Big Data & Google - not *for* us, but *from* us

DATA - What journalists need to know about private surveillance

Albrecht Ude [Journalist | Researcher | Research Trainer]

#dataharvest13 – Brussels – May 4th, 2013

Big Data & Google - not *for* us, but *from* us

what Google knows about us,

how this data tracking works,

and what to do against.

Albrecht Ude – to introduce myself

www.ude.de .

a very old homepage.

eine-woche-ohne.de .

An invitation not to use any Google-services for a week

(showing alternatives)

recherche-standards.wikispaces.com .

A set of standards of online research.

Google …

We all know Google,

used to call it a search engine.

Which is missleading,

as Google Inc. is world's biggest advertising company

and to keep this running,

it is world's biggest private data collector.

Google …

Just to know, the most frequented sites on WWW:

- ✓ 1:   google.com,
- ✓ 3:   Youtube (belonging to Google Inc.)
- ✓ 12:   Blogspot.com (ditto),
- ✓ 14:   Google India,
- ✓ 19:   Google Japan,
- ✓ 23:   Google Germany.

Source: Alexa.com "Top Sites - The top 500 sites on the web" (global)
http://www.alexa.com/topsites .
[as by Feb. 05th, 2013]

Google …

Among the 20 top sites on the web,
5 belong to Google, including the very top.

I.e. in Germany, Google has over 90% of the web search market,
so is a monopolist.

Google …

Google Inc. has just been 'decorated'
with the German **BigBrotherAward** (April 12th)

www.bigbrotherawards.de/2013/.hoard .

(Category "Global Data Hoarding"),
as the company is presumably the biggest private data collector

http://www.bigbrotherawards.de/2013-en/.hoard .

Google …

Eric Schmidt (Google CEO 2010)

cited in WSJ:

"I actually think most people don't want Google to answer their questions," he elaborates. *"They want Google to tell them what they should be doing next."*

In: Google and the Search for the Future. The Web icon's CEO on the mobile computing revolution, the future of newspapers, and privacy in the digital age.
By Holman W. Jenkins jr. - Wall Street Journal, The Weekend Interview, 14.08.2010
http://online.wsj.com/article/SB10001424052748704901104575423294099527212. html .

Google …

To anticipate / advise / order someone what to do next,

obviously, you need some information …personal information.

How does Google get these?

Where does Google obtain data?

✓ Google services which need an account
  (Mail, Docs, Google+)

✓ Google websites free to use
  (Websearch, News, Maps etc.)

✓ Other websites using Google services
  (Analytics, AdSense, DoubleClick)

✓ Devices by Google
  (Chrome, Android)

… and others like street view, wifi scanning, in-game advertising.

Where does Google obtain data?

Some sources are collecting data in depth,

others are like fishing with a dragnet.

Users of websites delivering data to Google Analytics or tracking advertisements normally don't realise Google is tracking *them*.

How does Google obtain data?

✓by http

✓by "browser fingerprint"

✓by cookies

✓by the user itself

✓by user's interaction

✓by search parameters

Where does Google obtain data?

by http

- ✓ Date
- ✓ Time
- ✓ IP adress
- ✓ Referer (webpage you've been on before clicking)
- ✓ Browser type & version
- ✓ Operation system & version
- ✓ Client's hostname and preferable mail server
  (often contaning domain names)

http://tools-on.net/ .

Where does Google obtain data?

by "browser fingerprint"

- ✓User Agent
- ✓HTTP_ACCEPT Headers
- ✓Browser Plugins installed (normaly many = highly individual)
- ✓Time Zone
- ✓Screen Size and Color Depth
- ✓System Fonts (normaly many = highly individual)
- ✓Cookies Enabled y/n
- ✓Supercookies [limited test]

https://panopticlick.eff.org/ .

Where does Google obtain data?

by cookies

✓ Services used

✓ Users preferences (to keep them)

Cookies are used to connect single visits into chain of visits ► a profile.

By cookies, Google can establish the user's history (of search, usage, behavoir, services used …)

The longest time of living of a Google-cookie on my computer:
Sunday, January 17th, 2038, 20:14:05

Where does Google obtain data?

by the user itself

many kinds of personal information
- ✓ Name,
- ✓ Adress,
- ✓ Date of birth,
- ✓ Photograph,
- ✓ Relations to other users,
- ✓ … .

Where does Google obtain data?

by user's interaction

- ✓time needed to type in a query (incl. typos)
- ✓time spent on the search result page
- ✓time between selecting different results for the same query
- ✓average amount of search requests per day
- ✓average amount of search requests per topic (to finish search)
- ✓usage of enter key or mouse to send a search request
- ✓usage of keyboard shortcuts to navigate through search results
- ✓usage of advanced search commands (how often)

[incomplete, see here for more:
http://www.rene-pickhardt.de/google-uses-57-signals-to-filter/ .

Where does Google obtain data?

by search parameters

These parameters are contained in the Google URL.
I.e. in "search redirect", when you think to click an external link

Looking for "pizza", you see (even on mouseover):
www.pizza.be/ .

And this is the link really:
**http://www.google.be/url?sa=t&rct=j&q=pizza&source=web&cd=8
&cad=rja&ved=0CGYQFjAH&url=http%3A%2F%2Fwww.pizza.be%2F&ei=
7SmEUeSyNo_G7Aa5-
oCoAg&usg=AFQjCNFdjO3rVKs_PUaAmVvZvks7KWSwSg&bvm=bv.4596008
7,d.ZWU** .

Where does Google obtain data?

by search parameters - in detail:

```
http://www.google.be/url?
sa      =    t&
rct     =    j&
q       =    pizza&
source=      web&
cd      =    8&
cad     =    rja&
ved     =    0CGYQFjAH&
url     =    http%3A%2F%2Fwww.pizza.be%2F&
ei      =    7SmEUeSyNo_G7Aa5-oCoAg&
usg     =    AFQjCNFdjO3rVKs_PUaAmVvZvks7KWSwSg&
bvm     =    bv.45960087,d.ZWU .
```

Where does Google obtain data?

by search parameters - in detail:

Some of these are easy to understand,

in case of others, usage and encoding remain unclear.

Yet clear is: they contain information on *you*!

What does this mean?

The German journalist Sebastian Jabbusch
used the Google Dashboard to review the data collected about himself:

"My private data retention"
(Meine private Vorratsdatenspeicherung)

http://jabbusch.tumblr.com/post/17208499220/meine-private-vorratsdatenspeicherung .

What to do against?

Use other search engines

- ✓ Bing & Yahoo (same Index, other input masks),
- ✓ Yandex & DuckDuckGo
- ✓ Ask & Exalead (rather small)
- ✓ Ixquick (metasearcher, not collecting your data)

If you need Google results

- ✓ try Startpage (metasearcher by Ixquick, not collecting your data)

What to do against?

Use browser addons to protect your data
https://addons.mozilla.org/ .

Especially helpfull against tracking are
  ✓Ghostery

  ✓NoScript

Block out Adwords, Adsense, Analytics and Doubleclick.

A good recommendation by the German Privacy Foundation:
http://www.privacyfoundation.de/service/firefox_erweiterungen/ .

What to do against?

Manage your cookies
Configure your browser to delete cookies after each session

Use other services
- ✓Look for other mail providers, map services, cloud storage ...
- ✓Hands of Google Chrome
- ✓Free your Android

    http://www.freeyourandroid.com/ .

Thank you for your patience!

Albrecht Ude

albrecht@ude.de

www.ude.de

https://privacybox.de/aude.msg