Project Title: *Database Management Current Industry Practices for Safeguarding Medical Personally Identifying Information.*

Charles Henderson, [chasehenderson@smu.edu](mailto:chasehenderson@smu.edu)

Nicole Bartholow, [nbartholow@smu.edu](mailto:nbartholow@smu.edu)

Project Problem:

One of the most sensitivetypes of personally identifying information (PII) databases are medical records.  In the digital age it is vital to quickly be able to pull patient data and compare it with other relational databases to determine intelligent disease clusters and anomaly detection.

Therefore it would be highly inefficient to protect patient PII through purely carbon data storage but even with the most sophisticated network security; PII is vulnerable to exposure when it is being pulled by databases and mixed with other datasets. Being able to reconnect anonymous data back to a patient is relevant problem, because each additional piece of data puts patients at risk for identification.  In the wake of the European Union’s General Data Protection Regulation we anticipate more regulation, and, ideally the methodology can change as the definitions of PII change, as well.

With a concise analysis of tactics to safeguard medical PII data across database queries, this project seeks to establish a methodology for securing personal data through socially responsible data definition and manipulation.

Research Methodology:

Through a combination of scholarly sources and interviews with leading database managers at the University of Utah Health Services, we will create an inclusive, though not wholly comprehensive, analysis of required security measures when querying medically significant datasets that contains PII data.

The initial steps require us to create a data model that would reflect a simplified medical record for a patient, including a subset of data we will seek to protect.   We will then define, for this project, what data is necessary to identify disease clusters or patterns and to establish the line between PII and that data. We should also define basic security roles that allow for differentiation of the information to be viewed.  We will then define the aggregation methods and architecture for the query-able data. As part of this question, we also need to establish the timeliness required for the data to be fresh and relevant for disease clusters or other patterns we seek beneficial enough to identify.  Finally, we will validate our approach.

Related Works:

Mohammed, Noman, et al. "Secure and private management of healthcare databases for data mining." *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*. IEEE, 2015.

Dubovitskaya, Alevtina, et al. "Secure and trustable electronic medical records sharing using blockchain." *AMIA Annual Symposium Proceedings*. Vol. 2017. American Medical Informatics Association, 2017.

Guo, Cheng, et al. "Fine-grained database field search using attribute-based encryption for e-healthcare clouds." *Journal of medical systems* 40.11 (2016): 235.

Tang, Huanrong, Ning Tong, and Jianquan Ouyang. "Medical Images Sharing System Based on Blockchain and Smart Contract of Credit Scores." *2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN)*. IEEE, 2018.

## IOM (Institute of Medicine). 2010. Clinical data as the basic staple of health learning: Creating and protecting a public good: Workshop Summary. Washington, DC: The National Academies Press

Kakon, Yarden Z. "“Hello, My Name is User# 101”: Defining PII Under the VPPA." *Berkeley Technology Law Journal* 33.4 (2019): 1251.

Manheim, Karl M., and Lyric Kaplan. "Artificial Intelligence: Risks to Privacy and Democracy." *Yale Journal of Law and Technology* (2018).

Houser, Kimberly A., and W. Gregory Voss. "GDPR: The End of Google and Facebook or a New Paradigm in Data Privacy." *Rich. JL & Tech.* 25 (2018): 1.

Boris Lubarsky , “Re-Identification of “Anonymized” Data,”1 GEO. L. TECH. REV. 202 (2017)

Other Interesting Article:

Christina Farr, “Why tech companies keep hitting the same wall when they try to get into health care” CNBC, MAY 30 2019

Research Plan and Schedule:

|  |  |
| --- | --- |
| Date | Milestone |
| June 7 | Complete Refined Project Proposal - Our understanding and agreement of the appropriate project scope of this topic has swung significantly over the last few days, from deeply theoretical to relatively basic to impossibly complex.  Pieces of each of these interpretations are reflected in this proposal. Our first objective is that we must agree on the scope of the project. All other intermediate milestones below will be adjusted will be defined by this scope. |
| June 9 | Establish connection to test DB instance, if needed |
| June 16 | Design and Create data model for patient - including separation of private versus needed data |
| June 19 | Checkin -  first draft of expanded Overview, DB instance (if needed),  Identify graphs |
| June 20 | Establish Slide Overview for initial presentation |
| June 23 | Complete Deck for Project Initial Presentation |
| June 25 | Project Initial Presentation |
| July 14 | Project Draft |
| August 6 | Project Presentation |
| August 18 | Project Final Paper |

Resources Needed: Not yet solidified