

DSCI-534: Biomedical Data Privacy Issues and Solutions Units: 4 FALL 2023

Instructor: Prof. Tatyana Ryutov Time: 2:00-5:20pm Days: Mondays Office: ONLINE Contact Info: tryutov@usc.edu Course website: https://piazza.com/usc/fall2023/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 (piazza.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, 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 are selected from academic literature, various periodicals and other sources.

Hours of Instruction

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

Class 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 will be affected.

Pop out questions (about 6) will be asked during each lecture. Responses will be submitted using Google forms. The students will have 72 hours to submit their responses for each lecture. Failure to submit the responses on time will result in a deduction of the class participation score.

Semester project

Each student must complete a 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 will be made available and reviewed in class. This self-proposed semester-long project allows students to select either a research-oriented or implementation-oriented direction. The project can be conducted in groups of two students or individually.

Work on the project will consist of several phases:

- Project Proposal (*due by week 5*): The project proposal (2 pages) should include a description of the topic, what the student intends to do and how, contain preliminary references.
- Written Project Status Report (*due by week 9*): A summary (6 pages) 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 20-30 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.

The exam can only be taken on the scheduled date and at the scheduled starting time. Accommodations for students with letters from DSP will be provided, though the exam will still need to be taken on the scheduled date and start time. There are no makeup exams. If you miss an exam due to a documented illness or an emergency, official written documentation will need to be submitted to instructor as soon as possible. Approval will be based on the instructor's discretion.

Grading Breakdown (preliminary)

Artifact	Weight
Final Exam	25%
Homework Assignments	40%
Class Participation	10%
Semester Project	25%
TOTAL	100%

Grading Timeline

None of the items in this class are auto graded. Assignments and the final exam will typically be graded within 7 days of the due date. Final project deliverables will typically be graded within 5 days of the due date. The class participation grades will typically be graded within 3 days after the end of classes.

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).

Diversity, Equity, and Inclusion (DEI) Statement

Our classroom is a place to expand our knowledge and experiences safely, while being respected and valued. We proactively strive to construct a safe and inclusive learning environment by respecting each other's dignity and privacy. We treat one another fairly and honor each member's experiences, beliefs, perspectives, abilities, and backgrounds, regardless of race, religion, language, immigration status, sexual orientation, gender identification, ability status, socio-economic status, national identity, or any other identity markers.

Disruptive or insulting remarks, gender or racial slurs, or other forms of bullying, intimidation or hate speech and other disrespectful language or behavior will not be tolerated. We welcome your thoughts on how we can improve our learning environment.

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 will 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
	Course introduction	Required:	
08/21		• [Warren] S. Warren and L. Brandeis. "The right to privacy" 1890.	

Lec1	 Course introduction Why do we need a course on data privacy? Data privacy definition Privacy frameworks Models of data protection 	 [Benthall] S. Benthall, S. F. Gürses, H. Nissenbaum: "Contextual Integrity through the Lens of Computer Science", 2017. [Shvartzshnaider] Shvartzshnaider, Y.; Apthorpe, N.; Feamster, N.; and Nissenbaum, H. I, "Going Against the (Appropriate) Flow: A Contextual Integrity Approach to Privacy Policy Analysis", 2019. Optional: [McGraw] D. McGraw, J. Dempsey, L. Haris, and J. Goldman. "Privacy as an enabler, not an impediment: building trust into health information exchange", 2009. [Tene] O. Tene and J. Polonetsky. "Privacy in the age of big data: a time for big decisions", 2012. [Nissenbaum] Nissenbaum, H. "Privacy as contextual integrity", 2004. 	
08/28 Lec2	 Legal aspects of privacy Privacy law and policy in the United States and the European Union HIPAA GINA 	 Required: [UDHR] Universal Declaration of Human Rights, https://www.un.org/en/about-us/universal-declaration-of-human-rights [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. 	
09/4 09/11 Lec3	 No class, University Holiday Ethical principles and privacy A Primer on Ethical Theory Code of Ethics Ethical issues related to incidental and secondary findings 	 Required: [Andrade] Gabriel Andrade, "Medical ethics and the trolley Problem", 2019. [Green] Green, R. C., et al. "ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing", 2013. [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", 2013. Optional: [PrivacyGenome] "Privacy and Progress in Whole Genome Sequencing" Presidential commission for the Study of Bioethical Issues, 2012. [Jarvik] Jarvik, G.P., et al. "Return of genomic results to research participants: the floor, the ceiling, and the choices in between", 2014. 	HW1 due
09/18 Lec4	 Genomic basics Basics of DNA Genetic variations Genomic data processing Sequencing technologies Variant calling, annotation and interpretation Standard sequence formats (e.g. FASTQ, BAM, VCF, etc) 	 Required: What is DNA? (Structure and Function of DNA), https://www.youtube.com/watch?v=T6_wKPAbf2k Genes and the Genome, https://www.youtube.com/watch?v=wv1TQXBQ6wQ Alleles/Dominant/Heterozygous/Phenotypes and more! https://www.youtube.com/watch?v=zNEtVaNQ0s8 What are DNA Mutations?, https://www.youtube.com/watch?v=3jwDl7nYBPM Mutation and Haplotype, https://www.youtube.com/watch?v=MeuQLee5YqI Next Generation Sequencing (NGS) - Data Analysis, https://www.youtube.com/watch?v=I4BAfRekohk&list=PLTt9kKfqE_0Gem 8hlcJEn7YcesuuKdt_n&index=4 The SAM/BAM Format Specification Working. "Sequence Alignment/Map Format Specification." 2019. 	Project Proposal due

		 How are Proteins Made? https://www.youtube.com/watch?v=ubdoUqmNF98 Next Generation Sequencing - A Step-By-Step Guide to DNA Sequencing, https://www.youtube.com/watch?v=WKAUtJQ69n8 	
9/25 Lec5	 Genomic privacy Direct-to-Consumer Genomics and Legal Rights Genomic (Big) Data Sharing Informed consent Diversity, equity and inclusion in genomic research 	 Required: [Rodriguez] Rodriguez, L. L., et al. "The complexities of genomic identifiability", 2013. Erlich, Y. & Narayanan, A. "Routes for breaching and protecting genetic privacy." Nature Reviews Genetics 15(6): 409-421. 2014, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4151119/ [Paltoo] Paltoo, D. N., et al. "Data use under the NIH GWAS data sharing policy and future directions", 2014. [Bentley] Amy R. Bentley, Shawneequa Callier, and Charles N. Rotimi, "Diversity and inclusion in genomic research: why the uneven progress?", 2017. NIH Genomic Data Sharing Policy, 2014, http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html Pereira, S., et al. (2014). "Open access data sharing in genomic research." Genes 5(3): 739-747, 2014, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4198928/ Optional: 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. Callier, S., Husain, R., & Simpson, R. "Genomic data-sharing: what will be our legacy?" Frontiers in Genetics, 5 (34): 1-3. 2014, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3942643/ N. A. Garrison, "Genomic Justice for Native Americans: Impact of the Havasupai Case on Genetic Research", 2012, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5310710/ [Rebbeck] Timothy R. Rebbeck, et. al."A Framework for Promoting Diversity, Equity, and Inclusion in Genetics and Genomics Research", 2022. 	
10/2 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. B. Blobel, R. Nordberg, J. Davis, and P. Pharow. "Modelling privilege management and access control". International Journal of Medical Informatics. 75: 597-623, 2006, https://www.sciencedirect.com/science/article/abs/pii/S13865056050017 47 	HW2 due
10/9 Lec7	 Information security tools Least privilege and separation of duties Role based access control Attribute based access control 	 47 Required: "Security in Computing", Charles P. Pfleeger, Shari Lawrence Pfleeger, Jonathan Margulies, 2015. Chapter 2.2 [Sandhu] R. Sandhu, David F. Ferraiolo, D. Richard Kuhn, "The NIST Model for Role-Based Access Control: Towards a Unified Standard", 2000. [Hu] V.C. Hu, R. Kuhn, D.F. Ferraiolo, and J. Voas." Attribute-based access control", 2015. 	

	 Monitoring and auditing 	 Optional: [Nweke] Nweke LO, Yeng P, Wolthusen SD, Yang B. "Understanding attribute-based access control for modelling and analysing healthcare professionals' security practices", 2020. [Chen] Chen Y, Malin B. "Detection of Anomalous Insiders in Collaborative Environments via Relational Analysis of Access Logs", 2011. 	
10/16 Lec8	 Basics of Cryptography Encryption basics Encryption tools for authentication, data confidentiality, integrity and non-repudiation Attribute-based encryption 	 Required: Charles P. Pfleeger, Shari Lawrence Pfleeger, Jonathan Margulie, "Security in Computing", 2015. Chapter 2.3 Optional: Selected chapters from Kratikal Academy, "Cryptography: Data and Application Security", 2017. [Akinyele] Joseph A. Akinyele, et al. "Self-Protecting Electronic Medical Records Using Attribute-Based Encryption", 2010. 	Project status report due
10/23 Lec9	 Cryptographic tools for privacy protection Homomorphic encryption Cryptographic methods for secure multiparty computation Application of blockchain 	 Required: [LindellI] Yehuda Lindell et al. "Secure Multiparty Computation for Privacy-Preserving Data Mining", 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", 2017. 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. [Kim] Miran Kim and Kristin Lauter. "Private genome analysis through homomorphic encryption", 2015. "Secure Genome-wide Association Analysis using Multiparty Computation", Hyunghoon Cho, David J. Wu, Bonnie Berger, Nature Biotechnology 36, 2018. 	
10/30 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", 2004. [Ruch] P. Ruch, et al. "Medical document anonymization with a semantic lexicon", 2000. [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", 2002. [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] Sean M. Randall , et. al, "Privacy-preserving record linkage on large real world datasets", 2014. Optional: [Lee] H.J. Lee, et al. "A hybrid approach to automatic de-identification of psychiatric notes", 2017. 	HW3 due
11/06	Anonymization	Required: • [Friedman] "Providing k-Anonymity in Data Mining" by Friedman et al.	

	amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays", 2008.	
 Privacy Preserving Data Analysis How can we combine data to reveal results, but no individual records? Differential privacy 	 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 	
Emerging frameworks and technologies		
 Game theoretic approaches to privacy Dynamic encryption Watermarking 		
Project Presentations		HW4 due
Project Presentations		Final project report due
	 How can we combine data to reveal results, but no individual records? Differential privacy Emerging frameworks and echnologies Game theoretic approaches to privacy Dynamic encryption Watermarking Project Presentations 	 How can we combine data to reveal results, but no individual records? Differential privacy Differential privacy TBD Game theoretic approaches to privacy Dynamic encryption Watermarking

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" <u>policy.usc.edu/scampus-part-b</u>. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, <u>policy.usc.edu/scientific-misconduct</u>.

Students with Disabilities

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me as early in the semester as possible. Your letter must be specific as to the nature of any accommodations granted. DSP is located in STU 301 and is open 8:30 am to 5:30 pm, Monday through Friday. The telephone number for DSP is (213) 740-0776.

Learning Experience Evaluations

Learning Experience Evaluations will be completed during the last day of class. This will be your opportunity to provide feedback about your learning experience in the class. This feedback helps the instructor determine whether students are having the intended learning experiences for the class. It is important to remember that the learning process is collaborative and requires significant effort from the instructor, individual students, and the class as a whole. Students should provide a thoughtful assessment of their experience, as well as of their own effort, with comments focused on specific aspects of instruction or the course. Comments on personal characteristics of the instructor are not appropriate and will not be considered. For this feedback to be as comprehensive as possible, all students should complete the evaluation.

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 <u>dsp.usc.edu</u>

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.