**CS 6440 Student Project Proposal**

|  |  |
| --- | --- |
| **Project Title:** CDC Population Health Informatics Framework: FHIR Identity Masking and Re-Linking Services | |
| **Mentor 1 Name:**  John Loonsk MD FACMI | **Mentor 1 Email:** John.Loonsk@cgifederal.com |
| **Mentor 2 Name:** Jason Bonander MA | **Mentor 2 Email:** jbonander@cdc.gov |
| **Project Background:** *EHRs and other health information technology offer opportunities to advance chronic disease management and prevention. Challenging these opportunities, though, are the myriad ways clinical care and public health are organized, the different ways that EHRs and other HIT systems store and exchange data, and the correlated, limited ability to have applications and services that be used in multiple settings. These issues beg for a framework or architecture that has descriptions of involved functions and systems, national standards for data and technical transactions, and a sense for how all these systems fit together. For these reasons and others, the CDC Center for Chronic Disease Prevention and Health Promotion is developing a Population Health Informatics Framework.*  Data linking and deduplication plays a critical role in connecting different population data sources. The flip side of data linking are protective confidentiality and security tools and approaches. Identified and de-identified data are no longer adequate differential privacy categories. Trusted third parties, trust agreements, specific pseudonymization methods, selective data release, and data perturbation need to be integrated as confidentiality tools and services. | |
| **Project Objective(s):** Help develop RESTful FHIR services that protect patient identity and confidentiality through the support of dynamic identity masking, pseudonymization and relinking | |
| **Description of the Solution:**   1. Provide pseudonymization services that deliver non-competitive, random identifiers to support patient identity masking and the re-identification of data sets 2. Enable the re-linking of patient identity, when appropriate, with a high security service, compatible with two factor authentication and encryption of data at rest 3. Services may be offered from a non-governmental, trusted third party that has identity access and may offer limited data set and other non-identified data sets to eligible users who do not, in fact, need patient identity 4. An approach to re-linking pseudonymized data and offering data updates to data recipients should be considered 5. This project is not about probabilistic identity matching which is offered by many commercial products | |
| **Desired Student Skills/Background:**   * Ability to read and make enhancements to existing code * Back-end RESTful Service development, web servers * Interest in security and confidentiality issues | |
| **Data Requirements and Potential Sources:**   * Synthetic patient data can be utilized * Use of standard FHIR resources with patient identity data elements | |
| **Other Comments:**   * The CDC may or may not advance the concept of this project * For all of these reasons any external communications about this project need to be approved by the advisor and the CDC * Any code and intellectual property developed for this project will be considered to be freely available to the CDC in perpetuity | |