December 3, 2008

VIA FACSIMILE (404-639-7395)
AND EMAIL (FOIARequests@cdc.gov)
Lynn Armstrong
Attn: FOIA Office, MS-D54
1600 Clifton Road, N.E.
Atlanta, GA 30333

RE: Freedom of Information Act Request and Request for Expedited Processing

Dear Ms. Armstrong,

This letter constitutes a request to the U.S. Centers for Disease Control and Prevention ("CDC") under the Freedom of Information Act ("FOIA"), 5 U.S.C. § 552, and is submitted on behalf of the Electronic Privacy Information Center ("EPIC") for documents in the possession of the CDC regarding Google Flu Trends.

Background

On November 11, 2008, Google announced the launch of "Google Flu Trends" a Google utility for locating geographic areas where people are searching for the word "flu" and other medical terms.[1] Google Flu Trends is an extension of Google Trends, a technology that analyzes search queries submitted by Google users.[2]

The CDC is involved with the Google Flu Trends project. The search engine company stated that it "shared our preliminary results with the Epidemiology and Prevention Branch of the Influenza Division at CDC throughout the 2007-2008 flu season."[3] On November 19, 2008, Google and the CDC jointly published an academic paper concerning Flu Trends.[4]

---

[1] Jeremy Ginsberg and Matt Mohebbi, Google Tracking flu trends, Google Blog, Nov. 11, 2008, http://googleblog.blogspot.com/2008/11/tracking-flu-trends.html.
[2] *Id*; *see also* About Google Trends, http://www.google.com/intl/en/trends/about.html (last visited Dec. 2, 2008).
[3] Jeremy Ginsberg and Matt Mohebbi, Google Tracking flu trends, Google Blog, Nov. 11, 2008, http://googleblog.blogspot.com/2008/11/tracking-flu-trends.html.
[4] Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, & Larry Brilliant, *Detecting Influenza Epidemics Using Search Engine Query Data*, Nature, Nov. 19, 2008, *available at* http://www.nature.com/nature/journal/vaop/ncurrent/pdf/nature07634.pdf.

To implement Flu Trends, Google "compared [internet users'] aggregated [search] queries against data provided by the [CDC], and found that there's a very close relationship between the frequency of these search queries and the number of people who are experiencing flu-like symptoms each week."[5] Google alleges that "Google Flu Trends can never be used to identify individual users," but there are no clear legal or technological privacy safeguards to prevent the disclosure of individual search histories concerning the flu, or related medical concerns, such as "AIDS symptoms," "ritalin," or "Paxil." [6] Furthermore, the company refused to publish any techniques adopted to protect the privacy of search queries for Google Flu Trends, despites calls for disclosure form privacy advocates.[7] The CDC received data from Google pursuant to the Flu Trends project, but has not published any techniques adopted to protect the privacy of Flu Trends data.

Google Flu Trends analyzes user searches submitted to the search engine. Typically, Google collects several pieces of data before returning a search query result. This information includes six elements: 1) the search query itself; 2) the Internet Protocol (IP) Address of the searcher; 3) the date and time of the query; 4) the requested URL; 5) the browser and operating system being used; and 6) a unique cookie ID assigned to the browser.[8] In addition, Google can retain a unique account identifier that tracks a user's activity across different computers. It is possible to use a database containing user search data to sort by time and location, to locate and identify the source of search queries, and to build individual profiles. These pieces of information are stored on Google's servers, and users do not have the ability to control the data after it is submitted to the search engine.

For Flu Trends, Google uses current user search data, as well as years of historic user data, including data for "all weeks between September 28, 2003 and

---

[5] Ginsberg, *supra* note 1.

[6] Google Flu Trends | How does this work?, http://www.google.org/about/flutrends/how.html (last visited Dec. 2, 2008); *see also* Editorial, *Google Flu Trends: When the Government Knows You're Sick*, N.Y. Times, Nov. 12, 2008, *available at* http://theboard.blogs.nytimes.com/2008/11/12/google-flu-trends-when-the-government-knows-youre-sick/.

[7] Letter from EPIC and Patient Privacy Rights to Google CEO Dr. Eric Schmidt, Nov. 12, 2008, *available at* http://epic.org/privacy/flutrends/EPIC_ltr_FluTrends_11-08.pdf; e-mail from Mike Yang, Google Senior Product Counsel to EPIC and Patient Privacy Rights, Nov. 17, 2008, *available at* http://www.patientprivacyrights.org/site/DocServer/Google_s_response_to_EPICPPR.pdf.

[8] Privacy FAQ – Google Privacy Center, http://www.google.com/privacy_faq.html#serverlogs (last visited Dec. 2, 2008).

March 11, 2007."[9] The search data is used to generate estimates of flu activity on a state by state basis. But Google says that Flu Trends could be used to provide data on smaller groups of users, which could increase the likelihood that individuals will be identified and linked to medical searches. Flu Trends "may be capable of providing [flu] estimates for large cities and metropolitan areas with high internet penetration, providing even more local influenza surveillance. We hope to explore this topic as well," Google said.[10]

Technical experts have questioned the efficacy of the "anonymization" techniques Google uses for search engine data.[11] Even absent IP Address data, cookies, or unique account identifiers, individual search histories can often be easily matched to users.[12] Medical data is particularly resistant to "anonymization."[13]

Documents Requested

EPIC requests copies of the following agency records in the possession of the Centers for Disease Control and Prevention:

1. all records regarding data received by the CDC from Google concerning Google Flu Trends;

2. all records regarding steps taken by Google and/or the CDC to identify individuals whose Internet searches were used as part of Google Flu Trends;

3. all records regarding steps taken by Google and/or the CDC to conceal the identity of individuals whose Internet searches were used as part of Google Flu Trends;

4. all communications between the CDC and Google regarding the identification or de-identification of influenza patients;

---

[9] Ginsberg, *supra* note 4, at 3.
[10] Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, & Larry Brilliant, *Detecting Influenza Epidemics Using Search Engine Query Data* (draft) at 9, *available at* www.google.org/about/flutrends/manuscript.pdf.
[11] Chris Soghoian, Debunking Google's log anonymization propaganda, Cnet, September 11, 2008 http://news.cnet.com/8301-13739_3-10038963-46.html.
[12] Malin, B., Sweeney, L. and Newton, E., *Trail Re-identification: Learning Who You are From Where You Have Been*, Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory Technical Report, LIDAP-WP12, February 2003.
[13] Malin, B. and Sweeney, L., *Re-Identification of DNA through an Automated Linkage Process*, Journal of the American Medical Informatics Association. Washington, DC: Hanley & Belfus, Inc., Nov 2001; 423-427.

Request for Expedited Processing

This request warrants expedited processing because it is made by "a person primarily engaged in disseminating information …" and it pertains to a matter about which there is an "urgency to inform the public about an actual or alleged federal government activity." 5 U.S.C. § 552(a)(6)(E)(v)(II).

EPIC is "primarily engaged in disseminating information." *American Civil Liberties Union v. Department of Justice*, 321 F. Supp. 2d 24, 29 n.5 (D.D.C. 2004).

Moreover, there is particular urgency for the public to obtain information about the CDC's involvement in Google Flu Trends. Flu Trends uses sensitive medical data that could be linked to individual users. No clear legal or technological privacy safeguards prevent the disclosure of individual search histories concerning the flu, or related medical concerns. Neither Google nor the CDC has described measures adopted to safeguard the privacy of users' medical search data. Google has provided data to the CDC, and continues to do so. The public should be informed of the CDC's ongoing role in Google Flu Trends.

Request for "News Media" Status

EPIC is a non-profit, educational organization that routinely and systematically disseminates information to the public. EPIC is a representative of the news media. *EPIC v. Dep't of Defense*, 241 F.Supp. 2d 5 (D.D.C. 2003).

Based on our status as a "news media" requester, we are entitled to receive the requested records with only duplication fees assessed. Further, because disclosure of this information will "contribute significantly to public understanding of the operations or activities of the government," as described above, any duplication fees should be waived.

Thank you for your consideration of this request. As provided in 5 U.S.C. § 552(a)(6)(E)(ii)(I). I will anticipate your determination on our request for expedited processing with ten (10) calendar days.

Respectfully submitted,

_____

John A. Verdi
Director, EPIC Open Government Project