

November 12, 2008

Dr. Eric Schmidt, CEO
Google Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043

Dear Dr. Schmidt,

We are writing to you regarding the Google Flu Trends initiative, a new web tool that may make it possible to detect flu outbreaks before they would otherwise be reported. This is an important use of user-generated data and could substantially support public health efforts in the United States and around the world.

At the same time, there is an obvious privacy concern. Search histories reveal personal information, and medical inquiries are particularly sensitive. A search for “flu symptoms” could as easily be a search for “AIDS symptoms” or “Ritalin” or “Paxil.”

In the aggregate, the data reveals useful trends and should be available for appropriate uses. But if disclosed and linked to a particular user, there could be adverse consequences for education, employment, insurance, and even travel. The disclosure of such information could also have a chilling effect on Internet users who may be reluctant to seek out important medical information online if they are concerned that their search histories will be revealed to others.

So, the question is how to ensure that Google Flu Trends and similar techniques will only produce aggregate data and will not open the door to user-specific investigations, which could be compelled, even over Google’s objection, by court order or Presidential authority. At present, the Google web site says only this about Privacy and Google Flu Trends:

Google Flu Trends can never be used to identify individual users because we rely on anonymized, aggregated counts of how often certain search queries occur each week.

Would you agree to publish the technique that Google has adopted to protect the privacy of search queries for Google Flu Trends? As you know, there is considerable debate as to what constitutes “anonymized” data, a point that was made clear in the recent dispute with Viacom regarding access to user data obtained by Google and is part of the ongoing debate about the status of IP addresses.

Please also consider that there is ample historical precedent for this concern. Census data, the quintessential form of aggregate data, was used during the Second

World War to identify and then displace Japanese Americans. The Department of Homeland Security sought information from the US Census about Muslim Americans in the United States after 9-11.

If Google has found a way to ensure that aggregate data cannot be reidentified, it should publish its results.

/s/

Marc Rotenberg, President
EPIC
Washington, DC
epic.org

/s/

Dr. Deborah Peel, President
Patient Privacy Rights
Austin, TX
patientprivacyrights.org