The "Invisible Web" Undermines Health Information Privacy

Jalees Rehman

Correspondence to:

Jalees Rehman, M.D.
University of Illinois at Chicago, College of Medicine
Departments of Medicine and Pharmacology
835 South Wolcott Ave, Room E403, Chicago, IL 60612, USA

Email: jalees.rehman[at]gmail[dot]com

Scientific Blog: The Next Regeneration http://www.scilogs.com/next_regeneration/

Personal Blog: Fragments of Truth https://thenextregeneration.wordpress.com/

The "Invisible Web" Undermines Health Information Privacy

"The goal of privacy is not to protect some stable self from erosion but to create boundaries where this self can emerge, mutate, and stabilize. What matters here is the framework — or the procedure — rather than the outcome or the substance. Limits and constraints, in other words, can be productive — even if the entire conceit of 'the Internet' suggests otherwise."
—Evgeny Morozov in *To Save Everything, Click Here: The Folly of Technological Solutionism* (1)

We cherish privacy in health matters because our health has such a profound impact on how we interact with other humans. If you are diagnosed with an illness, it should be your right to decide when and with whom you share this piece of information. Perhaps you want to hold off on telling your loved ones because you are worried about how it might affect them. Maybe you do not want your employer to know about your diagnosis because it could get you fired. And if your bank finds out, they could deny you a mortgage loan. These and many other reasons have resulted in laws and regulations that protect our personal health information. Family members, employers, and insurances have no access to your health information unless you specifically authorize it. Even healthcare providers from two different medical institutions cannot share your medical information unless they can document your consent.

The recent study *Privacy Implications of Health Information Seeking on the Web* (2), conducted by Tim Libert at the Annenberg School for Communication (University of Pennsylvania), shows that we have a for more nonchalant attitude regarding health privacy when it comes to personal health information on the Internet. Libert analyzed 80,142 health-related webpages that users might come across while performing online searches for common diseases. For example, if a user uses Google to search for information on HIV, the Center for Disease Control and Prevention (CDC) webpage on HIV/AIDS (http://www.cdc.gov/hiv/) (3) is one of the top hits, and users will likely click on it. The information provided by the CDC will likely give solid advice based on scientific results, but Libert was more interested in investigating whether visits to the CDC website were being tracked. He found that by visiting the CDC website, information of the visit is relayed to third-party corporate entities such as Google, Facebook, and Twitter. The webpage contains "Share" or "Like" buttons, which is why the URL of the visited webpage (which contains the word "HIV") is passed on to them – even if the user does not explicitly click on the buttons.

Libert found that 91% of health-related pages relay the URL to third parties, often unbeknownst to the user, and in 70% of the cases, the URL contains sensitive information such as "HIV" or "cancer," which is sufficient to tip off these third parties that you have been searching for information related to a specific disease. Most users probably do not know that they are being tracked, which is why Libert refers to this form of tracking as the "Invisible Web" that can only be unveiled when analyzing the hidden http requests between the servers. Here are some of the most common (invisible) partners that participate in the third-party exchanges:

Entity	Percent of health-related pages
Google	78
Facebook	31
Twitter	18
Amazon	16
Experian	5

What do the third parties do with your data? We do not really know because the laws and regulations are rather fuzzy here. We do know that Google, Facebook, and Twitter primarily make money by advertising so they could potentially use your information and customize the ads you see. Just because you visited a page on breast cancer does not mean that the "Invisible Web" knows your name and address, but companies do know that you have some interest in breast cancer. It would make financial sense to send breast cancer-related ads your way: books about breast cancer, new herbal miracle cures for cancer, or even ads by pharmaceutical companies. It would be illegal for your physician to pass on your diagnosis or inquiry about breast cancer to an advertiser without your consent, but when it comes to the "Invisible Web," there is a continuous chatter going on in the background about your health interests without your knowledge.

Some users won't mind receiving targeted ads. "If I am interested in web pages related to breast cancer, I could benefit from a few book suggestions by Amazon," you might say. But, we do not know what else the information is being used for. The appearance of the data broker Experian on the third-party request list should serve as a red flag. Experian's (4) main source of revenue is not advertising but amassing personal data for reports such as credit reports that are then sold to clients. If Experian knows that you are checking out breast cancer pages, then you should not be surprised if this information will be stored in some personal data file about you.

How do we contain this sharing of personal health information? One obvious approach is to demand accountability from the third parties regarding the fate of your browsing history. We need laws that regulate how information can be used, whether it can be passed on to advertisers or data brokers, and how long the information is stored.

Here is the Privacy Policy Summary for WebMD (5), a commonly visited health information portal:

> We may use information we collect about you to:
> - Administer your account;
> - Provide you with access to particular tools and services;
> - Respond to your inquiries and send you administrative communications;

3

- Obtain your feedback on our sites and our offerings;
- Statistically analyze user behavior and activity;
- Provide you and people with similar demographic characteristics and interests with more relevant content and advertisements;
- Conduct research and measurement activities;
- Send you personalized emails or secure electronic messages pertaining to your health interests, including news, announcements, reminders and opportunities from WebMD; or
- Send you relevant offers and informational materials on behalf of our sponsors pertaining to your health interests.

Users are provided with instructions for how they can opt out of the tracking and receiving information from the (undisclosed) sponsors, but it is unlikely that the majority of users read the privacy policy pages of the various health-related websites. It is even less likely that users will go through the cumbersome process of requesting that all their information be kept private and not be passed on to corporate sponsors.

Perhaps one of the most effective solutions would be to make the "Invisible Web" more visible. If health-related pages were mandated to disclose all third-party requests in real-time, such as in pop-ups ("Information about your visit to this page is now being sent to Amazon") and ask for consent in each case, users would be far more aware of the threat to personal privacy posed by health-related pages. Such awareness of health privacy and potential threats to privacy are routinely addressed in the real world, and there is no reason why this awareness should not be extended to online information.

Acknowledgment:

An earlier version of this article was first published on May 25, 2015, at 3QuarksDaily.com.

References

1. Morozov, E. (2014). *To Save Everything, Click Here: The Folly of Technologoical Solutionaism.* PublicAffairs

2. Libert, T. (2015, March). Privacy implications of health information seeking on the web. *Communications of the ACM, 58*(3), 68-77. doi:10.1145/2658983

3. HIV/AIDS. (n.d.). Retrieved from http://www.cdc.gov/hiv/

4. Experian. (n.d.). In *Wikipedia.* Retrieved from https://en.wikipedia.org/wiki/Experian

5. Privacy Policy Summary. (n.d.). Retrieved from http://www.webmdhealthservices.com/privacy-policy/