DSCI-534: Biomedical Data Privacy Issues and Solutions

Units: 4
FALL 2022

Instructors: Dr. Tatyana Ryutov, Prof. David W. Craig
Time: 2:00-5:20pm
Days: Mondays
Office: ONLINE
Contact Info: tryutov@usc.edu
Course website: https://piazza.com/usc/fall2022/dsci534

Course Description

Privacy concerns in healthcare, current law and regulations, existing and emerging technologies shaped by ethics, privacy considerations, medical implications. Special attention given to genomic data.

Learning Objectives

After successfully completing this course, the students will be able to:

- comprehend the significance of privacy of medical data in healthcare;
- analyze privacy laws and governing regulations;
- identify the fundamental concepts and key issues of genomic privacy;
- apply the existing privacy preserving methodologies; and
- approach complex biomedical data privacy problems from these angles:
    - Data Vulnerability: Demonstrate how seemingly private information, can be discovered (or exploited) using automated strategies.
    - Data Protection: Select privacy protection technologies that provide formal computational guarantees of privacy in disclosed datasets.
    - Technology Policy Design: Apply privacy protection technologies that complement policy regulations.

Recommended Preparation

Prior experience with information security, public policy, and legal frameworks is not required for this course. Basic understanding of engineering and/or technology principles; basic programming skills at the level of DSCI 549 or DSCI 510 is preferred. Some background in informatics, data science, or computer science will be valuable.

All key concepts and relevant methodology will be reviewed and introduced throughout the course, however students should be comfortable learning about basics of human genetics, precision medicine, various cryptographic methods, and statistics.

Course Notes

This course will be conducted online, using a combination of synchronous and asynchronous methods. The remote learning format of this interdisciplinary course will eliminate the need for students attending different schools (e.g., Viterbi and Keck) to travel between campuses. Therefore, the course will continue to be offered online after the COVID restrictions are lifted.

Grading type: letter. Piazza (poizza.com) will be used for posting copies of lecture slides, announcements, assignments, and intra-class communication. Blackboard (blackboard.usc.edu) will be used for posting of grades, lecture recordings, homework submission, quiz submission, exam submission. Zoom (usc.zoom.us) will be used for lectures and office hours.

Technological Proficiency and Hardware/Software Required

Students must provide their own laptop. The laptop specifications take into consideration that students will be creating, streaming, and downloading audio and video, communicating using video-conferencing applications, and creating and storing large multimedia files.

Required Readings and Supplementary Materials

There is no primary textbook for this course. Reading assignments will be selected from various periodicals and other sources.

Hours of Instruction

Once weekly for 200 minutes including two 10-minute breaks.

Description and Assessment of Assignments

Class assignments
Exercise 1 (***due by week 3***). Explore databases publicly accessible through the National Center for Biotechnology Information (NCBI) and other Web sites. The purposes of this lab is to:

- Introduce the nature and the characteristics of genomic sequences (protein & DNA);
- Explain how to retrieve or understand known genetic risk factors;
- Illustrate the ethnic differences in genetic profiles and disease risks;
- Demonstrate how accessible information on genes is to the general public; and
- Increase students' awareness of computational genetic resources available to them.

Exercise 2 (***due by week 6***): Hands-on exercise to assess one's risks for cancer, cardiovascular diseases, and other debilitating diseases given the genotyping data.

Exercise 3 (***due by week 10***): Hands-on exercise to determine the relationships of individuals given their genotyping data.

Exercise 4 (***due by week 14***) : Hands-on exercise: utilize a demo version of LUBA-PrivET to explore genomic data protection policies. This lab will allow students to experiment with different privacy policies and model interaction between data owner and data consumer.

An optional programming project will be offered for students proficient in programming.

Class participation
Pop out questions (about 6) will be ascked during each lecture. Responses will be sumitted using Google forms. If the student does not submit the response then this will result in a deduction of class participation.

Semester project
The semester project gives each student the opportunity to apply the concepts taught in the course. The project proposal (2 pages) should include a description of the topic, what the student intends to do and how, contain preliminary references.

Each student must complete an independent project on a data privacy issue in biomedicine. Projects should investigate a topic of interest to the student, and must demonstrate analysis and critical thought. Students may design their own project or choose from a predefined set of topics. A list of sample project topics (e.g., kin privacy breach affecting individuals whose relatives publicly share genomic data; abuse of DNA to stage crime scenes) will be made available and reviewed in class.

Work on the project will consist of several phases:

- Project Proposal (***due by week 4***): A one-pager that describes the topic, what the student intends to do and how, contains preliminary references.
- Written Project Status Report (***due by week 11***): A summary of the progress that was made.
- Final Project Presentation: Showcase of research methods and results.
    - ***The last two lectures*** will be devoted to individual student presentations. Students will be assigned specific date and time to present their findings by means of a 15-20 minute power point presentation. Presentations will be assessed based on the peer review: each student will complete a brief survey providing their thoughts and reactions to the presentations.
- Final Project Report (***due on the last day of class***): This will be in the form of a conference-style paper. It will summarize the research area, the methodology, experience, and contributions of your work.

Final examination
The final exam will be a two hour written test administered via the USC Blackboard. The exam format will be a combination of short answers and esseys. Final exam date and time: refer to the final exam schedule in the USC Schedule of Classes at classes.usc.edu.

Grading Breakdown

Assessment Tool (assignments)	% of Grade
Final Exam	25%
Homework Assignments	40%
Class Participation	10%
Semester Project	25%
TOTAL	100%

Assignment Submission Policy
Assignments and semester project will be submitted electronically via Blackboard. Assignments will be accepted after the deadline with the following grade penalties. Cumulative of 10% times number of days late:

- 1 day late: lose 10%
- 2 days late: lose 30% (10% + 20%)
- 3 days late: lose 60% (30% + 30%)
- Greater than 4 days late not accepted

No personal emergencies will be entertained (with the exception of the USC granted emergencies, in which case official documents need to be shown).

Additional Policies
Class notes policy: Notes or recordings made by students based on a university class or lecture may only be made for purposes of individual or group study, or for other noncommercial purposes that reasonably arise from the student's membership in the class or attendance at the university. This restriction also applies to any information distributed, disseminated, or in any way displayed for use in relationship to the class, whether obtained in class, via e-mail or otherwise on the Internet, or via any other medium. Actions in violation of this policy constitute a violation of the Student Conduct Code and may subject an individual or entity to university discipline and/or legal proceedings. Again, it is a violation of USC's Academic Integrity Policies to share course materials with others without permission from the instructor.

Participation: Students are expected to actively participate in this course. Participation includes:

- Careful reading and viewing of assigned materials by the date due
- Regular, substantive contributions to discussions and in-class questions
- Active engagement with online content

Course grades for students who do not contribute to the course through active participation may be affected.

Course Schedule: A Weekly Breakdown
Class sequence, dates, reading assignments, and topics are subject to change as the semester proceeds. Any revisions will be noted and announced in class.

Week	Topics	Readings	Delivarable
08/22 Lec1	**Course introduction** • Why do we need a course on data privacy? • Data privacy definition	Required: • **[Warren]** S. Warren and L. Brandeis. "The right to privacy." Harvard Law Review. 1890; V. IV, No. 5.	

	• Privacy frameworks • Models of data protection • Legal aspects of privacy • Privacy law and policy in the United States and the European Union	• **[Benthall]** S. Benthall, S. F. Gürses, H. Nissenbaum: "Contextual Integrity through the Lens of Computer Science". Trends Priv. Secur. 2(1): 1-69. 2017. Optional: • **[McGraw]** D. McGraw, J. Dempsey, L. Haris, and J. Goldman. "Privacy as an enabler, not an impediment: building trust into health information exchange". Health Affairs. 2009; 28(2): 416-427. • **[Tene]** O. Tene and J. Polonetsky. "Privacy in the age of big data: a time for big decisions". Stanford Law Review (Online). 2012; 64: 63. • **[Nissenbaum]** Nissenbaum, H. 2004. "Privacy as contextual integrity". Washington Law Review, 79, 119–157.	
08/29 Lec2	**Health Insurance Portability and Accountability Act (HIPAA) & Genetic Information Nondiscrimination Act** • HIPAA • GINA **Ethical principles and privacy** • A Primer on Ethical Theory • Code of Ethics • Ethical issues related to incidental and secondary findings	Required: • **[HIPAA]** U.S. Department of Health and Human Services Summary of the Privacy Rule of the Health Information Portability and Accountability Act (HIPAA). • **[GINA]** The Genetic Information Nondiscrimination Act of 2008. • **[Green]** Green, R. C., et al. (2013). "ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing." Genetics in Medicine, 15(7): 565-574. • **[Christenhusz]** Christenhusz, G. M., Devriendt, K., & Dierickx, K. 2013. "To tell or not to tell? A systematic review of ethical reflections on incidental findings arising in genetics contexts." European Journal of Human Genetics, 21(3): 248-255. Optional: • **[HIPAA-96]** Health Insurance Portability and Accountability Act of 1996 • **[PrivacyGenome]** "Privacy and Progress in Whole Genome Sequencing" Presidential commission for the Study of Bioethical Issues, 2012. • **[Jarvik]** Jarvik, G.P., et al. (2014). "Return of genomic results to research participants: the floor, the ceiling, and the choices in between." American Journal of Human Genetics, 94(6):818-26.	
09/05	**No class, University Holiday**		
09/12 Lec3	**Genomic basics** • The Human Genome Project • Clinical exome sequencing • Somatic cancer testing • Incidental findings & ACMG 59 genes • Considerations for interpreting germline and somatic genetic variants **Genomic data processing** • Sequencing technologies	Required: • **Lee H**, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M, Fox M, Fogel BL, Martinez-Agosto JA, Wong DA, Chang VY, Shieh PB, Palmer CG, Dipple KM, Grody WW, Vilain E, Nelson SF. Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. JAMA. 2014 Oct 18. doi: 10.1001/jama.2014.14604.) • Richards S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405-24. doi: 10.1038/gim.2015.30. Epub 2015 Mar 5. • The SAM/BAM Format Specification Working. "Sequence Alignment/Map Format Specification." 2019. http://samtools.github.io/hts-specs/SAMv1.pdf. • Li, H. et al. *Bioinformatics*, 15(16), 2078-2079 (2009).	HW1 due

	• Variant calling, annotation and interpretation • Standard sequence formats (e.g. FASTQ, BAM, VCF, etc)	• Danecek, P. et al. *Bioinformatics*, 27(15), 2156-2158 (2011). Optional: • Posey JE, et al. Molecular Diagnostic Experience of Whole-Exome Sequencing in Adult Patients. Genet Med. 2016 Jul;18(7):678-85. doi: 10.1038/gim.2015.142. Epub 2015 Dec 3. • Chong JX, et al. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am J Hum Genet. 2015 Aug 6;97(2):199-215. doi: 10.1016/j.ajhg.2015.06.009. Epub 2015 Jul 9. Review.	
09/19 Lec4	**Genomic privacy** • Genomic data ownership • Limitations of legal and regulatory frameworks to protect genomic privacy • Genomic privacy, your relatives, and the implications	Required: • Rodriguez, L. L., et al. (2013). "Research ethics. The complexities of genomic identifiability." Science, 339(6117): 275-276. • Erlich, Y. & Narayanan, A. (2014). "Routes for breaching and protecting genetic privacy." Nature Reviews Genetics 15(6): 409-421. • Belluz, J. (2014). "Genetic testing brings families together: and sometimes tears them apart." Vox, September 9. Optional: • Belluz, J. (2014). "23andMe reverses its decision to move to more lax privacy settings." Vox, September 16. • Sulmasy, D. P. (2014). "Naked bodies, naked genomes: the special (but not exceptional) nature of genomic information." Genetics in Medicine. • Lee, S. S. J. & Borgelt, E. (in press) "Protecting posted genes: social networking and the limits of GINA." American Journal of Bioethics. • Botkin JR, Belmont JW, Berg JS, et al. ASHG Statement: Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. American Journal of Human Genetics 2015 Jul 2;97(1):6-21.	Project Proposal due
9/26 Lec5	**Medical and genomic (Big) Data Sharing and Informed Consent** • Genomic databases, data sharing initiatives, Genomic Data Brokers • Direct-to-Consumer Genomics and Legal Rights • Types of consent	Required: • Paltoo, D. N., et al. (2014). "Data use under the NIH GWAS data sharing policy and future directions." Nature Genetics, 9: 934-938. • NIH Genomic Data Sharing Policy, http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html, Release Date: August 27 2014. • Allen, A. (2014). "Can Big Data and patient informed consent coexist?" Politico, September 3. • Callier, S., Husain, R., & Simpson, R. (2014) "Genomic data-sharing: what will be our legacy?" Frontiers in Genetics, 5 (34): 1-3. • Pereira, S., et al. (2014). "Open access data sharing in genomic research." Genes 5(3): 739-747. Optional: • Case study: Arizona Board of Regents v. Havasupai Tribe, included lack of informed consent, violation of civil rights, unapproved use of data, and violation of medical confidentiality/re-identification.	
10/03 Lec6	**Basics of Information security** • Information security and privacy goals • Authentication and identification • Access control models • Mandatory and discretionary controls	Required: • "Security in Computing", Charles P. Pfleeger, Shari Lawrence Pfleeger, Jonathan Margulies, 2015. Chapters 1; 2.1 Optional: • Selected chapters from Brian Walker, "Cyber Security Comprehensive Beginners Guide to Learn the Basics and Effective Methods of Cyber Security", 2019.	HW2 due

		• B. Blobel, R. Nordberg, J. Davis, and P. Pharow. "Modelling privilege management and access control". International Journal of Medical Informatics. 2006; 75: 597-623. (https://www.sciencedirect.com/science/article/abs/pii/S1386505605001747)	
10/10 Lec7	**Information security tools** • Least privilege and separation of duties • Role based access control • Attribute based access control • Monitoring and auditing	Required: • "Security in Computing", Charles P. Pfleeger, Shari Lawrence Pfleeger, Jonathan Margulies, 2015. Chapter 2.2 • **[Sandhu]** R. Sandhu, E. Coyne, H. Feinstein, and C. Youman. "Role-based access control models". IEEE Computer. 1996; 26(2): 38-47. • **[Hu]** V.C. Hu, R. Kuhn, D.F. Ferraiolo, and J. Voas. Attribute-based access control. IEEE Computer. 2015; 48(2): 85-88. Optional: • **[Nweke]** Nweke LO, Yeng P, Wolthusen SD, Yang B. "Understanding attribute-based access control for modelling and analysing healthcare professionals' security practices". J Adv Comput Sci Appl. 2020;11(2):683–690. • **[Chen]** Chen Y, Malin B. Detection of Anomalous Insiders in Collaborative Environments via Relational Analysis of Access Logs. CODASPY. 2011:63-74.	
10/17 Lec8	**Basics of Cryptography** • Encryption basics • Encryption tools for authentication, data confidentiality, integrity and non-repudiation • Attribute-based encryption	Required: • "Security in Computing", Charles P. Pfleeger, Shari Lawrence Pfleeger, Jonathan Margulies, 2015. Chapter 2.3 Optional: • Selected chapters from Kratikal Academy, "Cryptography: Data and Application Security", 2017. • **[Akinyele]** "Self-Protecting Electronic Medical Records Using Attribute-Based Encryption", Joseph A. Akinyele, Christoph U. Lehmann, Matthew D. Green, Matthew W. Pagano, Zachary N. J. Peterson, Aviel D. Rubin	
10/24 Lec9	**Cryptographic tools for privacy protection** • Homomorphic encryption • Cryptographic methods for secure multiparty computation • Application of blockchain	Required: • **[Lindelll]** "Secure Multiparty Computation for Privacy-Preserving Data Mining", Yehuda Lindell, Benny Pinkas. Journal of Privacy and Confidentiality Vol1. 2008. • **[Nakamoto]** S. Nakamoto. "Bitcoin: A Peer-to-Peer Electronic Cash System". 2008. • **[Kuo]** T. Kuo, H. Kim, L. Ohno-Machado. Blockchain distributed ledger technologies for biomedical and health care applications. Journal of the American Medical Informatics Association. 2017; 24(6), 1211-1220. Optional: • **[Halevi]** S. Halevi , "Homomorphic Encryption", IBM Research, 2017 • **[Blatta]** M. Blatta, A. Guseva, Y. Polyakova, and S. Goldwasse "Secure large-scale genome-wide association studies using homomorphic encryption", 2020 • "Secure Genome-wide Association Analysis using Multiparty Computation"., Hyunghoon Cho, David J. Wu, Bonnie Berger, Nature Biotechnology 36, 2018	

10/31 Lec10	**De-identification of biomedical data; Re-identification; Big Data record linkage; Inference and prediction of personal information** • De-identification: detect and suppress "identifiers" from unstructured data (e.g., clinical narratives • Identifiability concerns associated high-dimensional data • Privacy-Preserving Record Linkage	Required: • **[Gupta]** D. Gupta, M. Saul, and J. Gilbertson. "Evaluation of a deidentification (De-Id) software engine to share pathology reports and clinical documents for research". American Journal of Clinical Pathology. 2004; 121: 176-186. • **[Ruch]** P. Ruch, et al. "Medical document anonymization with a semantic lexicon." Proceedings of the 2000 American Medical Informatics Association Annual Fall Symposium (AMIA). 2000; 729-733. • **[Berman]** Berman JJ. "Concept-match medical data scrubbing: how pathology text can be used in research", 2003. • **[Tiara]** Taira R, Bui A, Kangarloo H. "Identification of patient name references within medical documents". In Proceedings of the 2002 AMIA Annual Fall Symposium. 2002; 757-761. • **[Malin]** Bradley Malin and Latanya Sweeney , "How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems". 2004 • **[Randall]** "Privacy-preserving record linkage on large real world datasets" Sean M. Randall , Anna M. Ferrante, James H. Boyd, Jacqueline K. Bauer, James B. Semmens Optional: • **[Lee]** H.J. Lee, et al. "A hybrid approach to automatic de-identification of psychiatric notes". Journal of Biomedical Informatics. 2017; 75: S19-S27.	HW3 due
11/07 Lec11	**Anonymization** • Formal models of anonymization • k-anonymity • l-diversity • t-closeness • Strategies for anonymization of DNA data • De-anonymization attacks	Required: • **[Friedman]** "Providing k-Anonymity in Data Mining" by Friedman et al. • **[Sweeney]** "Achieving k-anonymity privacy protection using generalization and suppression", Sweeney et al. • **[Machanavajjhala**] "l-diversity: Privacy beyond k-anonymity" by Machanavajjhala et al. • **[Li]** N. Li, T. Li, S. Venkatasubramanian,. "t-Closeness: privacy beyond k-anonymity and l-diversity". Proceedings of the 23rd IEEE International Conference on Data Engineering. 2007: 106-115 • **[Lin]** Z. Lin, M. Hewett, and R. Altman. "Using binning to maintain confidentiality of medical data". Proc AMIA Symp. 2002; 454-458. Optional: • [**Ohno-Machado**] "Protecting patient privacy by quantifiable control of disclosures in disseminated databases" by Ohno-Machadoa et al. • **[Homer]** N. Homer, et al. "Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays". PLoS Genetics. 2008; 4(8): e1000167.	Project status report due
11/14 Lec12	**Privacy Preserving Data Analysis** • How can we combine data to reveal results, but no individual records? • Differential privacy **Emerging frameworks and technologies**	Required: • **[Nissim]** K. Nissim, et al. "Differential privacy: a primer for a non-technical audience". White paper of the Privacy Tools for Sharing Research Data Project, Harvard University. 2017. • **TBD**	

	• Game theoretic approaches to privacy • Dynamic encryption • Watermarking		
11/21 Lec 13	Project Presentations		HW4 due
11/28 Lec 14	Project Presentations		Final project report due
	Final Exaamination will be held at the University's appointed time.		

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
studenthealth.usc.edu/counseling
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org
Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention and Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call
studenthealth.usc.edu/sexual-assault
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED)- (213) 740-5086 | Title IX – (213) 821-8298
equity.usc.edu, titleix.usc.edu
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants. The university prohibits discrimination or harassment based on the following *protected characteristics*: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in

applicable laws and governmental regulations. The university also prohibits sexual assault, non-consensual sexual contact, sexual misconduct, intimate partner violence, stalking, malicious dissuasion, retaliation, and violation of interim measures.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
usc-advocate.symplicity.com/care_report
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office of Equity and Diversity |Title IX for appropriate investigation, supportive measures, and response.

The Office of Disability Services and Programs - (213) 740-0776
dsp.usc.edu
Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Support and Advocacy - (213) 821-4710
uscsa.usc.edu
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101
diversity.usc.edu
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
dps.usc.edu, emergency.usc.edu
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call
dps.usc.edu
Non-emergency assistance or information.