A Risk Analysis Perspective on Biomedical Data Privacy

Emerging programs, such as the Precision Medicine Initiative, are promising to collect highly detailed biological, clinical, and mobile data from millions of individuals. These programs aim to make such big data collections available to a wide range of users (including the public), without eroding participants' trust in the research enterprise. For the past several decades, one of the hallmarks of privacy protection has been the "de-identification" of data; however, an increasing number of detective-like investigations have demonstrated how seemingly anonymous records can be linked back to the named individuals from whom the data was solicited. These demonstrations have rattled policy makers and data managers, leading to calls for new types of legislation, as well as new approaches to data privacy protections that are rooted in perturbation designed to meet statistically-rigorous definitions. Yet, in this seminar, I will illustrate how such attacks rely on adversarial frameworks that assume the recipients are armed with infinite motivation, capability, and lack regard for constitutional and contractual agreements. I will further show how models of reasonable (or rational) adversaries, with a focus on game theory, enable pragmatic privacy risk analysis in data sharing, such that substantial quantities of data can be disseminated with sufficiently minimized risk. This seminar will rely upon examples from my experience with helping establish the country's largest de-identified biorepository tied to an electronic medical record system (EMR) and directing a privacy research program for a NIH-sponsored consortia of academic medical centers conducting genomics research with EMR data.