Application of a System Engineering Structure for Private Information Sharing and Graph Theoretic Analysis of Private Health Information across Public Health Jurisdictions

J. C. Smart
Office of the Senior Vice President for Research
Georgetown University
Washington, DC 15213
smart@georgetown.edu

Joshua A. Ripple
OSVPR
Georgetown University
jar285@georgetown.edu

Abstract

Movement of people across jurisdictions complicates maintenance of accurate records on individuals for informing public health policy decisions and improving care. Contextual issues hinder the tracking of individuals across local, state, and federal jurisdictions by limiting sharing of private information across jurisdictions. To address this problem, we developed and implemented a prototype "black box" to analyze health records across a small subset of public health jurisdictions in the United States. A transient graph space was created in dynamic memory within a security isolated computational container so that attempts to access or tamper assure information irretrievability. Private health information could be deposited in this container only by the cooperating, authorized jurisdictions. A simple but reliable, verified graph-based algorithm that was unable to be altered after locking of the black box at the onset of the matching experiment was used. Output from this embedded system was restricted to graph vertex identifier and associated algorithm rankings so that only minimally sufficient information for identification of individuals by source jurisdictions was exposed. Follow-up manual validation of this system by participating jurisdictions revealed an algorithm matching accuracy of greater than 90%. The total number of accurate, verified matches discovered by the algorithm greatly exceeded the number of matches determined by the jurisdictions through current manual administrative procedures. The reduction in time and increase in accuracy of identification of people across jurisdictional boundaries will enable the participating government agencies to obtain a closer to real-time view of the status of diseases within their served populations and their associated public health needs.