Privacy Is An Increasing Illusion

By Chuck Dinerstein — August 30, 2018

Courtesy wokandapix [1]

Dr. Wells, my colleague and friend, and I frequently disagree about medical privacy issues. While we both agree that context matters, Jamie's context encompasses a more substantial number of situations than mine, I think medical privacy does not extend to the over-the-counter parts of the pharmacy, she does. We both agree on the "sanctity" of hospitals and offices. I mention this as my disclaimer because I think, given digital memory and a bit of statistical analysis, privacy is more of an illusion than a reality. The ability to re-identify data is rarely discussed and is the reason privacy is blurred.

De-identifying health data

The HIPAA privacy rule defines protected health information (PHI), information about an individual's physical and mental health, their care, and the payments for that care past, present and future. When the data is aggregating to look at trends or create predictive models, when the individual is lost in the summation healthcare data is not protected. Regulators have identified eighteen ways we commonly identify ourselves, which fall under the protection of HIPAA only when it is attached to our healthcare records.

- Names
- Biometric identifiers like finger and voice prints
- Full face photographic images
- Any geolocation smaller than your home state or the first three digits of your zip code, email addresses
- Any elements of dates other than the year, or your age when you're over 89.
- Phone and Fax numbers
- Health record, health plan or account numbers

- Certificate or license numbers
- Vehicle identifiers including serial numbers and license plates
- Device identifiers and serial numbers
- Web identifiers including URLs and Internet Protocol (IP) addresses
- Any other unique identifying number, characteristic or code.

Removing them from healthcare data de-identifies the healthcare record. Sort-of

Reidentifying data

Many of those identifiers are publicly available, remember without a health record attached they are fair game. We frequently share all of these identifiers with friends and as part of commerce, living in our digital world. One does not need to travel to the “Dark Web” or illicit sources to find this type of data aggregated into lists and available for free or for sale. Unicity, the uniqueness of our behavior, is a new term for me and quantifies the ability (or risk) in reidentifying a dataset. The greater the unicity, the easier it is to reidentify a specific individual within the data.

In what is now a classic study [1], researchers used de-identified credit card data for 1.1 million people, in 10,000 stores over a three-month period. Using just four pieces of “outside” data they could identify 90% of the shoppers. If the aggregation of the data is expanded, the categories are said to be more coarse. For example, rather than looking at each store as an individual site, we group them into Walmarts and Targets we make reidentification harder, but not impossible. The researchers found that coarser categories reduced unicity slowly and that reidentification only requires a few more outside data points. For this shopping data, knowing that your target was a woman increased the ability to identify them by 20% compared to men. Bottom line, given computational resources, and the plethora of our behavioral data which we freely provide through our transactions and phones, de-identification of data is often possible despite whatever safeguards we believe we have put in place.

The increasing calls for data transparency in science, healthcare, and commerce considered on its own merits is critical and essential; but when viewed in the context of an increasing ability to use unprotected information to unlock protected information it is more problematic. This context can be singularly focused when considering aggregated population data in the service of public health programs. I will leave the ending thought to the researchers,

> *“Finding the right balance between privacy and utility is absolutely crucial to realizing the great potential of metadata.”*

The first step in that process is understanding our new computational abilities so that we can understand why Dr. Wells and I both feel that the boundaries of HIPAA are blurring and that a civil discussion would be better than simply the shock and awe of “data-breaches” that are quickly forgotten.

[1] Unique in the shopping mall: On the reidentifiability of credit card metadata Science DOI:

10.1126/science.1256297

Source URL: https://www.acsh.org/news/2018/08/30/privacy-increasing-illusion-13247
Links
[1] https://pixabay.com/en/data-letters-scrabble-information-2355696/