. (PDF)

- *"Implications of Patient Privacy in Sorrell v. IMS Health."*
  Health Patient Privacy Summit,
  Washington, DC June 13, 2011. (video, more)

- *"The Science of Privacy Technology."*
  American Association for the Advancement of Science (AAAS),
  Topical Lecture at Annual Meeting, St. Louis, MO. February 17, 2006.

- *"De-Identifying Health Data."*
  Health Canada (Federal Department of Health),
  Ottawa, Ontario Canada. December 1, 2005.

- *"Strategies for De-Identifying Patient Data for Research."*
  Electronic Health Information and Privacy Conference,
  Ottawa, Ontario Canada. November 30, 2005.

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

- *"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.

- *"Identity Management: Dealing with Disclosure."*
  6th CACR,
  Toronto, Canada, November 3, 2005.

- *"Privacy Principles for Ubiquitous Technologies"*,

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

- "*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.

- "*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)

- "*Privacy Technologies for Large Research Databases*"
Spectrum Health and Michigan State University,
Grand Rapids, MI, September 23, 2005. (Slides and Abstract)

- "*Biometrics Alone Won't Do: Developing Holistic Identity Management Solutions*"
Biometrics Symposium 2005,
Arlington, VA, September 19, 2005. (Slides and Abstract)

- "*Risk Assessments of PIN Technologies for Domestic Violence Shelters,*"
National HMIS Conference,
St. Louis, Missouri, September 13, 2005. (Slides and Abstract)

- "*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)

- "*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)

- "*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)

- "*Privacy Technology: Artificial Intelligence to Save the World,*"
AAAI Spring Symposium (Invited talk).
Stanford. Palo Alto, CA. March 23, 2005. (Slides, References and Abstract).

- "*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).

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

- "*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)

- "*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)

- *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)

- "*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)

- "*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)

- "*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)

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

- "*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)

- "*Risk Assessments of PIN Technologies [identity management] for Domestic Violence Shelters (Updated).*" Housing and Urban Development. Washington, DC November 18, 2005.
- "*Privacy Principles for Ubiquitous Technologies*", Testimony before the Eurpoean Commission, Brussels, Belgium, October 25, 2005.
- "*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)
- "*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)
- "*Outsourcing and Offshoring Privacy Risks in light of HIPAA*", HIMSS Advocacy Day on the Hill. Washington, DC, 2004.
- "*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.
- "*Privacy-preserving Bio-terrorism Surveillance*", Annual Meeting of the American Association

for the Advancement of Science, Seattle, WA 2004.
- "*Vulnerabilities of Genomic Databases: the risks are real*", Annual Meeting of the American Association for the Advancement of Science, Seattle, WA 2004.
- "*The De-identification Provision of HIPAA*", American Health Lawyers. Atlanta, GA, 2002.
- "*Genetic Privacy*", U.S. Senate Committee Congressional Briefing, Washington, DC, 2000. Hosted by Senator Frist (R) and Senator Rockefeller (D).
- "*Virtual Fingerprints in Medical Data*" and *Technology for Taming Data*, National Library of Medicine, Board of Regents, Washington, DC, 1999.
- "*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.
- "*Anonymity, Technology and Policy*", Massachusetts Institute of Technology, Congressional Staffer Retreat, Cambridge, MA, 1999.
- "*Public-use files: production, linking and anonymity*", United States Health Care and Finance Administration, 1998.
- "*Re-identification of Health Data*", Testimony before the Massachusetts State Healthcare Committee, 1997.

Highly Distinguished Talks (Selected)

- "*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)
- "*Privacy and the Future of Genomic Research*", Pacific Symposium on Biocomputing, Kauai, HI, 2003. Keynote speaker.
- "*Innovations Possible With Privacy-Enabled Web and Medical Data*", Blue Cross Blue Shield Association of Michigan, Detroit, MI, 2002. Received Privacy Leadership Award.
- "*Medical Privacy in Data and the Role of the Physician*", IPA, Las Vegas, NV 2000.
- "*Rendering Medical Records Sufficiently Anonymous*", California Health Information Association, Sparks, NV, 2000.
- "*Data Privacy Legislation Trends and the Future of Psychiatric Notes*", American Psychiatric Association, Rancho Valencia, CA, 2000.
- "*The Compatibility of Detecting Outbreaks While Protecting Privacy*", State Territorial Epidemiologists, New Orleans, LA, 2000.
- "*No Data Privacy in Canada Either*", Stakeholders Meeting, Canadian Privacy Council, Ottawa, Canada, 2000.
- "*Technology's Challenge to the Privacy of Medical Data*", California Medical Association, Executive Retreat, Palm Springs, CA, 1999.
- "*Data Linkage, Privacy and Fraud Detection*" and "*Federal Data Sharing Concerns*", Defending Cyberspace, Washington, DC, 1999.
- "*Privacy Concerns for Collecting and Sharing Registry Data*", North American Association of Central Cancer Registries, Chicago, IL, 1999.
- "*Technology's Attack on Privacy*", California HealthCare Foundation, Oakland, CA, 1999.
- "*The Identifiability of Data*", Commission on Cancer, American College of Surgeons. Chicago, IL, 1998.

Distinguished Talks (Selected)

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

- *"Privacy-Preserving Surveillance"*, United States Department of Homeland Security: Directions for the Data Sciences Research Agenda. Alexandria, VA. September 2004. (Slides and References)
- *"Privacy Technology: New Directions for Computer Science"*, CERIAS Seminar Series, Purdue University. West Lafayette, IN. September 2004. (Slides and References)
- *"Using Non-Traditional Data Sources to Detect Outbreaks Early"*, Centers for Disease Control and Prevention. Atlanta, GA, 2000.
- *"Using Publicly and Semi-Publicly Available Data to Detect Bioterrorist Attacks"*, The Pentagon, Washington, DC, 2000.
- *"Towards a Bioterrorist Surveillance System"*, The U.S. Department of Defense, DARPA, Washington, DC, 2000.
- *"A Trilogy of Talks: Exploiting Existing Data Collections to Gain Strategic Information"*, Centers for Disease Control and Prevention. Atlanta, GA, 2000.
- *"A Trilogy of Talks: Protecting Privacy when Sharing Data"*, Centers for Disease Control and Prevention. Atlanta, GA, 2000.
- *"A Trilogy of Talks: Bioterrorism Surveillance"*, Centers for Disease Control and Prevention. Atlanta, GA, 2000.
- *"Public-use files: production, linking and anonymity"*, University of Michigan, Ann Arbor, MI, 2000.
- *"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.
- *"Informed Consent Provides No Privacy Protection in Data"*, American Psychiatric Association, Washington, DC, 2000.
- *"Privacy? Going, going…"*, Smith College, Northampton, MA, 2000.
- *"Privacy Challenges to the Digital Government Initiative"*, National Institute for Statistical Sciences, Durham, NC, 1999.
- *"Data Linkage Problems"* and *"Disclosure Control"*, American Statistical Association, Joint Statistical Meeting, Baltimore, MD, 1999.
- *"Anonymity Concerns are not a Matter of Security"*, Association of Computing Machinery, Computers, Freedom and Privacy Conference, Washington, DC, 1999.
- *"Computational Disclosure Control: Problems & Solutions"*, National Research Council, Washington, DC, 1999.
- *"Medical Data and Privacy Protections in Today's Technologically-Empowered Society"*, Maryland Medical Society, Baltimore, Maryland, 1998.

Other Non-conference Talks (Selected)

- *"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)
- *"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)
- *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)
- *"Algorithms for Distributed Privacy-Preserving Surveillance"*, Carnegie Museum of Science, Pittsburgh, PA, 2004.
- *"Data Detectives and Privacy Protectors"*, Berkman Center for Internet and Society, Harvard Law School, Cambridge, MA, 1999.
- *"Data Mining, Data Linkage, Privacy Protection and Instant Epidemiology"* and *"Learning

- *about Entities*", Center for Disease Control and Prevention (CDC), Atlanta, GA, 1999.
- "*Problems and Solutions in Releasing Medical Data*", New York University Medical School, New York, New York, 1998.
- "*Computational Solutions to Releasing Personal Information in the Age of Data Mining*", Carnegie Mellon University, Center for Automated Learning and Discovery, Pittsburgh, Pennsylvania, 1997.
- "*The Difficulties of Generating and Using Anonymous Data*", University of Pittsburgh, Medical Informatics Program, Pittsburgh, Pennsylvania, 1997.
- "*The Information Challenge to Privacy, Confidentiality and De-classification*", Carnegie Mellon University, The Heinz School for Public Policy, Pittsburgh, Pennsylvania, 1997.
- "*The Risks of De-Identified and Anonymous Data*", SRI International, Database Security, Menlo Park, California, 1997.
- "*Disclosing Anonymous Data in a Globally-Networked Society*", Stanford University, Database Group, Palo Alto, California, 1997.
- "*Yesterday Transistor Logic, Today Digital Computers and Tomorrow Atomic Automata*", Harvard Summer School, Computer Science, Cambridge, Massachusetts, 1997.
- "*Problems and Solutions in Releasing De-identified Patient Records*", National Library of Medicine, Annual Trainee Meeting, 1997.
- "*How to Disclose Anonymous Medical Data*", Massachusetts General Hospital, Laboratory of Computer Science, Boston, Massachusetts, 1997.
- "*Who Gets Access to What Kind of Patient Data*", Beth Israel Hospital, BIRT Lecture, Boston, Massachusetts, 1997.
- "*Generating Anonymous Data on-the-Fly*", Carnegie Mellon University, The Heinz School of Public Policy, Pittsburgh, Pennsylvania, 1997.
- "*How to De-Identify Medical Text*", Massachusetts General Hospital, Laboratory of Computer Science, Boston, Massachusetts, 1996.
- "*Replacing Personally-Identifying Information in Medical Records, the Scrub System*", Massachusetts Institute of Technology, MasterWorks, Cambridge, Massachusetts, 1996.
- "*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)

Media References to My Work (Selected from more than 425)

- Sample of Media reaction to my paper, *Discrimination in Online Ads*, February 2013. (NBC News, CBS WBZ-TV, Boston Globe, ABC news, BBC, Huffington Post, Technology Review, Wired)
- *New York Times*, May 2011, "Data Privacy, Put to the Test" by Natasha Singer (article, archive)
- *Scientific American*, July 2007, "A Little Privacy, Please" by Chip Walter (text)
- *Pittsburgh Post-Gazette*, December 26, 2005, "The Thinkers: Data privacy drives CMU expert's work" (text)
- *CBS News, Denver*, "Angel Protects Those Who Might be Targets for ID Theft," October 20, 2005 (text, video)
- *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)
- *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)
- *ABC News*, June 13, 2004, "University Web Site Watches Public Spies". (text)
- *CBS News*, Associated Press, March 15, 2004, "Privacy Safeguards Quietly Killed". (text)
- *CNN*, October 11, 2003, "Privacy researcher: Public Web cams troublesome". (text)
- *USA Today*, August 25, 2003, "CameraWatch, Who watches the watchmen?". (text)
- *CBS News*, Associated Press, November 4, 2002, "Germ Patrol: Like Never Before". (text)
- *Computer World*, October 14, 2002, "Privacy Algorithms." (text)

- *Pittsburgh Post-Gazette*, August 15, 2002, "Seeing profits in privacy, recent transplant launches firm to protect individual identities." ([more](#))
- Wall Street Journal, February 2001. Author Glen Simpson. Privacy of Census Data.
- Consumer Reports, August 2000, Section: Consumer Interest: Health and Medical Records Privacy. "Who knows your medical secrets?" pages 22-26. http://www.consumerreports.org
- Newsweek, October 16, 2000. "It Doesn't Take Much to Make You Stand Out".
- Newsday, November 21, 2000 Section: Health & Discovery, "When Medical Data Goes Public", pages C8-C10, Author Earl Lane (Washington Bureau), http://library.newsday.com/
- *National Journal's Technology Daily*, July 14, 2000, "Senator predicts three-pronged medical privacy plan." ([text](#))
- Wall Street Journal, February 2001. Author Glen Simpson. Privacy of Census Data.
- 20/20 Television News Magazine [20 second clip regarding pharmacy data]
- *New York Times*, February 16, 2000.
- *ZDNet*, April 7, 1999, "Free speech and privacy forever." ([text](#))
- *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

- [The Politics of Personal Data](#). The course examines legal, political, social, commercial and technical struggles for control over personal data in our globally connected data-rich world. Case studies include data sharing mandated by the state, traded for personal services, and controlled by individuals. Analyses demonstrate ways to think about clashes and the interplay between technology design and policy.
  2013-present *Harvard University*

- [Data Science to Save the World](#). This course explores ways to create and use technology to assess and solve real-world societal, political and governance problems. Case studies include twitter, online ads, and surveillance.
  2013-present *Harvard University*

- **Privacy and Technology**. The course examines how privacy is affected by recent developments in computer technology and critically examines popular concepts of privacy. Engages students in a rigorous analysis of technologies to understand the policy and ethical issues.
  2011-present *Harvard University*

- [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.
  2007-2009 *Carnegie Mellon University*

- [Privacy Technology](#). This course prepares students to conduct research in privacy technology and to develop privacy-enhancing technology in industry.
  2006-2009 *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.
  2004-2009 *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.
  2000-2006 *Carnegie Mellon University*

- Java Programming Courses: I, II, and 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.
    - Data Structures and Algorithms in C++ (graduate and undergraduate)
    - Data Structures and Algorithms in Pascal (graduate and undergraduate)
    - Great Ideas in Computer Science (undergraduate)
    - Computer science and programming in C (graduate and undergraduate)
    - Computer Architecture (graduate)

Past Students

MS or PhD Thesis Committee Service

- Xiaoqian Jiang.
  Ph.D. thesis, Predictions for Biomedical Decision Support, 2010. Committee Chair. Graduate Advisor. Carnegie Mellon University (ISR). On to postdoc at UCSD Medical School.

- [Ralph Gross]().
  Ph.D. thesis, Face De-identification Using Multi-Factor Active Appearance Models, 2008. Committee Chair. Graduate Advisor. Carnegie Mellon University (ISR). On to start-up company related to thesis, and part-time faculty at Carnegie Mellon University

- [Bradley Malin]().
  PhD. thesis, Trail Re-identification and Unlinkability in Distributed Databases, 2006. Committee Chair. Graduate advisor. Carnegie Mellon University (ISR). On to tenure-track Assistant Professorship at Vanderbilt University.

- Samuel Edoho-Eket
  Masters thesis, Detecting Bio-terrorist Attacks and Naturally Occurring Outbreaks Over a Distributed network While Protecting Privacy and Confidentiality: the PrivaSum Protocol, 2004. Thesis Advisor. Graduate Advisor. Carnegie Mellon University (INI Masters Program). On to national security agency.

- Ralf Holzer
  Masters thesis, Email Alias Detection Using Network Analysis, 2005. Thesis Advisor. Carnegie Mellon University (INI Masters Program). On to Deloitte Consulting.

- [Brooke Singer](#)
  Masters thesis, Personal Information Made Public, 2002. Co-Thesis Advisor. Carnegie Mellon University (School of Art Masters Program). On to national recognition as an artist, and faculty member at SUNY.

Masters Research

- Edo Airoldi
  Masters degree 2003. Statistical Machine Learning. Graduate Advisor. Carnegie Mellon University (ISR). Now tenure-track Assistant Professor at Harvard University.

- Laurie Jones
  Masters degree 2010. Privacy Technology for Generating Public Use Files. Graduate Advisor. Carnegie Mellon University (ISR).

- Yiheng Li
  Masters degree 2003. Generalized Association Rule Learning. Graduate Advisor. Carnegie Mellon University (ISR). On to UC Davis School of Law.

- Elaine Newton
  Work on face de-identification. Graduate Advisor. Carnegie Mellon University (EPP). On to NIST.

- Wanhong Xu
  Masters Degree 2010. Learning from Inferential Disclosures of Personal Information. Graduate Advisor. Carnegie Mellon University (ISR). On to technical team at Facebook.

- William Malloy
  Masters degree 2001. Electronic Disease Surveillance. Carnegie Mellon University (MISM). On to Captain in the U.S. Army.

Undergraduate Research

- William Gronim
  Extracting Names from Websites Containing Lists of People, 2006. Research Advisor. Carnegie Mellon University (CSD). On to elite technical team at Amazon.com.

- Daniel Yurovsky
  Contactless Capture of Hand Geometry Using Low Resolution Cameras, 2007. Research Advisor. Carnegie Mellon University (CSD). On to graduate school at Indiana University.

- Ben Vernot
  Linking Databases and Extracting Names, 2003. Research Advisor. Carnegie Mellon University (CSD). On to graduate school in biology at Carnegie Mellon University.

Last modified October 2015.