AGENCIA ESPAÑOLA DE PROTECCIÓN DE DATOS

Statement on Internet Search Engines

Spanish Data Protection Agency

1st December 2007

1. The importance of search engines in information society

Technology developments have opened up new possibilities of creating and accessing information on the Internet, and this situation requires that we consider the repercussions of technology on the rights of individuals—in principle, such repercussions are neutral.

The Spanish Data Protection Agency (AEPD) has taken the initiative of analysing the privacy policies of the largest global companies providing search engine services on the Internet. To this end it collected information from Google, Microsoft and Yahoo!, said information being completed via meetings with the global privacy controllers of these corporations in Europe and the United States.

We must underscore the huge importance, owing to the volume of data that is processed and the characteristics of that processing, of search engine services on the Internet (hereinafter, search engines), the main function of which is to provide lists of results relating to a search, such as addresses and files stored on web serves, by entering key words, thus arranging all of the information available on the Internet and making it more accessible. In addition, search engines usually provide customised services, allowing people to register by giving an email address and a password.[1]

In Spain, the fact that these are sensitive issues was shown by the recent appearance of the Director of the AEPD before the Constitutional Commission of the Congress on 28th November 2007, where some of the conclusions on the rendering of search engine services were disclosed.

Speaking before the Commission, the Director of the AEPD stated that the development of search engines, as well as that of other technologies, "puts in check the traditional criteria regarding the guarantee of privacy and requires an urgent update and adaptation." [2]

The acknowledgment of new rights guaranteeing that citizens may benefit from the access, use and advantages of technology has become a part of the present-day political and social agenda.

Nevertheless, along with this, there is an intensifying demand of the necessary balance such as to afford guarantees against the new risks posed by Internet services such as search engines and electronic mail.

[1] In a number of reports, the AEPD has considered the IP address as personal data.
[2] Intervention of Artemi Rallo, Director of the AEPD, in his appearance before the Constitutional Commission of Congress on 28th November 2007.

There is also an ever-increasing awareness of the need of establishing international standards defining a consensus of rules to guarantee privacy on the Internet. The AEPD is committed to this task.

In effect, Internet search engines process personal data that are either furnished by users themselves or derived from the use of the service, as well as those obtained from the processing and publication of personal information drawn from other websites in its search engine function.

The success of free search engines depends on the service provider's ability to furnish users with the most relevant search results. Hence, the importance of the arrangement of the websites that appear associated to the search and the linking of it to sponsored advertising messages, which is the usual business model of these companies.

These information society services are subject not only to the guarantees of the Organic Act 15/1999, on Data Protection (LOPD), but also to those of Act 34/2002 on information society services and electronic commerce (LSSI), which comprises within this concept *"the provision of search tools or links to other Internet sites."*

2. Personal data processed by search engines

In rendering these search engine services a wide range of personal data are processed (IP addresses, logs, data furnished by the users themselves, etc.). In addition, profiles can be made regarding user habits and the information is kept by the companies over long periods.

Within these services, it is possible to distinguish between the processing of user data and the processing of third-party data accessible via search engines.

Regarding the processing of user data, the most important of these is the processing of the so-called 'server logs', which can be defined as files containing information on the user such as search parameters, the IP address allocated to the user's computer, the date and time of the search and the cookie identifier, as well as the user's preferences, the characteristics of the surfer, the result of the search, the advertising that has been displayed with the outcome of a specific search and the user's clicks.

The data stored in cookies are also processed. These cookies are sometimes installed by the search engine and stored in a user's computer. Although the contents of these cookies may vary, they usually include information on the user's operating system, browser and the identification thereof. These cookies are used to improve service quality by analysing user trends and preferences. It should be noted that users may configure their browser so that it does not

accept cookies. In addition, it must be stressed that the life of cookies may last up to 30 years.

Regarding the processing of third-party data obtained on the Internet, special attention should be attached to the information stored in the cachet. Sometimes it is possible to access the textual contents of a page even though said page is no longer accessible on the Internet, because, although certain contents of same have been deleted, these are not always automatically deleted on the cachet.

3. Other services at the disposal of users

Occasionally, search engines require that users register themselves to enjoy other services aside from the search engine service, such as electronic mail, personal websites, historical search records and other services of the so-called Web 2.0 (for instance, web logs). Thus, there could be a record of the activities carried out by a user on the web, allowing user profiles to be made which could be used by a company, and it might happen that the user is not aware or sufficiently informed of that circumstance.

Last of all, it is possible for Internet search engines to use third-party services to collect personal information on users, such as information provided by affiliates.

Among customised services, it is important to mention the electronic mail service. Electronic mail services using websites as an interface are known as 'web mail' services (for instance, Yahoo! Mail, Hotmail, Gmail, etc.). Users can access their web mail from any location and they do not need to connect to any given Internet service provider in the way they have to when using a normal email account. It is usually free of charge, but to obtain their account users often have to disclose personal data to the service provider. Web mail uses the http protocol instead of POP and SMTP to receive and send email. In fact, messages are displayed on a classic HTML page.

This feature allows mail service providers to incorporate customised advertising on the HTML page on which the message is displayed (graphically, apart from the message itself). Web mail is largely dependent on sponsors and it displays a large amount of advertising.

Also, since web mail systems are based on the HTTP protocol, they may be vulnerable to so-called 'web bugs', which can discover the identity of a person's electronic mail by means of cookies and embedded HTML labels.

Another problem posed by electronic mail is the filtering or scanning of its contents. Normally it is filtered to prevent the entry of viruses, unwanted mail (spam), to detect specific contents or to list related advertising.

4. The legitimacy behind using and conserving personal data of users and third parties

4.1 The purposes contended by service providers

According to the information furnished by search engine services, the main purposes associated to the obtainment, use and conservation of personal data are the following:

A) Allowing search engine services to be rendered and enhancing them in view of user experience

The success of these services depends on the capacity to furnish users with the most relevant search results, incorporating an increasing number of information sources and data of a varying nature (for instance, images).

For such purpose, one of the most relevant circumstances affecting user experience is the seasonality of their habits, which requires that the processing of information spans a minimum period of one year and one month of information for analysis purposes.

B) Guaranteeing service security

Search engine services are exposed to computer attacks that can affect both the service provider and the users.

Among these are massive attacks aimed at causing a system overload and the use of malicious codes.

The increasing levels of sophistication and the innovations in these attacks make it necessary to have the appropriate means and time to ward them off.

C) Detecting other fraudulent activities

There is also the possibility of third parties fraudulently trying to alter the order of the search results in order to increase the expenses borne by their competitors or the payments to be made by their advertisers, by increasing the number of hits on their website via the sponsored advertisement or result.

The blockage of these fraudulent activities is likewise linked to seasonality, allowing time trends to be discovered in connection with these behaviours.

D) Conservation of information at the disposal of relevant authorities

The information obtained by search engine service providers may be requested by the authorities in order to investigate illegal activities, in conformity with the powers with which they are vested.

In order to comply with these possible requests, service providers are of the understanding that the conservation of the data they use is justified.

E) Financial and auditing procedures.

Search engine services are largely financed by means of associated advertising.

In order to adequately comply with their obligations and develop the financial and auditing processes, the necessary data must be conserved over periods that cannot be below 13 months.

4.2 The legal grounds for using the data

As described above, search engines process information that, according to the LOPD, is considered personal data, and their services are also regulated by the LSSI.

In consequence, said information processing must comply with the system of guarantees that both of these Acts establish for the benefit of the citizens and the recipients of information society services.

The above makes it necessary, first of all, to analyse the legitimacy, that is, the legal grounds justifying the processing of data by Internet search engines.

In this respect, it should be borne in mind that the access to and the use of search engine services is generally done voluntarily and free of charge.

The efficiency that the users of these services expect, along with the fact that the decision of using them is voluntary, allows us to consider that, in principle, the processing of user data is justified by the legal relationship, understood in an ample sense, which is established between users and the service provider. Said legal relationship would justify the processing of personal data pursuant to Art. 6.2 of the LOPD.

Additionally, in cases where the users of services voluntarily choose to register as users of customised services, the legitimacy for processing their data would be specifically grounded on the consent given by the registered users.

From what has been set out, it can be concluded that the legitimacy of processing user data affects the purposes that are directly related to the search engine service, namely the rendering and enhancement of the service, security,

the detection of fraud and invoicing. However, any use of personal data must be proportionate, in every case, to the purpose justifying that use.

On the other hand, a justification based on the need to conserve information to have it at the disposal of the relevant authorities requires a few explanations. Nothing can be objected to the fact that a relevant authority may ask a service provider for information that is necessary for meeting the legal requirements of a country abiding by the rule of law.

However, this does not mean that the service provider may legitimately conserve data over long periods, beyond what is necessary for rendering services, just in case in the future said information might be requested by the relevant authorities.

In Spain the obligation of retaining and conserving data for investigation purposes is based on specific laws (particularly Act 25/2007, dated 18th October, on the conservation of data relating to electronic communications and public communication networks), but search engine service providers, which are information society services, are not among those subject to said regulations. Because of this, the grounds that are advocated to conserve data must be rejected.

4.3 Legitimacy for processing third-party data

Search engines also allow information to be obtained on third parties that are not necessarily voluntary users of the services.

This situation takes place in cases where the information on those persons is available on websites, which makes it possible for search engines to capture and facilitate access to it via the listing they furnish as a part of their services.

In these cases, the legitimacy for disseminating personal information on the Internet should be required, first of all, of those who make it accessible on the web.

The lawfulness or otherwise of this dissemination of information may be grounded on varying circumstances, for instance the exercise of the fundamental rights linked to freedom of information; the compliance with legal obligations requiring dissemination of the information, including dissemination by electronic means, the unequivocal consent of those involved, and other circumstances.

Insofar as Internet search engines basically just list the information, the legitimacy for processing it would lie, in principle, with those enabling access to that information.

But there are two circumstances that must not be neglected:

- Internet search engines carry out information processing of their own, distinct from that of the websites to which they facilitate access;

- The legitimacy for processing the personal data of those making personal information accessible cannot be extrapolated in all cases to the search engines giving access to that information.

As far as Internet search engines are concerned, the legitimacy of those rendering this service can initially be found in the LSSI.

As suggested by the Stated Purpose of the LSSI, this regulation purports to establish an appropriate legal framework such as to generate in all the players involved on the Internet "the necessary confidence for using this new medium."

The provision of instruments for searching, accessing and gathering data or links to other Internet sites is called an intermediation service in the LSSI. This Act acknowledges a legitimate interest in rendering the service, initially excluding any liability on the part of the provider for the information they supply to the recipients of their services. However, it imposes upon them the obligation of collaborating "to prevent certain unlawful services or contents from continuing to be disseminated," which is something that can happen when the rights recognised by the LOPD are harmed.

5. Real, efficient information for users

The processing of personal data not only requires grounds legitimating it—it goes in hand with a number of additional guarantees that are supplementary to the lawful use of said data.

According to the LOPD, the legitimate processing of personal data by Internet search engines is subject to a pre-condition both when it is based on the existence of a legal relationship and when it is based on the consent of those who voluntarily register in order to use those services: the individuals whose data are being processed must be informed of what data are going to be used, by whom, and what purpose and to whom their data may be given.

And in the event (something usual in the rendering of these services) that devices installed in users' computers are used, allowing information to be obtained on how the computers are used, the LSSI requires that users be aware of that circumstance and, if they want to (even though it may entail the risk of fewer benefits when using search engine services), they shall be able to deactivate them in a simple manner, free of charge.

This information requirement is of essential importance in the use of search engine services on the Internet, because of the following reasons, among others:

- First of all, because the efficiency of search engine services is so high that users, largely unaware of the processing of personal data that is involved, tend to play down its importance;

- Secondly, because although search engine services have detailed privacy policies in relation to the use of the personal data of users, they could very well be considered as something virtual or fictional.

This inefficiency is due to a number of circumstances: on the one hand, search engine service providers do not sufficiently stress their own privacy policies on their homepages. On their homepages and on the information that is accessible via an icon (hyperlink), they do not even highlight the most relevant consequences of the use of personal data, particularly, they do not stress the possibility and consequences of deactivating the devices installed in computers via which information is obtained.

On the other hand, because the information on privacy policies is so complex and unintelligible that users, even though they may suspect the possible risks involved in the use of their data, are unlikely to obtain conclusions regarding those risks when accessing that information.

Some of these service providers contend, as an excuse for these information deficits, that it would hinder access to the search engine homepages and that it would 'obscure' the rendering of the service.

Other Internet search engines underscore in a clearer fashion the access to their privacy policies, despite the risk that this might diminish the efficiency of their business development.

The circumstances described allow it to be concluded that the users of search engine services do not have, in a clear and accessible manner, information on the consequences that using those services may have regarding their personal data.

This conclusion is particularly relevant in respect of the purpose for which the information is used, the use of devices installed in their computers, and the possible making of patterns of their habits based on the queries they make.

A number of consequences arise from this conclusion: the legitimacy for the processing of personal data conducted by Internet search engines may be questionable due to the information deficits that have been set out; because of this, it is necessary to urgently consider new information mechanisms allowing

users to effectively know about the use of their personal data; and there is the need to bring together the various search engine privacy policies so that, while allowing the search engine services to be rendered, they minimise consequences for user privacy.

6. Limiting the conservation of personal data and making them anonymous

A description has been given above of the purposes justifying the processing of data by search engine service providers; those purposes may also render legitimate the conservation of the information.

However, this legitimacy must be qualified; because once the information is no longer necessary for said purposes it must be cancelled. Likewise, once the purposes justifying the use of the data may be attained without identifying a specific user, said data must be made anonymous in such a way that the information that is conserved is no longer linked to specific users.

The rendering anonymous of the data must be irreversible in order to prevent the link from being re-established at a later stage (in this respect, it would be acceptable to partially eliminate the IP address, whereas replacing same and the cookie by a single identifier would not guarantee that outcome).

On the other hand, search engine service providers do not all coincide in terms of establishing the period during which said data are to be conserved in order to meet those purposes, ranging from 13 to 18 months.

7. Guaranteeing user and third-party rights

Although the processing of user data may be legitimate considering the needs inherent to search engine services, this does not exclude the fact that they must guarantee, at the data subject's request, the exercise of the rights of access, rectification, cancellation and opposition that users are acknowledged by the LOPD.

However, in the field of search engine services it is particularly relevant to protect the rights of the people whose data may be accessed as a result of the searches that are conducted.

These people may not be users of the search engine service but, despite this circumstance, their data may be known by anyone as a result of the capturing of the information available on the Internet by the search engines and the listings that they offer users to facilitate a selective access to information.

Although the initial incorporation of this personal information on the web "may be legitimate at source, its universal and secular conservation on the Internet may be disproportionate." [3]

People must have at their disposal reaction instruments in order to avoid, on their own initiative, to be subject to a global exhibition.

The LOPD is meant to provide citizens with those reaction instruments, mainly by way of the rights of cancelling data and opposing the processing of same, the purpose of these rights being to prevent the processing from taking place or to have it stopped.

Meanwhile the LSSI defines search engine services and intermediation services in which the service provider is not responsible, in principle, for the information contents to which access is provided, although said provider must delete those contents or prevent access when requested by a relevant body questioning the lawfulness thereof.

In this way, the LSSI has reinforced the possibility of exercising these rights, allowing the AEPD to set forth those requirements.

In respect of search engine services, the exercise of these rights involves a number of peculiarities.

Insofar as the activity of search engines focuses mainly on associating the terms of the search to the websites on which said information appears, the exercise of the rights of cancellation and opposition should be associated to a correlative exercise of those rights against those responsible for those websites, being the persons who originally allow access to the personal information.

Otherwise, if the search engine cancels the data or prevents access to same, a subsequent check within the search engine would again allow access to the information.

However, there are occasions when those responsible for a website may be under the obligation of maintaining the information (for instance, because it is required by law), but this same obligation does not apply to those responsible for the search engine service.

In these cases, the former must maintain the information on the Internet, whereas the latter may not be subject to an equivalent requirement.

[3] Intervention by Artemi Rallo, Director of the AEPD, in his appearance before the Constitutional Commission of Congress on 28th November 2007.

Here is where citizens' right to oppose the processing of their data may be considered a virtual right if there is a legitimate and grounded reason [for such opposition] in relation to their specific personal situation.

The person responsible for a website may be prevented, owing to a legal requirement, from processing the data. But this obligation may not be imposable on the person responsible for the search engine, who must take measures not only to stop processing the information but also to prevent future access to the information via its service.

The AEPD has been defining, via a number of decisions,[4] criteria for protecting the right of cancellation of the information available on the Internet and, specifically, the appropriateness of the right of opposition in respect of search engine services.[5]

8. Electronic mail services

One of the discussions carried out with the search engines providing email services has to do with the filtering of their contents. What has been found from the meetings with the companies that AEPD has spoken to is that these companies scan emails for the following purposes:

- **To prevent viruses**, checking whether the attachments contain known viruses. Most service providers include this antivirus control in order to protect their systems and those of their users;

- To weed out spam, which may hinder the capacity to provide a good service;

- **To detect a specific content** such as illegal material and withdraw same;

- **To include customised advertising** depending on the contents of the message. Some entities automatically filter the contents of emails in order to offer products related to same as a way of financing their service.

The Article 29 Working Party (WP 29), in which AEPD participates, has ruled on this matter in its Opinion 2/2006 (WP 118), analysing the cases where search engines may scan the contents of emails and establishing that communications are only to be filtered in order to prevent viruses and spam, in order to adopt the adequate technical and management means to preserve the security of their services, as established in Article 4 of Directive 2002/58/EC on privacy in telecommunications, which is transposed in Article 34 of Act 32/2003, the Spanish General Telecommunications Act.

[4] TD/00266/2007 and TD/00299/2007
[5] TD/00463/2007

Even so, service providers must inform about their email screening practices and offer their subscribers the possibility of deciding on filtering to prevent spam.

Aside from these cases in which emails are scanned due to service security, the Spanish Data Protection Agency considers that the interception of contents is not in conformity with Spanish Law.

CONCLUSIONS

1. Internet search engines are information society services subject to the guarantees of the Organic Law on Data Protection (LOPD), as well as to those of the Information Society Services and Electronic Mail Act (LSSI).

2. Internet search engines process and retain large volumes of data on the users to whom they offer their services.

3. The processing of said information may allow the activities carried out by a user on the web to be registered, thus making it possible to establish user profiles that can be used by a company without the user being aware or sufficiently informed of that circumstance.

4. There are significant differences in the privacy policies of the various search engines (in aspects such as the criteria for retaining personal data and information policies), which need to be brought together and unified in order that they guarantee user privacy and minimise the risks for that privacy.

5. The information included in search engine privacy policies on the use of the personal data of users is inefficient, it is not sufficiently highlighted, and there are serious doubts as to whether it is understandable for the general community of Internet users.

6. There is an urgent need to develop new information mechanisms that are clear and sufficiently visible, allowing users to effectively know that their data will be used when they use search engine services.

7. It is necessary to limit the use and conservation of personal data. Once the information is no longer necessary for the purposes of the service, it must be cancelled.

Likewise, from the moment when the purposes justifying the use of the data may be attained without identifying a specific user, the data must be made anonymous so that the information that is conserved cannot be linked to specific users.

8. Search engine services are under the obligation of allowing the rights of cancellation and opposition to be exercised by individuals whose data are listed on other websites in their search engine function.

Although the initial incorporation of this personal information on the web may be legitimate at source, the universal maintenance of said information on the Internet may be disproportionate.

9. Communications may only be filtered for preventing viruses and spam, in order to preserve service security. Aside from these cases, the interception of contents is considered not in conformity with Spanish law.

10. It is necessary to establish international standards to define agreed rules for guaranteeing privacy on the Internet.