5063/00/EN/FINAL
WP 37

Working Document

Privacy on the Internet
- An integrated EU Approach to On-line Data Protection-

Adopted on 21st November 2000

The Working Party has been established by Article 29 of Directive 95/46/EC. It is the independent EU Advisory Body on Data Protection and Privacy. Its tasks are laid down in Article 30 of Directive 95/46/EC and in Article 14 of Directive 97/66/EC. The Secretariat is provided by:

The European Commission, Internal Market DG, Unit Free flow of information and data protection.
Rue de la Loi 200, B-1049 Bruxelles/Wetstraat 200, B-1049 Brussel - Belgium - Office: C100-2/133
Internet address: www.europa.eu.int/comm/dg15/en/media/dataprot/index/htm

CHAPTER 1: INTRODUCTION

This document aims at offering an integrated EU approach regarding the issue of on-line data protection. The word "integrated" underlines the fact that this analysis mainly departs from the texts of both the general data protection directive (Directive 95/46/EC) and the privacy and telecommunications directive (Directive 97/66/EC) but also takes into account and brings together all opinions and documents adopted by the Working Party up to now on certain critical issues which are related to this issue[1].

The Working Party has stated in several occasions in the past, when discussing the priorities for future work, the necessity of dealing with data protection issues related to the use of Internet. In order to deal with these issues in a systematic and efficient way the so-called Internet Task Force (ITF) was created in 1999. The main purpose of the ITF is to bring together resources and expertise from different national Data Protection Authorities with the aim of contributing to the uniform interpretation and application of the existing legal framework in this field.
The ITF has drafted several papers that have been adopted by the Working Party during the last two years. From the beginning of 2000 on the ITF has intensified the frequency of its meetings with a view to achieving as result a synthesis paper that can serve as reference for addressing present and, to the extent possible, future Internet privacy issues.

The main objective of this document is to offer a first approach to the issue of on-line privacy that can serve to raise awareness concerning the privacy risks related to the use of the Internet and that can at the same time offer guidance in the interpretation of both directives in this field. The Working Party is aware of the fact that privacy is high on the list of web users' concerns[2]. It is therefore especially important for the Working Party to address this issue while being aware of the fact that some controversial issues, raising particular debate, might require future work.

- This document is not intended to be exhaustive in itself but aims to cover the most typical situations which Internet users can face when using any of the services available in the Net (such as e-mail, surfing, searching, newsgroups, etc.) Because of its general character, it does not deal with specific issues that might deserve further study by the Working Party in the future, such as for instance, the control of the e-mail in the working

[1] In particular: Opinion 1/98: Platform for Privacy Preferences (P3P) and the Open Profiling Standard (OPS) adopted by the Working Party on the Protection of Individuals with regard to the Processing of Personal Data on June 16, 1998; Working document: Processing of Personal Data on the Internet, adopted by the Working Party on 23 February 1999, WP 16, 5013/99/EN/final; Recommendation 1/99 on Invisible and Automatic Processing of Personal Data on the Internet Performed by Software and Hardware, adopted by the Working Party on 23 February 1999, 5093/98/EN/final, WP 17; Recommendation 2/99 on the respect of privacy in the context of interception of telecommunications, adopted on 3 May 1999, 5005/99/final, WP 18; Opinion n°3/99 on public sector information and the protection of personal data, adopted by the Working Party on 3 May 1999; Recommendation 3/99 on the preservation of traffic data by *Internet Service Provider*s for law enforcement purposes, adopted on 7 September 1999, 5085/99/EN/final, WP 25; Opinion 1/2000 on certain data protection aspects of electronic commerce, Presented by the Internet Task Force, Adopted on 3rd February 2000, 5007/00/EN/final, WP 28; Opinion 2/2000 concerning the general review of the telecommunications legal framework, presented by the Internet Task Force, adopted on 3rd February 2000, WP 29, 5009/00/EN/final; Opinion 5/2000 on the use of public directories for reverse or multi-criteria searching services (reverse directories), WP 33, adopted on 13th July 2000 and Opinion **7**/2000 on the European Commission Proposal for a Directive of the European Parliament and the Council concerning the processing of personal data and the protection of privacy in the telecommunications sector of 12 July 2000 COM (2000) 385, adopted on 2 November 2000, WP 36..

[2] This is pointed out in a six-month study just released by the Markle Foundation. See Article by AARON, D., *A Euro-American proposal for privacy on the Net*, Washington Post, August 2, 2000.

place. This working document is based on the present state of the art of the Internet, which is by nature a very dynamic and evolving phenomenon.

In order to facilitate the reading, the working document deals first with the basic technical description and the general legal issues. After that, all different Internet services are addressed separately covering in each chapter both the technical and legal issues at stake. A specific chapter is dedicated to privacy-enhancing measures and technologies that can be used to increase the privacy of the Internet users. A last chapter deals with the conclusions.
A glossary of technical terms has been included at the end of the document to enable the readers to understand the technical concepts used in the text of the document. All words printed in italics are contained in the glossary.

The ITF has deliberately chosen to keep a certain degree of overlap in the text of this document. This has been done in order to make possible selective reading of the document by readers who are especially interested in one chosen topic. For this purpose some additional – sometimes repetitive - descriptions have been kept in the text to facilitate the consultation of the different chapters as such.

The work of the Internet Task Force has been co-ordinated by Peter HUSTINX, chairman of the Dutch Data Protection Authority. The consolidated version of the working document has been prepared by a Drafting Group appointed within the ITF and composed of Diana ALONSO BLAS (from the Dutch Data Protection Authority) and Anne-Christine LACOSTE (from the Belgian Data Protection Authority). The work done by the Drafting Group included in particular the structuration and the checking of the coherence of the whole document, the integration and further development of additional legal issues and technical information as well as comments received from other delegations, the development of the glossary of technical terms and the conclusions of the paper.

Delegates from the Data Protection Authorities of six countries have been involved in the work of the Internet Task Force at different stages of its work, preparing papers that have served as a basis for a number of chapters, commenting on the contributions of other members of the ITF and contributing to the discussion during the five meetings of the ITF in 2000.
In particular, the following persons deserve being mentioned: Anne-Christine Lacoste and Jean-Marc Dinant (Belgium), Ib Alfred Larsen (Denmark), Marie Georges (France), Angelika Jennen and Sven Moers (Germany), Emilio Aced Félez (Spain) and Diana Alonso Blas, Ronald Hes and Bernard Hulsman (the Netherlands). The ITF would like to thank Christine Sottong-Micas (Secretariat of the Article 29 Data Protection Working Party, European Commission) and Karola Wolprecht (trainee session 1999/2000 at the European Commission) for their help and assistance.

CHAPTER 2: INTERNET TECHNICAL DESCRIPTION

I. Basics

The Internet is a network of computers communicating with each other on the basis of the Transport Control Protocol/Internet Protocol (TCP/IP)[3]. It is an international network of interconnected computers, which enables millions of people to communicate with one another in "cyberspace" and to access vast amounts of information from around the world[4].
Historically speaking, the ancestor of the Internet is the ARPAnet military network (1969). The basic idea was to build a trans-US digitised network enabling computers operated by the military, defence contractors and universities conducting defence-related research to communicate with one another by redundant channels even if some portions of the network were damaged in the war[5].

The first electronic mail programs appeared in 1972. In 1985, The American National Science foundation built the NSFNET network to link together six U.S. supercomputer centres. In the late 1980s, this network was transferred to a group of universities called MERIT. The network then became more and more open to non-academic institutions and to non-US organisations. In 1990, Tim Berners Lee, working at the CERN in Geneva, designed the first browser and implemented the concept of *hyperlink*, and since then a variety of new services and functionalities have been continuously added.

It is however necessary to bear in mind that TCP/IP is still the core *protocol* used for data transmission over the Internet and that all services rely on it. This *protocol* was designed to be very simple to set up and is independent of any specific computer or operating system.

On the Internet, every computer is identified by a single numerical IP address of the form A.B.C.D. where A, B, C and D are numbers in the range of 0 to 255 (e.g. 194.178.86.66).

A *TCP/IP network* is based on the transmission of small packets of information. Each packet includes the IP address of the sender and of the recipient. This network is connectionless. It means that, unlike the telephone network for instance, no preliminary connection between two devices is needed before communications can start. It also means that many communications are possible at the same time with many partners.

The *DNS (Domain Name System)* is a mechanism for assigning names to computers identified by a IP address. Those names are in the form of <names>. top level domain where <names> is a string constituted by one or many substrings separated by a dot. The top level domain can be a generic domain like “com” for commercial websites or “org” for non-profit organisations, or a geographical domain like “be” for Belgium. *DNS* has to be paid for and companies or individuals wanting a domain name have to identify themselves. Some public tools on the Net make it possible to retrieve the link between

[3] The technical aspects described in this work have been drastically simplified to make them comprehensible to a non-expert. For more details, see: *Communication from the Commission to the Council and the European Parliament, The organisation and management of the Internet*, International and European Policy Issues, 1998- 2000, COM (2000) 202 final, 11 April 2000.
[4] See Reno v. ACLU decision (June 26 1997), Supreme Court of the United States, available at www2.epic.org/cda/cda_decision.html
[5] See Reno v. ACLU decision (June 26 1997)

the domain name and the company as well as between the IP address and the domain name. A domain name is not in itself necessary for connecting a computer to the Internet. Domain names are dynamic. One single Internet computer can have one or many domain names – or even none at all - but one specific domain name always refers to one particular IP address.

A limited amount of IP addresses exist at the present time. This number depends on the length of the field assigned to the IP address in the *protocol*;[6]. The IP addresses are assigned in Europe through an international procedure[7] to Internet Access Providerswho then reassign them to their clients, organisations or individuals. By using a publicly available search tool like, for instance, http://www.ripe.net/cgi-bin/whois it is possible to identify the party responsible for a particular IP address allocation. Typically, this will be:

- the manager of a Local Area Network linked to the Internet (e.g. an SME or a public administration). In this case, he/she will probably use a fixed IP addressing scheme and keep a list of correspondence between people's computers and IP addresses. If this person is using the *Dynamic Host Configuration Protocol* (DHCP[8]), the *DHCP* program will typically keep a logbook containing the Ethernet card number. This unique world-wide number identifies a particular computer in the LAN.

- an Internet Access Provider which has a contract with an Internet subscriber. In this case, the IAP will typically keep a log file with the allocated IP address, subscriber's ID, date, time and duration of the address allocation. Furthermore, if the Internet user is using a public telecommunications network (mobile or terrestrial phone), the number called (and date, time and duration) will be registered by the phone company for billing purposes.

- the Domain Name Holder which might be a company's name, the name of the employee of a company or a private citizen.

In these cases, this means that, with the assistance of the third party responsible for the attribution, an Internet user (i.e. his/her civil identity: name, address, phone number, etc.) can be identified by reasonable means.

A *router* is an important device which provides routes for *TCP/IP networks*. This means that the TCP/IP route is dynamic, depending on the failure or overloading of some routers or links. It can also be used as a *firewall* between an organisation and the Internet. It can especially guarantee that only authorised IP addresses can originate from a particular *ISP*.

It is important to note that the speed of transmission is the single most valuable criterion for routing in TCP/IP networks. With information circulating at almost the speed of light, it can be more efficient to route TCP/IP packets from London to Madrid via New York if

[6] The upgraded version (IPversion6) of the IP addressing system is currently being developed based on numbers that are 128 bits long.

[7] The Internet Corporation for Assigned Names and Numbers (ICANN) is the non-profit corporation that was formed to assume responsibility for IP address space allocation (http://www.icann.org). In Europe the addressing space is managed by the RIPE organisation (Réseaux IP Européens) (http://www.ripe.net). For more details about the evolving process of Internet Domain Names, see the Commission communication referred to in footnote 2.

[8] The Dynamic Host Configuration *Protocol* (DHCP) is an Internet protocol for automating the configuration of computers that use TCP/IP. DHCP can be used to automatically assign IP addresses. (http://www.dhcp.org)

there is a traffic jam in the network in Paris. Some tools allow the net user to know the route between two points, but this can theoretically change every second, even during the transfer of a single web page.

More sophisticated protocols using TCP/IP

Some *protocols* are designed to provide certain services in addition to TCP/IP. Basically the most widely used *protocols* are:

- the HTTP (HyperText Transport Protocol used for surfing,
- the FTP (File Transfer Protocol) used to transfer files,
- the NNTP (News Network Transport Protocol) used to access newsgroups,
- the SMTP (Simple Mail Transport Protocol) and POP3 *protocols* (to send and receive e-mails).

Layers and protocol hierarchy in an Internet communication process

<table>
<tr><td>HTTP
used for surfing and searching</td><td>SMTP
used for sending e-mail</td><td>POP3
used for downloading e-mails from a mail-server to client</td><td>NNTP
used for transferring newsmessages</td><td>FTP
used for downloading or uploading files</td><td>etc.
many other high level protocols in use or being developed</td></tr>
<tr><td colspan="6">TCP/IP</td></tr>
<tr><td>PPP
used by modems on phone lines</td><td>X-75
used by terminal adapter on ISDN lines</td><td>ADSL
used by an ADSL modem on standard phone lines</td><td>ETHERNET
used by LAN cards on a Local Area Network</td><td colspan="2">etc.
Many other low level protocols in use or being developed</td></tr>
</table>

- These *protocols* are necessary because the TCP/IP *protocol* only permits the transmission of bulk information from one computer to another. The computer delivering a service is called a SERVER. The computer using a service is called a CLIENT. To provide a technical service, both the client and the server use the same *protocol*, i.e. the same communication rules. The Internet is often referred to as a client/server network. It is important to note that whatever the service used, the TCP/IP *protocol* is always used by every service mentioned above. This means that every threat to privacy linked to the TCP/IP *protocol* will be present when using any service on the Web.

- In order to avoid any misunderstandings with the general meaning of the word "service", the term *protocol* will be used in this paper to designate HTTP, FTP, NNTP and other services available on the Internet.

A *proxy server* is an intermediary server between the Internet user and the Net. It acts as a *Web cache*, dramatically improving the rate of display of information (e.g. the display of web pages). Many large organisations or Internet Access Providers have already implemented this solution. Each page, image or logo downloaded from outside by a member of an organisation is stored in a cache on the proxy server and will be instantaneously available to another member of this organisation.

II. ACTORS INVOLVED IN THE INTERNET

It should be noted that a company or an individual can play different roles regarding the Internet, and may thus concurrently perform various data processing operations (e.g. logging connections as a telecoms operator, and storing visited websites as an ISP), with all this entails concerning the application of privacy principles.

Telecommunications operator

In Europe, the telecoms infrastructure used to be *de facto* the monopoly of traditional telecommunications operators. This situation is however evolving. Furthermore, this monopoly is often reduced to the cables or optical fibres, while for wireless communications and emerging technologies like *WAP*, *UMTS*, etc, competition is emerging between national carriers.

The traditional telecommunications operator is still, however, an important actor since it provides the data communications between the net user and the Internet Access Provider (IAP).

The telecommunications operator processes traffic information for billing purposes, such as the calling number and its location (for mobiles), called number, date, time and duration of the communication[9].

Internet Access Provider

The IAP provides, normally on a contractual basis, a TCP/IP connection to:

- Individuals using a *modem* or a terminal adapter (ISDN). In this case the subscriber will receive a IP address for the duration of his/her connection and this address will probably change the next time he/she dials up. This is called a dynamic IP address.

In the case of a connection by ADSL or via video cable, the IP address will usually be static, as far as those connections are permanent.

In order to obtain a connection, the individual[10] has to conclude a contract (where the subscription is free) and give his/her name, address and other personal data. Typically the user will receive a user identification name (UserId that may be a pseudonym) and a password so that nobody else can use his/her subscription. At least for security reasons, Internet Access Providers usually seem to systematically "log" the date, time, duration and dynamic IP address given to the Internet user in a file. As long as it is possible to link the logbook to the IP address of a user, this address has to be considered as personal data.

- Organisations using a dialup connection or, more often, a line leased to the company's office. This leased line will normally be provided by the traditional telecoms operator. The connection can also be established via a satellite line or a terrestrial radio system. The IAP will give IP addresses to the company and use a *router* to ensure that the addresses are observed.

IAPs own one or more leased lines (twisted pair, optical fibre, satellite link) connected to other bigger IAPs.

[9] The processing and storage time of such data is subject to strict legal conditions, as explained later.
[10] A small enterprise may also of course conclude such a contract, but such cases will not be considered in this paper.

Internet Service Provider

The *Internet Service Provider (ISP)* provides services to individuals and companies on the Web. It owns or hires a permanent TCP/IP connection and uses servers permanently connected to the Internet. Classically, it will offer web hosting (web pages stored on its web server), access to newsgroups, access to an FTP server and electronic mail. This involves one or more servers using the HTTP, NNTP, FTP, SMTP and POP3 *protocols*.

Firms playing the role of IAPs will frequently offer the services of *ISPs*. This is why the generic term *ISP* is often used to include both IAPs and *ISPs*. But, from a conceptual viewpoint, the roles are different. Namely, the IAP, being a gate to the Internet, will route all traffic from the Internet subscriber, while the *ISP* will only be aware of what happens on its servers[11]. In this report, when the term *ISP* is used, it generally includes IAP's. The term IAP is only used when it is clear that it deals only with Internet access; in all other cases the generic term *ISP* is used.

From a technical viewpoint, it is the presence of servers equipped with *protocols* that will be decisive in gathering personal data. In the case of HTTP servers generally, a logbook or logfile is systematically created by default and may contain all or some of the data present in the HTTP request header (browser chattering) and the IP address. The logbook is standard practice and is created by each server.

The user

The Internet user can be an individual accessing the Net from home, generally using a temporary TCP/IP connection (and thus a dynamic IP address) via a *modem*, a terminal adapter (ISDN), or a permanent connection (thus static IP address) through ADSL, cable TV, etc. Connection via a mobile phone, whilst generally more expensive, is also possible.

Should a subscriber give a false identity or use the identity of another user (typically by giving someone else's UserId and password), it is still possible to trace back the owner of the line to which a particular IP address has been given by comparing this information with the information contained in the IAP logbook. This is, in fact, what the police does when tracing criminal intrusions into computers linked to the Internet.

The same applies if the individual is using a LAN or an Intranet.

The user can also be an organisation, a public administration or a company which uses the Internet not only to provide or to look for information but also to collect data for the purpose of its tasks or activities (administrative procedures, selling of goods or provision of services, publication of directories, small ads, sending out questionnaires, etc.)

III. SERVICES AVAILABLE ON THE INTERNET[12]

Anyone with access to the Internet may use a wide variety of communication and information retrieval methods. The most common are electronic mail (see Chapter 4), newsgroups and chat rooms (see chapter 6) and the World Wide Web (see Chapter 5).

All these methods can be used to transmit text; most can transmit sound, pictures and moving video images. Taken together, these tools constitute a unique medium, known to

[11] This paper will not deal with *ISP*s as content providers although some of them provide content in certain circumstances (for instance, some ISPs have their own portal site).

[12] For a detailed description of these services see decision Reno v. ACLU (June 26, 1997).

its users as "cyberspace", available to anyone, anywhere in the world, with access to the Internet.

e-mail

E-mail enables an individual to send an electronic message to another individual or to a group of addressees. The message is generally stored electronically on a server, waiting for the recipient to check his/her mailbox, and sometimes making its arrival known through some type of prompt.

Newsgroups

Newsgroups are used to share information or express opinions about specific matters. They serve groups of regular participants but others may read their postings too. There are thousands of such groups, each serving to promote the exchange of information or opinion on a particular topic. About 100 000 new messages are posted each day.

Chat rooms

Two or more individuals wishing to communicate directly can enter a chat room to engage in real-time dialogue by typing messages that appear almost immediately on the others' computer screens.

World Wide Web

The best known category of communication over the Internet is the World Wide Web, which allows users to search for and, retrieve information stored in remote computers. In plain terms, the Web consists of a vast number of documents stored in different computers all over the world.

Navigating the Web is relatively straightforward. A user may either type the address of a known page or enter one or more keywords into a commercial "search machine" in an effort to locate sites on a subject of interest. Users generally explore a given web page or move to another by clicking a computer "mouse" on one of the page's icons or links. The Web is thus comparable, from the reader's viewpoint, either to a vast library including millions of readily available and indexed publications or a sprawling mall offering goods and services (see Chapter 7).

Any person or organisation with a computer connected to the Internet can "publish" or collect information (see Chapters 6, 7 and 8). Publishers or those who collect data include government agencies, educational institutions, commercial entities, interest groups and individuals. Those may either make their material available to the entire pool of Internet users, or restrict access to a selected group.

IV. Privacy risks[13]

Privacy risks inherent in the use of the TCP/IP protocol

Due to the fact that the Internet has, from the very beginning, been considered as an open network, there are many characteristics of communication *protocols* which, more by accident than design, can lead to an invasion of the privacy of Internet users.

As far as the TCP/IP protocol is concerned, there are three characteristics which appear to constitute a potential invasion of privacy.

[13] The French CNIL has in its website a section called "vos traces" where Internet users can view the traces they leave behind when using the Internet. This section is available in French, English and Spanish. See www.cnil.fr

- The **route** followed by TCP/IP packets is dynamic and follows the logic of performance. In theory, it may change during the downloading of a web page or the transmission of an e-mail, but in practice it remains largely static. In telecommunications, performance is linked more to the congestion of the network than to the physical distance between telecommunications nodes (*routers*). This means that the "shortest" way between two towns located in the same EU country may pass through a non-EU country which may or may not have adequate data protection[14]. The average Internet user has no reasonable means of changing this route, even if he/she knows which route is followed at a particular moment.

- Due to the fact that the translation between the Domain Name and the numerical IP address occurs via a **DNS server**, whose function is to ensure this translation, this DNS server receives, and can keep trace of, all the names of the Internet servers the Internet user has tried to contact. In practice, those DNS servers are mainly maintained by Internet Access Providers, who have the technical capability to know much more than that, as will be described in the next chapters.

- The **ping** command, available on all operating systems, allows anyone on the Internet to know if a particular computer is turned on and connected to the Internet. It is a command which involves typing the letters PING followed by the IP address (or the corresponding name) of a selected computer. The user of the "pinged" computer will usually not be aware that and for which reasons somebody has tried to find out if he/she was connected at a given moment.

It should be noted that permanent Internet connections via cable and ADSL present the same risks.

Even if these data-processing operations are legitimate and, depending on circumstances, unavoidable for the smooth operation of the Internet network, the Internet user should be made aware of the fact that these operations are taking place and of available security measures.

Privacy risks inherent in the use of high level protocols

This section focuses on three characteristics that are almost always present when implementing the HTTP *protocol* in the most frequently used browsers. It has to be noted that a combination of these characteristics can have serious consequences for the privacy of Internet users.

HTTP is of strategic importance insofar as it is the main *protocol* used on the Web and can offer services like electronic mail and discussion fora, which up to now had usually been provided by specialised high level protocols such as POP3, SMTP or NNTP[15].

The browser's chattering

It is generally known that typing "http://www.website.org/index.htm" means something like "show me the page named "index.htm" on the server www.website.org by using the HTTP *protocol*. One might think that only the IP address of the surfer and the file he/she wants to see are transmitted to the website. This is, however, not the case.

[14] See Chapter 2 for more details on this issue.
[15] See DINANT, Jean-Marc, Law and Technology Convergence in the Data Protection Field? *Electronic threats to personal data and electronic data protection on the Internet*, ESPRIT Project 27028, Electronic Commerce Legal Issues Platform.

The following table lists some of the data systematically transmitted in the HTTP header while making an HTTP request (Automatic browser chattering) and thus available to the server:

HTTP Var.	***Opera 3.50***	***Netscape 4.0 Fr***	***Explorer 4.0 UK***
GET	GET /index.html HTTP/1.0	GET /index.html HTTP/1.0	GET /index.html HTTP/1.0
User-Agent:	Mozilla/4.0(compatible; Opera/3.0; Windows 95) 3.50	Mozilla/4.04 [fr] (Win95; I ;Nav)	Mozilla/4.0 (compatible; MSIE 4.01; Windows 95)
Accept :	image/gif, image/x-xbitmap, image/jpeg, /	Image/gif, image/x-xbitmap, image/jpeg	image/gif, image/x-xbitmap, image/jpeg, image/pjpeg, application/vnd.ms-excel, application/msword, application/vnd.ms-powerpoint, /
Referer :		Where.were.you/doc.htm	Where.were.you/doc.htm
Language :		Fr	fr-be

The technical definition of those fields can be found in the RFC 1945 for HTTP 1.0 or in the RFC 2068 for HTTP 1.1. The following remarks can be made in this respect:

- The first line is the only one which is indispensable.
- In the "Accept" line, every browser mentions that the Internet user is using Windows 95. One could wonder why. Netscape adds that the browser version is a French one. Every browser gives its own name, version and sub-version identification.
- While describing the accepted formats, Microsoft informs every site that the Internet user's computer has Powerpoint, Excel, and Word installed on it.
- Opera does not disclose the referring page.
- Opera does not reveal the language spoken. Netscape reveals that the Internet user is French-speaking. Microsoft reveals that the Internet user is a French-speaking Belgian.

Invisible hyperlinks

Hyperlinks are the added value of the Internet. They make it possible to browse from one continent to another simply by a mouse click. What is hidden to the eyes of the common user is that classical browsing software makes it possible for the HTTP request to include a command to download images for inclusion in the HTML page code. Those images do not need to be located in the same server as the one which has received the original call for a particular web page.

In this case, the HTTP_REFERER variable contains the referring page reference, i.e. the main page in which the images will be located. In others words: if a website includes in its web page in HTML an invisible link to an image located on the website of a cybermarketing company, the latter will know the referring page before sending the advertising *banner*. When doing a search on a search engine, the name of the web page includes the keywords typed.

Cookies

Cookies are pieces of data that can be stored in text files that may be put on the Internet user's hard disk, while a copy may be kept by the website. They are a standard part of HTTP traffic, and can as such be transported unobstructed with the IP-traffic.

A *cookie* resides on a user's hard drive and contains information about the individual that can be read back by the website that deposited it or by anyone else with an understanding of that website's data format. A *cookie* can contain any information the website wants to include in it: pages viewed, advertisements clicked, user identification number and so on[16]. In some cases, they may be useful for providing a certain service through the Internet or to facilitate the surfing of the Internet user. For instance, certain custom websites rely on *cookies* to identify users each time they return, so users do not have to log into the website each time they check their news.

The SET-COOKIE is placed in the HTTP response header[17], namely in invisible *hyperlinks*. If a duration is stipulated[18], the cookie will be stored on the Internet user's hard disk and sent back to the website originating the cookie (or to other websites from the same sub domain) for that duration. This sending back will take the form of a COOKIE field involved in the browser chattering described above.

By putting together the browser chattering and invisible *hyperlinks*, a cybermarketing company can, by default, know all the keywords typed by a particular Internet user into the search engine on which this company is advertising, the computer, operating system, browser brand of the Internet user, the user's IP address, and the time and duration of HTTP sessions. These raw data make possible, if combined with other data available to the company, to infer new data like[19]:

1. The country where the Internet user lives.
2. The Internet domain to which he/she belongs.
3. The sector of activity of the company employing the Internet user.
4. The turnover and size of the employing company.
5. The function and position of the surfer within this company.
6. The Internet Access Provider.
7. The typology of websites currently visited.

The *cookie* allows a permanent and unique identifier to be sent systematically with every information request, whereas the IP address remains a relatively weak identifier because it can be hidden by proxies and is not reliable, due to its dynamic character for Internet users accessing the Internet by *modem*. Many cybermarketing companies have already done such invisible profiling[20].

---

[16] See the book by HAGEL III, J. and SINGER, M., *Net Worth: the emerging role of the informediary in the race for customer information*, Harvard Business School Press, 1999, p. 275.

[17] Technically speaking, it is also possible to implement cookies in JavaScript or in the <META-HTTP EQUIV> fields located in the HTML code.

[18] Cookies with no fixed duration are called "session cookies" and disappear when the browser is unloaded or when the socket closes.

[19] GAUTHRONET, Serge , "On-line services and data protection and the protection of privacy" European Commission, 1998, p.31 and 92 available at http://europa.eu.int/comm/dg15/en/media/dataprot/studies/servint.htm

[20] For DoubleClick alone, about 26 million Internet users in March 1997 (GAUTHRONET, op. cit., p. 86) and more than one billion cybermarketing banners downloaded each month outside the US (ibid., p. 96). Presently more than 500,000,000 advertising banners sent each day for one single cybermarketing company. See http://www.doubleclick.net/company_info/investor_relations/financials/analyst_metrics.htm

Privacy risks linked with implementation of the HTTP protocol in common browsers

The combination of browser chattering, invisible *hyperlinks* and *cookies* provide the means for invisible profiling of every individual Internet user who uses a browser installed by default. This profiling is not "per se" linked to the HTTP *protocol*, as defined by the W3C[21]. Furthermore, the HTTP 1.1 *protocol* definition has explicitly drawn the attention of the industry to possible privacy issues while implementing the HTTP *protocol*[22]:

– "*Having the user agent describe its capabilities in every request can be both very inefficient (given that only a small percentage of responses have multiple representations) and a potential violation of the user's privacy*" [page 68]
– " *It may be contrary to the privacy expectations of the user to send an Accept-Language header with the complete linguistic preferences of the user in every request*" [page 98]
– *" The client SHOULD not send the From header[23] field without the user's approval, as it may conflict with the user's privacy interests or their site's security policy. It is strongly recommended that the user is able to disable, enable, and modify the value of this field at any time prior to a request.*" [page 118]

"*HTTP clients are often privy to large amounts of personal information (e.g. the user's name, location, mail address, passwords, encryption keys, etc.), and SHOULD be very careful to prevent unintentional leakage of this information via the HTTP protocol to other sources. We very strongly recommend that a convenient interface be provided for the user to control dissemination of such information, and that designers and implementers be particularly careful in this area. History shows that errors in this area are often both serious security and/or privacy problems, and often generate highly adverse publicity for the implementer's company.*" [page 143][24]

V. Some economic considerations

The Internet has experienced extraordinary growth over the last years. The number of "host" computers - those that store information and relay communications - rose from about 300 in 1981 to approximately 9 400 000 in 1996. Roughly 60% of these hosts are located in the United States. About 40 million people used the Internet in 1996 and about 200 million were expected to use it by 2000[25]. It is expected that half of the European population will be connected to the Internet by 2005[26].

In many European countries, Internet subscription is free for individuals but the subscriber has to pay for the line to the telecommunications operator. The IAP or *ISP* will be remunerated by a retroconnect fee paid back by the telecommunications operator (TO) on the basis of the duration of the local call made by the Internet subscriber. This means that even in cases where a user has free subscription to the Internet he/she will still

[21] The World Wide Web Consortium is a non-profit organisation hosted by Inria (France), MIT (USA) and the University of Keio (Japan). The members of this consortium are notably Microsoft, AOL, Netscape, and Center for Democracy and Technology (http://www.w3.org/Consortium/Member/List). This consortium produces non-mandatory but *de facto* standardisation intended to guarantee the interoperability of computers on the Internet.
[22] http://www.w3.org/Protocols/rfc2068/rfc2068 . The page numbering in brackets refers to the W3C's numbering.
[23] "From header" field is used for naming the referring page.
[24] The word "privacy" is mentioned 18 times in the RFC 2068.
[25] See Reno v. ACLU decision (June 26, 1997).
[26] European Commission press release, *Commission welcomes new legal framework to guarantee security of electronic signatures*, 30 November 1999.

have to bear the expenses of the telephone lines used. This will benefit both to the IAP/ISP and telecom operators.

Software producers will also benefit from the use of the Internet because, even if they make their products freely available to the consumer (freewares, browsers, etc.), they receive a remuneration for the use of their softwares by website servers.

Direct marketing is one of the major rental activities on the Web. Cybermarketing companies place advertising banners on web pages, often in such a way that the collection of personal data remains widely invisible to the data subject. Thanks to the use of invisible links in combination with browser chattering and cookies, unknown marketing companies are able to profile Internet users on a one-to-one basis. One single cybermarketing company could send about half a billion personalised advertising banners on the Web every day. Direct marketing companies finance many search engines.

By putting an invisible *hyperlink* to cybermarketing companies on their own web pages, common websites (and search engines in particular) will instruct common browsers like Netscape and Internet Explorer to open an independent HTTP connection with the cybermarketing company's HTTP server. As explained before, the browser will automatically chat various data while doing the HTTP request, namely: the IP address, the referring page (in the case of a search engine, this variable contains the keywords typed by the searcher), the brand, version and language of the browser used (e.g. Internet Explorer 4.02, Dutch, type and OS used: Windows 2000, Linux 2.2.5, Mac OS 8.6 and so on) and, last but not least, the identifying *cookie* (e.g. UserId=342ER432) which might already have been placed by the cybermarketing company through previous invisible *hyperlinks.*

The average Internet user is generally unaware of the fact that while typing an URL (Unified Resource Locator), many banners that he/she will see as a result do not originate from the website he/she is visiting. Nor are users aware of the fact that, while downloading one ad *banner*, their browser will systematically transmit a unique ID, IP address, and complete URL of the webpage they are visiting (this includes keywords typed on search engines and the name of press Articles they are reading on line). All those data can be merged to build a global profile of a citizen surfing from one site to another, thanks to the unique Id stored in the *cookie*.

The capture of user information in on-line environments is considered to have economic and strategic importance. The following paragraph taken from a famous American publication[27] illustrates this idea: *Too many businesses, including many of the leading-edge entrepreneurial companies emerging on the Internet, have not focused enough on the value of customer profiles. The winners and losers of this new era will be determined by who has rights to on-line customer profiles.*

It is worth mentioning that the collection of Internet users data is usually free of any costs for the company, as consumers often provide the information themselves, e.g. by filling in forms. Websites often use loyalty programs like games, questionnaires, newsletters, that involve the provision of personal information by the visitor of the website.

Recent cases confirm the increasing value attached by businesses to consumer profiles. Lists of customers are being sold or shared, most often through mergers of IT companies which thus increase the detail and number of profiles they can use.

There will eventually be acquisitions that are based on consumer data, where the primary asset that's being bought is the consumer data. (…) Consumer data right now is the currency of e-commerce in a lot of ways. Those are valuable customers because

[27] See the book "Net Worth" (op cit), page xiii (preface).

they've shown that they're buyers, and they've bought from a competing store. (…) Names in a database save a company from spending marketing dollars to acquire a customer -- usually about $100 per customer[28].

Customer data have also been offered for sale when Internet companies go bankrupt. A company selling toys recently included the sale of its customer profiles as part of the company's liquidation. These customer profiles were collected from users under the privacy policy that no information would ever be shared with a third party without the express consent of the user. The profiles include names, addresses, billing information, shopping behavioural information and family profiles with the names and birth dates of children.

TRUSTe, which had approved the company's privacy policy, advised on August 8, 2000 that it had filed an objection with the United States Bankruptcy Court, to the Federal Trade Commission (FTC) consent agreement with the company on the conditions for liquidating the assets[29].

A comprehensive data protection policy must take account of a balanced choice between economic interests and human rights. Two big issues remain unresolved.

- Nowadays a large volume of individual data on many Internet users has been collected on the Internet without the prior knowledge and/or consent of the data subject, mainly due to the invisible side-effects of Internet technology. It is foreseeable that, in the next few years, more and more personal data will be exchanged for material gain[30], but how far can the Internet user go in doing this? What kind of personal data can be shared by the data subject itself, for how long and under what circumstances?
- If the funding of particular websites (e.g. search engines) comes mainly from the cybermarketing industry, there may be a temptation to use personalised profiling to ensure that services which were previously free exclude people who do not have sufficient income, have not responded to hundreds of advertising banners or wish to preserve their privacy.

VI. Conclusions

- The Internet was conceived as an open network at world level (www) through which information could be shared. It is however necessary to find a balance between the "open nature" of the Internet and the protection of the personal data of the Internet users.
- Enormous amounts of data on Internet users are collected on the Internet while often users are not aware of this fact. This lack of transparency towards the Internet users needs to be addressed in order to achieve a good level of personal data and consumers' protection.
- Protocols are technical means that in fact determine how data are to be collected and processed. Browsers and software programmes also play an important role. In some cases they include an identifier that makes possible to link the Internet user to his/her activities in the Net. It is therefore the responsibility of those involved in the design and development of these products to offer users privacy-compliant products In that sense it is important to mention that article 14 of the draft telecoms directive of 12

[28] Quoted from M. HALPERN and HARMON, *E-mergers trigger privacy worries* by Deborah KONG, http://www.mercurycenter.com/svtech/news/indepth/docs/consum012400.htm
[29] http://www.truste.org/users/users_investigations.html
[30] See for instance the discussion on infomediaries in Chapter 9.

July 2000 declares that, where required, the Commission shall adopt measures to ensure that technical equipment incorporates the necessary safeguards to guarantee the protection of personal data and privacy of users and subscribers.

CHAPTER 3: APPLICATION OF DATA PROTECTION LEGISLATION

I. General legal considerations

The point of departure for the legal analysis of all the different phenomena to be carried out in the following chapters is the fact that both data protection directives (Directive 95/46/EC and 97/66/EC) apply in principle to personal data processed on the Internet[31].

All legal considerations included in this document are based on the interpretation of these Directives as well as on the documents adopted by the Working Party and in some cases (if so indicated) the jurisprudence of the European Court of Human Rights.

Personal data on the Internet

As has been already mentioned in this paper, Internet Access Providers and Managers of Local Area Networks can, using reasonable means, identify Internet users to whom they have attributed IP addresses as they normally systematically "log" in a file the date, time, duration and dynamic IP address given to the Internet user. The same can be said about *Internet Service Providers* that keep a logbook on the HTTP server. In these cases there is no doubt about the fact that one can talk about personal data in the sense of Article 2 a) of the directive[32].

In other cases, a third party can get to know the dynamic IP address of a user but not be able to link it to other data concerning this person that would make his/her identification possible. It is obviously easier to identify Internet users who make use of static IP addresses.

The possibility exists in many cases, however, of linking the user's IP address to other personal data (which is publicly available or not) that identify him/her, especially if use is made of invisible processing means to collect additional data on the user (for instance, using *cookies* containing a unique identifier) or modern *data mining* systems linked to large databases containing personally-identifiable data on Internet users.

Therefore, even if it might not be possible to identify a user in all cases and by all Internet actors from the data processed on the Internet, this paper works on the basis that that the possibility of identifying the Internet user exists in many cases and that large masses of personal data to which the data protection directives apply are therefore processed on the Internet.

Application of the directives

As the Working Party has already stated on previous occasions, the general data protection directive 95/46/EC applies to any processing of personal data falling within its scope, irrespective of the technical means used. Personal data processing on the Internet therefore has to be considered in the light of this directive[33].The general directive thus applies in all cases and to all the different actors that we have dealt with in the first part of this chapter (technical description).

---

[31] See WP 16, Working document: *Processing of Personal Data on the Internet*, adopted by the Working Party on 23 February 1999, 5013/99/EN/final.

[32] See also recital 26 of the preamble to the directive.

[33] The expression "the directive" refers in this paper to Directive 95/46/EC.

The specific directive 97/66/EC on the protection of privacy and personal data in the telecommunications sector particularises and complements the general directive 95/46/EC by establishing specific legal and technical provisions. Directive 97/66/EC applies to the processing of personal data in connection with the provision of publicly available telecommunications services in public telecommunications networks in the Community. Internet services are telecommunications services. The Internet thus forms part of the public telecommunications sector.

Directive 95/46/EC applies to all matters that are not specifically covered by Directive 97/66/EC, such as the obligations on the controller and the rights of individuals or non-publicly available telecommunications services[34]. Personal data voluntarily provided by the Internet user during his/her connection to the Internet would always fall under the scope of application of this Directive.

In the following table, an attempt is made to define cases in which the specific directive 97/66/EC applies and those where directive 97/66/EC applies, by stating the most relevant principles. It should, however, be taken into account that some overlapping will occur when actors play several roles at the same time.

Actor	**Task**	**Possible processing of personal data**	**Relevant provisions of the telecoms directive:**
Telecoms provider Ex. AT&T	-Connecting Internet users and *ISP*s	- Logging Internet user-*ISP* connections -Transfer *CLI* of Internet user to *ISP*	-Telecoms directive, especially: confidentiality of the communications, traffic and billing data and presentation and restriction of calling line and connected line identification.
Internet Service Provider[35] Ex. World Online	-Providing the requested Internet service -Transfer request from Internet user to *proxy* server (cache) -Transfer request from Internet user to website -Transfer reply from *proxy* server to Internet user -Transfer reply from website to Internet user	-Logging incoming *CLI*'s - Allocation of IP-address to a session - Possibility of storing lists of visits to websites, sorted by IP address -Exchanging data with requested websites - Logging of sessions (login and logout times, and amounts of transferred data) -Extracting information from headers and content.	- Telecoms directive, in particular: confidentiality of the communications, traffic and billing data
Portal service Ex. Yahoo, AOL, Macropolis	- Selection of supply of information -Providing information (content provider), and sometimes services or goods	- Logging of requests to sites behind the portal - Possible logging of visits to the site - Logging of referring pages, keywords typed (chattering data) - Posting *cookies* on hard disk of Internet user. - Profiling	-Telecoms directive (applicable to the *ISP* hosting the *portal site*)
Regular website/homepage	-Providing information (content provider), and	- Possible logging of visits to the site	

[34] See recital 11 of Directive 97/66/EC.

[35] In principle the term *Internet Service Provider* as used in this paper also includes Internet Access Providers (see definition in the glossary of terms). This paper only refers to Internet Access Providers when dealing with issues that only apply to them.

Ex www.coe.int	sometimes services or goods	- Logging of referring pages, keywords typed (chattering data) - Posting *cookies* on hard disk of Internet user. - Profiling	
Providers of additional services Ex. Nedstat Doubleclick Banners	-Customising webpages	-Profiling (by merging the *clickstream* of several websites)	-Not always a telecommunications service, so the telecoms directive only applies in some cases.
Providers of *routers* and connecting lines (often owned by telecoms providers)	- Connecting *ISP*'s	- Directing data from Internet user to IP website. -Risk of illegal interception	- Telecoms directive: in particular, security and confidentiality of communications

In deciding whether both directives are applicable or not, the key question is obviously to determine if the service concerned can be considered as a "telecommunications service", as defined in Article 2 d) of Directive 97/66/EC: *transmission and routing of signals via telecommunication networks.*

If the specific telecoms directive is applicable, it is necessary to apply the specific rules contained in it.

Telecoms provider

There is no doubt that connecting Internet users to an *ISP*, providing Internet services to Internet users and routing requests and replies from Internet users to website servers and back are telecommunications services. So, Directive 97/66/EC applies to telecommunications providers, *Internet Service Provider*s and providers of *routers* and lines for Internet traffic.

Internet Service Providers (also including Access Providers)

The same can be said about *Internet Service Provider*s; there is no doubt that the specific telecoms Directive applies to their activities.

An interesting case concerns those institutions or persons which have direct access to the Internet without the help of an *ISP*. These institutions are in fact acting as *Internet Service Providers* connecting their own private network to the Internet.

Article 3 of Directive 97/66/EC defines its scope of application by specifying that it concerns publicly available telecommunications services in public telecommunications networks in the Community. In the above-mentioned case, there is not a public network but a private network for a given group of users. It can therefore be concluded that these services, whilst falling within the definition of telecommunications services, cannot be considered as publicly available services and do not therefore fall within the scope of application of Directive 97/66/EC.

It is important to mention that in such cases the provisions of the specific Directive would apply again if information is sent to somewhere outside the private network.

Obviously, the provisions of the general data protection directive are fully applicable in these cases.

Regular websites

Normally a website is hosted by an *ISP*. This means that the person responsible for a website (for instance the website of the Council of Europe) rents some storage capacity from an *ISP* for storing its website and making it available. It also means that the *ISP* replies to Internet users' requests for webpages on behalf of the Council of Europe.

Consequently, the person who "runs" the website (in this case the Council of Europe) only decides on which information will be made available on the website, but does not him/herself carry out any kind of operation *involving the transmission or routing of signals on telecommunications networks*.

Where goods or services can be ordered through a website, the person responsible for the site will provide those services/goods. The telecommunications services as such will not normally be provided by the person responsible for the site, but by the *ISP*.

It can therefore be said that websites are subscribers to the telecommunications services (transmission) of the webhosting *ISP* but do not themselves carry out any of these services. Directive 97/66/EC is applicable to the *ISP*'s as such but not to the websites, to which the general directive applies.

Portal services

A *portal site* provides an ordered overview of weblinks. The Internet user can easily visit selected websites of other content providers via the *portal* visited.
A *portal site* is hosted by an *ISP*. In some cases the *portal site* belongs to the *ISP* (for instance worldonline.nl); in others the *ISP* hosts the *portal site* for a third party which provides the content.

In both cases it is the *ISP* which provides the telecommunications service as defined in Article 2 of Directive 97/66/EC and to whom this directive applies – it does not apply to the content-provider.

Additional services

The providers of additional services do not, in all cases, fall within the scope of application of the privacy and telecommunications directive.

Some of these service providers (like Nedstat) process data which they collect from websites and then sell back to the owners of the websites. The data they process come from the Internet but their activity does not in principle involve the *transmission or routing of signals on telecommunication networks.* They do not therefore play an essential role in the communications process between the Internet user and the website. If the data they process only consist of aggregated non-identifiable data, it could even be said that they do not come under the general directive as no personal data would be involved.

Actors like Doubleclick, Engage or Globaltrash place advertisements in requested pages. Normally there is a contractual agreement binding these advertisers to the *ISP* hosting the webpages in which banners are placed.

For this purpose, technically speaking every time a website is accessed it contacts the advertiser (*hyperlink* by automatic means) so that this can place banners on the requested pages.

Additionally, the advertiser can place *cookie* files on the hard disk of the Internet user in order to construct profiles of visitors to the site, so that customised banners can be placed on the webpage[36].

It is unclear whether the core activities of Doubleclick, Engage and other advertisers can be regarded as a telecommunications service or not. It appears that they do not transmit and route signals as defined in Article 2 of the telecoms directive. They provide content information to be placed on the requested webpages, making use of the available telecommunications infrastructure and networks.

This is, in any case, a good example of a situation in which the existing definition of telecommunications services is difficult to apply to Internet-related services.

II. The revision of the telecoms directive: the definition of "electronic communication services"

The European Commission announced in 1999 in a communication[37] its intention to carry out a general review of the existing legal framework for telecommunications at European level. Within the framework of this general review of the legal framework for telecommunications, the existing directive on the processing of personal data and the protection of privacy in the telecommunications sector will also be revised and updated.
The Article 29 Working Party has already made public some thoughts concerning this revision in its opinion 2/2000, presented by the Internet Task Force and adopted on 3rd February 2000[38].
The text of the European Commission's Communication pointed out that the planned review would pay special attention to the terminology used by Directive 97/66/EC in order to make it clear that new services and technologies are covered by this Directive, thus avoiding possible ambiguities and facilitating the consistent application of data protection principles. In its opinion 2/2000, the Working Party welcomed such a re-examination of the terminology for these purposes.
The proposal for a directive on the processing of personal data and the protection of privacy in electronic communications services was published by the Commission on 12 July 2000[39]. The European Commission press release[40] underlines the fact that one of the objectives of the new package is to ensure the protection of the right to privacy on the Internet.

This proposal no longer refers to "telecommunications services", but to "electronic communications services". The explanatory memorandum to the proposal mentions that this change was necessary to align the terminology with the proposed directive establishing a common framework for electronic communications services and networks[41].

The term "electronic communications services" is not defined in the proposed privacy and telecommunications directive but in Article 2 b) of the proposed directive establishing a common framework for electronic communications services and networks.

---

[36] The book "Net Worth" (op cit.) mentions in page 275: "Because *cookies* can also be used to match browsing habits and preferences, they are increasingly being used to target advertisements to specific people. Indeed, Doubleclick, Globaltrash and ADSmart are examples of companies that use *cookies* to target advertisements to consumers at their enabled websites."
[37] Document COM (1999) 539.
[38] Opinion 2/2000 concerning the general review of the telecommunications legal framework, presented by the Internet Task Force, adopted on 3rd February 2000, WP 29, 5009/00/EN/final.
[39] Document COM (2000) 385.
[40] Commission proposes overhaul of rules for electronic communication, Brussels 12 July 2000, IP/00/749.
[41] COM (2000) 393.

The new definition reads as follows: *Electronic communications services means services provided for remuneration which consist wholly or mainly in the transmission and routing of signals on electronic communications networks, including telecommunications services and transmission services in networks used for broadcasting, but excluding services providing, or exercising editorial control over, content transmitted using electronic communications networks and services.*

The new definition is actually based on the same core idea as the previous one (the transmission and routing of signals on electronic communications services) but the inclusion of a list of examples of services included and excluded from the definition is very helpful as it sheds light on the discussions outlined in the previous section.

It can be concluded from the list included in the new definition that those who provide content transmitted using electronic communications networks and services will not fall within the scope of application of the revised privacy and telecommunications directive. This is confirmed by the preamble to the proposed directive establishing a common framework for electronic communications services and networks (recital 7) in which it is stated that *it is necessary to separate the regulation of transmission from the regulation of content.* It is, however, stated that this separation should not overlook the links existing between them.

The main consequence of this separation is that additional services such as DoubleClick or those which provide content to a *portal* or a website (but not host them) are covered not by this directive, but only by the general one. It also means that *Internet Service Provider*s are covered by the specific directive insofar as they act as Access Providers and provide connection to the Internet, and are only covered by the general directive when acting as content providers[42].

The advantange of the clear separation between regulation of content and transmission is the clarity that it brings with it. In practice,however, it will be less easy to work with such a separation; think for instance about the case of an *Internet Service Provider* that also provides content, by hosting its own *portal site*. This *ISP* will then have to apply the general directive to all its activities and the specific directive (which entails specific obligations) to the activities in which it plays the role of access provider.

Another interesting aspect of the new definition of "electronic communications services" is the reference to the fact that the service should be provided for remuneration. Neither the preamble nor the explanatory memorandum refer to the inclusion of this term or give any guidance as to how to interpret it. This could be intepretated as meaning that Free Access Providers (FAPs) would fall outside the scope of application of the revised privacy and telecommunications directive, as they do not receive remuneration (or at least not financial) from Internet users.

This interpretation is however not correct since it has been made clear in the jurisprudence of the European Court of Justice, when dealing with services in the sense of article 50 (ex article 60) of the EC Treaty[43], that the remuneration does not necessarily has to be paid by the recipient of the service; it can for instance also be paid by advertisers.

In the case of the FAPs those who place advertisements or banners in the Internet pages are the ones who in fact offer a remuneration to the FAPs. It is therefore clear that these services fall under the definition of electronic communications service and therefore under the scope of the directive.

[42] This aspect is not considered in this paper.
[43] Case C-109/92 Wirth [1993] ECR I-6447 , 15.

It would however be desirable to clarify this issue in the text of the directive since no every reader of the text is aware of the interpretation of this term given by the European Court of Justice. This could be done for instance in the preamble to the directive.

III. Other legal provisions applicable

There are also a number of other Community regulations that deal with some aspects related to the Internet. The following instruments can be mentioned: Directive 1999/93/EC on a Community framework for *electronic signatures*[44], Directive 97/7/EC on the protection of consumers in respect of distance contracts[45] and Directive 2000/31/EC on certain legal aspects of information society services (Directive on electronic commerce)[46].

However, most of these regulations do not lay down extensive specific rules for data protection and, in most cases, leave the regulation of this matter to the specific Directives. For instance, the electronic commerce Directive lays down, in Recital 14, that *"the protection of individuals with regard to the processing of personal data is solely governed by Directive 95/46/EC and Directive 97/66/EC which are fully applicable to information society services (...) and therefore it is not necessary to cover this issue in this Directive"*, and in Article 1.5 b) that *"this Directive shall not apply to questions relating to information society services covered by Directives 95/46//EC and 97/66/EC"*.

Recital 14 of the e-commerce Directive underlines the fact that *the implementation and application of this Directive should be made in full compliance with the principles relating to the protection of personal data, in particular as regards unsolicited commercial communications and the liability of the intermediaries. This Directive can not prevent the anonymous use of open networks such as the Internet.*

Nevertheless, Article 8 of the *electronic signature* Directive enacts some specific data protection rules for certification service providers and national bodies responsible for accreditation or supervision. This Article obliges the Member States to ensure that certification service providers and national bodies responsible for accreditation or supervision comply with the requirements of the general data protection directive. Furthermore, this provision states that certification service providers who issue certificates to the public may only collect personal data directly from the data subject, or after the explicit consent of the data subject, and only insofar as it is necessary for the purposes of issuing and maintaining the certificate. The data may not be collected or processed for any other purposes without the explicit consent of the data subject.

The third paragraph of Article 8 of this directive is especially important. It declares that, without prejudice to the legal effect given to pseudonyms under national law, Member States shall not prevent certification service providers from giving a pseudonym in the certificate instead of the signatory's name.

The preamble to this Directive (recital 24) emphasises the importance of certification service providers observing data protection legislation and individual privacy in order to increase user confidence in electronic communications and electronic commerce.

---

[44] Directive 1999/93/EC of 13 December 1999 on a Community framework for *Electronic signature*s, Official Journal of the European Communities, 19 January 2000, L 13/12 to 13/20.

[45] Directive 1997/7/EC of 20 May 1997 on the protection of consumers in respect of distance contracts, Official Journal of the European Communities, 4 June 1997, L 144.

[46] Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce), Official Journal of the European Communities, 17 July 2000, L 178/1 to 178/16.

IV. Application of national data protection legislation and its international effects

Article 4 1. a) and b) of the Directive provide for the application of national provisions of a Member State where:

- "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;
- the controller is not established on the Member State's territory, but in a place where its national law applies by virtue of international public law".

The Directive specifies that the notion of establishment implies the real and effective exercise of activity through stable arrangements, and that the legal form of such establishment (branch or subsidiary with a legal personality) is not a determining factor in this respect.

As stated in article 4 1. c) of the Directive, data collected using automated or other equipment located in the territory of the EU/EEA are subject to the provisions of Community data protection law.

Recital 20 of the Directive provides further explanation: "the fact that the processing of data is carried out by a person established in a third country must not stand in the way of the protection of individuals provided for in this Directive; whereas in these cases, the processing should be governed by the law of the Member State in which the means used are located, and there should be guarantees to ensure that the rights and obligations provided for in this Directive are respected in practice".

While the interpretation of the notion of "equipment" or "means" has given rise to debate about their extent, some examples undoubtedly fall within the scope of application of Article 4.

This will be the case, for example, for a text file installed on the hard drive of a computer which will receive, store and send back information to a server situated in another country. Such text files, named cookies, are used to collect data for a third party. If the computer is situated in an EU country and the third party is located outside the EU, the latter shall apply the principles of the national legislation of that Member State to the collection of data via the means of the cookie.
In such a case, according to article 4 2., the controller will also have to designate a representative in the territory of the Member State, without prejudice to legal actions which could be initiated against the controller himself..

V. Conclusions

- Large masses of personal data to which the data protection directives apply are processed on the Internet.
- The general directive applies in all cases while the specific directive applies to telecommunications services. Determining when it is a telcommunication service is sometimes difficult due to the terminology used in Directive 97/66/EC.
- The revision of the legal framework for telecommunications has helped to clarify the scope of application of the privacy and telecommunications directive. Some aspects might however need some additional clarifications, especially the reference to the need to include remuneration in the definition of electronic communications services.

The interpretation given to this text by the European Court of Justice should be explained in the preamble to the directive to avoid any possible misunderstanding concerning the scope of application of the directive .

- The European data protection legislation has to be applied to data collected using automated or other equipment located in the territory of the EU/EEA.

CHAPTER 4: ELECTRONIC MAIL

I. Introduction

It is not easy to describe the technical basics of e-mail in a few words. This is mainly due to the following facts:

• There are some official *protocols* but, as with the HTTP *protocol*, the degree of privacy risk will depend on the way in which these *protocols* are actually implemented. There are thousands of different e-mail client or server programs and it appears very difficult to draw overall conclusions, since no reliable data are available on the use of such programs.

• The invisible processing operations performed by those programs are, as indicated by the word "invisible", not easy to detect and these programs are becoming so large and complicated that it is almost impossible to be sure that all functionalities, even the most concealed, are listed.

As a result, the following description cannot be considered to be exhaustive, and will not always be representative of what happens daily on tens of millions of personal computers connected to the Internet all over the world.

II. Actors

Several actors are involved in the process of handling an e-mail, and data protection issues need to be considered by each of these actors and at every step of the process. The actors are[47]:

- The sender of a message
- The recipient of a message (holder of an e-mail address)
- The e-mail service provider (Mail Server which stores the e-mail sent to a user until the user wants to get it)
- The software supplier of the e-mail client program for the sender
- The software supplier of the e-mail client program for the recipient
- The software supplier of the mail server program

III. Technical description

Basically, a user who wants to make use of e-mail needs the following:

• An "e-mail client" which is a program installed on the user's pc.

• An e-mail address (an e-mail account)

• A connection to the Internet

[47] The telecoms operator is not specifically involved in the e-mail process but plays a key role in conveying the signals that make every form of electronic mail communication possible. This actor has specific security obligations arising from the directives.

The process of sending an e-mail

A wide variety of "e-mail clients" are available, but they all need to follow the Internet standards. Sending an e-mail basically consists of the following steps:

- The user creates a message in his/her "e-mail client" and fills in the address field of the addressee with the appropriate e-mail address.
- By pressing the "send" button in the e-mail client, the e-mail will be transferred to the mail server of the correspondent (usually an organisation) or to the mailbox at the user's e-mail account by an *ISP*.
- If the e-mail is delivered to the mail server of the organisation, this mail server will transmit the e-mail either directly to the receiver or to a mail relay server ("outbound relaying").
- The e-mail may pass through several mail relay servers until it reaches the mail server of the receiver.
- The receiver is either directly connected to the mail server (e.g. in a local area network) or he/she needs to establish a connection in order to obtain the mail.

E-mail addresses

An electronic mail address has two parts separated by a "@" character, for example john.smith@nowhere.com or subs34219@nowhere.org

- The right part identifies the host where the recipient has an account. It is in fact a DNS name referring to the IP address of the mail server.
- The left part describes the unique identification of the recipient. It is the name by which the recipient is known by the e-mail service. There is no technical obligation at all for this identifier to be the actual name of the recipient. It can be a pseudonym chosen by the recipient or a random code arbitrarily given by the mail server during the process of registering the recipient.

From a technical point of view, identification is not necessary to send a mail. In fact it appears to be just like the real world where anybody can send a letter without giving his or her name. When *spamming*, the sender will not usually use an e-mail account but access the SMTP *protocol* directly. This will allow him/her to remove or change his/her e-mail address.

E-mail protocols

Two *protocols*, in addition to the TCP/IP *protocol*, are used for e-mail:

1. The first is called Simple Mail Transport Protocol (SMTP) and is used to SEND a mail from a client to the mail server of the recipient. The mail is not sent directly to the recipient's client computer because this computer is not necessarily swirched on or properly connected to the Internet when the sender decides to e-mail. This means that to receive a mail, the Internet user must have a mailbox (an account) on a server. This also means that the mail service provider has to store the message and wait until the addressee fetches it.

2. The second is called POP protocol and is used by the recipient to establish a connection with the mail server to check if there is some mail for him/her. To do so, the recipient has to provide his/her mailbox name and a password so that nobody else can read his/her mail.

Usually, e-mail client programs include both *protocols* because an Internet user wishing to send mail also probably wishes to receive an answer.

IV. Privacy Risks

A number of issues raise specific privacy risks.

Collection of e-mail addresses

As stated above, the e-mail address is indispensable in establishing a connection. It is also, however, a valuable source of information which includes personal data on the user. It is therefore useful to find out about different methods of collecting e-mail addresses.

E-mail addresses can be collected in several ways:

- The provider of the "e-mail client" software, which is purchased or obtained free of charge, could ask the user for registration.
- It is also possible to build a code into the client's software which will transmit his/her e-mail address to the software provider without his/her knowledge (invisible processing).
- In some browsers, there have been reports of security holes which allow a website to know the e-mail addresses of visitors. This can be done via a malicious active content using, for example, a *JavaScript.*
- Some browsers can also be configured to send the e-mail address as an anonymous password when opening FTP connections (this, however, is not usually a default setting).
- The e-mail address can be requested by various websites in various situations (e.g. on commercial sites in a purchase order, for registration before entering a chat room, etc.).
- E-mail addresses could be collected in public spaces on the Internet in various other ways[48].
- The e-mail could be intercepted during the transmission of a message.

Traffic Data

It is essential to draw a distinction between the content of an e-mail and traffic data. Traffic data are those data needed by the *protocols* to carry out the proper transmission from the sender to the recipient.

Traffic data consist partly of information supplied by the sender (e.g. e-mail address of the recipient) and partly of technical information generated automatically during the processing of the e-mail (e.g. date and time sent, type and version of "e-mail client").

All or part of the traffic data is placed in a header, which is transmitted to the recipient along with the message itself. The transmitted parts of the traffic data are used by the recipient’s mail server and "mail client" to handle the incoming mail properly. The recipient could use the transmitted traffic data (e-mail properties) for analysis purposes (e.g. to check the routing of the e-mail through the Internet).

[48] Further investigations on *spam* and e-mail address collecting have been carried out by the French Data Protection Authority, better known as CNIL. See especially the CNIL report on Electronic Mailing and Data Protection, October 14, 1999, available at the CNIL website: www.cnil.fr

The following items are normally considered to be included under the definition of "traffic data":

- e-mail address and IP address of sender
- type, version and language of the client agent
- e-mail address of receiver
- date and time of sending the e-mail
- size of the e-mail
- character set used
- subject of the mail (this also gives information about the content of the communication)
- name, size and type of any attached documents
- list of SMTP relays used for the transmission

In practice traffic data are normally stored by the e-mail servers of the sender and the recipient. They could also be stored by the relay-servers in the communication path through the Internet.

As traffic data is not formally defined in Directive 97/66/EC, attention should be drawn to the fact that personal data which are not needed for carrying out the communication or for billing purposes but are generated during the transmission, could be wrongly considered by some Internet actors as traffic data, which they think they can store.

The Article 29 WP dealt with some of the privacy problems related to traffic data in Recommendation 3/99 on the preservation of traffic data by *Internet Service Providers* for law enforcement purposes[49]. The Working Party considers that the most effective means of reducing unacceptable risks to privacy whilst recognising the need for effective law enforcement, is that traffic data should in principle not be kept only for law enforcement purposes and that national laws should not oblige telecommunications operators, telecommunications services and *Internet Service Provider*s to keep traffic data any longer than is necessary for billing purposes.

The Spring 2000 Conference of European Data Protection Commissioners in Stockholm emphasised in its official declaration the fact that, "where traffic data are to be retained in specific cases, there must be a demonstrable need, the period of retention must be as short as possible and the practice must be clearly regulated by law".

E-mail content

The confidentiality of communications is protected by Article 5 of Directive 97/66/EC. Under this provision, no third party should be allowed to read the contents of e-mail between two parties. If the e-mail content is stored at relay-servers during transmission, it should be deleted as soon as it has been forwarded.

If a relay-server is not able to forward the e-mail, it could be stored for a short and limited period on that server, until it is returned to the sender together with an error message stating that the e-mail could not be delivered to the recipient.

The contents of an e-mail are stored at the mail-server until the user's "e-mail client" asks for it to be delivered. In some cases the user can choose to leave the e-mail stored at the mail-server even if he/she has got his/her own copy. If the user has not exercised this

[49] Recommendation 3/99 on the preservation of traffic data by *Internet Service Provider*s for law enforcement purposes, adopted on 7 September 1999, 5085/99/EN/final, WP 25.

choice, the mail must be deleted as soon as the mail server can be sure that the recipient has received it.

If a virus scan is carried out in form of content scanning, it should be set up automatically only for this purpose. The contents must not be analysed for any other purpose and must not be displayed to anybody, even if a virus has been found.

Another privacy risk associated with e-mail is related to the inability of a user to easily and effectively remove an e-mail message that has either been sent or received as the operation of the delete function will not necessarily expunge a mail from the system. It can in that case be relatively easy for another user of the same machine or a system manager in the case of a networked machine to retrieve a message that the original user intended to delete and believes has been removed from the system. This issue is obviously not confined to e-mail but it is particularly significant in this context. In order to address this issue systems should be designed so that the operation of the delete function actually expunges information from the system.

Hardware and software can be used to monitor the traffic on a network. This is called *sniffing*. The *sniffing* software is able to read all the data packets on a network thus presenting in clear text all communication which is not encrypted. The simplest form of *sniffing* can be carried out using an ordinary pc connected to a network using commonly available software.

If *sniffing* is carried out at central knots or junctions in the Internet this could allow for large-scale interception and surveillance of e-mail content and/or traffic data by choosing certain characteristics, typically the presence of keywords. *Sniffing*, as a general and exploratory surveillance activity, even if conducted by government agencies, can only be allowed if it is carried out in accordance with the conditions imposed by Article 8 of the European Convention on Human Rights.

In this context, it is interesting to note the current concerns expressed world-wide about possible monitoring of international communications and the "Echelon" satellite interception system in particular. Global surveillance is today a hot item on the European Parliament agenda [50]. In a report to the Director-General for Research at the European Parliament[51] on the development of surveillance technology and the risk of abuse of economic information, it is said that the “Echelon” system has been in existence for more than twenty years. According to this report, Echelon makes heavy use of the NSA[52] and GCHQ[53] global Internet-style communications networks to let remote intelligence customers talk to computers at each collection site and receive results automatically.
Another controversial surveillance system is Carnivore which, according to the information published by EPIC[54], monitors traffic at the facilities of Internet service providers in order to intercept information contained in the electronic mail of criminal suspects. EPIC states that Carnivore can reportedly scan millions of e-mails each second and is capable of enabling law enforcement agents to intercept all of an ISP customer's digital communications. Serious questions have been raised in the American Congress, in the media and in the privacy community about the legality of Carnivore and its potential for abuse. In response to the public uproar over Carnivore, Attorney General Janet Reno

[50] For more information, see the European Parliament Committee on Citizens' Freedoms and Rights, Justice and Home Affairs: http://www.europarl.eu.int/committees/en/default.htm See also EPIC Alert 7.07, 20 April 2000.
[51] Report Interception Capabilities 2000, May 1999.
[52] National Security Agency, USA.
[53] British counterpart of the NSA.
[54] EPIC Alert 7.15, August 3, 2000.

announced on July 27 2000 that the technical specifications of the system would be disclosed to a "group of experts" to allay public concerns.

The discussion about global surveillance of communications is also on the agenda in the Council of Europe. The Committee of Experts on Crime in Cyberspace released its "Draft Convention on Cyber-crime" on April 27 2000[55]. This convention would facilitate the collection of information by requiring companies that provide Internet services to collect and store information for law enforcement agencies. It would require international exchange of such information between governmental authorities in different fields of jurisdiction, even with those which are not parties to the European Convention of Human Rights or to other instruments of the Council of Europe or the EU in the field of data protection. So far, no requirement on substance to protect the fundamental right to privacy and personal data in third countries receiving personal data about EU citizens is foreseen nor basic principles for meeting fundamental human rights standard such as necessity or proportionality are provided for.
Without wishing to comment on the text of the draft convention at this point, the Working Party would, however, like to reiterate the point of view stated by the European Data Commissioners in a statement made during the Stockholm conference in April 2000. This statement reads as follows: *The Spring 2000 Conference of European Data Protection Commissioners notes with concern proposals that ISPs should routinely retain traffic data beyond the requirements of billing purposes in order to permit possible access by law enforcement bodies.*
The Conference emphasises that such retention would be an improper invasion of the fundamental rights guaranteed to individuals by Article 8 of the European Convention on Human Rights. Where traffic data are to be retained in specific cases, there must be a demonstrable need, the period of retention must be as short as possible and the practice must be clearly regulated by law.

The Article 29 Working Party has dealt with the privacy aspects of interception of communications in its recommendation 2/99[56]. In this recommendation, the Working Party points out that each interception of telecommunications, defined as a third party acquiring knowledge of the content and/or traffic data relating to private telecommunications between two or more correspondents, and in particular of traffic data concerning the use of telecommunications services, constitutes a violation of an individual's right to privacy and of the confidentiality of correspondence. It follows that interceptions are unacceptable unless they fulfil three fundamental criteria, in accordance with Article 8 (2) of the European Convention for the Protection of Human Rights and Fundamental Freedoms of 4 November 1950[57], and the European Court of Human Rights' interpretation of this provision: a legal basis, the need for such a measure in a democratic society, and conformity with one of the legitimate aims listed in the Convention[58].

[55] The text of the draft treaty is available at: http://conventions.coe.int/treaty/en/projets/cybercrime.htm
[56] Recommendation 2/99 on the respect of privacy in the context of interception of telecommunications, adopted on 3 May 1999, 5005/99/final, WP 18.
[57] It should be stressed that the fundamental guarantees recognised by the Council of Europe on the interception of telecommunications create obligations for Member States regardless of the distinctions made at European Union level according to the Community or intergovernmental nature of the fields addressed.
[58] Council of Europe Convention No 108 also stipulates that interception may be tolerated only when it constitutes a necessary measure in a democratic society for the protection of the national interests listed in Article 9 (2) of that Convention and when it is strictly defined in terms of this purpose.

V. Analysis of special issues

Webmail

E-mail systems that use web pages as an interface are collectively referred to as "*Webmail*" (e.g. Yahoo, HotMail, etc.). *Webmail* can be accessed from everywhere and the user does not need to make a connection to a specific *ISP*, as when using an ordinary e-mail account.

Webmail is normally free of charge, but in order to obtain a free account users are often required to supply the provider with personal data. From the investigations carried out by Data Protection Authorities it appears to be the case that many *Webmail* providers sell or share personal data for marketing purposes.

Webmail uses the HTML *protocol* (instead of POP) to read and check the e-mail. In fact the messages are delivered on a classical HTML page. This feature allows the mail service provider to include (graphically speaking, outside the message itself) personalised advertising on the HTML page where the message is presented. *Webmail* is heavily sponsored and many banner advertisements are displayed.

As *Webmail* systems are based on the HTTP *protocol* they can be vulnerable to so-called "Web Bugs", that is, an attempt to unmask the e-mail identity of a person using embedded HTML tags and *cookies*.

Webmail providers should not include invisible *hyperlinks* into webpages where the e-mail account is part of the URL. Otherwise, by doing this they will help transmit the e-mail address of the data subject to the advertising company. This is another way in which the user's privacy is invaded by invisible processing.

Directories

There are several services on the Internet supplying directories of e-mail addresses. These public directories are subject to the same rules as those applicable to telephone directories and other publicly available data, as will be explained in Chapter 6. Within the existing legal framework, users must be given at the very least the right to opt out of having his/her data processed, in accordance with Directive 95/46/EC (Article 14) and Directive 97/66/EC (Article 11).

It should be noted that the draft revised directive concerning the processing of personal data and the protection of privacy in the telecommunications sector harmonises the obligations of data controllers in this respect, and provide for an opt-in right in directories to be exercised by data subjects. The Working Party considers this an important improvement.

Spam

"*Spam*" can be defined as the practice of sending unsolicited e-mails, usually of a commercial nature, in large numbers and repeatedly to individuals with whom the sender has had no previous contact[59]. The Article 29 Working Party has already dealt with this issue in its opinion 1/2000 on certain aspects of electronic commerce[60].

The problem from the citizen's point of view is threefold: firstly, the collection of one's e-mail address without one's consent or knowledge; secondly, the receipt of large amounts of unwanted advertising; and thirdly, the cost of connection time.

[59] See CNIL report on Electronic Mailing and Data Protection, October 14, 1999.
[60] Opinion 1/2000 on certain data protection aspects of electronic commerce, Presented by the Internet Task Force, Adopted on 3rd February 2000, 5007/00/EN/final, WP 28.

E-mail addresses can be collected in public directories or by means of different techniques. For instance the e-mail address can be delivered by the user him/herself when buying goods or services via the Internet. In other cases, e-mail addresses supplied by the user to one supplier can be sold by that supplier to a third party.

In the opinion of the Working Party, the rules of the data protection directive provide a clear answer to the privacy issues raised by *spam* and give a clear picture of the rights and obligations of those involved. Two situations should be distinguished:

- If an e-mail address is collected by a company directly from a person with a view to electronic mailing by that company or a third party to which the data are disclosed, the original company must inform the person of those purposes at the time of collecting the address[61]. The data subject must also, as a bare minimum, be given at the time of collection and at all times thereafter the right to object to this use of his/her data by easy electronic means, such as clicking a box provided for that purpose, by the original company and later on by the companies which have received data from the original company[62]. Certain national laws implementing the relevant directives even require the company to obtain the data subject's consent. The requirements of the e-commerce Directive's Article on unsolicited commercial communications complement these rules at a technical level by imposing the obligation to consult a register on the service provider, without detracting in any way from the general obligations applicable to data controllers.
- If an e-mail address is collected in a public space on the Internet, its use for unsolicited electronic mailing would be contrary to the relevant Community legislation - for three reasons. Firstly, it could be seen as "unfair" processing of personal data under the terms of Article 6(1)(a) of the general directive. Secondly, it would be contrary to the “purpose principle” in Article 6(1)(b) of that directive, in that the data subject made his/her e-mail address public for a quite different reason, for example participation in a newsgroup. Thirdly, given the cost imbalance and the nuisance to the recipient, such mailing could not be regarded as passing the balance of interest test in Aticle 7(f)[63].

A particular feature of electronic commercial mailings is that while the cost to the sender is extremely low compared to traditional methods of direct marketing, there is a cost to the recipient in terms of connection time. This cost situation creates a clear incentive to use this marketing tool on a large scale, and to disregard data protection concerns and the problems caused by electronic mailing.
The cost of unsolicited e-mail is borne both by the recipient and by the Internet Mail provider of the recipient (it can be the *webmail* server or the *ISP* of the recipient).
The mail server has to store unsolicited e-mails for a while. The recipient has to pay[64] to download a message that he/she does not want and loses time in sorting received messages and throwing away unsolicited mails, especially when *spamming* messages are not identified as such in the subject line (typically by putting an “ADV:” advertisement

[61] Directive 95/46/EC, Article 10
[62] Directive 95/46/EC, Article 14.
[63] That provision (one of several possible legitimate grounds for processing) requires data processing to be "necessary for the purposes of legitimate interests pursued by the controller . . . except where such interests are overridden by the interests of fundamental rights and freedoms of the data subject".
[64] The telecoms operator if the user is using a *modem*. Otherwise, if the user is using a leased line, even if the cost does not rise immediately (it is a flat fee) due to a spam message, it is clear, from a macro economic viewpoint, that the traffic overheads linked to massive spam are charged to the *ISP*s with subsequent consequences on the price of leased lines.

code in the first characters of the subject line). It is estimated that *spam* (also known as unsolicited electronic junk mail) now constitutes ten percent of all world-wide e-mail[65].

VI. Confidentiality, security aspects

E-mail offers the same possibilities for communication as traditional mail, so the same rules apply as to the secrecy of the correspondence.

Everyone has the right to send a mail to everybody else without that mail being read by a third party. Article 5 of Directive 97/66/EC, which covers communications and related traffic data for example sent by e-mail, lays down obligations as to the confidentiality of communications. In addition to these obligations, Article 4 of the same directive obliges the providers of telecommunications services to take appropriate technical and organisational measures to safeguard the security of their services and to inform users about a particular risk of a breach of security and any possible remedies, including the costs involved**.**

In the off-line world, everyone has the possibility of sending a letter anonymously or under a pseudonym. In order to be able to send anonymous e-mail, the user can obtain an anonymous e-mail address from several providers of such a service.

From the user's point of view, a number of issues are relevant depending on the type of e-mail:

- Confidentiality, which is protection of the transmitted data to prevent eavesdropping. One possible way to guarantee confidentiality is *encryption* of the message to be sent.

Encryption and decryption are based on programs supplementing ordinary e-mail programs (plug-ins) or e-mail programs and browsers offering these facilities. The strength of the *encryption* depends on the algorithms and key length used.

- *Integrity* which is a guarantee that information is not altered accidentally or on purpose. *Integrity* can be obtained by calculating a special code on the basis of the text and transmitting this special code which is encrypted along with the text itself. The receiver can then decrypt the code and, by re-calculating the code, check if the message has been modified.

- *Authentication* which guarantees that a user is who he/she claims to be. *Authentication* can be verified by exchanging *digital signature*s based on *digital certificate*s. These certificates do not need to mention the real name of the user. They can mention pseudonyms, as stipulated in Article 8 of the *electronic signature* directive[66].

VII. Privacy-enhancing measures[67]

Two kinds of tools deserve mention in this chapter: e-mail filters and anonymous e-mail[68].

1) E-mail filtering screens a user's incoming e-mail and only lets through e-mails that he/she has indicated he/she would like to receive. These systems are largely used to screen out *spam*.

[65] See the book "Net Worth" (op cit), page 3.
[66] Directive 1999/93/EC of 13 December 1999 on a Community framework for *Electronic signature*s, Official Journal of the European Communities, 19 January 2000, L 13/12 to 13/20.
[67] See Chapter 9 on privacy-enhancing measures for more details.
[68] See the book " Net Worth" (op. cit), page 275 and following.

Nowadays several companies provide tools that Internet users can install on their computer to screen out unwanted e-mail. In addition, several e-mail packages allow users to filter messages as they are received at the desktop.
The most effective filters are those that allow in only certain e-mails. Although this system works for those who have an unchanging network of e-mail correspondents, it would be cumbersome for the bulk of the population because each new e-mail partner would have to be approved.
The more common filtering technologies allows all e-mail in except for e-mail from certain domain names or e-mail addresses or with keywords in the subject line. However, persistent senders frequently change domain name or e-mail address in order to get around these filters, especially because web-based e-mail accounts are often free and easy to join and leave at any time. Finally, it is difficult to effectively filter by using keywords because the likelihood of error is quite high.

2) Anonymous e-mail allows users to offer their e-mail address on-line without having to give away their identity[69]. This service is currently available free of charge on the Internet through a collection of companies providing "remailer" services.

With these services, the remailer strips off a user's identity for delivered e-mail. Replies to the anonymous e-mail go to the remailer, who then matches the anonymous address with the actual e-mail address and delivers the e-mail response securely to the customer.

VIII. Conclusions

From the data protection viewpoint, the following issues regarding e-mail need to be addressed:

Invisible processing performed by "mail clients" and SMTP relays

The data subject should be given the opportunity to remain as anonymous as possible, especially when taking part in discussion fora. It appears to be the case that the e-mail addresses of participants to these fora are very often sent together with the content of the message[70]. This is not in line with Article 6 of Directive 95/46/EC, which limits the processing of information to that which is necessary for a legitimate purpose[71].

Preservation of traffic data by intermediaries and mail service providers

According to Article 6 of Directive 97/66/EC, traffic data must be erased as soon as the communication has ended. The Directive provides for a limited number of exceptions to this principle, for example if further processing is necessary for billing purposes[72].

Interception

The interception of e-mail (communication and related traffic data) is illegal, unless authorised by law in specific cases in accordance with the European Convention of

[69] This paper also refers to this kind of service in Chapter 6 (publications and fora), in its section V on privacy-enhancing measures.
[70] For further details, see Chapter 6 below.
[71] This principle is further developed in Recommendation 1/99 on Invisible and Automatic Processing of Personal Data on the Internet Performed by Software and Hardware, adopted by the Working Party on 23 February 1999, 5093/98/EN/final, WP 17.
[72] See also Recommendation n°3/99 on the preservation of traffic data by Internet Service Providers for law enforcement purposes, adopted by the Working Party on 7 September 1999.

Human Rights and Directive 97/66/EC. In every case, large scale sniffing must be prohibited. The principle of specificity, which is the corollary of forbidding all exploratory or general surveillance, implies that, as far as traffic data are concerned, the public authorities may only have access to traffic data on a case-by-case basis, and never proactively and as a general rule[73].

Storing and Scanning of e-mail content

The content of e-mail has to be kept secret and must not be read either by any intermediary or by the Mail Service Provider, even for so called "network security purposes". If anti-virus scanning software is used to scan attached documents, the software installed must offer sufficient guarantees regarding confidentiality. If a virus is found, Service Provider should be able to warn the sender of the presence of the virus. Even if this is the case, the e-mail service provider is not allowed to read the content of the message or attachments.

The Article 29 Working Party strongly recommends encrypting the content of e-mails. This is particularly important when it contains sensitive personal data. User-friendly tools for encrypting the content of e-mail messages should be available from providers of e-mail services at not additional cost. At the same time, providers should offer users the opportunity to download e-mails from the mail server of the provider to the client of the user through a secure connection. The need for *integrity* and *authentication* should be considered as well.

Unsolicited e-mails (spam)

If an e-mail address is collected by a company directly from a person with a view to unsolicited electronic mailing by that company or a third party to which the data are disclosed, the original company must inform the person of those purposes at the time of collecting the address. The data subject must also be given at the time of collection and at all times thereafter the right to object to this use of his/her data by easy electronic means, such as clicking a box provided for that purpose, by the original company and later on by the companies which have received data from the original company.
If an e-mail address is collected in a public space on the Internet, its use for electronic mailing would be contrary to the relevant Community legislation.

E-mail directories

As in the case of telephone directories, the data subject must presently have at least the ability to opt out, in accordance with the above mentioned principles of purpose limitation (Article 6.1b of Directive 95/46/EC) and the right to opt out of directories (Article 11 of Directive 97/66/EC). Furthermore, the data subject should have the possibility to join a special directory of e-mail addresses not to be used for direct marketing purposes.
It is important to bear in mind that this right to opt out shall be changed into an opt-in right in the current version of the proposal for a Directive on the protection of privacy in the telecommunications sector; this constitutes a substantial improvement for the data subjects.

[73] See in this context the Working Party recommendation 2/99 on the respect of privacy in the context of interception of telecommunications, adopted on 3 May 1999, 5005/99/final, WP 18.

CHAPTER 5: SURFING AND SEARCHING

I. Introduction

Perhaps the most common activity of Internet users is visiting websites for the purpose of collecting information. This involves passively viewing the content of a webpage. It is also possible to interact with websites in a more active way. Often the Internet user has to click through via a *hyperlink*, push on an advertisement on the screen (banner) or fill in further information on a form. All of these activities will be collectively referred to as 'websurfing'. In practice this is done by means of a web browser that connects the Internet user to a webserver somewhere on the Internet.

From a data protection perspective, three major questions can be asked:

- What information on the Internet user's activities is generated during websurfing?
- Where is this information stored?
- What information is requested for services delivered by websites?

The last issue concerns personal data that an Internet user willingly discloses and the corresponding conditions, but will not be discussed here, as this chapter focuses on the personal data inherent in the (technical) process of websurfing. The subsequent steps in the websurfing process are sketched out, and an indication given of the personal data generated.

II. Technical description and actors involved

The process of websurfing

• Telecoms providers. In order to contact a website an Internet user generally contacts the Internet by a telephone connection to an *Internet Service Provider (ISP)*. The telecom provider logs the call to the *ISP*.

• Internet Access Provider. The entry point to the *ISP* is the network access server. This server generally records the Calling Line Identification of the connection. Most IAP's log the login name, login and logout times and the amount of data transferred during a session. It should be noted that in some cases the telecoms provider is also the IAP.

• Allocation of the IP address. Once the contact with the IAP has been established, the IAP allocates a dynamic IP-address for the duration of the Internet user's session[74]. Henceforth all communication during a session is to and from this IP-address. The IP number is carried with all the packets transmitted in all subsequent stages of communication. It should be noted that the allocated IP number is always within a certain range of numbers allocated to the respective IAP. Hence external parties can easily retrieve the IAP from which IP-packets originate[75] [76].

After this, the Internet traffic is sorted at the *ISP* by the so-called port number, which specifies the service and corresponding *protocol*. A request to visit a website is generally done through the HTTP *protocol*. At the *ISP* this traffic is recognised by a corresponding

[74] Sometimes static IP-addresses are used for the same user over a long period. Static IP-addresses are often used when alternative access technologies (ADSL, cable, mobile) are used. Since these are becoming more widespread, the relative use of static IP-adresses is growing.

[75] In some cases, other parties, such as universities, organisations or companies may themselves play the role of ISP.

[76] To some extent, IP-addresses are also allocated geographically.

port number. It may also be transferred directly to a *router* which connects the Internet user with the external websites required.

The request is often transferred to a dedicated *proxy* server. This server logs the request for a certain website. The *proxy* server contains a copy of the content of the most frequently visited websites. If the website requested by the Internet user is in the *proxy* server, this server only needs to prompt the respective website for an update of any changes since the moment the copy was stored in the *proxy*. This measure strongly reduces the amount of data to be exchanged between the *ISP* and the website, since it only communicates the changes instead of the full pages. The *proxy* server may store a detailed list of the visits to websites connected to an IP-address at a given time. These can be linked to an individual user by the IP-address and the logging of the session times.

• *Routers*. On the path between the *ISP* and the website visited, the traffic generally passes through several *routers* that direct the data between the IP-address of the Internet user and the IP-address of the website. With regard to the storage of personal data, these *routers* are considered as neutral elements, even though dedicated facilities could be applied to intercept the Internet traffic at these points.

• Regular websites. Once the connection with the website has been established, the website collects information on the visiting Internet user. All requests are accompanied by the destination IP-address. The website also knows from which page an Internet user has been transferred (the previous page reference, or URL, is known). The information on website visits is generally stored in the 'Common Log File'. All the above mentioned information can be used to create, by means of a log analyser, accumulated information on the traffic to and from a website and the activities of visitors.

Upon connection with a website, some additional information is collected in the communication between the most common browser software used by Internet users and the websites visited. This is often referred to as 'chattering data.' It generally includes the following items[77]:

- Operating system
- Type and version of browser
- *Protocols* used for websurfing
- Referring page
- Language preferences
- *Cookies*

The website has additional gathering power if it posts so-called *cookies*[78]. These are pieces of data that can be stored in text files which may be put on the Internet user's hard disk, while a copy may be kept by the website. They are a standard part of HTTP traffic, and can as such be transported unobstructed with the IP-traffic. A *cookie* can contain a unique number (GUI, Global Unique Identifier) which allows better personalisation than dynamic IP-adresses. Such *cookies* extend the capability of websites to store and 'personalise' information on their visitors. The *cookie* may be re-read on a regular basis by the site to identify a Internet user and recognise him/her when he/she visits again, check possible passwords, analyse the path during a session and within a site, record transactions, such as Articles purchased, customise a site etc.

Cookies can differ in nature: they can be persistent but can also have a limited duration, when they are called "session *cookies*". In some cases, they may be useful for providing

[77] For more details, see Chapter 2 above.
[78] In this case we refer to persistent cookies, i.e. cookies that persist for longer than one session.

a certain service through the Internet or to facilitate the surfing of the Internet user. For instance, certain custom websites rely on *cookies* to identify users each time they return, so users do not have to log into the website each time they check their news.

The privacy implications of the use of *cookies* should however not be underestimated. This issue will be dealt with in the legal analysis section of this chapter.

- *Portal sites*

Because of the growing complexity of the Internet, Internet users often connect to a website via a so-called *portal* site, which provides an overview of weblinks in an ordered way.
Often such *portals* contain links to commercial sites, and could be compared to a shopping mall hosting many stores. The *portal* sites collect information in the same way as websites in general, but may also store information on visits to all the sites 'behind' the portal.
A *portal site* is always hosted by an *Internet Service Provider* and in some cases can belong to the *ISP*. In such cases, the *ISP* has the possibility of collecting data on a user's visits to sites " behind" this *portal* and can therefore create a complete profile of the user.
The Dutch Data Protection Authority (Registratiekamer) concluded in a report[79] about the Internet and privacy, based on investigations into 60 *ISPs* in the Netherlands, that it is possible for the content provider (in this case the *ISP* that owns a *portal*) to know how many advertisements have been placed, how often a user has visited an e-shop, which products he/she has bought and how much he/she has paid for them.

- *Providers of additional services*

The data collected by websites is sometimes (automatically) transferred to a third party to the original communication (e.g. companies specialised in the analysis of web statistics, such as Nedstat). The purpose can be to create accumulated statistical data on visits to the website, which is sold back to the owner of the respective websites. Advertisement banners generally collect information on the websites visited by a person by means of *cookie*-files. Service providers like DoubleClick or Globaltrash accumulate the information on website visits to all the different sites on which they put advertisements. A profile of the Internet users' preferences can be compiled with these data, and subsequently used to customise webpages.

Surfing from the perspective of the Internet user

A PC installed with browser software will in many cases, after starting up, automatically load a selected starting page from the web. This starting page may contain *hyperlinks* that can be activated to visit other websites or search engines. While browsing, the browser programme of the Internet user sends a request to a server (that can be located anywhere in the world) to transmit a specified webpage (marked by its URL) that is hosted by this webserver. By clicking on a *hyperlink* the Internet user in fact downloads the requested webpage to his/her computer.

After having connected to his/her *ISP*, the Internet user generally chooses one of the following approaches when surfing:

- Directly addressing the website required by entering the URL, such as www.amazon.com. The URL also contains the *protocol*.

[79] See the Registratiekamer report (ARTZ, M.J.T. and VAN EIJK, M.M.M.), *Klant in het web: Privacywaarborgen voor Internettoegang*, Achtergrondstudies en verkenningen 17, June 2000, available at: www.registratiekamer.nl This report underlines the fact that in the Netherlands almost each access provider has its own homepage that is also used as *portal* to start surfing.

• Reaching the website via a referring (*portal*) site that contains *hyperlinks* towards other sites. These *portal* services are becoming more popular as the number of webpages is growing and Internet users need more guidance to find interesting material.
• Retrieving relevant sites by first entering a query to a website using a search-engine. Search engines use indexing by means of keywords. The user enters one or more keywords and initiates the search. The search engine then searches for the titles of the corresponding sites and their URL addresses in its own index database. The search engine has the power to assemble personal profiles as it accumulates the search terms entered by an Internet user and the websites consequently visited. The personalisation is often done by means of *cookies*. Several search engines also offer more personalised services whereby an Internet user is required to provide information on personal preferences in order to get, for example, regular updates of websites on a certain topic[80].

Overview of the most relevant data generated and stored in different parts of the websurfing process

	Data generated and/or stored	Remarks
1.Telecoms provider	Traffic data of connection to *ISP*	May be the same party as *ISP*
2. *ISP*: Network Access Server	*CLI*, IP-address, session data	
3. *ISP: Proxy*	Webpages visited by IP-address at a certain time	
4. *Routers*	IP-address	
5. Websites	IP-address Previous page URL Session data (time, type of transaction) Names and sizes of files transferred *Cookies*	Assembled in the 'Extended Common Log File'
6. *Portals*	Collective information about visits to the websites it refers to *Cookies*	Possibility of creating full profiles of users (communication and behavioural data of the user available to the *ISP*)
7. Service Providers (incl. search engines)	Collected log analysis from websites Data/profiles from websites accumulated via *cookies* Search engines: keywords entered by the Internet user	e.g. NedStat e.g. DoubleClick

III. Privacy risks

Millions of Internet users around the word often surf the World Wide Web or search for information on the Internet. These activities are, however, not risk-free from a privacy point of view.

[80] In this context it is relevant to mention the Common Position on search engines adopted by the International Working Group on Data Protection and Telecommunications adopted at the Hong Kong meeting on the 15th of April 1998, available at: http://www.datenschutz-berlin.de/doc/int/iwgdpt/pr_en.htm

In the context of the Internet, a lot of information is collected and processed in a manner which is invisible to the data subject. The Internet user is sometimes not aware of the fact that his/her personal data have been collected and further processed and might be used for purposes that are unknown to him/her. The data subject does not know about the processing and has no freedom to decide on it.[81]

Additional risks exist when data collected during the surfing activities of Internet users can be linked with other existent information on the same user. The fear of such a connection of personal data concerning Internet users has been very present in the discussion on the merger between Internet advertiser DoubleClick and market research firm Abacus Direct.

It was feared that, should the two firms merge, the DoubleClick database containing data on Internet usage habits would be cross-referenced with the Abacus Direct database containing real names and addresses, as well as detailed information on customer buying habits[82].

This merger took place in November 1999. According to the information provided on the Doubleclick website[83], name and address information volunteered by a user on an Abacus Alliance website were to be linked by Abacus through the use of a match code and the DoubleClick *cookie* with other information about that individual.
Information in the Abacus Online database includes the user's name, address, retail catalogue and online purchase history, and demographic data. The database also includes the user's non-personally-identifiable information collected by websites and other companies with which DoubleClick does business.
According to DoubleClick, no link has been made up to now between the DoubleClick and the Abacus databases.

New monitoring software

New monitoring technologies are becoming available to *ISP*s which will generate far more information about traffic patterns and content preferences than existed in the public switched telecommunications network (PSTN). Such technologies promise to deliver the Internet equivalent of PSTN call-detail records, and more.
These kinds of software programs are popularly known as E.T. applications *"because once they have lodged in the user's computer and learned what they want to know, they do what Steven Spielberg's extra-terrestrial did: phone home"*[84].
To given an example, Narus, a private software company in Palo Alto, Californa (USA), offers software to *ISP*s that 'monitors the data stream and parses each packet to extract packet header and payload information[85]. Narus claims to work closely with key partners, including Bull, Cisco and Sun Microsystems. This software can be used for the identification and measurement of Internet telephony and other applications (eg, the web, e-mail or IP fax), but it also aims to monitor potentially billable content within the IP

[81] The Article 29 Working Party has already dealt with this topic in its recommendation 1/99, adopted on 23 February 1999: Recommendation 1/99 on Invisible and Automatic Processing of Personal Data on the Internet Performed by Software and Hardware, adopted by the Working Party on 23 February 1999, 5093/98/EN/final, WP 17
[82] See EPIC alert 6.10, 30 June 1999. The same concern was already raised during the case of Harriet M. Judnick v.s. DoubleClick at the Superior Court of the State of California.
[83] www.doubleclick.net:8080/privacy_policy/ This merger is discussed in detail in Chapter 7 on electronic transactions on the Internet.
[84] See the cover-page story of Time magazine by COHEN, Adam on 31 July 2000: *How to protect your privacy: who's watching you? They're called E.T. programs. They spy on you and report back by "phoning home". Millions of people are unwittingly downloading them.*
[85] http://www.narus.com

traffic (eg copyrighted material requiring a royalty or on-demand use of an application, or audio clips). The Narus software reports to *ISP*s in real time on the top websites visited as well as the types of content viewed and downloaded[86].
Alexa[87] is a tool that can be added to a browser to accompany the user while surfing, by providing additional information about the site visited (about the registered site owner, ratings and reviews of the site) and making suggestions on related sites. In return for providing this service to users, Alexa has complied one of the largest databases on patterns of web usage. Amazon paid 250 million US dollars in stock for Alexa in early 1999. In its privacy policy, Alexa states that it collects information on web usage which remains anonymous, by using their web usage logs and *cookie* data.

Amongst other products produced by Alexa is the zBubbles program, an on-line shopping tool that collects surfing data on the user in order to offer product recommendations, comparative shopping advice, etc. According to the information published by Time Magazine[88], zBubbles also sends information back to Alexa when users are not shopping. This product is designed to be installed on the screen during the whole duration of the navigation session, even though most users are not shopping all the time.

Another interesting example of monitoring software is Radiate, formerly known as Aureate. Radiate is an advertising company that works with the makers of *shareware*. It is reported[89] that Radiate's advertisements came with E.T. software that embedded themselves in 18 million people's computers and used their Internet connection to report back on what advertisements people were clicking on. The original version of Radiate's software, which still resides in countless computers, was written to keep phoning home even after the *shareware* that put it there was deleted. Users needed a special tool to delete the file, which the company provided on its website later on.

Presently hundreds of E.T. applications exist. More than 22 million people are believed to have downloaded them[90]. E.T. monitoring software programs are again an example of technologies that process personal data on users without their knowledge (invisible processing): most computer users have no idea that these software programs have been placed in their computers.

Often the makers of these E.T. applications say that, although they are able to collect data about computer users, they do not connect them to individuals. This does not, however, offer sufficient guarantees to the user since, given the commercial value of individualised data, companies that collect them could change their policies at any time. The potential risk of data misuse is still there[91].

IV. Legal analysis

The point of departure for the legal analysis of surfing and searching phenomena on the Internet is that both data protection directives (Directive 95/46/EC and 97/66/EC) apply in principle to the Internet[92].

---

[86] See PALTRIDGE, Sam, *Mining and Mapping Web Content*, in: Info, The Journal of policy, regulation and strategy for telecommunications, information and media, vol. 1, no. 4, August 1999, p. 327-342
[87] http://www.alexa.com
[88] As mentioned in the Article by COHEN, A. in Time Magazine (op cit).
[89] As mentioned in the Article by COHEN, A. in Time Magazine (op cit.).
[90] As mentioned in the Article by COHEN, A. in Time Magazine (op cit.).
[91] As mentioned in the Article by COHEN, A. in Time Magazine (op cit.).
[92] See WP 16, Working document: *Processing of Personal Data on the Internet*, adopted by the Working Party on 23 February 1999, 5093/98/EN/final.

Main provisions of the general directive 95/46/EC: Finality principle, fair processing and information to the data subject

Three of the issues dealt with in the general directive deserve special attention in this chapter: the finality principle, the principles of fair processing and the information to be given to the data subject.

Information to the data subject

On the Internet, data flows happen very quickly and the traditional rules concerning information to the data subject about the processing and the finality are often ignored. In some cases, Internet users are not fully aware of the existence or capacities of the software or hardware through which the processing takes place (for instance *cookies* or E.T. software applications).

The Working Party has dealt with these cases in its recommendation 1/99[93]. In this recommendation, the Working Party underlined the fact that a condition for legitimate processing of personal data is the requirement that the data subject be informed and thus made aware of the processing in question. Internet software and hardware products should provide Internet users with information about the data that they intend to collect, store or transmit, and the purpose for which these are required.
Internet software and hardware products should also enable the data user to easily access any data collected on him/her at any later stage.
The speed of data flows on the Internet cannot be used as an excuse for not fulfilling the obligations of the general directive. In fact, the Internet is a medium that makes it possible to provide quick and simple information to the data subject. Whenever personal data are going to be collected, essential information[94] should be given to the individual in a way which should ensure a fair collection of personal data, i.e., depending on the situation, either directly on the screen or form where the collection takes place, or through a box prompt on the screen (for instance in case of sending of cookies). The occasion should be given to the individual to click somewhere if he/she does not agree to this processing or if/she wishes to have additional information.
Some websites post a privacy policy in which information is given about the data they process, the finalities of the processing and the way in which a data subject can exercise his/her rights. This is however not always the case and, even when privacy policies are posted, they do not always contain all the necessary information.

While being very much in favour of posting accurate and complete privacy policies, the Working Party strongly encourages the provision of information to the data subject directly on the screen or using information boxes prompting on the screen at the point when data are collected without requiring the data subject to take any positive action to access this information, as Internet users do not always read the privacy policies of all the sites they visit when surfing from one to another.

In order to play a serious information role, privacy policies should not be too long, have a clear structure and provide accurate information about the data policy of the site in clear and understandable terms. The work of the OECD in this field (privacy policies generator

[93] Recommendation 1/99 on Invisible and Automatic Processing of Personal Data on the Internet Performed by Software and Hardware, adopted by the Working Party on 23 February 1999, 5093/98/EN/final, WP 17.
[94] The information in the box should at least contain details on who controls the processing, the finalities of the processing and, where applicable, the right to object to the processing.

or privacy wizard) could help achieve these goals, although using the generator does not in itself guarantee compliance with the European Directives.

In practice, privacy policies are on their own unlikely to be sufficient as it is often the case that the posted privacy policies do not contain sufficient information from a data protection point of view. A recent study carried out in the USA by EPIC[95] of the privacy policies of the top 100 e-commerce sites showed that few high-traffic websites offered appropriate privacy protection. In fact, not a single one of them fulfilled important elements of the Fair Information Practices investigated in the survey[96].

Finality principle

The information to be provided to the data subject should in all cases contain ample and clear facts as to the finality of the processing. Article 6 of the general directive prohibits further processing of the data for a non-compatible use.

This principle is especially important for websites collecting information from Internet users about their surfing behaviour, for software programs authorised by the user to monitor their Internet behaviour for a specific purpose but not for other (unknown) purposes, and also for *Internet Service Provider*s.
Navigation data on Internet users should in principle only be collected by *Internet Service Provider*s insofar as they need to provide a service to the user, in this case to visit the sites he/she so wishes. *Internet Service Provider*s sometimes cite the need to keep these data in order to be able to monitor the performance of their systems. It is, however, not necessary to keep identifiable data for that purpose, since it is possible to measure and monitor the performance of a system on the basis of aggregated data.

A recent Registratiekamer report[97] concluded that when *ISP*s keep traffic data at individual level on users, they do not do so in their role as access provider. This information is especially interesting for them for their activities as content providers. It should, however, be made clear that this is a totally different purpose.

It would be useful if the purpose limitation principle could be embedded in technical means. This should also be seen as a form of Privacy-Enhancing Technology[98].

Fair processing

Article 6 of the general directive contains a number of principles aimed at guaranteeing the fair processing of personal data. One of them is the finality or purpose limitation principle, to which the previous paragraphs referred.

This Article also specifies that personal data should be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data are collected. This means that once data are anonymised so that it is no longer possible to link the data to the data subject, they can be used for other purposes - for instance, to measure the performance of the service offered by an *ISP* or to compile a survey of the number of visitors to a website.

Leading search machines keep query logs consisting of a record of queries and other information, including the terms used[99]. The terms used are of interest to businesses

[95] Survey "Surfer Beware III: Privacy Policies Without Privacy Protection", see EPIC alert 7.01, 12 January 2000. Available at www.epic.org/reports/surfer-beware.html
[96] The American Fair Information Practices serve as basic guidelines for safeguarding personal information in the USA.
[97] *Klant in het web: Privacywaarborgen voor Internettoegang* (op cit.)
[98] See Chapter 9 below.

trying to select *meta-tags* for webpages and for gauging on-line demand for content related to a particular product, company or brand name. If no link exists between the query log and the identity of the Internet user who entered the key word, there are no legal obstacles to hinder keeping these aggregate data.

If they are not anonymised, data on searching and surfing on the Internet should not be kept once the Internet session has finished. This aspect will be explained in more detail when dealing with the provisions of the specific privacy and telecommunications directive on traffic data.

When considering the fairness of the purpose of data processing, Article 7 of the Directive should also be taken into consideration. This Article sets out several conditions for fair processing, including the consent of the individual and the balance between the legitimate interest of the data controller and the fundamental rights of the individual. This balance of interests should always be borne in mind by the data controller when collecting data from an Internet user.

Main provisions of the specific privacy and telecommunications directive

As can be seen in the table presented in Chapter 3, there are some provisions of the telecommunications directive which are especially relevant to surfing and searching on the Internet.

Even if the title of Directive 97/66/EC refers to the telecommunications sector in general, it is clear that the terminology used in the text itself is chosen on the basis of ISDN technology. Most of the provisions of this directive use terms such as “calls”, which allude to traditional and ISDN telephony and make it more difficult to apply to Internet services. Nevertheless, it is usually possible to include Internet services within the scope of application of the directive although, as can be seen from the following paragraphs, some difficulties have to be faced.

Many of these terminology problems are, however, solved in the text of the proposal for a revised directive of 12 July 2000[100]. In this proposal, a number of definitions are updated to ensure that all the different types of transmission services for electronic communications are covered, regardless of the technology used.

The references to the term "calls" are now limited to cases in which the legislator specifically wishes to refer to telephone calls, as is made clear by the inclusion of a definition of this word in Article 2 e)[101]. In all other cases, the new text refers to "communications" or " communications services".

The following paragraphs will comment on the most relevant provisions of Directive 97/66/EC. Where useful, this paper will refer to the changes introduced by the new proposal for a revised directive.

Article 4: Security

Providers of telecommunications services should offer adequate security measures which take into account the state of the art. These measures should be proportional to the risks involved in the specific situation.

---

[99] See PALTRIGDE, S., *Search engines and content demand*, in *Mining and Mapping Web Content*, in: Info, The Journal of policy, regulation and strategy for telecommunications, information and media, vol. 1, no. 4, August 1999, p.330-333.

[100] COM (2000) 385.

[101] "Call" shall mean a connection established by means of a publicly available telephone service allowing two-way communication in real time.

This provision is especially relevant for the providers of *routers* and connecting lines as these facilities carry massive amounts of information.

In the new proposal, this Article remains unchanged except for the replacement of the term "telecommunications service" by "electronic communications services".

Article 5: Confidentiality

National regulations shall ensure the confidentiality of communications. They shall in particular prohibit listening, tapping, storage or other kinds of interception or surveillance of communications, by parties other than users, without the consent of the users concerned.[102]

There are several actors involved in surfing and searching activities on the Internet to whom this Article applies: providers of *routers* and connecting lines, *Internet Service Providers* and telecommunications providers generally.

In principle, this Article refers to the content of the communication. The distinction between traffic data and content is not, however, easy to apply in the context of the Internet, and certainly not when referring to surfing. Surfing data could in principle be regarded as traffic data. However, the Working Party thinks that surfing through different sites should be seen as a form of communication and as such should be covered by the scope of application of Article 5.

The surfing behaviour of an Internet user (navigation data) visiting different websites can in itself reveal a lot about the communication taking place. By knowing the names of the websites visited, one can in most cases gain a fairly accurate picture of the communication which has taken place. Furthermore, it is then straightforward for anyone armed with the traffic data to visit the site and see exactly what content was accessed.

The Working Party thinks, therefore, that the surfing data of an Internet user should receive the same level of protection as “content”. This form of communication should therefore remain confidential. In this sense *clickstreams* can be considered as falling within the scope of application of this Article.

The new proposal for a revised directive defines "traffic data" in Article 2.1c): *"traffic data" shall mean any data processed in the course of or for the purpose of the transmission of a communication over an electronic communications network.* Navigation data would therefore fall within this definition and be considered as traffic data.

The revision of this Directive has brought major improvements by extending the scope of Article 5 to cover not just the content of the communication but also the related traffic data. By giving equal protection to content and related traffic data the (sometimes difficult) distinction between these concepts becomes less important. The Working Party welcomes this improvement.

Article 6: Traffic and billing data

Traffic data must be erased or made anonymous upon *termination of the call*. In order to interpret this Article in an Internet context, it is necessary to define what can be considered as traffic data and what can be seen as the content of the communication.

[102] See in this respect the Working Party recommendation 2/99 on the respect of privacy in the context of interception of telecommunications, adopted on 3 May 1999, 5005/99/final, WP 18.

This Article seems to be closely related to circuit-switched telecommunications connecting two or more communicating parties. Traffic data are created in the process of establishing and maintaining this connection. This makes the application of this Article in an Internet context especially difficult.

The following applies to Internet traffic: packets that are transmitted are 'wrapped' in several '*protocol*' headers (for instance, TCP-header, IP-header and Ethernet-header). These *protocol* headers are read in every knot (*router*) a packet passes through, to decide where the packet is to be sent next. There does not, however, seem to be any need for every interlying knot to store any header information after a packet has been transmitted.

Processing of header information (which might also include data on the content of the packets) should be considered as traffic data in the sense of Article 6 of Directive 97/66/EC and should therefore be made anonymous or erased once these data are no longer needed to maintain the communication; in other words, as soon as the website is accessed by the Internet user.

There is no doubt that data such as the session login data (login and logout times, amount of data transferred, time of starting and ending the session and so on) should be included within the scope of application of Article 6.

The list of websites visited by an Internet user (surfing behaviour) must in all cases be considered as traffic data (and possibly be given the same protection as content). Above all, this list should in principle be erased *upon termination of the Internet session*.

It is interesting to note that a record of a user's own surfing activities is kept in his/her personal computer. This can be a problem when several people share the same computer.

The Working Party has in the past given its views on the issue of *ISP*s preserving traffic data for enforcement purposes[103]. This recommendation states that traffic data which are not necessary for billing should not in principle be kept. In the case of free *ISP*s, there would then be no need to keep traffic data as they do not need it for billing any longer than they need for their normal operations.

The revised directive replaces the terms "upon termination of the call" by "upon completion of the transmission", which makes things much clearer. Surfing behaviour should therefore be erased once the Internet connection has ended.

The new text introduces the possibility of further processing for the provision of value-added services or for marketing one's own electronic communications services if the subscriber has given his/her consent. The term "value added service" is not, however, defined in the context of this proposal; the Working Party feels it is necessary to clarify what this definition should include in order to guarantee the limitation of the purpose and limit new risks to privacy. Similarly, the Working Party recommends that a "necessity test" be included concerning the possibility of processing traffic data for the provider's own marketing[104].

Article 8: Calling and connected line identification

There are no calling lines on the Internet to be identified or not. There is no separate routing channel by which the identity of the calling party can be indicated before the connection has been established.

[103] Recommendation 3/99 on the preservation of traffic data by *Internet* Service Providers for law enforcement purposes, adopted on 7 September 1999, 5085/99/EN/final WP 25

[104] See opinion 7/2000 of the Working Party, adopted on 2 November 2000, WP 36.

On the Internet, the IP-address cannot be separated from the communication (the packets), so the concept of *CLI* is not directly applicable.

Technically speaking, it is not possible to provide Internet-related telecommunications services without transmitting and using the IP address used by the Internet user during a session.

It can therefore be concluded that Article 8 of the telecoms directive cannot be applied to IP addresses in the same way as it is applied to telephone numbers.

The proposal for a revised directive of 12 July 2000 follows this line of thinking. The wording of this Article remains practically unchanged and refers to "calls", a concept that is reserved for telephone services in the new text.

V. Privacy-enhancing measures

Privacy protection when web surfing can be made effective in several ways. Here are some options for improving the privacy protection of the user[105].

Firstly, many personal data retrieval methods are based on the use of *cookies*. The browser software used by the Internet user makes it possible to refuse the posting of *cookies* on his/her hard disk, on a case-by-case basis or systematically. It should be noted, however, that more and more websites are only offering a full service if the *cookie* function is enabled.

On 20 July 2000, Microsoft announced that it was introducing a beta security patch for the next version of Internet Explorer to allow for better management of web *cookies*[106]. The test version of the patch should be available to the public by the end of August.
According to preliminary information, the patch will offer several features to allow users to control *cookies* more effectively. The browser will be able to differentiate between first-party and third-party *cookies* and the default setting will warn the user when a persistent third-party *cookie* is being posted. Persistent third-party *cookies* are heavily used by Internet advertisers, such as DoubleClick or Engage, to track computer users' activities. In addition, the new functionality will allow Internet users to delete all *cookies* with a single click and will make information about security and privacy more easily accessible. The security patch does not, however, increase consumer control over the use of first-party *cookies* prevalent on commercial websites.
The *cookie* management features follow on the heels of other recent security patches issued by Microsoft to correct data leak issues. In May 2000, the company released a patch for the popular Outlook program that would turn off *cookies* in e-mail messages.
It is however regrettable that this technology still does not enable the site originating the cookie to indicate immediately the finality for which the cookie will be used.

Secondly, the *ISP* may positively contribute to the Internet user's privacy by limiting the personal data stored to the minimum required to establish communication and maintain technical performance. In particular, in many cases it is possible for the *ISP* to hide the IP-number of an Internet user from a website by referring to that site from a special *proxy* server. In that case only the masqueraded IP number allocated by the *proxy* server is transmitted, while the address of the Internet user is kept with the ISP. Such services are, however, rarely offered as a standard service.

Thirdly, it is possible for some *portal* sites to act as *trusted parties* which guard the user's personal data. Such 'infomediaries' may act as vigilantes who only supply

[105] See Chapter 9 on privacy-enhancing measures for more details.
[106] EPIC Alert 7.14, July 27, 2000.

personal data to websites that respect the Internet user's privacy, or they may 'barter' the personal data submitted for certain benefits with the full information and consent of the Internet user[107]. This last option should however be viewed with caution.

The most rigorous method is for the Internet user to choose services that intentionally hide his/her IP-address from the websites visited. Some 'anonymiser' websites and corresponding dedicated software products are available to hide the Internet user's IP-address by redirecting the communication across dedicated servers that substitute the IP-address with another.

The existence of new software monitoring E.T. programs obviously raises new questions about possible ways of protecting against these programs. One possible – but not easily workable – protection method[108] would be to physically segment computer hard drives into public and private areas, so that downloads do not have access to information which people want to keep confidential. In any case, extreme care is recommended when downloading applications from the Internet or from e-mail.

VI. Conclusions

- It is necessary to provide anonymous access to Internet to users surfing or searching in the Net. Therefore, the use of *proxy* servers is higly recommended.
- The increasing use of monitoring software is a trend that should be considered and given the necessary attention as it can have serious consequences for the privacy of the Internet users.
- Some of the concepts and definitions used in the present wording of the telecoms directive are not easy to apply in the context of Internet-related services.

  -The traditional separation between content and traffic data cannot be easily applied to Internet activities, particularly not in the context of surfing. On one hand, the concept of traffic data should be broadly interpreted to include header data as well as all login data. On the other hand, surfing behaviour data should be given the same level of protection as content data.

  - The provisions on *CLI* would also need to be reviewed in the context of the Internet.
- The revision of this directive has led to a big improvement on the first of these points by extending the scope of Article 5 to include not just the content of the communication but also the related traffic data, thus giving equal protection to both. The Working Party welcomes this improvement. The second problem has also been solved by making it clear that this provision only applies to telephone calls and not to the Internet.

  The revision of the directive has greatly increased its clarity by adapting the terminology to the present broader context, thus facilitating interpretation of the existing provisions. The Working Party would however like to point out that the concept of " value added services" needs further specification in order to exclude too broad an interpretation.

[107] See the book "Net Worth" (op cit.) for more details.
[108] As suggested by Cheswick, chief scientist at Lucent technologies, in the Article by COHEN, A. in Time Magazine (op cit.).

CHAPTER 6: PUBLICATIONS AND FORA

I. Introduction

Publications and fora available on the Internet share a common feature in that they make personal data publicly available, with (e.g. public discussion fora) or without (e.g. directories) the participation of the person concerned. The reasons for publishing personal data vary depending on the context. The Internet user can disclose some information because he/she is asked to do so in order to access a chat room, for example, or the information can be published by a third party, such as a public administration, for administrative reasons.
The fundamental question raised by this disclosure of information is the application of privacy principles to data publicly available on the Web. Contrary to a wide spread opinion, the protection of the data protection legislation still applies to data made public. This chapter will pay particular attention to the reasons and the necessity for each publication of personal data, to the purpose of the publication and to the risks of misuse of those data.

II. Technical description

Public discussion fora

The technical aspects of data processing on public discussion fora vary depending on the nature of the forum. Two main kinds of fora can be distinguished: newsgroups and Chats.

Newsgroups

Newsgroups are fora classified by subject, where all data sent by users are stored for a fixed period of time, in order to allow contributions or answers of users on a specific subject.

A question or Article includes a "title" and a "body". The link between an Article and the answer to that Article is a "thread".

Messages are transferred to newsgroup servers using specific *protocols*. The usual processing *protocol* for news is NNTP (News Network Transfer Protocol), although some newsgroups also use the HTTP *protocol*. NNTP processes permanent connections between newsgroups servers, and updates messages automatically. Messages are kept by a newsgroup server on a hard drive, which can be consulted by any person connected. News is presented in HTML format.

Each server compares its list of Articles in every discussion group with the others, and exchange new Articles with them. Such comparisons result in millions of exchanges of data on the Internet.

Given the number of groups, users only store a selected list of groups, and the consultation software only presents the titles of news items, leaving downloading of the body of the Articles to the initiative of interested users.

Chats

There are three main kinds of Internet chat: Internet Relay Chat (IRC), Webpage (*Java*) Chat, and ICQ (I seek you) Chat

1. IRC is the original chat medium on the Internet. It uses a *protocol* allowing users to communicate in real time publicly in a forum with an undefined number of people, or

privately with only one correspondent. Chat rooms depend on the subjects discussed, like newsgroups, but differ in that the channels are cancelled at the end of a discussion.

Due to delays in the transmission of information on the main IRC, independent networks have been created. The main networks are EfNet, UnderNet and DalNet.

2. Webpage Chat makes it possible to chat without a separate program: the only tool required is a recent Internet web browser. There are two kinds of webpage chat: the dedicated webpage chat, available on most of the web portal search sites, and webpage chat set up by an individual on his/her own homepage. While webpage chat is simple to use, it also has limited capabilities: it is only possible to exchange text, and it is not possible either to change colours or send sounds, or to send or receive files, to run scripts or customise anything about the chat interface, unlike IRC.

3. ICQ is a tool which informs the user who is on-line at any time. It informs the user when pre-defined persons (on a personal contact list) log on, and allows him/her to contact them, chat and send messages to them while still surfing the Net – provided all participants are using ICQ. The program can be told to set the user as invisible, away or not available.

Publications and directories

Publications and directories are usually available on the Internet in the form of a database, offering search criteria in order to obtain information on one or several individuals.

The source of information for telephone directories is traditionally the official national directory edited, depending on the country, by the main telecoms operator or an ad hoc company responsible for its compilation, on the basis of the list of telephone subscribers.

E-mail directories are compiled using various means, from the voluntary inscription of Internet users on a list presented by an *ISP*, to uncontrolled collection of e-mails on websites such as newsgroups.

Other forms of publications, such as lists provided by public bodies, are drawn up depending on the subject. They can include, for example, the case-law of a country, with the dates of judgements, courts, location, perhaps even the names of the parties, the judge, and a summary of the case.

Most Internet databases offer several search criteria, allowing personalised access to the information and results structured in different ways. In a telephone directory, a search could be started from a name or telephone number, in a case-law database the criteria could be the date of a judgement, the name of a party, etc.

III. Privacy risks

Public discussion fora

The main risk in terms of privacy[109] results from the accessibility of the personal data disclosed by the Internet user. The accessibility of data can lead to further collection and utilisation for purposes which are not always clearly foreseen by the person participating in the public forum. Nor is the person always aware of the details usually published together with the content of the contribution made on the forum.

[109] The Spanish Data Protection Authority (Agencia de Proteccion de Datos) has addressed this issue in its document "Recomendaciones a los usuarios de Internet" (Recommendations to Internet users), available in Spanish and English on its website: www.agenciaprotecciondatos.org

In the case of newsgroups, for example, the e-mail address of the contributor is usually published together with the name or pseudonym of the person posting the message[110]. Some chat fora display the IP address of a participant's computer, as well as his/her pseudonym. Some *Internet Service Providers* allow for the possibility of attending a forum without being identified by the other participants but also, on the other hand, the possibility of attending but allowing other participants to read a specific profile drawn up by the person concerned.

The personal information available on-line varies from one forum to the other. A general rule is that in order to access a chat room, a detailed identification list is completed at the request of the *Internet Service Provider*, which usually includes the e-mail address, birth date, country, sex and sometimes certain preferencies of the person.

From a technical point of view, the provision of such detailed information is not, however, necessary for the smooth operation of the newsgroup or chat service, in the sense of Article 6 of Directive 95/46/EC.

This registration information could, moreover, lead to further utilisation of the data by the *ISP*, and could be combined with additional details on the person collected on-line in chat rooms.

Two of the main purposes for using the data collected and/or published are:

1. to control the nature of the content broadcast. This is done to ensure that inappropriate content is not made available and/or to establish liability if any of the content proves to be illegal[111]. For that purpose, and in order to keep the content identifiable, data traces are often kept whenever material is contributed, without pre-selection, even though only the e-mail address and possibly the name of the contributor would be sufficient.

2. the compilation of lists of personal data. Personal data can be collected on the Web by means of software which can search the network and draw together all the available data about a named person. The Working Party quoted in its recommendation 3/97[112] from a newspaper Article explaining *how one could compile a detailed biography of a randomly selected individual using such software and exploiting information from all the discussion groups in which the person participated*, including, for example, his/her address, telephone number, place of birth, workplace, favourite holiday destination and other personal interests. These data can be collected and further processed for different purposes, such as direct marketing, but also credit rating, or selling the data to insurance companies or employers. Some Internet sites already offer publicly available search tools which make it possible to find all the messages contributed in newsgroups by one person on the basis of his/her name or e-mail address[113].

[110] The e-mail address often includes the name of the Internet user in its first part, especially when the address is automatically defined by an IAP using the registered name of the user. Most of the time however, the user can change the content of that part of the address and, for example, use a pseudonym. It is also possible to ask for a second address, for which the IAP will allow the user to choose the name.
[111] Perhaps to avoid that liability falling on the service provider responsible for the fora.
[112] Recommendation 3/97 on anonymity on the Internet, adopted by the Working Party on 3 December 1997.
[113] See, for example, the Internet site of Deja: "http://www.deja.com/home_ps.shtml?", which provides a "powersearch tool" offering several search criteria including the author of newsgroup messages. The site mentions that it has the most extensive database of newsgroup contributions on the web.

Publications and directories

The on-line availability of personal information taken from public registers or other publicly available sources such as directories, raises similar questions to those mentioned above. They relate to the further possible use of personal data on a worldwide level for a purpose different from the one for which they were first made publicly available[114].

As has already been stressed, the computerisation of data and the possibility of carrying out full-text searches creates an unlimited number of ways of requesting and sorting information, with Internet dissemination increasing the risk of collection for improper purposes. Furthermore, computerisation has made it much easier to combine publicly available data from different sources, so that a profile of the status or behaviour of individuals can be obtained. In addition, particular attention should be paid to the fact that making personal data publicly available serves to fuel the new techniques of *data warehousing* and *data mining*[115]. Using these techniques, data can be collected without any advance specification of the purpose, and it is only at the point of actual usage that the various purposes are defined[116].

Several specific cases can be mentioned to illustrate this area of concern:

- Whilst case-law databases are public legal documentation instruments, their publication in electronic form on the Internet, providing wide search criteria on court cases, could lead to the creation of information files on individuals. This would be the case if the databases were consulted in order to obtain a list of court judgements on a specific individual rather than to find out about case law.

- Specific information on an individual can also be obtained by combining the data included in separate electronic databases. Names of people not entitled to vote could be obtained in this way by combining the population registers with the electoral rolls.

- Address directories on the Internet usually provide search criteria on individuals not just by name, but also by address and by telephone number. Individuals do not foresee such reverse searches when they consent to the publication of their address in the "paper" telephone directory. The availability of data in electronic form means it could be used for different purposes: e.g., direct marketing, by selecting categories of persons living in the same area (perhaps to sell alarm systems in residential areas), or the identification and filing of a person who telephones a firm for a simple - and to his mind - anonymous request for information.

Publications on the Internet can lead to other forms of collecting personal information, targetting not just personal information included in a chat, a public register or a directory, but also direct information provided in a personal web page. Automatic indexing of those pages by search robots can lead to the compilation of files which include personal information from those pages, and the possible marketing and *spamming* of the author of these pages or of persons contributing to them.

---

[114] See on this subject the contribution of Mr. Marcel PINET, member of the CNIL, at the International Conference of Data Protection Commissioners organised in Santiago de Compostela, Spain, in September 1998, available at www.cnil.fr, onder Internet-Initiatives.

[115] *Data mining* and *data warehousing* involve "digging through tons of data" to uncover patterns and relationships contained within, for example, the business activity and history of an organisation; data warehousing is supposed to provide support for decision-making. Processing the vast amount of information is done with the aid of software allowing easy connection between related information in the database. See the Registratiekamer report (BORKING, J., ARTZ, M. and VAN ALMELO, L.), *Gouden bergen van gegevens: over datawarehousing, datamining en privacy,* Achtergrondstudies en verkenningen 10, September 1998, available at www.registratiekamer.nl

[116] Opinion n°3/99 on public sector information and the protection of personal data, adopted by the Working Party on 3 May 1999.

IV. Legal analysis

Public fora

There are plans to impose obligations upon *Internet Service Providers* in order to limit the risks of unlawful collection of personal data released in chat rooms or newsgroups.

The Council of Europe Recommendation No R (99) 5 for the Protection of Privacy on the Internet[117] offers as a guideline to *Internet Service Providers* that they should inform users of the privacy risks present in the use of Internet before they subscribe or start using the services. Such risks may cover *data integrity*, confidentiality, the security of the network or other risks to privacy, such as the hidden collection or recording of data.

The registration form to be completed by individuals requesting access to a public forum must comply with the provisions of Article 6 of Directive 95/46/EC on the fair processing of personal data, which states that personal data must be collected for a legitimate purpose, and that no unnecesary or irrelevant data may be collected for that purpose.

The legitimate nature of the purpose can be determined with reference to Article 7 of Directive 95/46/EC, which provides, in particular, for the explicit consent of the individual to the processing of his/her personal data, and for the balance between the legitimate interest of the data controller and the fundamental rights of the individual (Article 7 a. and f.)

Users must be informed in a clear and visible way about that purpose, the quality of the data collected and the possible storage period for the data. If the user is given no clear indication of the conditions for processing the data, the absence of a reaction may not be regarded as implicit agreement to further processing of those data by the data controller (e.g. for marketing purposes).

It must be emphasised that service providers do not necessarily need to know the precise identity of the user at all times. Before accepting subscriptions and connecting users to the Internet, they should inform them about the possibility of accessing the Internet anonymously or making use of a pseudonym and using its services anonymously[118].

This principle has been recognised by the Working Party in its recommendation 3/97 on anonymity on the Internet[119]. While there is no possible doubt about the legitimacy of anonymity in situations such as the sharing of personal experiences (victims of sexual offences or persons suffering from alcohol dependency) or political opinions, the Working Party has stressed that the need for anonymity on the Internet goes much further than these specific cases, *because identifiable transactional data by its very existence creates a means through which individual behaviour can be surveyed and monitored to a degree that has never been possible before.*

The control of newsgroups and chats in order to ban inappropriate content should be exerted in accordance with the principle of proportionality laid down in Article 6 of Directive 95/46/EC where the identification and collection of all personal data contributed in a public forum is considered as disproportionate compared with other existing means of control. Other possibilities have been proposed, such as contract

[117] Recommendation of the Committee of Ministers to Member States adopted on 23rd February 1999. Available at www.coe.int/dataprotection/

[118] S. LOUVEAUX, A. SALAÜN, Y. POULLET, *User protection in the cyberspace: some recommendations*, CRID, p. 12, available at http://www.droit.fundp.ac.be/crid/**.**

[119] Recommendation adopted by the Working Party on 3 December 1997.

solutions providing for "content quality", or the involvement of a moderator whose role would be to monitor contributions for illegal and harmful content.

In addition to these fundamental principles, it should be added that the preservation of traffic data by *Internet Service Providers* is very strictly regulated, as it is for telecommunications operators. As a general rule, traffic data must be erased or made anonymous as soon as the communication ends (Article 6 paragraph 1 of Directive 97/66/EC). Telecommunications operators and *Internet* Service Providers are not allowed to collect and store data for law enforcement purposes only, unless required to do so by a law based on specific reasons and conditions[120].

Publications and directories

The Working party has reiterated[121] that European data protection legislation applies to personal data made publicly available, and that those data still need to be protected.

The essential principle applicable to public personal data is the principle of finality or purpose limitation, according to which personal data are collected for specific, explicit and legitimate purposes and must not be subsequently processed in a manner which is incompatible with these purposes (Article 6.1b) of Directive 95/46/EC)

The Working Party has also underlined that personal data made publicly available do not constitute a homogeneous category which can be dealt with uniformly from a data protection point of view: while there may be public access to data, such access may be subject to certain conditions (such as proof of legitimate interest), and to restrictions as to further utilisation (such as utilisation for marketing purposes).

The publication of personal data on the Internet might lead to further processing of the data which the data subject might not expect. Articles 10, 11 and 14 of Directive 95/46/EC stipulate in this respect that the data subject has the right to be informed about the usage of his/her personal data. The data subject shall also be informed about his/her right to object to the processing of personal data for marketing purposes, by simple and effective means.

The idea of a "one-stop shop" to object to the processing of personal data on a single list might offer an interesting solution to the difficulties encopuntered by individuals in objecting to each data processing operation, given their proliferation at national and international level[122].

If the intended purpose of the processing is incompatible with the original purpose, the balance between the right to privacy and the interests of the data controller shall be struck by the imposition of stricter conditions upon the data controller. The latter shall obtain the consent of the data subject or be able to invoke a legal or statutory basis for the processing.

It is, however, not always clear whether the data controller is obliged to respect the data subject's right to object, or obtain his/her consent in order to be able to process data.

The regulation of Internet directories in different countries is an example of such different approaches. The question is whether consent is required before a directory is

[120] Recommendation n°3/99 on the preservation of traffic data by *Internet Service Providers* for law enforcement purposes, adopted by the Working Party on 7 September 1999.
[121] Opinion n°3/99: see above.
[122] This could be particularly useful regarding the dissemination of directories on the Internet. Complaints handled by data protection authorities often relate to the publication of data from a specific country when the person concerned has been registered in an opposition list, but only in his/her own country.

made available in electronic form a directory when it presents different search criteria from those originally foreseen in a paper directory.

Some countries (such as Spain and Belgium) consider that extended search criteria lead to the possibility of processing personal data for purposes which are not compatible with the original purpose, and that no such processing should therefore be allowed without the prior information and explicit consent of the data subject. In other countries (e.g. United-Kingdom) compliance with the right to object foreseen in the Directive appears to be considered in principle sufficient, although it will depend on the fact of whether or not there is a legal duty to publish the information in the directory .

These interpretations of the legal texts lead to differences in the level of protection in EU countries and to practical conflicts when, for example, a directory including personal data on citizens of a country with a higher protection is put on the Internet from a country with a less protective policy.

Such conflicts have been discussed at European level and a common interpretation of the texts by the Working party has lead to an official position which recommends the harmonised application of the principle by EU Member States[123].

Article 12 of the proposal for a revision of Directive 97/66/EC[124] lays down that the individual has the right to decide without cost whether and which of their data are to be included in public directories, for which specified purpose and to what extent. This constitutes a positive step in the right direction, and has been given full support by the Working Party.

V. Privacy enhancing measures

In addition to the legal provisions mentioned above, there are technical solutions which can improve the protection of personal data at different levels.

As a general principle, the Working party points out that browser software should be configured by default in such a way that only the minimum amount of information necessary for establishing an Internet connection is processed[125].

Anonymity on public fora

With regard to anonymity on the Internet and on public fora in particular, the notion of "pseudo-identity" could offer an alternative solution to the question of the balance between legitimate control of abuses and the protection of personal data. Such an identity would be attributed to an individual through a specialist service provider. The anonymity would then be respected in principle, but a link could be reconstructed with the real identity of the individual by the specialist service provider in specific cases, e.g. suspicion of criminal activity. As for e-mail, anonymous remailers either give the user an anonymous address, to which other people can send their mail, which is then forwarded to the real address of the user (sometimes referred to as a pseudonymous server), or they post or mail the sender's message without any trace of his/her name or address[126].

[123] Opinion 5/2000 on the use of public directories for reverse or multi-criteria searching services (reverse directories), WP 33, adopted on 13th July 2000.
[124] In its public version of 12 July 2000, COM(2000) 385.
[125] Recommendation n° 1/99 of the Working party on invisible and automatic processing of personal data on the Internet performed by software and hardware, adopted on 23 February 1999.
[126] These remailers are called Cypherpunk (for the first generation) or Mixmaster (for the second generation, using more advanced techniques) remailers. Well-known anonymous servers on the web were "anon.penet.fi" or "alpha.c2.org". It appears however that both have closed. A new one is "Nym.alias.net". Anonymous messages can also be sent through an HTML document. In this case, the message and the final recipient are sent unencrypted to the WWW server used.

Systematic indexation of data

Tools also exist to ensure that authors of personal pages are not subject to systematic indexation of their pages and the collection of their personal data without them being aware of it. The aim of the *Robot exclusion protocol* is to prevent all or some of the pages of a website being automatically indexed by a search engine[127]. This protocol is identified by most search engines on the Web. The file "robots.txt" inserted in the Internet address contains instructions aimed at search robots stating that some robots are not welcome or that only some identified pages on the site may be read and indexed.

As only a service provider is able to insert a so-called "Robot exclusion *protocol"* in the site address, authors of personal web pages hosted by a service provider can, if they cannot get the service provider to agree to insert such a protocol, include a Robots *Meta-tag* on every page they do not wish to be indexed. The disadvantage of such *Meta-tag* robots is that they are not yet recognised by all search engines on the Internet.

On-line access to public information

The last subject dealt with in this chapter concerns on-line access to public information which is nevertheless still subject to privacy protection rules.

Technical solutions applied to such databases can help limit illegal use of the information they contain:

- Search criteria must be defined in such a way that data can only be used in accordance with the original purpose. The Working Party insisted in its Recommendation of 13 July 2000 on reverse directories that "the data controller (…) has to implement technical and organisational measures which are appropriate to the risks represented by the processing and the nature of the data protected (see Article 17 Directive 95/46/EC). This means for example that the database should be designed in a way that prevents possible fraudulent uses, such as the unlawful modification of search criteria or the possibility of copying or accessing the whole data base for further processing. Search criteria must, for example, be sufficiently precise to only allow for the presentation of a limited number of results per page. The result should be that the purpose to which the subscriber has consented, is also guaranteed by technical means."[128]

- The on-line consultation of databases can be restricted by, for example, limiting the field of the query or the query criteria. It should be impossible to collect a large volume of data using a wide query such as the first letters of a name. It could also be made technically impossible to request court judgements, for example, based on the name of an individual, or to request the name of a person based on his/her telephone number.

For this purpose, technical tools should be configured and used according to the legal principles described in this chapter.

---

[127] Opinion n° 3/99, see above.

[128] The International Working Group on Data Protection in Telecommunications had adopted a similar recommendation on reverse directories at its meeting in Hong Kong on the 15th of April 1998: *if the reverse directories are not forbidden by law, they are services which require the express consent given voluntarily. At least the right to object and the right of access generally recognized by existing national and international rules on the protection of personal data shall be guaranteed; It is in any case necessary to endow the persons with the right to be informed by their provider of telephone or e-mail service, at the time of the collection of data concerning them, or if they have already subscribed , by a specific means of information, of the existence of services of reverse search and - if express consent is not required - of their right to object, free of charge, to such a search.* The whole text of this recommendation is available: http://www.datenschutz-berlin.de/doc/int/iwgdpt/pr_en.htm

VI. Conclusions

In theory, the legal provisions and technical means available offer valuable protection to the data subject as regards the public availability of some of his/her personal data on the Internet. “The principle of finality, according to which personal data cannot be processed for a purpose incompatible with the purpose originally specified, is of major importance with regard to data made public under specific circumstances.

Particular attention shall also be given to the principle of limitation of the period of storage of personal data. Those data should be erased after a reasonable period, in order to avoid the constitution of profiles that gather e.g. messages sent by an individual to a newsgroup during several years.

Those individuals shall be made aware of the duration period foreseen for the storage and the availability on line of such public data.

At the present time, problems reside mainly in the lack of informationfor both data subjects and data controllers about the legal provisions to be observed.

In order to improve the situation, the main objective is to step up efforts to achieve greater transparency on the Internet and to harmonise the interpretation of fundamental principles concerning the data subject's control of his/her data.

Directive 97/66/EC, in its revised version of 12 July 2000, offers a welcome opportunity to harmonise some of these issues.

CHAPTER 7: ELECTRONIC TRANSACTIONS ON THE INTERNET

I. Introduction

Electronic commerce can be defined as "any form of transaction in which the actors interact electronically rather than by physical exchanges or direct physical contact".[129] This definition covers transactions involving the purchase of goods or services and also those used to improve the quality of services, or the provision of new services by private or public organisations.

Given the above definition and bearing in mind that the main purpose of this chapter is to study Internet-related issues, it will focus on transactions that occur through the Internet, leaving aside any other form of interaction carried out by private or public networks.

The impact of electronic transactions is expected to be felt world-wide, as electronic commerce is, by definition, global and enables every company (regardless of size or turnover) to offer and sell its products throughout the world.

Electronic transactions allow organisations to be more efficient and flexible, to work more closely with suppliers and to fulfil the needs and expectations of their customers in ways they had previously only dreamt of.

However, a huge amount of information is needed to achieve all these goals and this could entail the invasion of essential areas of individual privacy.

II. Actors

The main actors involved in electronic transactions are:

- the user, in the context of Directive 95/46/EC, the natural person who wants to buy a product or demands a service [130],
- the telecoms operator, who is not specifically involved in e-commerce transactions but plays a key role in conveying the signals that make every form of electronic transmission of data possible. This actor has specific security obligations arising from the directives.
- the *Internet Service Provider (ISP)* providing access to the Internet,
- the electronic merchant - the entity which offers products or services through the Internet,
- the financial platform needed in most cases and involving both the merchant's bank and the consumer's bank and a payment gateway dealing with the necessary technical aspects to authorise the financial operation and the payment. This payment gateway deals with all the connections among financial institutions enabling the exchange of electronic money by ensuring that all the actors meet the necessary requirements to accomplish the transaction.
- *Trusted Third Parties*. In the most complex and secure cases these are needed to authenticate the parties and provide strong enough *encryption* to ensure the confidentiality of the transaction.

Three different models for electronic transactions can be identified, depending on the forms of trading and the actors or operators involved[131].

[129] Information Society Project Office of the European Commission, *Electronic Commerce - An Introduction* (http://www.ispo.cec.be/ecommerce/answers/introduction.html)

[130] Most electronic commerce transactions (around 90%) are carried out nowadays between companies, i.e. legal persons, which are not covered by Directive 95/46/EC (see Articles 2 a) and 3.1)

1) Online delivery of intangible goods and services. Mainly used by software houses and communications enterprises for which the Internet infrastructure is ideal for the remote real-time distribution and sale of their products. These range from software, video films, games and on-line music to subscriptions to on-line journals, magazines or technical support programmes.

In this case, apart from the obvious savings gained by direct access to the consumers, thus avoiding any dependency on intermediaries, there is a great advantage for companies which engage in this type of commerce. They can obtain precise and accurate knowledge of the final consumer, his/her hobbies, interests and buying patterns.

This category also covers most of the services offered by public sector organisations, such as on-line self-assessed tax payments or returns, electronic applications or requests for welfare payments and follow-up actions.

2) Electronic ordering of tangible goods. This category includes many different types of companies. First of all, large enterprises using the Internet to obtain direct access to the consumer. IT hardware manufacturers or retailers have been the first to use this commercial channel, which is easy to understand due to the nature of the Internet user. Nowadays, an increasing number of enterprises sell clothing, perfumes, books, CD's, flight tickets, etc.

The Internet gives small and medium-sized companies the opportunity to develop new commercial activities on a scale which would be unattainable using their traditional resources. In fact, as some observers have noted, there is a big difference between the initial investment needed to offer a hundred thousand music CDs through an electronic shop on the Internet, and trying to do the same by opening a shop in a city centre.

Furthermore, all electronic commerce sites delivering tangible goods ultimately depend on a logistical organisation to deliver the items to the final consumer at his or her home address. These logistical organisations are currently investing in Internet technologies to support the electronic ordering and tracing of shipments between partner companies and between the logistical company and the final consumer, so that all the participants can find out in real-time where the ordered goods are and when they are expected to arrive. In this context, it is quite possible that certain distributors and logistical experts will decide to merge in the near future to make use of the key information possessed by the logistical companies on the distribution process (collection and delivery addresses mainly).

3) Commercial networks and shopping malls. On-line commerce does not exclude traditional distributors with no substantive knowledge of the new technologies. They have the option to join a structure called Internet malls which give them with the chance to combine their wares in the showcase of an electronic shopping mall. In malls, shops are classified according to categories and visitors use an internal search system to find a list of sites offering the desired product. Advertising banners could be targeted on the basis of keywords typed or shops visited, and the Internet mall provides a secure payment infrastructure for its members.

Depending on their role, Internet shopping malls often collect very detailed and accurate information about the visitors and buyers (shops visited, interests, buying patterns, addresses, personal details and payment information) that can be of great interest in establishing customer profiles when developing advertising or marketing strategies[132].

---

[131] The following classification has been taken from the study by the Commission of the European Communities *"On-line services and data protection and the protection of privacy"*. It can be found at http://europa.eu.int/comm/internal_market/en/media/dataprot/studies/serven.pdf

[132] The way in which this information is collected is explained in more detail in Chapter 5 on surfing and searching

The role of these malls may change in the future if they are integrated into wider sites, the so-called *portals*, which are web "supersites" providing a variety of services including web searching, news, white and yellow pages directories, free e-mail, discussion groups, online shopping and links to other sites.

These modern portals are offering ever greater opportunities for shopping on a world-wide basis, through both classified advertisements and search engines. And nothing prevents these portals from offering, some time in the near future, their own secure payment platforms and intelligent user agents who can search the Web, negotiate prices (including even the privacy terms of a commercial engagement)[133] and conclude agreements on behalf of the consumer.

III. Secure payments

The growing importance of electronic commerce means that payment systems are needed for the sale of goods and services. Concerns about the security risks of sending credit card details over the Internet and the possibility of confidential personal information being disclosed to unauthorised third parties, are two of the limiting factors on the expansion of electronic commerce.

Several methods have been, and are still being, developed to address these concerns. Nowadays, the most common is the Secure Sockets Layer (SSL)[134], which is implemented in the most popular browsers and establishes a secure channel between the consumer and merchant computers. This is achieved by means of *encryption* and *digital certificates*.

The basic operation procedure of SSL works in the following way. Before the merchant's computer (server) can begin a secure connection with the consumer's computer (client), the client needs to ensure that it is connected to a secure server. To verify the identity of the server, the server's *digital certificate* is used. After the server is authenticated, the client and server can encrypt data to each other and ensure the *integrity* of that data, including the credit card number used in the transaction and any other personal details.

It should be noted that SSL does not enable the customer to have control over the subsequent use or processing of his/her personal data made by the merchant, and that the *authentication* of the client is not mandatory, making fraud through the misuse of somebody else's identity a possibility.

In order to deal with these difficulties and provide a totally reliable framework for electronic commercial transactions, some credit card companies have jointly developed a new protocol with the support of the main software developers. The protocol is called Secure Electronic Transactions (SET) and provides for confidential transmissions (using *encryption*), *authentication* of the parties (cardholder, issuer, merchant, acquirer and

---

[133] Opinion 1/98: Platform for Privacy Preferences (P3P) and the Open Profiling Standard (OPS) adopted by the Working Party on the Protection of Individuals with regard to the Processing of Personal Data on June 16, 1998. (http://europa.eu.int/comm/dg15/en/media/dataprot/wpdocs/index.htm). See also the book by HAGEL III, J. and SINGER, M., *Net Worth: the emerging role of the informediary in the race for customer information*, Harvard Business School Press, 1999 and the report *Intelligent software agents and privacy*, by J. BORKING, B.M.A. VAN ECK and P. SIEPEL, Registratiekamer in cooperation with the Ontario Information and Privacy Commissioner, Achtergrondstudies and verkenningen, January 1999, available at www.registratieakamer.nl

[134] A complete description of the SSL system can be consulted in http://developer.netscape.com/tech/security/ssl/howitworks.html and http://home.netscape.com/eng/server/console/4.0/help/app_ssl.htm

payment gateway via digital certificates) and *integrity* and non-revocation of payment instructions for goods and services (through *digital signatures*)[135].

As the aforementioned system is not particularly suitable when a large number of small-value transactions are required, an alternative method called electronic money or e-cash is being developed. The general principle is to download money onto the hard disk of a computer (or, in the near future, on to the chip of a smart card). Each time an on-line payment is made, the user transfers money units (tokens) from his/her computer or smart card to the account of the trader or service provider. There are several competing technologies in this area. The most interesting from the point of view of protecting personal information are the completely anonymous payment systems based on a blind signature mechanism[136]. These mechanisms could prevent the tracing of transactions, as the bank that "signs" the e-cash does not link the consumer to a specific transaction.

IV. Privacy Risks

Regardless of the type of transaction executed or the payment system used, the essential difference between the physical world and the electronic world is that, in the former, there are a lot of activities that can remain anonymous (looking at showcases, walking through various shops, examining different products and, if you pay in cash, purchasing goods), whereas in the latter everything can be recorded, added to previous or freshly generated information and processed almost without cost to produce enriched information on every individual. And it can be done not only without the consent of the citizen concerned, but even without his or her knowledge. Moreover, with the current *data warehouse* and *datamining*[137] technologies, enormous amounts of information can be processed, not only in order to select individuals that meet some requirements or criteria but also to discover hidden relationships among apparently unconnected data, thus making explicit some patterns of behaviour which could be used to take commercial or administrative decisions regarding certain citizens.

In most cases, when a data subject carries out a purchase or engages in a service such as a subscription, it is mandatory to supply personal details to the merchant or service provider in order to authenticate the buyer, give payment guarantees or provide a physical or electronic address for the delivery of the goods or services. So, unless you pay using e-cash or use privacy-enhancing technologies to hide your IP address and buy an intangible good, anonymityis seldom a possibility at the present time on the Web.

---

[135] Using SET during a transaction, the parties involved communicate by means of two pairs of unique and asymmetrical encryption keys: public encryption keys for signing the documents relating to a transaction, i.e. the purchase offer, and private keys including a digital signature for the actual transaction, i.e. the payment instruction, which ensures the integrity of the transmission and that the order will not be revoked. It operates as a dual signature: the two keys interact in such a way that a payment cannot be valid unless the purchase offer is accepted by the merchant, while the actual order is not honoured unless the payment is approved by the financial institution. The trader has no knowledge of the payment instructions while the bank does not have access to the contents of the order. For a detailed functional description of the complex SET protocol, see SET Secure Electronic Transaction Specification Book 1: Business Description that can be found at http://www.setco.org/download.html. See also GARFINKEL, S., *Web security and commerce*, Oreilly associates, June 1997, chapter 12: Understanding SSL and TLS.

[136] For a theoretical discussion of how these systems work see, CHAUM, David "A Cryptographic Invention Known as a Blind Signature Permits Numbers to Serve as Electronic Cash or to Replace Conventional Identification. The Author Hopes It May Return Control of Personal Information to the Individual" http://www.eff.org/pub/Privacy/chaum_privacy_id_Article, which appeared in *Scientific American*, August 1992

[137] See the Registratiekamer report (BORKING, J., ARTZ, M. and VAN ALMELO, L.), *Gouden bergen van gegevens: over datawarehousing, datamining en privacy*, Achtergrondstudies en verkenningen 10, September 1998, available at www.registratiekamer.nl

This chapter will therefore concentrate on the risks associated with the secondary unauthorised use of personal data and those related to a breach of confidentiality or impersonation.

1. One of the more usual secondary uses of personal data is advertising. Once the individual has been identified, whether he or she supplied the information when logging into the server or by means of other technological devices such as *cookies*, previous information about the individual is used to customise advertisements depending on his/her habits, interests, *clickstream* or buying patterns. And not only ads referring to the website owner of services or offers, but also those issued by third parties which have agreements to support the financial cost of running the server by displaying its publicity.

The paradigms of Internet advertising are the techniques used by publicity agences such as DoubleClick. DoubleClick activities are based on supplying advertising space on the Net and making it easy for advertisers to choose the space that will provide a suitable base for their communication activities. The other key element in DoubleClick's success is the IT technology which makes it possible to isolate identification criteria and offer advertisers tools for the individual targeting of users. This technology uses a database containing data on several million Internet users, thus ensuring that only the desired target audience will be contacted by their advertising campaigns.

To achieve this, DoubleClick collects and processes personal data which make it possible to identify users, describe their habits and determine, in real time, those elements of the population that are likely to meet the targeting criteria of current advertising campaigns. DoubleClick assigns a unique identification number to every user that visits one of the websites in the DoubleClick network and posts a *cookie*, which is later used to identify the user when he/she logs onto another DoubleClick site and, according to his/her data, to customise the most suitable ad for him/her. Even if the visitor does not accept the *cookie*, his/her profile can still be created, especially if he/she has a static IP address.

The personal data recorded in the DoubleClick database are: the permanent part of the IP address, i.e. the Net address, domain, country, state (US), postcode, SIC code (Standard Industrial Classification System code, US), size and turnover of the company (optionally), operating system, version number, service provider, identification number (assigned by DoubleClick), referencing of browsing activities (collection and analysis of the sites visited by the user) [138].

DoubleClick merged with Abacus Direct Corporation on November 23, 1999. Abacus, now a division of DoubleClick, will continue to operate Abacus Direct, the direct mail element of the Abacus Alliance. In addition, it was announced that Abacus has begun building Abacus Online, the Internet element of the Abacus Alliance.
According to information placed on the Doubleclick website, Abacus Online portion of the Abacus Alliance will enable U.S. consumers on the Internet to receive advertising messages tailored to their individual interests[139].
With regard to the aforementioned merger, a California citizen filed a complaint in the Superior Court of the State of California seeking an injunction against DoubleClick for unlawful, misleading and deceptive business practices on the Internet which violate the Privacy Rights of the General Public. The complaint also stated that DoubleClick misleads and has misled the General Public " *(...) into a false sense of privacy and security regarding their Internet use, while deceptively acquiring, storing and selling millions of Internet users' most private and personal information for profit. (...) When an*

[138] As mentioned in the study *On-line services and data protection and privacy*, by GAUTHRONET, S. and NATHAN, F., published by Commission of the European Community. Available at http://europa.eu.int/comm/internal_market/en/media/dataprot/studies/serven.pdf
[139] www.doubleclick.net:8080/privacy_policy/

Internet user visits a participating website, a uniquely identified cookie is placed in his or her computer. Then, when that user visits a website that has information about the user's identity (...), the user's identity is linked with the identifying cookie. The defendants, through using the Abacus database, are then able to obtain a potentially vast amount of personal information about the user. In addition, the Internet user's buying habits, responses to advertising and the websites he or she visits are tracked and recorded"[140].

DoubleClick affirms that, following the public reactions to this project of linking their database with the database of Abacus, the effective steps to lauch such matching have not been taken up to now.

Another example of how personal data can be processed in a way which the ordinary Internet user does not expect, is the work carried out by SurfAid, a small company which is part of the IBM Global Services division based in Somers (New York)[141]. This company receives the access log files of its customers on a daily basis and pre-processes these files to find out the route followed by visitors to the client website. Then, some powerful *datamining* tools are used to explore the client's file, which in some cases contains more than one hundred and fifty millions hits, and produce a daily report accessible to the client. Afterwards, the clients can use *OLAP* programs to break down and analyse the information.

2. Another risk that individuals face when conducting electronic transactions, is the breach of confidentiality of the information transmitted. Since the Internet is an open public network with well known *protocols* focusing more on sharing information than on protecting its confidentiality or security, it is not very difficult for anyone with some technical knowledge to find a number of software tools to intercept and disclose the data transmitted on the Internet. It is also possible to impersonate a company or institution to obtain information that might later be used to commit some kind of fraud or crime.

3. There is a new form of trade developing: mobile e-commerce, which is based on the third generation of cellular phones and other handheld devices that can have secure access to e-mail and web pages using a new protocol[142]. Consequently, location and traffic data as well as travelling patterns might be added to the transactional and browsing data to produce an even more accurate profile of the consumer. And, when one takes into account the mergers and concentrations among telecom companies, service providers, *portals* and content companies, the possibility of aggregation, integration and joint processing increases exponentially.

As a simple example of what may happen in the near future, it is foreseeable that advertisements could follow people everywhere through their cell phones or personal digital assistants. "It's a global positioning type of targeting and it's not that far away" a spokesperson for DoubleClick has announced [143].
Another example is the joint project between Yahoo! and CellPoint Systems AB to co-market a person-to-person locator using cell phones. The Yahoo! Find-A-Friend system can be used to obtain information such as: "John is close to Piccadilly Circus, about 3.2km north-west of you" by using the resources of the GSM cell phone network. Even though consent is required to be part of this scheme, the example shows the new

[140] Harriet M. Judnick v.s. DoubleClick, Inc.
[141] WATTERSON, Karen, *La minería de datos ya es una tendencia dominante*; DATAMATION (Spanish Edition), February 2000
[142] Wireless Application Protocol (WAP).
[143] Jane Weaver, MS NBC, 16/04/2000

capabilities present in new telecommunications technologies which allow people to be traced through mobile devices [144].

V. Legal analysis

First of all, it should be remembered that, as was explained in detail in Chapter 3, the data protection rules contained in Directive 95/46/EC and Directive 97/66/EC apply to the Internet and to the personal data processed in electronic transactions[145]. The following paragraphs will focus on those aspects of these legal texts which are especially relevant to the field of electronic transactions.

Lawfulness of the processing: finality principle (Articles 5-7 of Directive 95/46/EC)

The first aspect to consider is the fair and lawful collection and processing of data, including the finality and proportionality principles. In the context of electronic transactions, it is important to consider the fact that personal data might be collected in a way that is invisible to the data subject. The Working Party has frequently stated its concern about all kinds of processing operations presently being performed by software and hardware on the Internet without the knowledge of the person concerned and which are therefore "invisible" to him/her[146].
When personal data are collected from the Internet user, clear information should be given to the data subject about the purpose of the processing, and on the recipients or categories of recipients of this information, so that he/she is able to decide whether to carry out the transaction under the said conditions.

In addition, secondary uses of personal data should also be made explicit and consent must be obtained, should the secondary uses not be considered compatible with the main purpose. Examples of incompatible secondary uses are the communication of transactional data to third parties to allow them to establish buyer profiles for their advertising campaigns[147], or to use *datamining* tools to extract behaviour patterns from the list of names of websites visited by an Internet user.

It should also be noted that the data subject's consent to process his/her personal data in the framework of a commercial electronic transaction is not required for the collection of the data necessary to accomplish the transaction. This in itself is a legitimate ground to process the personal data of the user required for this purpose, as stated in Article 7 b) of the directive. Any other related data, including invisible data which are in no way needed to achieve the transaction, can only be processed on the basis of other legitimate grounds listed in Article 7 of the directive - i.e. unambiguous consent, compliance with legal regulations, vital interest of the data subject or legitimate interests of data controllers which are not overridden by the fundamental rights or freedoms of the data subject. This is also valid for government transactions since the legitimacy of the collection and processing of personal data by public bodies arises from legal regulations[148].

---

[144] For further information, see http://www.cellpt.com/v2/000504.htm

[145] Processing of personal data on the Internet. Working document adopted by the Working Party on the Protection of Individuals with regard to the Processing of Personal Data on February 23, 1999. (http://europa.eu.int/comm/dg15/en/media/dataprot/wpdocs/index.htm)

[146] Recommendation 1/99 on Invisible and Automatic Processing of Personal Data on the Internet Performed by Software and Hardware adopted by the Working Party on the Protection of Individuals with regard to the Processing of Personal Data on February 23, 1999. (http://europa.eu.int/comm/dg15/en/media/dataprot/wpdocs/index.htm)

[147] Directive 95/46/EC, Article 14 (b)

[148] See also Chapter 6 for a discussion on the purpose specification principle applied to publicly available data.

A secondary use that is frequently mentioned by data controllers of personal websites is the technical maintenance and dimensioning of the IT equipment. This is obviously a legitimate concern in order to offer a good service to customers, but one which can be fully met with unidentifiable data, since only aggregate figures are needed to dimension the computers and telecommunication lines. Data controllers may only keep personal data for technical reasons if this is strictly necessary for this purpose and one of the legitimate grounds for processing data is applicable in this case.

Information to the data subject (Article 10 of Directive 95/46/EC)

Furthermore, clear information must be provided by the data controller to the data subject, including the identity of the controller, the purposes of the processing, the recipients of the information, whether answers are obligatory or voluntary and the possible consequences of any failure to reply, and the existence of the right of access and the right to rectify the data concerning the data subject. In the case the data subject is entitled to object to the processing, he/she should be made aware of that.

The information should be given to the data subject either directly on the screen where the information is collected or through a box prompt, as explained in chapter 5.

It is very easy for websites to provide the data subject with this information and to ascertain that the data subject has at least had the opportunity to read it by displaying it as a mandatory part of the transaction process, before any decision has been made by the consumer. In order to be completely sure that the clauses displayed have not been modified later, they can include an electronic signature of the clauses created with the merchant's private key. In this way the user has proof of which conditions he/she agreed to. This idea seems to implement Article 10, paragraph 3, of the e-commerce directive which states that *contract terms and general conditions provided to the recipient must be made available in a way that allows him to store them and reproduce them*[149].

Preservation of personal/traffic data (Article 6 of Directive 95/46/EC and Article 6 of Directive 97/66/EC)

Article 6.1 e) of the directive contains an obligation not to keep identifiable data longer than required for the purpose for which the data were collected.

With regard to traffic data, the strict limitations imposed by Article 6 of Directive 97/66/EC must be observed: traffic data must be erased or made anonymous once the communication (in this case the electronic transaction) has been completed.

The Working Party has addressed the specific issue of the preservation of traffic data by *Internet Service Provider*s for law enforcement purposes in its recommendation 3/99[150]. This recommendation underlines the fact that in principle traffic data should not be kept only for law enforcement purposes and that national laws should not oblige telecommunications operators, telecommunications services and *Internet Service Provider*s to keep traffic data for a period of time longer than necessary for billing purposes[151].

[149] Directive 2000/31/EC of 8 June 2000.

[150] See http://europa.eu.int/comm/dg15/en/media/dataprot/wpdocs/index.htm

[151] See also in this respect the official declaration of the European Data Protection Commissioners in Stockholm above mentioned, according to which, *where traffic data are to be retained in specific cases, there must be a demonstrable need, the period of retention must be as short as possible and the practice must be clearly regulated by law*.

Automated individual decisions (Article 15 of Directive 95/46/EC)

As mentioned before, data related to transactions cannot be kept indefinitely. This is especially the case when data are intended to be used in automated decisions concerning individuals (such as refusing a request or denying the completion of a purchase) based on previously stored data.

If this is the case, appropriate guarantees should be given to the data subject[152]. These guarantees include the right for every person not to be subject to a decision which significantly affects him or her and which is based solely on the automated processing of data, unless agreed under a contract or authorised by law, and the right to know the logic involved in any automatic processing of data concerning the data subject.

Rights of the data subjects (Article 12 of Directive 95/46/EC)

It is also mandatory to establish clear and efficient procedures to allow data subjects to exercise their rights of access, rectification, erasure or blocking. When data subjects exercise their rights, the controller shall provide them with transparent information about whether there is (or not) personal data registered in the data controller's files and, if this is the case, which data are being processed, the source of these, the purposes of the processing, the categories of data concerned and the recipients or categories of recipients to whom the data are intended to be disclosed. This information should be made available in an intelligible form and, in the context of electronic transactions, it is recommended that the information be given through the on-line connection established, providing the data subject has not asked to receive it in any other standard way.

A very important issue regarding access to data related to, or collected through, electronic transactions is the data subject's right to obtain information not just on the basic or primary data, but also on the derived or consolidated information. This means that, if some type of profiling, classification or division into categories has been carried out, or data obtained from third parties has been added, this processed information should also be made available to the individual, as specified in Article 12 a) of the Directive.

Obligations of the data controller: confidentiality and security (Articles 16 and 17 of Directive 95/46/EC and 4 and 5 of Directive 97/66/EC)

When it comes the issues of confidentiality and security, data controllers must take appropriate measures to protect the information supplied by their customers against unauthorised access or disclosure - in particular, when the processing involves the transmission of data on a network, as is the case with electronic transactions on the Internet. These measures must take into account the risks to security and confidentiality, the nature of the data and state of the art technology.

Applicable law (Article 4 of Directive 95/46/EC)

Another issue causing concern regarding electronic commerce on the Internet is the law applicable to the processing of personal data collected from websites outside the EU/EEA. This raises a number of problematic issues which should be analysed on a case-by-case basis. That analysis should, however, bear in mind that the provisions of Directive 95/46/EC clearly apply to processing operations carried out using equipment wholly or partly located in the territory of the EU, even when the data controllers are located outside the Community[153].

[152] See also article 12.1 (a) 3rd paragraph of Directive 95/49/EC.
[153] For further details, see Chapter 3.

VI. Conclusions

- Clear and understandable information shall be offered to the data subject in full compliance with the information principle. More specifically, data protection information which is closely related to the fulfilment of the electronic transaction should be displayed as a compulsory step in the process of the electronic transaction in order to ensure that this information has been made available to the individual. This must be understood regardless of the information given to non-buyer website visitors. As a supplementary measure, a digital signature of the personal data processing conditions should be made available to the data subject so that he/she can check later that the clauses have not been modified.
- The proportionality principle must be fully observed. Only data which is required for the electronic transaction should be collected. In addition, the processing of any data (especially if the data are processed in a way invisible to the data subject) must be justified on the basis of one of the legitimate grounds in Article 7 of the directive.
- When the data subject decides not to give any more personal details than are required for the accomplishment of the electronic transaction, no discrimination should be exercised against him/her in the conditions offered for the transaction.
- No secondary processing must be carried out without the knowledge of the data subject, and full information on the logic involved in these processes must be provided to the data subject when access is sought. Furthermore, there must be unambiguous consent or some other legitimating criteria laid down in Directive 95/46/EC in order to make the processing lawful.
- Subject to existing legal regulations, *encryption* technology should be used to protect, as far as possible, the confidentiality of the electronic transactions and to guarantee the *integrity* of the messages by means of an *electronic signature*.
- Where neccesary, in order to secure transactions, it could be recommended to make use of *digital certificates* technology and, in particular, if a higher level of security is needed, the *digital certificates* could be stored in smart-cards.
- From the data protection perspective, the opportunity to use secure and anonymous payment methods is a key element for privacy on the Internet.
- The collection and processing of personal data using automated or other equipment located in the territory of the EU/EEA are subject to the provisions of Community data protection law.
- With regard to traffic data, the strict limitations imposed by Article 6 of Directive 97/66/EC must be observed and Recommendation 3/99 on the preservation of traffic data by *Internet Service Providers* for law enforcement purposes should be taken into account.

CHAPTER 8: CYBERMARKETING

I. Introduction

The Internet is not just a worldwide information platform, but also a worldwide market place where competing businesses try to attract potential customers. Success depends on reaching as many consumers as possible and especially those really interested in the product or service offered by the business. To achieve this, they use profiles and targeted advertisements which are based on these profiles and launched by banners placed on websites.

Another way of reaching consumers is electronic mailing, and sending large numbers of unsolicited e-mails repeatedly to e-mail addresses (i.e. individuals) found in public Internet spaces is often seen as the most effective way. This unpopular type of electronic mailing is called „spamming"[154].

In both cases, it is necessary to have personal data on the consumers. These data are ofen easily collected from the Internet. Many Internet users do not realise that while surfing they leave behind a large volume of data which can be used to make assumptions about their areas of interest, preferences and behaviour[155].

Targeted advertising can be acceptable to a certain extent, when it is in the consumer's interest. But, if the user does not know which data are collected and by whom, and for what purpose they will be used, he/she will lose control of his/her personal data. It is, therefore, wrong to collect these data without the user's consent and even without his/her knowledge.

II. Technical description

Online Profiling and Advertising[156]

Online profiling can be done in different ways:

- A website creates profiles by collecting data on its customers that are based on the interactions between the website and the customer. This is done by the use of *cookies*, which track the user's actions on the Web. Depending on how the user's browser is configured, he/she might not be aware of the fact that the website is placing a *cookie*. Using the customer's profile, the website will offer the customer products (e.g. books) or references to other websites that may be of interest to this user.
- In the field of "incentive cybermarketing", individuals may take part in a game or competition provided that they deliver personal data as an input for profiles. In this case, the collection of data is normally carried out with the knowledge of the individual and therefore subject to his/her permission[157].
- Network advertising companies (e.g. DoubleClick, Engage[158]) manage and deliver *banner* advertisements[159] (hereinafter referred to as *banner* ads) on a

[154] See Chapter 4: e-mail, section V. Analysis of specific issues, spam.

[155] See Chapter 5: surfing and searching for more details about the data generated during the surfing process.

[156] In this context it is important to mention the Common Position regarding Online Profiles on the Internet, adopted by the International Working Group on Data Protection in Telecommunications at the 27th meeting of the Working Group on 4/5 May 2000 in Rethymnon / Crete. The text of this recommendation is available at: http://www.datenschutz-berlin.de/doc/int/iwgdpt/pr_en.htm

[157] This will, however, only be the case when the website offers sufficient information to the user conerning the data processed, the finality of the processing, the identity of the controller, etc. See Article 10 of the Directive.

[158] For more details on the techniques used by such advertising companies see: *Privacy Risks* in Chapter 5, Surfing and Searching and Chapter 7, Electronic Transactions on the Internet.

contractual basis for numerous websites. The *banner* ads are placed on the requested website via an invisible *hyperlink* to the advertising company.

To provide the customer with the most „adequate“ *banner* ad, the network advertisers create profiles by using *cookies* set via the invisible *hyperlink*. Depending on the configuration of the browser, the user may be aware that the *cookie* is being placed and may or not give his/her consent. The customer's profile is linked to the identification number of the ad company's *cookie* so that it can be enlarged every time the customer visits a website which has a contract with the advertiser.

After having been analysed, the collected data can be supplemented with demographic data (age, gender etc.) and combined with other data characterising the group to which the user obviously - i. e. because of his/her online behaviour - belongs (e. g. interests, behaviour). This analysis and supplementation work can be carried out by special programs (especially *datamining* tools) which are available on the market.

The results of these procedures are very detailed profiles which allow the web enterprise or the network advertiser to predict the tastes, needs and purchasing habits of a consumer and, based on these assumptions, to deliver *banner* ads which match most closely the consumer‘s interests.

When the collected data, gathered through the identification number of the advertiser's cookie, are not linked to identifiable data[160] of a specific person, they can be regarded as anonymous. But under frequent circumstances, e.g. when the customer fills an order form on the web site where the advertiser has placed the banner ad, identifiable data could be linked or merged with existing data already placed on a cookie, and provide for an identifiable profile of the person concerned[161].

Electronic mailing

For a commercial mailing campaign, a company must obtain an extensive and appropriate list of e-mail addresses of potential users. As stated above, it is often quite simple to use the resources available on the Internet.

There are three different ways to collect e-mail addresses from the Internet[162]: direct collection from customers or visitors of websites, purchase or hire of lists provided by third parties[163] and collection from public spaces[164] such as public e-mail directories or e-mailing lists, news groups or chat rooms.

There are some tools available on the Internet to help collect e-mail addresses. These programs search websites or parts of the Usenet which have to be specified in advance by a list of URLs or keywords related to a predefined field of interest (e.g. sports, travel) and subsequently provide all e-mail addresses found on the sites/pages or in the fora. There are a number of services which work as list brokers in collecting e-mail addresses and selling or hiring the e-mailing lists at a very low price.

Furthermore, other tools specialise in sending e-mails as an "e-mail service provider“, i. e. without using an ISP or any other provider offering an e-mail service. These programs

---

[159] Banner advertisements are small graphic boxes which appear above, or are integrated into, the web site content.

[160] It shall be kept in mind that the definition of identifiable data under article 2 (a) of Directive EC/95/46 is very wide: “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”.

[161] See Chapter 3: application of data protection legislation; section I, General legal considerations: personal data on the Internet.

[162] See Chapter 4 on e-mail for more details about collection of e-mail addresses.

[163] These lists can also contain e-mail addresses collected from public spaces on the Internet.

[164] See Chapter 6 on publications and fora.

ensure on one hand that all e-mail spam filters installed by those providers are bypassed and on the other hand enable a fast and automatic operation. If required by the sender, he can use the host-spamming service, where a third party operates the spamming,which is also offered at a low price.

III. Legal Analysis

Different directives may apply to online profiling and electronic mailing.

The data protection directive

The general directive states that personal data must be collected fairly, for specified, explicit and legitimate purposes, and be processed in a fair and lawful manner in accordance with those stated purposes[165].
Processing must take place on legitimate grounds such as consent, contract, law or a balance of interests.[166] Furthermore, the individual has to be informed about intended processing which also includes transmission to third parties before that transmission takes place[167], and given the right to object to the processing of their personal data for direct marketing purposes[168]. The data subject must also have the right to access the data related to him/her and to rectify, erase or block these data[169].

The distance selling directive

The distance selling directive[170] requires that consumers, at the very least, are given the right to object to distance communications operated by means of electronic mail[171].

The specific privacy and telecommunications directive

Directive 97/66/EC gives national legislators the choice of implementing "opt in" or "opt out" rules for unsolicited commercial communications [172]. Cases where automatic calling machines or faxes are used for marketing purposes are subject to prior consumer consent[173]. The definition of automatic calling machines, which is very loosely worded, could easily be applied to electronic mail.

In July 2000, the European Commission produced a proposal for a new directive concerning the processing of personal data and the protection of privacy in the electronic communications sector to replace Directive 97/66/EC.
In this proposal, the article on unsolicited commercial communications explicitly includes electronic mail, which is only permitted in the case of subscribers who have given their prior consent.

[165] Directive 95/46/EC, Article 6.
[166] Directive 95/46/EC, Article 7.
[167] Directive 95/46/EC, Article 10.
[168] Directive 95/46/EC, Article 14.
[169] Directive 95/46/EC, Article 12.
[170] Directive 97/7/EC of the European Parliament and of the Council of 20 May 1997 on the protection of consumers in respect of distance contracts
[171] Directive 97/7/EC, Article 10.
[172] Directive 97/66/EC, Article 12 (2).
[173] Directive 97/66/EC, Article 12 (1).

The e-commerce Directive

The e-commerce directive[174] states that commercial electronic mails must be identified as such[175] and that opt-out registers, in which individuals not wishing to receive such electronic mails may register themselves, must be regularly consulted and observed[176].
Although neither the general directive nor the telecommunications directive explicitly refer to e-commerce, they have to be applied in this area: the recitals and Article 1 paragraph 5 b of the e-commerce directive make it clear that this directive is in no way intended to change the legal principles and requirements contained in the existing legislative framework. It follows that the implementation of the e-commerce directive must be completely in line with the data protection principles defined in the respective legislation. Therefore, national data protection legislation will continue to be applicable to companies responsible for the processing of personal data[177]. Furthermore, Member States may implement regulations embodied in the telecommunications directive and which go beyond the requirements of the e-commerce directive, i.e. commercial communications may be subject to the prior consent of the recipient[178].

IV. Conclusions

The rules laid down in the general directive, the e-commerce directive, the distance selling directive and the telecommunications directive are applicable to the use of electronic mailing for cybermarketing purposes.
Only the general directive applies to online profiling. Although it forms part of e-commerce, online profiling is not dealt with in the e-commerce directive. Furthermore, network advertising is not covered by the revised telecommunications directive either, as providers performing this service are explicitly excluded from the scope of this directive.
It is, therefore, possible to conclude the following :

Online Profiling and Advertising[179]

- Internet Service Providers must inform users about the intended processing of their data before these are collected[180]. This includes the type, scope and storage period and the purposes of the processing, i.e. use for profiling[181]. If the data are transmitted to third parties, this must also be explicitly mentioned.

  This information should also be given in cases where data are collected using pseudonyms or non-personalised identification numbers. In particular, users must be informed before any *cookie* used for profiling is placed. This should be done by a special box (prompt) which is activated even if the browser does not notify the user about the setting of the *cookie*.

---

[174] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market.
[175] Directive 2000/31/EC, Article 7.
[176] Directive 2000/31/EC, Article 7.
[177] Directive 95/46/EC, Article 4.
[178] Directive 97/66/EC, Article 12. Proposal for a new directive concerning the processing of personal data and the protection of privacy in the electronic communications sector, Article 13 on unsolicited commercial communications.
[179] These conclusions are based on the decision reached by the German Data Protection Authorities concerning a specific network advertiser. The *International Working Group on Data Protection in Telecommunications* adopted a *Common Position* which also reflects this decision. See http://www.datenschutz-berlin.de/doc/int/iwgdpt/pr_en.htm
[180] Directive 95/46/EC, Article 10.
[181] Directive 95/46/EC, Article 6.

- Users must, at all times and at the very least, be given the right to object to the processing of their data[182]. As a result, data collected during future use of the Internet may not be used to enrich an existing profile. This also applies in cases where the processing is subject to the user's prior consent.
- The personalisation of profiles must be subject to the informed prior consent of the individuals. They must have the right to withdraw their consent at any time and with future effect.
- Users must, at any time, be given the opportunity to access their profiles for inspection. They must also have the right to correct and erase the data stored[183].

Electronic mailing

- The enterprise collecting an e-mail address *directly from a user* with a view to electronic mailing performed by that enterprise itself or by a third party to which the e-mail address will be disclosed, has to inform the user by adequate technical means of those purposes at the time of collection[184].
- As long as Member States can choose between implementing opt-in or opt-out, enterprises sending commercial e-mails must ensure by adequate technical means that those e-mails can be identified as such by the recipient[185].
- As long as Member States can choose between implementing opt-in or opt-out, before sending commercial e-mails the enterprise must consult opt-out registers, where users indicate that they do not wish to receive commercial e-mails. These entries have to be respected in all cases[186]. The existence of international opt-out registers would be very beneficial.
- Collecting e-mail addresses *from public spaces on the Internet* and using them for commercial e-mailing goes against the relevant Community legislation, i.e. the general directive[187]. Firstly, this practice constitutes unfair processing of personal data[188]. Secondly, it goes against the purpose principle,[189] as persons publish their e-mail address for a specific purpose, e.g. to participate in a newsgroup, this purpose being quite different to that of commercial e-mailing. Thirdly, it cannot be regarded as passing the balance of interest test[190], in view of the fact that the addressee suffers in terms of time, cost and unreasonable disruption.
- Five Member States (Germany, Austria, Italy, Finland and Denmark) have adopted measures aimed at banning unsolicited commercial communications. In the other Member States, either an opt-out system exists or the situation is not fully clear. Companies in opt-out countries may target e-mail addresses not only within their own country but as well to consumers in Member States with an opt-in system. Moreover, since e-mail addresses very often give no indication of the country of residence of the recipients, a system of divergent regimes within the internal market does not provide a common solution for the protection of consumer's privacy. Opt-in is thus a well-balanced and efficient solution in order to remove obstacles to the provision of commercial communications whilst protecting the fundamental right of privacy of consumers.The Working Party thus welcomes and supports the proposal

[182] Directive 95/46/EC, Article 14.
[183] Directive 95/46/EC, Article 12.
[184] Directive 95/46/EC, Article 10.
[185] Directive 2000/31/EC, Article 7.
[186] Directive 2000/31/EC, Article 7.
[187] See *Opinion 1/2000 on certain data protection aspects of electronic commerce* presented by the Internet Task Force (WP 28).
[188] Directive 95/46/EC, Article 6 (1) (a).
[189] Directive 95/46/EC, Article 6 (1) (b).
[190] Directive 95/46/EC, Article 7 (f).

to address unsolicited electronic mail in the same way as automatic calling machines and facsimile machines. In all these situations, the subscriber has no human interface and supports parts or the whole of the costs of the communication. The degree of invasion into privacy and the economic burden are comparable.[191]

[191] See Opinion 7/2000 on the European Commission Proposal for a Directive of the European Parliament and the Council concerning the processing of personal data and the protection of privacy in the telecommunications sector of 12 July 2000 COM (2000) 385, adopted on 2 November 2000, WP 36.

CHAPTER 9: PRIVACY-ENHANCING MEASURES

I. Introduction

The EC data protection directive contains two principles which have direct consequences for the design and use of new technologies:
- its "finality" or "purpose" principle requires that personal data only be used where necessary for a specific legitimate purpose; in other words, personal data cannot be used without legitimate reason and the individual remains anonymous (Articles 6 (1) b and 7).
- its "data security" principle requires that controllers implement security measures which are appropriate to the risks confronting personal data in storage or transmission, with a view to protecting personal data against accidental or unlawful destruction and against 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 (Article 17).

The "finality" or "purpose" principle mentioned above is the underlying motive for the concept of Privacy-Enhancing Technologies (PETs). This concept refers to a variety of technologies that safeguard personal privacy, notably by minimising or eliminating the collection or further processing of identifiable data[192].
Privacy-Enhancing Technologies aim to hinder any unlawful forms of processing by, for instance, making it technically impossible for unauthorised persons to access personal data, so as to prevent the possible destruction, alteration or disclosure of these data.
The practical implementation of this concept requires organisational and technical solutions.

These technologies are often based on the use of a so-called identity protector[193]. An identity protector may be regarded as an element of the system that controls the release of an individual's true identity to various processes within the information system. Its effect is to cordon off certain areas of the system, which do not require access to true identity. One of the most important functions of the identity protector is to convert a user's actual identity into a pseudo-identity, an alternate (digital) identity that the user may adopt when using the system.
Several techiques can be used to introduce an identity protector into an information system; among others, *encryption* techniques involving *digital signature*s, blind signatures, digital pseudonyms and *Trusted Third Parties*.

II. Privacy-enhancing technologies

This section describes and analyses a number of Privacy-Enhancing Technologies[194].

Cookies killers

Two kinds of responses to solcing the privacy problems of cookies are analysed below. The first originated from the Internet industry itself and has been incorporated into the

[192] See the report by HES, R. and BORKING, J. (editors), *Privacy-enhancing technologies: the path to anonymity (revised edition)*, Registratiekamer, in cooperation with the Ontario Information and Privacy Commissioner, Achtergrondstudies en Verkenningen 11, The Hague, November 1998. Available at www.registratiekamer.nl
[193] See the PET report by the Registratiekamer (op cit.) - particularly page 7 and following - for more details.
[194] See also the EPIC online guide to practical privacy tools, available at www.epic.org/privacy/tools.html

main browsers in the market. The second came from various privacy activists or software houses. It consists of tools which make it possiboe to delete all or some *cookies*.

The cookie opposition mechanisms used by the industry

The only visible attempt to solve the problem of *cookies* is the *cookie* opposition mechanism used in common browsers since version 3. It is possible for an aware Internet user to parameterise the browser by choosing between three options:

– to accept every *cookie*
– to refuse every *cookie* or *cookie* not sent back to the originating server (Netscape)
– to be asked on a case-by-case basis

Cookie opposition mechanisms remain insufficient for many reasons:

1. Normally the default setting is the most privacy invasive (accepting all *cookies*) and the average Internet user does not know that the *cookie* is widely used, e.g. by cybermarketing companies to track keywords typed on search engines using invisible processing means.
2. The *cookie* blocking mechanism inhibits the reception of new *cookies* but does not prevent the systematic and invisible sending of *cookies* already received.
3. *Cookies* can be very different in nature: some cookies are useful and non-identifying (e.g. preferred language). Others are identifying but may be used in compliance with privacy regulations. In general, it can be said that session *cookies*[195] are much less privacy invasive than persistent *cookies*. Refusing all *cookies* might not be in the interest of the Internet user.
4. Several websites deny access to users that do not wish to accept *cookies*.
5. Several websites (or the websites invisibly hyperlinked) send trains of *cookies*, and a case-by-case approach will oblige the Internet user to refuse each of them one after the other, causing what has been called "click fatigue", which will lead the user to accept the cookie once and for all, so as not to be disturbed any more.
6. In some cases, the wording of the *cookie* warning[196] seems incomplete and might be misleading .
7. When installing a new browser, the first site (by default the website of the browser producer) to be visited can send a *cookie* before the user has had the opportunity to deactivate the *cookie* feature.

In July 2000, Microsoft announced that it was introducing a beta security patch for the next version of Internet Explorer that would allow for the better management of web *cookies*[197]. According to preliminary information, the patch will offer several features that will allow users to control *cookies* more effectively. The browser will be able to differentiate between first-party and third-party *cookies* and the default setting will warn the user when a persistent third-party *cookie* is being served. In addition, the new functionality will allow Internet users to delete all *cookies* with a single click and will make information about security and privacy more easily accessible. The security patch does not, however, increase consumer control over the use of first-party *cookies* prevalent on commercial websites.

[195] Cookies with no fixed duration will not be stored on the hard disk but only in the RAM memory.
[196] In MSIE 4.O UK, the cookie warning is worded as follows: *"In order to provide a more personalised browsing experience, will you allow this website to put information on your computer? If you click Yes, the website will save a file on your computer. If you click No, the current web page may not display correctly."* The Internet user has then to click on a new button to know the domain (not the sender !) of the cookie and its duration.
[197] EPIC Alert 7.14, July 27, 2000.

Independent programs

Cookie washer, *cookie* cutter, *cookie* master or *cookie* cruncher are some of the freeware or *shareware* programs that every Internet user can download and use on the Net[198]. Similar remarks to those above can be made here:

1. The Internet user has to process his/her own *cookies* files daily on a case-by-case basis because of the different nature of the *cookies*.
2. In the case of *shareware* programs, the Internet user sometimes has to pay to protect him/herself.
3. The *cookies* handling mechanism is not always user-friendly or easy to understand for an average Internet user.

Proxy servers

The *proxy* server is an intermediary server between the Internet user and the Net. It acts as a *web cache*, dramatically improving the performance of the Internet. Many large organisations or Internet Access Providers have already implemented this solution. Each page, image or logo downloaded from outside by an organisation's member is stored on a *cache* and will be instantaneously available to another member of this organisation.

In this case it is not necessary for every member of the organisation located before the proxy server to have his/her own IP address, because they do not directly access the Internet. Furthermore, the *proxy* server will not normally[199] transmit the IP address of the Internet user to the website and can filter the browser chattering. Due to the fact that a *proxy* server handles the HTTP protocol, the *cookies* stored in the HTTP header can therefore be easily removed, changed, or stored by the *proxy* server.

Anonymisation software

Anonymisation software allows users to interact anonymously when visiting websites, by first passing through an anonymising website that disguises their identity[200].

By stopping at an anonymising website before going anywhere else on the Internet, the user can allow personal data, such as the user's IP address, to be withheld from the receiving website. Anonymiser sites also block system data (such as the operating system and browser being used) from being sent to websites, block *cookies* from being deposited into browsers and block *Java* and *JavaScript*, which can access personal data in browsers.

The anonymiser[201] or the "zero knowledge system[202] are good examples of these.

The **Anonymiser** claims to:

- act as an intermediary between the user and the sites he/she visits, concealing the user's identity from invasive tracking measures
- block Internet programs embedded in the web page (*Java* and *JavaScript*) that may damage the user's computer or gather sensitive personal data.

[198] Some of these programs can be found on http://tucows.belgium.eu.net/cookie95.html .
[199] Unfortunately some proxies add the TCP-IP address of the PC they are working for to the HTTP header.
[200] See the book " Net Worth" (op. cit), page 273 and following.
[201] http://www.anonymizer.com/3.0/index.shtml
[202] http://www.zeroknowledge.com

The Anonymiser offers two services, anonymous surfing and anonymous e-mail, and one product, the Anonymising server. The Anonymising server enables anyone to create his/her own anonymising site.

The Internet user sometimes has to pay to take full advantage of anonymous services. He/she always has to connect to the Anonymiser website to use the anonymising services. It means that this service remains very vulnerable to surveillance by a third party. The Anonymiser can provide anonymous services such as surfing, mailing or file transfer.

Technically speaking, the Anonymiser acts as a *proxy* server and will hide HTTP browser chattering and the IP address of the surfer.

The main problem while using this service is that the Internet user has to trust a particular company, and that this company will be aware of everything the user is doing on the Web.

The **zero knowledge system** proposes software called "Freedom". This solution is based on at least three TCP/IP relays combined with heavy (at least 128-bit) *encryption*. Because the TCP/IP is used by every service on the Net, every service is thereby encrypted and anonymised. Each of the three TCP/IP intermediary stations knows only the TCP address of its predecessor. They keep no log book, so that even two relays put together are unable to trace back the information requested or retrieved. Of course, the routing of the information is dynamic and will be likely to change even during a very brief communication. A *cookie* management system seems to be integrated into Freedom.

Another example of this kind of services is offered by **privada.com.** This company offers services that support all network transaction types, including browsing, email, messaging, and soon, commerce. Privada's infrastructure is based on a system of compartmentalisation and encryption.
The user receives a CD-ROM or downloads a client application, PrivadaControl, from his or her ISP. PrivadaControl communicates with Privada network servers that reside at the ISP's premises and functions as the user's personal privacy firewall. PrivadaControl aims at protecting all user information and data from the point of transaction up through the network—ensuring the user's privacy from all parties, including Privada and the ISP.
Using PrivadaControl, the user creates a private, digital account that represents his or her activities online while completely dis-associating all personal user information from online activity. PrivadaControl appears to allow the user to create or delete digital identities, choose between them while interacting on-line, and set attributes and characteristics about themselves.
This system does not block all Java applets, cookies, or Active-X controls but allows the user to decide the level at which personalisation and web services may function. Cookies are placed on centralised servers within the Privada Network, not on the user's personal computer. Any log files or data mining efforts on the part of a website are associated with the user's online identity—not his real identity. Privada claims that users can easily remove any or all cookies that have been set.

The system proposed by **iPrivacy** is presented as permitting anonymous e-commerce, from surfing to shopping to shipping. It enables consumers to browse and search the net privately, purchase online privately and have them delivered without revealing the identity of the recipient. According to the company, not even them would be aware of the true identity of the consumers who use the services. As far as a transaction is concerned,

only the customer and the credit card user would know any personal information about the purchase made online[203].

E-mail filters and anonymous e-mail[204]

These systems have been already described in the e-mail chapter. The following is a summary of their main features.

- E-mail filtering screens a user's incoming e-mail and lets through only those e-mails that he/she has indicated that he/she would like to receive. These systems are largely used to screen out junk mail.

- Anonymous e-mail allows users to offer their e-mail address on-line without having to give away their identity[205]. This service is currently available free of charge on the Internet through a collection of companies providing "remailer" services. With these services, the remailer strips off a user's identity for delivered e-mail.

Infomediaries

An individual can also decide to make use of a so-called infomediary[206]. The infomediary has been described as follows: “An infomediary, or information intermediary, is a trusted person or web-enabled organisation that specialises in information and knowledge services for, about, and on behalf of a virtual community. The infomediary facilitates and stimulates intelligent communication and interaction among the members of the virtual community. It administers and cultivates a proprietary knowledge asset that contains content and hyperlinks that are of specific interest to the community. In accordance with the privacy constraints that are mandated by the virtual community, the infomediary gathers, organises and selectively releases information about the community and its members in order to fulfill the needs of the virtual community…”.

The infomediary is a new kind of business intermediary to help customers capture, manage and maximise the value of their personal data[207]. Consumers have shown that they are willing to release personal information if they can profit by doing so, but they increasingly recognise that they are selling their privacy cheaply to companies that are using it to promote their own interests. The returns of the information they divulge are, in simple terms, unsatisfactory[208].

Infomediaries could help consumers to strike the best bargain with the vendors, by aggregating their information with that of other customers and using their combined market power to negotiate with vendors on their behalf. They act as custodians, agents and brokers of customer information, marketing it to businesses (and giving them access to it) on the consumer's behalf while at the same time protecting their personal data against abuse.

The positive aspect of an infomediary is that, in many cases, it can purchase the desired goods or services and deliver them to the final consumer while leaving him/her cloaked

[203] http://www.iprivacy.com
[204] See the book " Net Worth" (op. cit), page 275 and following.
[205] This paper also refers to this kind of service in Chapter 6 (publications and fora), in its section on privacy-enhancing measures.
[206] http://www.fourthwavegroup.com/Publicx/1635w.htm
[207] One of the most complete studies about this new body is the book "Net Worth: the emerging role of the infomediary in the race for customer information"; HAGEL III, J. and SINGER, M., Harvard Business School Press.
[208] HAGEL III, J. and SINGER, M, op. cit.

in anonymity. The infomediary company can also provide intelligent agents to help subscribers accomplish their task.

Infomediary clients will theoretically have the option of remaining forever anonymous while they browse the Web and make purchases on-line. However, they will be encouraged not to do so because they will be paid a small fee by vendors every time they divulge who they are or what their e-mail address is. This fee may take the form of a monetary payment, or a discount in the price of the product sold.

Clients will also receive cash payments in return for providing selected vendors with access to their information profiles. The amount of the cash payment will depend on the privacy preferences of individual clients. Clients who choose to remain totally anonymous will forgo cash payments in return for assurances about their privacy. Clients who are comfortable with the controls imposed by the infomediary on access to their information and who see the value of selective disclosure to vendors, could generate cash payments for themselves.

In conclusion, it can be said that while an infomediary can play a positive role in protecting the personal data of users with whom they have a trust-based relationship, the basis of this business is the possibility of making profits by divulging or giving access to customers' personal data.

Depending on the circumstances and the nature of the infomediary, it can be both privacy-enhancing and privacy-invasive.

III. Other privacy-enhancing measures

Other techniques can also be used to improve the transparency of processing or facilitate the exercise of data subjects' rights. Examples include.

P3P

P3P stands for Platform for Privacy Preferences[209]. The objective of P3P is to allow websites to express their privacy preferences and users to exercise their preferences over these practices, so that users can take informed decisions about their web experiences and control the use of their information. The whole data protection community has followed the development of P3P with great interest.

In April 1998, the International Working Group on Data Protection in Telecommunications issued a common position on the Essentials for privacy-enhancing technologies (e.g. P3P) on the World Wide Web[210]. This paper sets out the essential conditions to be met by any technical platform for privacy protection on the World Wide Web, with the objective of avoiding the systematic collection of personal data:

1. Technology cannot in itself secure privacy on the web. It needs to be applied according to a regulatory framework.
2. Any user should have the option to browse the web anonymously. This also applies to the downloading of information in the public domain.
3. Before personal data, particularly those disclosed by the user, are processed by a website provider, the user's informed consent must be obtained. In addition, certain non-waivable ground rules should be built into the default configuration of the technical platform.

[209] The last working draft of the P3P protocol can be found on the W3C website at http://www.w3.org/TR/1999/WD-P3P

[210] This text is available at: http://www.datenschutz-berlin.de/doc/int/iwgdpt/priv_en.htm

Two months later, in June 1998, the Working Party also issued an opinion[211]. This opinion stressed the fact that a technical platform for privacy protection will not in itself be sufficient to protect privacy on the Web. It must be applied within the context of a framework of enforceable data protection rules providing a minimum and non-negotiable level of privacy protection for all individuals. This opinion also mentioned a number of specific issues that would be raised by the implementation of such a system within the European Union.

In order to investigate the application of P3P in the context of the European data protection directive and to foster communication between the EU data protection community and software developers, a joint-seminar was organised in September 1999. A high-level delegation from the World Wide Web Consortium and members of the Internet Task Force participated in this seminar. This seminar showed that a good number of issues still needed to be addressed.

Once these issues are resolved, P3P could play a positive role if applied within an adequate framework. The main positive aspects of P3P are the following[212]:

- P3P can help standardise privacy notices. While this in itself does not offer privacy protection, it could, if implemented, greatly advance transparency and be used to support efforts to improve privacy protection.
- P3P can support the growth of privacy choices, including anonymity and pseudonimity.

The limitations[213] of P3P should, however, be taken into account:

- P3P can not protect the privacy of users in countries with insufficient privacy laws: it does not have the ability to create public policy, nor can it demand that its specifications be followed in the marketplace.
- P3P cannot ensure that companies follow privacy policies. In fact, P3P cannot guarantee that the site is doing what it claims to do. The sanctions for failure to comply with a declaration of intent can only be set by law or through membership of a self-regulatory body.

The labelling of privacy

The labelling consists of a quality stamp put on a website. Over the years, various privacy labels have appeared, with TRUSTe[214], Privaseek[215], the Better Business Bureau[216], WebTrust[217] being examples of such labelling systems. These American organisations aim at operating at international level, also in Europe, what is already the case for some of them. At the same time, similar initiatives with international purposes are taken in Europe, such as for instance L@belsite in France.

A privacy label is granted to companies that fulfill a number of requirements specified by the labelling organisation. This organisation can exercise some kind of control over

[211] Opinion 1/98 on Platform for Privacy Preferences (P3P) and Open Profiling Standard (OPS), adopted on 16 June 1998, WP 11, XV D/5032/98.
[212] See the Article by CAVOUKIAN, A. and GURSKI, M. (Information and Privacy Commissioner Ontario) and MULLIGAN, D. and SCHWARTZ, A. (Center for Democracy Technology), *P3P and privacy: an update for the Privacy Community*, available at: wysiwyg://16/http://www.cdt.org/privacy/pet/p3pprivacy.
[213] See the previous footnote.
[214] http://www.truste.org
[215] http://www.privaseek.com
[216] http://www.bbbonline.org/businesses/privacy/index.html
[217] http://www.cpawebtrust.org/consumer/index.html

compliance with the privacy policies published by companies holding their label by carrying out periodical checks on the activities of these companies. In some cases, the labelling organisation also deals with complaints filed by data subjects concerning companies with this label on their websites.

The labelling of privacy raises a number of issues:

1. The first concerns the label content. The right of information, access, the data minimisation principle, the right of opposition, the principle of legitimacy and proportionality, and the obligation to notify the national data protection authority are some of the cornerstones of the European data protection principles. The main social risk would be the widespread dissemination of privacy labels throughout Europe, which could be misleading for users and data controllers. Although they may give this impression, not all labels seriously guarantee compliance with the aforementioned data protection principles.

2. The second problem lies in the control of website privacy practices. Many kinds of control can be envisaged. Some of the major concerns regarding this issue are:

- Who has the control, how, with what sort of mandate from the controlled company? In the worse case, it appears that the controller will primarily be the data subject him/herself, with all the problems this entails in identifying failures to observe posted privacy practices, proving these and reporting them to the label controller. Besides, not all labelling bodies can ensure that companies do as their policies say they will;
- Who will pay? Given that labelling is a private initiative which does not often benefit from government financial support, some labelling bodies will be under pressure from the companies they are supposed to control.
- What, if any, sanctions will be applied?

The possible privacy-enhancing effects of privacy labels should not, however, be underestimated, as they can help raise the awareness of Internet users about privacy. Some proposals can be made to address the above mentioned problems:

1. The label content: In order to guarantee that privacy labels are in line with European data protection legislation, a European standard for privacy labels could be agreed on by the Working Party. This standard should specify the requirements a label should fulfill[218].

Different labels could coexist as long as it is clear to Internet users which labels meet European standards.

2. The control of website privacy practices: The reliability of website privacy practices could be substantially improved by obliging websites with the quality stamp to undergo periodical audits. The European standard for privacy labels could include this requirement and determine possible ways of carrying out such compulsory controls: self-auditing using a standard checklist, third party audit, etc.

IV. Conclusions

- Recommendations should be issued to produce privacy-compliant browsers with the most privacy-friendly default settings;
- anonymous proxy servers can hide the IP address and could be offered as a free standard feature with an Internet subscription by every *ISP*;

[218] Some very interesting work in this field has been done by the French Data Protection Authority (CNIL). This work could serve as inspiration for the European standard. See www.cnil.fr

- websites should not deny access to users who do not want to accept *cookies*, unless those session *cookies* are indispensable in order to make the link between a user and his/her different purchases online, thus providing for adequate billing;
- the use of PETS should be encouraged, especially if installed by *ISP*s or other actors
- it appears that individuals need to be given more information about the existence of privacy-enhancing technologies. The public sector should take the necessary steps to raise awareness and support the development of these solutions, in addition to using and promoting them[219].
- A European standard for privacy labels could be agreed upon by the Working Party. This standard should include the obligation for websites to undergo periodical audits.

[219] In the Netherlands, a motion was approved during the Parliamentary discussion of the new data protection law in the Second Chamber in which the Government was requested to encourage the development and use of PET, and to encourage the public sector to take the initiative as a promotor of PET in its own processing of personal data. Motion number 31 by NICOLAÏ C.S., presented on 18 November 1999 regarding Bill 25 892 (Regels inzake de bescherming van persoonsgegevens, Wet bescherming persoonsgegevens), The Hague, Tweede Kamer, vergaderjaar 1999–2000, 25 892, nr. 31.

CHAPTER 10 : CONCLUSIONS

This document has dealt with a number of topics presented in separated chapters; each one of them includes conclusive remarks about specific issues. There are nevertheless common issues related to all Internet services described in this document, that deserve being dealt with in more general terms.
After a summary of the trends and privacy risks observed through all the different aspects of the Internet use, it is attempted to provide some guidelines and recommendations, considering actions that could be taken at various levels.

1. Trends and risks

The development of the Internet is exponential. A growing amount of services is available to the Internet user, from shopping online to participating in fora with people all around the world. Due to this complexity, it becomes more and more difficult to have an adequate overview of all possibilities offered to the user. Companies look for a way to attract the user and distinguish themselves from others by offering personalised and/or free services.
Personalisation of the services is dependent upon utilisation of personal data of the users, which companies try to obtain using different sources, such as encouraging the provision of such data by the users themselves in the framework of loyalty programs, free gifts or services, collection from public available sources, etc.
The profiles constituted are not only valuable for the companies who want to target a consumer, but have an economic value in themselves as they are often sold or hired to others.

The development of new technologies makes it easier today to follow an Internet user. For instance, when a consumer uses a mobile phone to connect him/herself to the Internet, data indicating his/her location can be generated.
When the user makes an Internet connection through new means such as ADSL or cable, he/she is assigned a static IP address that facilitates the tracking from session to session.
New generations of software and hardware offer new features increasing the capability to monitor the user's activities in real time, often without his/her knowledge. Numerous examples of invisible processing and E.T. software have been given all through this document.
In this context, it becomes difficult for the average user to remain anonymous while being on the Internet.

The combination of these developing capabilities brings with it new risks for the privacy of the Internet user, especially when data are concentrated in the hands of one or a limited number of controllers.
When these controllers make use of data mining technologies for example, they have the technical possibility not only of processing and reorganising the data but also to uncover new links and characteristics related to the data subject, who is usually not aware of this possibility and does not expect such a processing.
Such risks also arise from the fact that some data are preserved online for a very long period of time; for instance the messages posted to newsgroups and mailing lists are often kept several years and can be consulted using reverse search tools.

Such availability of personal data enables unexpected secondary use of those data, which is often incompatible with the purpose for which the data where originally collected.

2. Guidelines and recommendations

2.1. Raising the awareness of the Internet user

Given the increasing risks for the privacy of the Internet user, as described above, it is especially relevant to ensure that adequate means are put into place in order to ensure that the user gets all the information he/she needs to make an informed choice. Several actors have a role to play in the provision of this information to the user.

In the first place, any controller collecting personal data online has to give all necessary information to the data subject. This information, mentioned in article 10 of Directive 95/46/EC, shall be given in all cases at the occasion of the collection of data. Although having a privacy policy posted on the website is a good way of providing general information to the public, it is necessary to provide information to the data subject from which the data are being collected, in a simple and accessible way each time that data are collected, e.g. in the same screen where he/she has to fill in his/her data or through a box prompt.

Where the data controller is a private company, the compliance with these rules is not only important in legal terms, but also out of commercial self interest, as the trust and confidence of individuals will increase and might have an impact in the involvement of the individual with the company. As regards the development of e-commerce for instance, it is being observed that users are reluctant to engage in electronic transactions if they fear that their personal data will not be correctly protected and secured.

Where the controller is a public authority, the compliance with the data protection rules is a key element as the behaviour of such authority should be an example for the public in general. For instance, public authorities implementing e-government activities should build in privacy as one of the cornerstones of the system of exchange of data. Besides, even when they do not play a role of data controller, the responsibility of these authorities lies in the field of general education and information of the public.
In particular, data protection authorities are entrusted with the task of raising awareness, about the risks linked with the use of the Internet but also about the rights and obligations foreseen by the legislation. This can be done in several ways, such as publication of brochures, reports, press releases, practical recommendations included in the notifications forms, organisation or participation in conferences or seminars in these issues, directed to the different actors and sectors of the society.

Privacy association and advocates have traditionally been performing such public awareness activities, in a way that has sometimes led to significant improvements as regards the privacy compliance of Internet products.
In several countries of the European Union, it has been observed that consumers associations are also increasingly getting involved and interested in the privacy aspects of consumers activities. Such role can be particularly positive as it does not limit itself to the provision of information but also extends to the representation of consumers in their relation with companies or public authorities. Such associations can e.g. monitor the compliance of ISPs with the laws, or inform public authorities about the complaints they receive about a specific web site or Internet company.
Professional associations can also have a positive influence, informing new actors about their legal obligations.

All above-mentioned parties play a significant role in giving the consumer the information necessary in order to allow him to make a responsible choice. It is then up to the individual to make use of the means that are available to him/her to ensure the respect of his/her rights, and possibly to make clear that he/she will not accept services or products that are not in compliance with the existing legal framework.

2.2. Applying existing legislation in a coherent and co-ordinated way

On-line data protection can only be sufficiently guaranteed if the existent legal framework is complied with. Considering the international character of the network, it is essential that data controllers can rely on a coherent and co-ordinated interpretation and application of the European data protection rules. This is not only important for data subjects and controllers inside the EU but also for those outside the Union that also have to take this legal framework into consideration, in particular when they collect personal data using means located inside the Union. The Working Party has an important role to play in this context.

The Working Party has at several occasions identified some lacunae or controversial issues in the existing legislation and issued documents providing for common interpretation and possible solutions. Special attention has been paid to the revision of the Directive 97/66/EC, which has brought with it some significant improvements in the terminology used. Although the Working Party welcomes the fact that new issues have been taken into account in the draft Directive, some proposals have been made on specific points that could still be better addressed.

The Working Party is concerned about the fact that amendments of existing legislation will sometimes tend to stricter legal requirements as regards in particular the possibilities of surveillance on the web and the generalisation of identification requirements of users. The Working Party has recalled that, although other legitimate interests could be at stake, a balance should always be stroke between them and the protection of the personal data of the individual.

It should be emphasised that interpretation and application of the legislation is not only the task of public authorities but that the private sector can provide fruitful contribution by investing in the development of self regulation or codes of conduct addressing more specific issues raised in a particular sector.

2.3. Developing and using privacy compliant, privacy friendly and privacy enhancing technologies

As already stated, the processing of personal data on the Internet very much depends on the technical configuration of the hardware and software as well as on the protocols and technical standards used for the transmission of information.
It is therefore especially important to take into account privacy requirements at the earliest stage of developing all these tools; e.g. a browser should not transmit more information than necessary to establish a connection to a website. Those involved in the design and development of these technical tools are encouraged to consult the national Data Protection Authorities about the existing data protection legal requirements.
Moreover, in order to make clear to the general public which products are privacy-compliant, it would be useful to put in place a system of certification marks that would

allow an easy recognition of those products that comply with the data protection requirements.

Moreover, while new technologies are traditionally considered as a threat to privacy, it should be stressed that they also represent a useful tool in terms of safeguarding privacy.

Some of the existing technologies can first be used in order to improve the transparency and the friendliness of the information provided to the data subject, e.g. by giving users simple and accessible information at the moment of collection of personal data.
They can secondly be a useful tool in order to simplify the exercise of the rights of the data subjects, e.g. allowing a direct access on-line to the personal data of the individual or giving the possibility to oppose the processing.

Taking into account that the average user is not necessarily familiar with the technical aspects of using the Internet, and is not always in the situation of deciding himself on or even changing the configuration of the hardware and software used, it is crucial that the default settings of the products offer the highest level of privacy protection.

A number of additional tools, better known as “privacy enhancing technologies”, has been developed in order to help users safeguarding their privacy, notably by minimising or eliminating the collection or further processing of identifiable data and technically hindering any unlawful forms of processing. Examples of such tools are proxy servers, cookie killers, anonymisation software, pseudonymisation tools (in particular valuable for profiling), e-mail filters, etc. Possible new products might include smart cards containing a portable identity protector (PIP) which the individual will be able to insert in any machine from which he/she establishes an on-line connection.

From all the actors already mentioned in paragraph 2.1., the industry and the public sector are the first ones that should invest and encourage the development and implementation of privacy protective technologies. The user should be made aware of the existence of these means, which should be available without involving unreasonable costs.

2.4. Building trusted mechanisms for control and feedback

On-line data protection can only be effective if adequate means are in place to monitor and evaluate the compliance with the legal framework and technical requirements explained above.
For that purpose, even if data protection authorities are in charge of the control of enforcement in the first place, other actors are taking steps in the direction of self monitoring, as they have realised the impact of their privacy policy on the behaviour of the consumers towards them.
Data protection authorities can contribute to the development and well functioning of such self monitoring systems by providing guidance, e.g. in the form of checklists for self evaluation agreed at European level.

Furthermore, labels could be granted with a view of helping the consumer getting a trustworthy indication of the compliance of a data processing with EU Data protection legislation. The Working Party intends to take action in this field in order to ensure in particular that privacy labels are granted to web sites which are in line with European data protection legislation.

The Working Party invites all actors involved in Internet activities to consider this working document and to take the necessary steps to put its recommendations into practice.

The Working Party hopes that this working document will contribute raising the awareness and will promote public debate on this issue, which will certainly require further analysis and follow up in the future.

GLOSSARY OF TECHNICAL TERMS[220]

ADSL

ADSL (Asynchronous Digital Subscriber Line) is a telecommunication protocol that can be used on classical copper twisted peer lines. It permits to reach speed up to one mbps while the line remains simultaneously free for classical phone conversation. ADSL requires dedicated ADSL modems to be put at both ends of the local line.

Authentication

Verifying the identity of a user logging onto a computer system or verifying the *integrity* of a transmitted message.

Banner

Banner advertisements are small graphic boxes which appear above or are integrated into the website content.

Calling Line Identification (CLI)

When a call is made, this enables the called user to identify the calling user by presenting the number of the calling line.

Clickstreams

Information derived from an individual's behaviour, pathway, or choices expressed while visiting a website. They contain the links that a user has followed and are logged on the web server (the *ISP* computer for those who do not run their webserver).

Cookies

Cookies are pieces of data created by a webserver that can be stored in text files that may be put on the Internet user's hard disk, while a copy may be kept by the website. They are a standard part of HTTP traffic, and can as such be transported unobstructed with the IP-traffic. A *cookie* can contain a unique number (GUI, Global Unique Identifier) which allows better personalisation than dynamic IP-adresses. It provides a way for the website to keep track of a user's patterns and preferences.
The *cookies* contain a range of URLs (addresses) for which they are valid. When the browser encounters those URLs again, it sends those specific *cookies* to the Web server.
Cookies can differ in nature: they can be persistent but can also have a limited duration, the so-called session *cookies*.
You can have your browser disable *cookies* or warn you before accepting a cookie.

Data integrity

The process of preventing accidental erasure or adulteration in a database.

Datamining

[220] Some of these definitions have been taken from the following sources:
- http://www.techweb.com/encyclopedia
- http://webopedia.Internet.com
- Personal Data Privacy and the Internet: a guide for data users, Office of the Privacy Commissioner for Personal Data, Hong Kong, 1998.

This implies "digging through tons of data" to uncover patterns and relationships contained within the business activity and history. This is usually done with programs that analyse the data automatically.

Data warehouse

A database designed to support decision-making in an organisation. It can contain enormous amounts of data. For example, large retail organisations can have 100GB or more of transaction history. When the database is organised for one department or function, it is often called a data mart rather than a *data warehouse*.

Digital certificate

A *digital certificate* is an electronic document which contains two groups of information and which is intended as proof of identity in the electronic world. The first is the certificate information itself, including the name or a pseudonym of the natural or legal person requesting the certificate, its public key, the certificate's validity dates and the name of the Certification Authority (CA). The second piece is the Certification Authority's *digital signature*. The entire message is digitally signed by a Certification Authority which is trusted by many servers (CAs are a specific kind of *Trusted Third Parties*) and can verify the relationship between a natural or legal person and its public key.

Digital signature

A *digital signature* is a data string that is added to a message and guarantees its *integrity* by encrypting it (or a message digest) with the signatory's private key. Anybody who receives the signed message can check if it has been modified simply by decrypting the signature with the sender's public key and comparing the decrypted string with the original message or digest.

Domain Name Service (DNS)

The DNS (*Domain Name Service*) is a mechanism for assigning names to computers identified by a IP address. Those names are in the form of <names>. top level domain where <names> is a string constituted by one or many substrings separated by a dot.

Dynamic Host Configuration Protocol (DHCP)

The *Dynamic Host Configuration Protocol* (DHCP) is an Internet *protocol* for automating the configuration of computers that use TCP/IP. DHCP can be used to automatically allocate IP addresses. (http://www.dhcp.org)

Electronic signature

Data in electronic form that are attached to or logically associated with other electronic data and which serve as method of *authentication* (Article 2.1 of the *Electronic signature* directive).

Encryption

Encoding information and messages in such a way that they cannot, in principle, be read by someone other than the intended recipient who has access to a key or password. There are two main kinds of *encryption* systems.

- The Symmetric or Private Key system, which uses a secret key shared between both the sender and the receiver of a message, its main advantage being the speed of processing and its main drawback the difficulty of seccurely sharing keys among a great number of users.
- The Asymmetric or Public Key System, which uses a pair of keys, generated so that even knowing one of them, is almost impossible to guess the other. Messages encrypted using one of the keys are decrypted using the other. One of the keys is made public and used to encrypt the messages which every user decrypts with his or her secret private key. The Private Key is also used to sign messages digitally.

Firewall

A method for keeping a network secure. It can be implemented in a single *router* that filters out unwanted packets, or it may use a combination of technologies in routers and hosts. *Firewall*s are widely used to give users secure access to the Internet as well as to separate a company's public web server from its internal network. They are also used to keep internal network segments secure. For example, a research or accounting subnet might be vulnerable to snooping from within.

Hyperlinks

A pre-defined link between one object and another. The link is displayed either as text or as an icon. On World Wide Web pages, a text *hyperlink* is displayed as underlined text usually in blue, while a graphics hyperlink is a small graphical image.

Internet Service Provider (ISP)

A company that provides access and connections to the Internet to members of the public and companies.
Small *Internet Service Provider*s (ISPs) provide the service via *modems* and ISDN while the larger ones also offer private line hookups. Customers are generally billed a fixed rate per month, but other charges may apply. For a fee, a website can be created and maintained on the *ISP*'s server, allowing a smaller organisation to have a presence on the Web with its own domain name.
Large Internet services also provide proprietary databases, forums and services in addition to Internet access.
In this report, the term *ISP* generally includes IAPs. The term IAP is only used when it is clear that it deals only with Internet access; in all other cases the generic term *ISP* is used.

Java and JavaScript

Java is a full-blown programming language and is not intended for the casual programmer and certainly not the end user. *JavaScript* is a scripting language that uses a similar syntax as *Java*, but it is not compiled into bytecode. It remains in source code embedded within an HTML document and must be translated one line at time into machine code by the *JavaScript* interpreter. *JavaScript* is very popular and is supported by all Web browsers. *JavaScript* has a more limited scope than *Java* and deals primarily with elements on the web page itself.

Meta Tags

Meta Tags are HTML tags that provide information about a webpage. Unlike normal HTML tags, *Meta Tags* do not affect how the page is displayed. Instead, they provide information such as who created the page, how often it is updated, what the page is about and which keywords represent the page's content. Many search engines use this information when building their indices.

Modem

(**MO**dulator-**DEM**odulator) A device that adapts a terminal or computer to an analog telephone line by converting digital pulses to audio frequencies and vice versa. The term usually refers to 56 Kbps modems (V.90), the current top speed, or to older 28.8 Kbps modems (V.34). The term may also refer to higher-speed cable or DSL modems or to ISDN terminal adapters, which are all digital and not technically modems. A *modem* is an analog-to-digital and digital-to-analog converter. It also dials the line, answers the call and controls transmission speed. Modems have evolved at 300, 1200, 2400, 9600, 14400, 28800, 33300 and 56000 bps. Whatever the top speed, some lower speeds are always supported so that the *modem* can accomodate earlier modems or negotiate downwards on noisy lines.

OLAP

On**L**ine **A**nalytical **P**rocessing. Decision support software which allows the user to quickly analyse information that has been summarised into multidimensional views and hierarchies. For example, *OLAP* tools are used to perform trend analysis on sales and financial information. They can enable users to drill down into masses of sales statistics in order to isolate the most volatile products. Traditional *OLAP* products, also known as multidimensional *OLAP*, or MOLAP, summarise transactions into multidimensional views ahead of time. User queries on these types of databases are extremely fast, because the consolidation has already been done. *OLAP* places the data into a cube structure that can be rotated by the user, making it particularly suitable for financial summaries.

Portal site

A *portal site* provides an overview of weblinks in an ordered way. Via the visited *portal* the Internet user can easily visit selected websites of other content providers.
Modern portals are "supersites" that provide a variety of services including web searching, news, white and yellow pages directories, free e-mail, discussion groups, online shopping and links to other sites.

PPP

PPP (Point to Point Protocol) is a telecommunication protocol widely used to connect two computers by using their serial port or a modem put on it. It is the low layer protocol mainly used between the personal PC of a home user and the Internet Access Server of an Internet Service Provider while establishing a TCP/IP connection on classical phone lines.

Proxy server

The *proxy server* is an intermediary server between the Internet user and the Net. It acts as a Web cache, dramatically improving the performance of the Internet. Many large organisations or Internet Access Providers have already implemented this solution. Each

page, image or logo downloaded from outside by an organisation's member is stored on a cache and will be instantaneously available to another member of this organisation.

It is no longer necessary for every member of the organisation located before the *proxy* server to have his/her own IP address, because they do not directly access the Internet.

Protocol

In this context, a *protocol* is a set of technical rules that must be observed by two partners to exchange information. Protocols are organised into a hierarchy of so-called layers. Each layer is responsible for handling one particular aspect of the telecommunications process and provides basic functions to be used by the upper layers. Traditionally, on the Internet the TCP/IP *protocol* is always used as the intermediate layer. Ethernet (used in Local Area Networks), ADSL (used in phone lines), ATM (used by telecommunications operators), X-75 (used on ISDN lines), PPP (used on standard telephone lines) are some examples of lower-level protocols. On the other end of the scale, HTTP (for surfing), SMTP and POP (for e-mail), FTP (for transferring files) are higher-level protocols. This means that every potential privacy threat present in the TCP/IP *protocol* will be one of the weaknesses of the upper protocols. In basic terms, layers are a set of subprograms running on a computer linked to the Internet.

Router

A *router* is an important device which provides routes for *TCP/IP networks*. This means that the TCP/IP route is dynamic, depending on the failure or overloading of some routers or links. It can also be used as a *firewall* between an organisation and the Internet and guarantees that only authorised IP addresses can originate from a particular *ISP*.

Shareware

Software that can be downloaded from the Internet. It can normally be downloaded free for trial purposes, but a small amount must be paid to the software developers to be able to use it legally. Software which can be downloaded and used completely free of charge is known as *freeware.*

Sniffing

Sniffing software makes it possible to monitor the traffic and read all the data packets on a network, thus presenting in clear text all communications which are not encrypted. The simplest form of *sniffing* can be carried out using an ordinary pc connected to a network using commonly available software.

Spamming (or spam)

The sending in bulk of unsolicited advertising marketing material via e-mail.

TCP/IP network

A *TCP/IP* (Transport Control Protocol/Internet Protocol) *network* is based on the transmission of small packets of information. Each packet includes the IP address of the sender and of the recipient. This network is connectionless. It means that, unlike the phone network for instance, no preliminary connection between two devices is needed before communications can start. It also means that many communications are possible at the same time with many partners.

Trusted Third Parties[221]

A *Trusted Third Party* (*TTP*) can be described as an entity trusted by other entities with regard to security-related services and activities.

A *TTP* would be used to offer value-added services to users wishing to enhance trust and business confidence in the services they receive, and to facilitate secure communications between business partners. TTPs need to offer value as regards the *integrity*, confidentiality and successful performance of the services and information involved in communications between business applications. In addition, users will require *TTP* services to be available when they need them within the terms of the agreed service contract.

Typically, a *TTP* will be an organisation which has been licensed or accredited by a regulatory authority and provides security services, on a commercial basis, to a wide range of bodies, including those within the telecommunications, financial and retail sectors.

For example, a *TTP* could be used to support the provision of *digital signature*s to secure the *integrity* of documents. In addition, they could provide end-to-end *encryption* services to users, and incorporate, for example, a recovery or backup function for a key, to enable recovery if the key is lost (typically for documents and files that have been encrypted by employees) or to support a request for lawful interception.

The use of TTPs is subject to the fundamental requirement that the *TTP* is trusted by the entities it serves to perform certain functions.

UMTS

UMTS (Universal Mobile Telecommunications System) is a "third-generation"broadband, packet-based and wireless transmission protocol who will offer tranmission speed higher than 2 Mbps. This new broadband protocol will allow transmission to digital video with TV quality to mobile devices. Presently, the GSM network allow speeds about 11 Kbps, sufficient for the transmission of the voice but not for moving images[222].

WAP

WAP (Wireless Application Protocol) is a telecommunication protocol conceived by many mobile phone manufacturers. It permits access from a dedicated mobile phone to Internet services such a Mail, Chat, Web surfing.[223]

Web cache

A computer system in a network that keeps copies of the most-recently requested Web pages in its memory or on disk in order to speed up retrieval. If the next page requested has already been stored in the cache, it is retrieved locally rather than from the Internet. Web-caching servers sit inside the company's *firewall* and enable all popular pages retrieved by users to be instantly available. Since the content of web pages can change, the caching software is always checking for newer versions of the page and downloading them. Pages will be deleted from the cache after a set amount of non-activity.

Webmail

[221] Definition taken from the ETSI "Requirements for *TTP* services".
[222] See http://www.umts-forum.org/
[223] See for more information: http://www.wapforum.org

E-mail systems that use web pages as an interface (e.g. Yahoo, HotMail etc.). *Webmail* can be accessed from everywhere and the user does not need to make a connection to a specific *ISP*, as when using an ordinary e-mail account.

Done at Brussels, 21st November 2000

For the Working Party

The Chairman

Stefano RODOTA