Googling Data Protection: Don't be Evil

ABU BAKAR MUNIR* AND TEH TAI YONG†

LT [Keywords to Follow]

Introduction

The Opinion on Data Protection Issues Related to Search Engines (WP148) (the Opinion)[1] adopted by Art.29 Data Protection Working Party (Working Party)[2] on April 4, 2008 is probably the biggest blow to search engines in relation to data protection issues. The search engine company that is affected the most is Google Inc, a US-based corporation which has business across Europe and other parts of the globe.

Before the Working Party issued the Opinion, it wrote to Google on May 16, 2007[3] requesting Google to clarify several issues, i.e. legal justification for the storage of server logs and the storage period chosen, to what extent the anonymised data still contain significant information about the internet user, whether the anonymisation is reversible, etc. In this letter the Working Party highlighted the Resolution on Privacy Protection and Search Engines[4] adopted in London on November 2–3, 2006 by the 28th International Data Protection and Privacy Commissioners' Conference, which resolved, inter alia, that providers of search engines should offer their services in a privacy-friendly manner.[5]

Google responded to the Working Party's query via its letter by Peter Fleischer, Global Privacy Counsel, dated June 10, 2007,[6] stating Google's commitment to raise the bar on privacy practices for the benefit of its users and providing the justifications for data retention. Nearly 10 months after Google responded to the Working Party's letter, the Opinion was issued. The Common Position on Privacy Protection and Search Engines[7] adopted on April 15, 1998 and revised on April 6–7, 2006 by the International Working Group on Data Protection in Telecommunications and the Resolution on Privacy Protection and Search Engines adopted in London on November 2–3, 2006 by the 28th International Data Protection and Privacy Commissioners' Conference were reiterated by the Working Party.

The Opinion is a clear manifestation that there are grounds of disagreement between the Working Party and the search engine providers in relation to data protection issues. The Working Party highlighted some issues to be resolved by the industry, namely the retention period, further processing for different purposes, cookies, anonymisation and data correlation across services.

This article examines why Google seems to be the ''prime target'' when the Working Party raises the data protection issues with search engine providers. It analyses the grounds of disagreement between the Working Party and the search engine providers; whether an IP address falls under the definition of personal data and hence is subject to data protection laws; how long the retention period ought to be before personal data must be deleted or anonymised in an irreversible and efficient way, etc. In concluding, the article provides a brief review of the privacy policy adopted by Google, which has come under scrutiny in recent months.

Why Google?

Before we delve further into the issues relating to data protection, one may wonder why Google seems to be the prime target by the Working Party in this data protection saga. On this issue, Paczkowski asked sarcastically, ''Why haven't we heard anything about the Working Party's letters to those two companies? Insufficient postage for airmail?''[8] Not surprisingly, the Working Party's effort has been criticised as being Google-centric.[9] Recognising the criticisms, the Working Party decided to expand its examination to other major search engine providers, such as Yahoo and Microsoft.

* Professor of Law, Faculty of Law, University of Malaya Kuala Lumpur, Malaysia.
† Advocate and Solicitor, Klang, Malaysia.

1 Available at *http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_en.pdf* [Accessed August 13, 2008].

2 The Article 29 Working Party is an independent EU advisory body on data protection and privacy established pursuant to Art.29 of Directive 95/46. It is entrusted with the tasks laid down in Art.30 of Directive 95/46 and in Art.14 of Directive 97/66.

AQ1

3 Available at *http://ec.europa.eu/justice_home/fsj/privacy/news/docs/pr_google_16_05_07_en.pdf* [Accessed August 13, 2008].

4 Available •at *http://www.privacyconference2006.co.uk/index.asp?PageID=3*.

5 The 28th International Data Protection and Privacy Commissioners' Conference resolved that:

> ''1. Among other things, providers of search engines should inform users upfront in a transparent way about the processing of data in the course of using their services.
> 2. In view of the sensitivity of the traces users leave when using a search engine, providers of search engines should offer their services in a privacy-friendly manner. More specifically, they shall not record any information about the search that can be linked to users or about the search engine users themselves. After the end of a search session, no data that can be linked to an individual user should be kept stored unless the user has given his explicit, informed consent to have data necessary to provide a service stored (e.g. for use in future searches).
> 3. In any case, data minimization is key. Such a practice would also be beneficial for the providers of search engines in simplifying arrangements for meeting demands for user-specific information from third parties.''

6 Available at *http://64.233.179.110/blog_resources/Google_response_Working_Party_06_2007.pdf* [Accessed August 13, 2008].

7 Available • at *http://www.datensch utz-berlin.de/attachments/238/search_engines_en.pdf?1178095352*.

AQ2

8 John Paczkowski, ''European Data Protection Officials: Yahoo and Microsoft Have Search Engines?'', AllThingsDigital, June 12, 2007, available at *http://digitaldaily.allthingsd.com/20070612/google-eu-privacy/#comments* [Accessed August 13, 2008].

9 See discussion in Barry Schwartz, ''Google, Yahoo, Microsoft, & Other Search Engines Must Comply with EU Privacy Rules'', Search Engine Land, February 22, 2008, available at *http://searchengineland.com/080222-083116.php* [Accessed August 13, 2008].

Peter Hustinx, the European Data Protection Supervisor, stated that ''The use of the Internet the way Google is doing it could introduce tremendous privacy problems''.[10] However, Danny Sullivan argued:

> ''Both Google and Yahoo have 30 year cookies. So where's the letter for Yahoo from the Working Group? And isn't 14 years from Microsoft excessive? But if Google's going to get called out, why aren't the others?''[11]

It is submitted that there are two reasons why Google is perceived to be the target of the Working Party, compared to other search engine providers. The first is Google's massive market share in search engine business in Europe. According to the survey conducted in Europe by comScore Inc, Google sites dominated 79.2 per cent of the European search properties in March 2008.[12] Other search properties fall far below Google in the said survey, with eBay ranked as second (3.1 per cent), Yandex ranked as third (2.2 per cent), Yahoo! Sites ranked as fourth (2.0 per cent) and Microsoft Sites ranked as fifth (1.9 per cent).[13] Google dominated the searches in the European countries and its share of searches in January 2008 was as follows: Portugal (94 per cent); Spain (93 per cent); Switzerland (93 per cent); Finland (92 per cent); Belgium (92 per cent); Denmark (92 per cent); Austria (88 per cent); Italy (84 per cent); Netherlands (84 per cent); France (83 per cent); Norway (81 per cent); Sweden (80 per cent); Germany (80 per cent); Ireland (76 per cent); the United Kingdom (73 per cent).[14] Google's success as the leading search engine provider in the Europe is evident from the statistics.

Analysing the figures from the global perspective, Google topped the ranking in the survey search property conducted in August 2007 for 61 billion searches done worldwide. Google Sites conquered 37 billion of the searches (60.7 per cent), whereas Yahoo! Sites had 8.5 billion searches (14.0 per cent) and Microsoft Sites had 2.1 billion searches (3.5 per cent).[15] In the United States, Google also topped the ranking for search providers.[16] The survey conducted by Nielsen Company showed that Google Search ranked the first as top search provider in May 2008 with a 59.3 per cent share of searches, followed by Yahoo! Search (16.9 per cent) and MSN/Windows Live Search (13.3 per cent).[17] The other search providers (AOL, Ask.com, My Web, Comcast, AT&T WorldNet, Nex Tag, Dogpile.com) only managed to captured less than 4 per cent of the share of searches individually.[18]

Google Inc's net income of US$3.077 billion for the year 2006 increased to US$4.203 billion for the year 2007, i.e. an increase of 36.59 per cent in one year.[19] For the first quarter of 2008, the NASDAQ board company recorded a US$1.307 billion net income, compared to US$1 billion in the first quarter of 2007 and US$592.3 million for the first quarter of 2006—evidence that the weakening economy has not affected Google's business.[20]

The other reason is perhaps the rather ''lack of engagement'' approach taken by other search engine providers in dealing with the Working Party. Since the issue of data protection came under the limelight, Microsoft and Yahoo have not given much public response to the Working Party. Microsoft issued the following statement:

> ''Microsoft has a long-term commitment to providing customers with control over the collection, use and disclosure of their personal information. While we have not received formal communication from the Article 29 Working Party, we recognize that online search is creating legitimate concerns about privacy and are actively engaged with data protection authorities around the world to ensure that our practices meet the highest standards when it comes to protecting privacy''.

Meanwhile Yahoo issued a statement that:

> ''Our users' trust is one of Yahoo's most valuable assets. That's why maintaining that trust and protecting our users' privacy is paramount to us. Our data retention practices vary according to the diverse nature of our services.''[21]

Besides the above statements, the other search engine providers generally do not make many public statements regarding the issue of data protection raised by the Working Party.[22]

Engagement, actively on the part of Google, inactively by other search engines, was reflected in the Working Party's letter to Google dated May 16, 2007 where it stated:

10 See the discussion in Paul Meller, ''EU Data Protection Group Questions Other Search Engines'', InfoWorld, June 21, 2007, available at *http://www.infoworld.com/article/07/06/21/EU-questions-other-search-engines_1.html* [Accessed August 13, 2008].

11 Danny Sullivan, ''Google Responds to EU: Cutting Raw Log Retention Time; Reconsidering Cookie Expiration'', Search Engine Land, June 12, 2007, available at *http://searchengineland.com/070612-041042.php* [Accessed August 13, 2008].

12 http://www.comscore.com/press/release.asp?press=2208 [Accessed August 13, 2008].

13 *http://www.comscore.com/press/release.asp?press=2208* [Accessed August 13, 2008].

14 *http://www.techcrunch.com/2008/03/18/the-web-in-charts%e2%80%94google-vs-microsoft-yahoo-vs-china/* [Accessed August 13, 2008].

15 *http://www.comscore.com/press/release.asp?press=1802* [Accessed August 13, 2008].

16 According to Nielsen Company (an internet media and market research company), Google Search has maintained its position as the top search provider in the US: May 2008 (59.3% share of searches); April 2008 (62.0% share of searches); March 2008 (58.7% share of searches); February 2008 (58.7% share of searches); January (56.9% share of searches). See *http://www.nielsen-netratings.com/press.jsp?section=new_pr&theyear=2008&country=United%20States&themonth=5* [Accessed August 13, 2008].

17 Available at *http://www.nielsen-netratings.com/pr/pr_080619V.pdf* [Accessed August 13, 2008].

18 Available at *http://www.nielsen-netratings.com/pr/pr_080619V.pdf* [Accessed August 13, 2008].

19 *http://finance.google.com/finance?q=NASDAQ:GOOG* [Accessed August 13, 2008].

20 See Miguel Helft, ''Google Defies the Economy and Shows Profit Surge'', April 18, 2008, *New York Times*, April 18, 2008, at *http://www.nytimes.com/2008/04/18/technology/18google.html* [Accessed August 13, 2008]. Also see Miguel Helft, ''Profits Up 69% at Google, Exceeding Expectations'', *New York Times*, April 20, 2007, at *http://www.nytimes.com/2007/04/20/business/20google.html* [Accessed August 13, 2008].

21 See the discussion in Barry Schwartz, ''European Union to Question Data Retention Policies of Other Search Engines'', Search Engine Land, June 21, 2007, at *http://searchengineland.com/070621-144447.php* [Accessed August 13, 2008].

22 The website of Yahoo! Pressroom, *http://yhoo.client.shareholder.com/press/* [Accessed August 13, 2008], does not contain much information about the ongoing discussion about data protection raised by the Working Party in Europe. On the other hand, the website of Microsoft PressPass, *http://www.microsoft.com/presspass/default.mspx* [Accessed August 13, 2008], contains the statement of Thomas Myrup Kristensen, EU Internet Policy Director, Microsoft Europe before the Committee on Civil Liberties, Justice, and Home Affairs, European Parliament, ''Data Protection on the Internet (Google-DoubleClick and other case studies)'' on January 21, 2008, which addressed the issue of online advertising.

> ''The Article 29 Working Party continues to appreciate Google's ongoing engagement with the data protection community on a range of issues and in particular its readiness to consult with it in contrast with a relative lack of engagement by some of the other leading players in the search engine community.''

With the massive market share in Europe and also worldwide, and with the ''lack of engagement'' approach taken by other search engine providers, it is not surprising that Google seems to be the main target of the regulatory body. The reason is obvious—implementation or changes to be executed would be more effective as Google is the market leader. The others would naturally adopt the ''wait and see'' approach and await Google's reaction.

Are IP addresses personal data?

The application of the Data Protection Directive (Directive 95/46)[23] for search engine providers having business operations in Europe is not disputed by Google. By using the argument that the processing is carried out in the context of the activities of an *establishment* of the controller and that the processing makes *use of equipment* on the territory of the Member State, the Working Party in the Opinion stated:

> ''The combined effect of Articles 4(1)(a)[24] and 4(1)(c)[25] of the Data Protection Directive is that its provisions apply to the processing of personal data by search engine providers in many cases, even when their headquarters are outside the EEA.''

On this issue, Peter Fleischer responded:

> ''Google is a U.S. company and we respect U.S. laws—but we are also a global company, doing business across Europe and across the world, and we recognize the need to respect the laws of the countries in which we do business. We are therefore committed to data protection principles that meet the expectations of our users in Europe and across the globe.''[26]

Therefore, the real issue is whether internet protocol addresses (IP addresses) and cookies fall under the definition of ''personal data''. Article 2(a) of the Data Protection Directive states that:

> ''. . . '[P]ersonal data' shall mean any information relating to an *identified or identifiable* natural person ('data subject'); *an identifiable person* is one who can be identified, *directly or indirectly*, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity'' (emphasis added).

The existence of the second alternative of *identifiable* or *may be identified* is also seen in the data protection laws of various European countries, as follows:

> 1. Germany: s.3(1) of the Federal Data Protection Act[27] defines personal data to mean any information concerning the personal or material circumstances of an *identified or identifiable* individual;
> 2. Sweden: s.3 of the Personal Data Act 1998[28] defines personal data to mean all kinds of information that *directly or indirectly may be referable* to a natural person who is alive);
> 3. Netherlands: art.1(a) of the Personal Data Protection Act[29] defines personal data to mean any information relating to an *identified or identifiable* natural person;
> 4. Finland: s.3(1) of the Personal Data Act 523/1999[30] defines personal data to mean any information on a private individual and any information on his/her personal characteristics or personal circumstances, where these are *identifiable* as concerning him/her or the members of his/her family or household); and
> 5. Spain: art.3(a) of the Organic Law 15/1999 of December 13 on the Protection of Personal Data[31] defines personal data to mean any information concerning *identified or identifiable* natural persons).

It is worth noting that the definition of ''personal data'' under the UK Data Protection Act 1988[32] is stricter compared with the Data Protection Directive and the laws in the above European countries, where data has to be identified. Section 1(1) of the UK Data Protection Act 1988 states, inter alia, that:

> ''. . . '[P]ersonal data' means data which relate to a living individual who *can be identified*—
>
> (a) from those data, or
> (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,
>
> and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual'' (emphasis added).

On this issue, the Working Party is of the view that IP addresses fall under the definition of personal data. It stated in the Opinion that:

> ''An individual's search history is personal data if the individual to which it relates, is identifiable. Though IP addresses in most cases are not directly identifiable by search engines, identification can be achieved by a third party. Internet access providers hold IP address data. Law enforcement and national security authorities can gain access to these data and in some Member States private parties have gained access also

23 At *http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML* [Accessed August 13, 2008].
24 Data Protection Directive Art 4(1)(a) states that ''the processing is carried out in the context of the activities of an *establishment* of the controller on the territory of the Member State; when the same controller is established on the territory of several Member States, he must take the necessary measures to ensure that each of these establishments complies with the obligations laid down by the national law applicable''.
25 Data Protection Directive Art.4(1)(c) states that ''the controller is not established on Community territory and, for purposes of processing personal data makes use of equipment, automated or otherwise, situated on the territory of the said Member State, unless such equipment is used only for purposes of transit through the territory of the Community''.
26 Google's letter dated June 10, 2007 responding to the Working Party's query, available at *http://64.233.179.110/blog_resources/Google_response_Working_Party_06_2007.pdf* [Accessed August 13, 2008].
27 Available at *http://www.bdd.de/Download/bdsg_eng.pdf* [Accessed August 13, 2008].
28 Available at *http://www.sweden.gov.se/content/1/c6/01/55/42/b451922d.pdf* [Accessed August 13, 2008].
29 Available at *http://www.dutchdpa.nl/downloads_wetten/wbp.pdf?refer=true&theme=purple* [Accessed August 13, 2008].
30 Available at *http://www.tietosuoja.fi/uploads/hopxtyf.HTM* [Accessed August 13, 2008].
31 Available at *https://www.agpd.es/upload/Ley%20Org%E1nica%2015-99_ingles.pdf* [Accessed August 13, 2008].
32 Available at *http://www.opsi.gov.uk/Acts/Acts1998/ukpga_19980029_en_1* [Accessed August 13, 2008].

> through civil litigation. Thus, in most cases—including cases with dynamic IP address allocation—the necessary data will be available to identify the user(s) of the IP address.''

The Working Party also referred to the Opinion 4/2007 on the Concept of Personal Data (WP 136)[33] it adopted earlier, which states:

> ''. . . [U]nless the Internet Service Provider is in a position to distinguish with absolute certainty that the data correspond to users that cannot be identified, it will have to treat all IP information as personal data, to be on the safe side.''

Google, on the other hand, argues that ''A black-and-white declaration that all IP addresses are always personal data incorrectly suggests that every IP address can be associated with a specific individual''.[34] It further argues:

> ''In some contexts this is more true: if you're an ISP and you assign an IP address to a computer that connects under a particular subscriber's account, and you know the name and address of the person who holds that account, then that IP address is more like personal data, even though multiple people could still be using it. On the other hand, the IP addresses recorded by every website on the planet without additional information should not be considered personal data, because these websites usually cannot identify the human beings behind these number strings.''[35]

Peter Fleischer reiterated that whether or not an IP address is personal data depends on how the data is being used.[36]

Alma Whitten, Google's software engineer, gave an illustration of technical workings of IP addresses, which is useful for the discussion on whether IP addresses are personal data:

> ''An IP address is an address for a computer on the Internet, which exists to allow data to be delivered to that computer. When you enter a website's name—like *http://www.google.com*—that is actually a handy shortcut for the website's IP address—right now, one of Google's is *http://72.14.207.99/*. So when a website needs to send your computer something (for instance, your Google search results), it needs your IP address to send it to the right computer.
> The situation gets a bit more complex, though, because the IP addresses that people use can change frequently. For instance, your Internet service provider (ISP) may have a block of 20,000 IP addresses and 40,000 customers. Since not everyone is connected at the same time, the ISP assigns a different IP address to each computer that connects, and reassigns it when they disconnect (the actual system is a bit more complex, but this is representative of how it works). Most ISPs and businesses use a variation of this 'dynamic' type of assigning IP addresses, for the simple reason that it allows them to optimize their resources.
> Because of this, the IP address assigned to your computer one day may get assigned to several other computers before a week has passed. If you, like me, have a laptop that you use at work, at home, and at your corner café, you are changing IP addresses constantly. And if you share your computer or even just your connection to your ISP with your family, then multiple people are sharing one IP address.''

The authors are of the view that IP addresses are personal data. First, the Data Protection Directive consists of the secondary alternative of *identifiable*. This would mean that the person need not be identified by the information. So long as there is a possibility that the person could be identified, the information is sufficient to be personal data. Although IP addresses changes every time the users log on to the internet, it is possible for the person to be identified if time and place is provided. If it is a personal computer, it is easy to identify the person as the computer is usually accessed by the log-in of username and password. If it is a computer accessible by the public (for example in an internet café or public library), it is not impossible to identify a particular person at particular time as usage could be granted after registration with the operator or librarian.

Secondly, the requirement is that a person can be *identified directly or indirectly*. The person can be identified by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity. This provision should be read together with Recital 26 of the Data Protection Directive which states, inter alia, that account should be taken of *all the means* likely reasonably to be used either by the controller or by any other person to identify the said person. The test is very wide as *all the means* of identification should be taken into account. Hence it is a matter of tracing before the person can be identified by way of an IP address.

Thirdly, the principles of protection should not apply only when data is anonymised in an irreversible and efficient way. In such a case, the person is no longer identifiable because the ''link'' to the person is destroyed and can never be found again. Therefore IP addresses are personal data whenever they are anonymised in an irreversible and efficient way●.

AQ3

Hence it is submitted that the technical workings furnished by Google and the argument that in some context IP addresses *could* be used to trace a particular user in fact strengthen the argument by the Working Party. Therefore the only way to consider IP addresses not to be personal data is when the IP addresses are anonymised in an irreversible and efficient way. Otherwise, IP addresses should be considered as personal data and the principle of data protection should apply to search engine providers.

It should be noted that search engine providers also process and retain various data of internet users, such as query logs (content of the search queries, the date and time, source (IP address and cookies), the preferences of the user, and data relating to the user's computer); data on the content offered (links and advertisements as a result of each query); and data on the subsequent user navigation (clicks). Search engines may also process operational data relating to user data, data on registered users and data from other services and sources such as email, desktop searches and advertising on third-party websites.[37] As long as the information processed falls under the definition of ''personal data'', the search engine providers have to observe and comply with the principles under the Data Protection Directive.

33 Available *at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf* [Accessed August 13, 2008].

34 Alma Whitten, ''Are IP addresses personal?'', Google Public Policy Blog, February 22, 2008, at *http://googlepublicpolicy.blogspot.com/2008/02/are-ip-addresses-personal.html* [Accessed August 13, 2008].

35 Whitten, ''Are IP addresses personal?'', Google Public Policy Blog, February 22, 2008, at *http://googlepublicpolicy.blogspot.com/2008/02/are-ip-addresses-personal.html* [Accessed August 13, 2008].

36 Peter Fleischer, ''The European Commission's data protection findings'', Google Public PolicyBlog, April 7, 2008, at *http://googlepublicpolicy.blogspot.com/search?q=The+European+Commission%27s+data+protection+findings* [Accessed August 13, 2008].

37 See the Opinion adopted by the Working Party (WP148), available at *http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_en.pdf* [Accessed August 13, 2008].

Retention period of personal data

Retention of personal data remains one of the unresolved issues between the Working Party and the search engine providers. The Working Party in the Opinion stated:

> ''In view of the initial explanations given by search engine providers on the possible purposes for collecting personal data, the Working Party does not see a basis for a retention period beyond 6 months.''[38]

The six-month retention period is very much shorter compared to the current practice of the search engine providers. Google and MSN anonymise user data after 18 months, while Yahoo does the same after 13 months.[39] Darren Waters stated that the recommendation (the Opinion by the Working Party) is likely to be accepted by the European Commission and could lead to a clash with search giants like Google, Yahoo and MSN.[40]

On the issue of retention of user data, Google has reduced the previously established period of 18 to 24 months to 18 months as a response to the Working Party's concern. Google in its letter to the Working Party on June 10, 2007 stated:

> ''. . . [W]e also *firmly reject* any suggestions that we could meet our legitimate interests in security, innovation and anti-fraud efforts with any retention period shorter than 18 months.''[41]

However, after the Working Party adopted the Opinion on April 4, 2008, Google did not seem to embark on the ''firm rejection'' approach, but instead elects for a more diplomatic and friendly approach by stating that:

> ''The findings are another important step in an ongoing dialogue about protecting user privacy online—a discussion in which Google will continue to be engaged. It's also a debate in which we hope our users will participate.''[42]

Would there ever be common ground that might enable the Working Party and the search engine provider to agree on the retention period? To answer this, it is pertinent to look at the justifications furnished by the search engine provider and the response by the Working Party. Google publishes the justifications for retention of data on its Official Blog. The following is an extract from a post by Peter Fleischer, Google's global privacy counsel, on May 11, 2007[43] stating why Google remember information about searches (similar grounds were cited as justifications for retention of data in Google's replies to the Working Party's query in its letter dated June 10, 2007):

> ''i) Improve our services: Search companies like Google are constantly trying to improve the quality of their search services. Analyzing logs data is an important tool to help our engineers refine search quality and build helpful new services. Take the example of Google Spell Checker. Google's spell checking software automatically looks at your query and checks to see if you are using the most common version of a word's spelling. If it calculates that you're likely to generate more relevant search results with an alternative spelling, it will ask 'Did you mean: (more common spelling)?' We can offer this service by looking at spelling corrections that people do or do not click on. Similarly, with logs, we can improve our search results: if we know that people are clicking on the #1 result we're doing something right, and if they're hitting next page or reformulating their query, we're doing something wrong. The ability of a search company to continue to improve its services is essential, and represents a normal and expected use of such data.
>
> ii) Maintain security and prevent fraud and abuse: It is standard among Internet companies to retain server logs with IP addresses as one of an array of tools to protect the system from security attacks. For example, our computers can analyze logging patterns in order to identify, investigate and defend against malicious access and exploitation attempts. Data protection laws around the world require Internet companies to maintain adequate security measures to protect the personal data of their users. Immediate deletion of IP addresses from our logs would make our systems more vulnerable to security attacks, putting the personal data of our users at greater risk. Historical logs information can also be a useful tool to help us detect and prevent phishing, scripting attacks, and spam, including query click spam and ads click spam.
>
> iii) Comply with legal obligations to retain data: Search companies like Google are also subject to laws that sometimes conflict with data protection regulations, like data retention for law enforcement purposes. For example, Google may be subject to the EU Data Retention Directive, which was passed last year, in the wake of the Madrid and London terrorist bombings, to help law enforcement in the investigation and prosecution of 'serious crime'. The Directive requires all EU Member States to pass data retention laws by 2009 with retention for periods between 6 and 24 months. Since these laws do not yet exist,

38 WP148, available at *http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_en.pdf* [Accessed August 13, 2008].

39 Darren Waters, ''Search Engines Warned Over Data'', BBC News, April 7, 2008, at *http://news.bbc.co.uk/2/hi/technology/7335359.stm* [Accessed August 13, 2008].

40 Waters, ''Search Engines Warned Over Data'', BBC News, April 7, 2008, at *http://news.bbc.co.uk/2/hi/technology/7335359.stm* [Accessed August 13, 2008].

41 Google's letter dated June 10, 2007 responding to the Working Party's query, available at *http://64.233.179.110/blog_resources/Google_response_Working_Party_06_2007.pdf* [Accessed August 13, 2008].

42 Fleischer, ''The European Commission's data protection findings'', Google Public Policy Blog, April 7, 2008, at *http://googlepublicpolicy.blogspot.com/search?q=The+European+Commission%27s+data+protection+findings* [Accessed August 13, 2008]. Peter Fleischer reiterated Google's position by stating that: ''While the working party has welcomed our decision to anonymise data logs after 18 months as a positive privacy protective step, it suggested in findings released today that this period might still be too long. We believe that data retention requirements have to take into account the need to provide quality products and services for users, like accurate search results, as well as system security and integrity concerns. We have recently discussed some of the many ways that using this data helps improve users' experience, from making our products safe, to preventing fraud, to building language models to improve search results. This perspective—the ways in which data is used to improve consumers' experience on the web—is unfortunately sometimes lacking in discussions about online privacy.''

43 Peter Fleischer, ''Why does Google remember information about searches?'', The OfficialGoogleBlog, May 11, 2007, at *http://googleblog.blogspot.com/2007/05/why-does-google-remember-information.html* [Accessed August 13, 2008]. See also Peter Fleischer, ''How long should Google remember searches?'', The OfficialGoogleBlog, June 11, 2007, at *http://googleblog.blogspot.com/2007/06/how-long-should-google-remember.html* [Accessed August 13, 2008], which stated the following grounds as *legitimate interest* in retaining server logs:

> ''i) to improve our search algorithms for the benefit of users;
>
> ii) to defend our systems from malicious access and exploitation attempts;
>
> iii) to maintain the integrity of our systems by fighting click fraud and web spam;
>
> iv) to protect our users from threats like spam and phishing;
>
> v) to respond to valid legal orders from law enforcement as they investigate and prosecute serious crimes like child exploitation; and
>
> vi) to comply with data retention legal obligations.''

> and are only now being proposed and debated, it is too early to know the final retention time periods, the jurisdictional impact, and the scope of applicability. It's therefore too early to state whether such laws would apply to particular Google services, and if so, which ones. In the U.S., the Department of Justice and others have similarly called for 24-month data retention laws.''

The Working Party in the Opinion sums up the grounds provided by the search engine providers, namely (1) improving the service; (2) securing the system; (3) fraud prevention; (4) accounting requirements; (5) personalised advertising; (6) statistics; (7) law enforcement and legal requests. In dealing with these issues, the Working Party has categorically addressed each of the grounds provided by the search engine providers. The Working Party does not agree with the grounds of improving the service, accounting requirements and personalised advertising. It stated that service improvement cannot be considered to be a legitimate reason for storing data that has not been anonymised and it seriously doubted that personal data were really essential for accounting purposes. It also stated that for personalised advertising, the search engine providers must adhere to Data Protection Directive and the Working Party has clear preference for anonymised data.[44]

As for system security, the Working Party is of the view that personal data stored must be subject to strict purpose limitation. For fraud prevention, the length of time will depend on whether the data are indeed necessary. The search engine providers must comply with law enforcement and legal requests, but compliance should not be mistaken for a legal obligation or justification for storing such data solely for these purposes.[45]

Instead, the Working Party offered the view that there are three grounds which search engine providers may appeal to for different purposes, namely Art.7(a) (consent); Art.7(b) (performance of a contract); and Art.7(f) (legitimate interest) of the Data Protection Directive.[46]

The authors share the view of the Working Party. First, grounds such as improving the service, accounting requirements and personalised advertising are not criteria for legitimate processing under Art.7 of the Data Protection Directive. The search engine providers frequently claim that longer retention of data is essential to improve their service.[47] This seems to be the main reason for the search engine providers to justify longer retention. Although the users would benefit from better service, there is no strong reason for the search engine providers to retain data up to 18 months. With the fast-moving pace of society now, six months of search history would be sufficient. To an ordinary user, what was searched by a user 18 months ago may not have much relevance today.

Secondly, the search engine providers must comply with the principles under Art.6 of the Data Protection Directive.[48] Article 6.1(a) of the Data Protection Directive provides that personal data must be ''processed fairly and lawfully'' and Art.6.1(b) of the Directive provides, inter alia, that personal data must be ''collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes''. When a user conducts a search using the search engines, the purpose of providing the information to the search engine providers is to enable it to link to the sites the user wishes to visit. The user does not furnish the information so that the search engine providers can retain the same for 18 months in order to analyse his browsing behaviour or characteristics. Furthermore, Art.6.1(e) of the Data Protection Directive specifically provides that personal data must be ''kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed''. Very often, cookies are sent to the users' computer without their knowledge in order to trace the users' browsing history and pattern. These cookies could remain in the users' computer up to 30 years if not deleted by the users.

Thirdly, if the ground given is for security, prevention of fraud, law enforcement or legal requests, the purpose for which the data is retained must be strictly adhered to. The data retained under these purposes shall not be used by the search engine providers to improve their service or personalised advertising. That would be using different grounds to justify retention of the data generally, while the data is used for another purpose. Hence, if the search engine wishes to retain data for security, prevention of fraud, law enforcement or legal requests, they must establish a proper system with safeguards for retention to ensure such data is not used for

44 See the Opinion adopted by the Working Party (WP148), at *http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_en.pdf* [Accessed August 13, 2008].

45 See the Opinion adopted by the Working Party (WP148), at *http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_en.pdf* [Accessed August 13, 2008].

46 Data Protection Directive Art.7:

> ''Member States shall provide that personal data may be processed only if:
>
> (a) the data subject has unambiguously given his consent; or
>
> (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; or
>
> . . .
>
> (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1).''

47 Fleischer, ''The European Commission's data protection findings'', Google Public Policy Blog, April 7, 2008, at *http://googlepublicpolicy.blogspot.com/search?q=The+European+Commission%27s+data+protection+findings* [Accessed August 13, 2008]. Peter Fleischer stated: ''Today, a Google search is far more likely to provide you with the information you're looking for than it did a few years ago. This has not happened by accident. It is the result of our engineers painstakingly analysing the patterns in our server logs to improve the relevance of our searches.''

48 Data Protection Directive Art.6 states:

> ''Member States shall provide that personal data must be:
>
> (a) processed fairly and lawfully;
>
> (b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards;
>
> (c) adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed;
>
> (d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified;
>
> (e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Member States shall lay down appropriate safeguards for personal data stored for longer periods for historical, statistical or scientific use.''

other purposes. Google's example that the US Department of Justice called for 24-month data retention laws[49] is acceptable provided that there is a proper system in place to ensure data retained on such a ground is not misused for another purpose.

Fourthly, Directive 2006/24[50] (Data Retention Directive) and Directive 2002/58[51] (Directive on Privacy and Electronic Communications) have no application to search engine providers. Article 6 of the Data Retention Directive provides for period of retention of not less than six months and not more than two years. However, the Data Retention Directive (Art.1) and Directive on Privacy and Electronic Communications (Art.3) apply only to publicly available ''electronic communications services'' or ''public communications networks''. The definition of the said phrases is stated in Directive 2002/21[52] (Framework Directive).[53] As highlighted by the Working Party in the Opinion, Art.2(c) of the Framework Directive defines ''electronic communications service'' as meaning:

> ''. . . a service normally provided for remuneration which consists wholly or mainly in the conveyance of signals on electronic communications networks, including telecommunications services and transmission services in networks used for broadcasting, *but exclude services providing, or exercising editorial control over*, content transmitted using electronic communications networks and services; it does not include information society services, as defined in Article 1 of Directive 98/34/EC, which do not consist wholly or mainly in the conveyance of signals on electronic communications networks''.

The said definition explicitly excludes services providing or exercising editorial control over content. Search engine providers are not an electronic communications service and not subject to the Data Retention Directive. Therefore the retention periods stated in Art.6 of the Data Retention Directive do not apply to search engine providers. The retention period of data by search engine providers has to be justified by legitimate grounds.

Fifthly, a shorter data retention period would increase the efficiency of data management of the search engine providers. Analysis for statistics or use for accounting would have to be completed within the timeframe. After the allowed retention period, the search engine providers still could retain the information. The difference is that the personal data has to be anonymised, i.e. a person cannot be identified or identifiable from the information.

Sixthly, search engine providers should consider the criteria provided under Art.7 of the Data Protection Directive. The consent of the data subject must be obtained if the search engine providers wish to have a longer retention period. However, the consent must be unambiguously given by the data subject. Besides the issue of consent by the data subject, search engine providers could also consider other criteria such as performance of contract, compliance with legal obligation, protection of the vital interests of the data subject, public interest, legitimate interests, etc.[54]

Google's privacy policy: a brief review

At this juncture, it would be useful to conduct a review on Google's privacy policy available on its websites[55] against the Data Protection Directive. The privacy policy is divided into several sections.

Information we collect and how we use it

Generally, information is collected and processed not only for addressing the query by the users, but also to provide products and services to users (including the display of customised content and advertising); auditing, research and analysis in order to maintain, protect and improve their services; ensuring the technical functioning of the network; and developing new services.

This may contravene Art.6.1(b) and 6.1(c) of the Data Protection Directive, which require data to be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes, and adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed.

Cookies are sent to the users' computer and log information is recorded automatically without express consent by the users. This may contravenes Art.7(a) of the Data Protection Directive, which requires unambiguous consent on the part of the users. Users may elects to refuse all cookies but some Google features and services may not function properly if cookies are disabled. In this case, users are left with no choice but to accept cookies in order to subscribe to the features and services. The real issue is therefore the period of retention of cookies.

The period of retention of the information collected is not mentioned in the policy. This may contravene Art.6.1(e) of the Data Protection Directive, which requires the data to be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed.

Choices for personal information

Google will only seek consent from the users if it uses the information ''in a manner different than the purpose for which it was collected''. Article 7(a) of the Data Protection Directive requires personal data to be processed only if the user has given his unambiguous consent.

Google will not collect or use sensitive information for purposes other than those described in the policy and/or in the specific service notices, unless it has obtained the users' prior consent. That would mean that it can collect or use sensitive information without prior consent of the users if the purposes are described in the policy and/or in the specific service notices. Article 8 of the Data Protection Directive provides that processing of sensitive information is prohibited unless the user has given his explicit consent.

49 Fleischer, ''Why does Google remember information about searches?'' May 11, 2007, The OfficialGoogleBlog, at *http://googleblog.blogspot.com/2007/05/why-does-google-remember-information.html* [Accessed August 13, 2008].

50 See *http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:105:0054:0063:EN:PDF* [Accessed August 13, 2008].

51 See *http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:201:0037:0047:EN:PDF* [Accessed August 13, 2008].

52 See *http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:108:0033:0050:EN:PDF* [Accessed August 13, 2008].

53 Data Retention Directive Art.2 and Directive on Privacy and Electronic Communications Art.2 respectively state that the definitions in the Framework Directive shall apply.

54 Refer to Data Protection Directive Art.7.

55 Available at *http://www.google.com.my/intl/en/privacypolicy.html* [Accessed August 13, 2008].

Information sharing

Opt-in consent would only be obtained for sharing of sensitive personal information. The preferred position is that opt-in consent should be obtained in all cases if Google wishes to share the personal information (whether sensitive information or otherwise) with other party.

The policy provides that information may be shared with the subsidiaries, affiliated companies or other trusted businesses or persons for purpose of processing on Google's behalf based on its instructions and compliance with the policy any other appropriate confidentiality and security measures. Article 17.3 of the Data Protection Directive stipulates that the carrying out of processing by way of a processor must be governed by a *contract* or *legal act* binding the processor to the controller. Such issues are not addressed in the policy.

Information security

Google shall take appropriate security measures to protect against unauthorised access to or unauthorised alteration, disclosure or destruction of data. Article 17.1 of the Data Protection Directive provides that the controller must implement appropriate technical and organisational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing. The responsibilities imposed by the Data Protection Directive is higher, the controller must also protect the personal data against *accidental loss* and *unlawful forms of processing*.

Accessing and updating personal information

Users are provided with access to personal information and it is possible either to correct this data if it is inaccurate or to delete such data at the data user's request if it is not otherwise required to be retained by law or for legitimate business purposes. It is noted that such right of rectification is subject to the law of data retention or legitimate business purposes. Such a right must not be qualified by ''legitimate business purposes'', which is very wide and vague. Such qualification may be in contravention with Art.12(b) of the Data Protection Directive.

Concluding remarks

The Working Party makes several firm stands in the Opinion, namely anonymisation of personal data once they are no longer necessary for the purpose for which they are collected; reduction of the data retention period to not more than six months; cookies' lifespan should not be longer than demonstrably necessary, etc. Google has a massive search engine business in Europe. Its market share in the search engine business is unrivalled by any other companies globally, especially in that region. Understandably, Google wishes to keep open dialogue and good working relations with the Working Party because of the Working Party's influence in policy decision-making at the European Commission level.

Balancing business efficacy for better services and protecting privacy and data protection is not an easy task. Issues of privacy and data protection are getting more attention from lawmakers, businesses and consumers. Search engine providers must accept and respect such reality. Consumers judge the search engine providers' commitment to privacy and data protection through their privacy policy. The Opinion adopted by the Working Party is supported by legal grounds after careful consideration of the applicable laws. Search engine providers ought to digest the Opinion and comply with the rules and regulations, in Europe and elsewhere.[56] It will be interesting to know how Google deals with the latest response by the Working Party. Let us wait and see. It is crucially important is for Google to honour its Hippocratic oath for corporations or corporate motto, ''Don't Be Evil''.

56 The principles in the Data Protection Directive have been adopted in many jurisdictions.

THOMSON

SWEET & MAXWELL

Author: Please take time to read the below queries marked as AQ and mark your corrections and answers to these queries directly onto the proofs at the relevant place. DO NOT mark your corrections on this query sheet:

AQ1: I was unable to access the web page in the footnote—can you check the address or provide the date you last accessed it, please?

AQ2: I was unable to access the web page in the footnote—can you check the address or provide the date you last accessed it, please?

AQ3: This seems to contradict what is said in the next paragraph: ''Therefore the only way to consider IP addresses not to be personal data is when the IP addresses are anonymised in an irreversible and efficient way.'' Can you clarify, please?