



COMMENTS ON DEFINITION OF 'PERSONAL DATA'

The European Privacy Officers' Forum (EPOF) is composed of data protection compliance officers and internal legal counsel in charge of data protection in Europe from approximately forty multi-national companies. EPOF members deal with issues raised by the practical implementation of European data protection law on a daily basis. We note with interest that the Article 29 Working Party will consider as part of its 2006-2007 Work Programme key provisions of the Data Protection Directive 95/46/EC ("Directive") including the definition of 'personal data'. We also note that a general analysis on the definition of this important concept will also form part of the Working Party's consideration of data protection issues raised by the development of Radio Frequency Identification (RFID) technology.

EPOF members would like to take this opportunity to highlight to the Working Party practical issues raised by how 'personal data' is defined across the Member States. We concentrate specifically on the issue of 'identifiability' and where the dividing line is drawn between 'personal data' and 'anonymous data'¹, as a consequence of significant compliance issues faced by EPOF members in this area.

Personal data is very broadly defined in Article 2 of the Directive as "any information relating to an identified or identifiable natural person...". Where this definition is applied unqualified then it may be interpreted in such a way that data will remain 'personal' and subject to the full remit of the law if individuals remain *in any way* identifiable. We believe that the concept of personal data should rather be defined pragmatically, based upon the likelihood of identification, as per the qualifying principle present in Recital 26 of the Directive². In our view, it should not be the case that an organisation has to be sure that there is no conceivable method, however unlikely in reality, by which the identity of individuals can be established. This is a highly impractical approach, usually requiring considerable resource to be expended on disproportionate statistical analysis. The

¹ 'Anonymous data' can include both data that has been anonymised and data that never identified an individual. By 'anonymisation' we mean a process by which information directly or indirectly identifying an individual is removed from a collection of personal data. 'Pseudonymised data' (coded data) is information that relates to a specific individual, from which direct identifiers (usually name) have been removed, but to which a specific unique pseudonym or code has been attached. In the light of the arguments set out in this paper, pseudonymous data can be personal or anonymous depending upon the content of the data and the circumstances in which both the information itself and the associated 'key' are held.

² "... whereas, to determine whether a person is identifiable, account should be taken of all means likely reasonably to be used either by the controller or by any other person to identify the said person..."

responsibility of organisations is to ensure that effective safeguards are put in place to prevent the data from being processed in such a way that it leads to identification. The rights, freedoms, and legitimate interests of individuals can more than adequately be protected if data is processed in such a way that all means *likely reasonably* to be used to identify the said person will fail.

The concept of 'data minimisation', the requirement for organisations to only process data about identified or identifiable individuals where absolutely necessary for any particular purpose, is a fundamental principle of data protection. The anonymisation of personal information is a process which fundamentally supports this concept, allowing information about individuals to be processed without reference to identifying features. It is imperative therefore that the regulatory regime does not discourage data minimisation by making the process of anonymisation disproportionately difficult to accomplish, applying the full regulatory regime equally to personal data and to data collections that for all practical intents and purposes are not identifiable to individuals.

In making judgements about whether information is personal data, an organisation should consider the following factors:

1. How that data could be matched with publicly available information, analysing the statistical chances of identification in doing so;
2. The chances of the information being disclosed and being matched with other data likely held by a third party;
3. The likelihood that 'identifying' information may come into their hands in future, perhaps through the launch of a new service that seeks to collect additional data on individuals;
4. The likelihood that data matching leading to identification may be made through the intervention of a law enforcement agency, and
5. Whether the organization has made legally binding commitments (either through contract or through their privacy notice) to not make the data identifiable.

Considerations on all these issues are of course **contextual**, based upon an assessment on a case-by-case basis of the likely chances that identification may occur in any reasonably foreseen set of circumstances. In terms of 'reasonableness' or 'fairness', an additional aspect of this assessment may involve consideration as to the sensitivity of the information and any potential harm that could arise for individuals if data is later made identifiable.

However, some Member States, such as Belgium, Sweden and France, have interpreted data protection law to mean that if someone *can* be identified from certain data, no matter how technically or legally difficult it is to ascertain the identity of the physical person from such data, then the data is deemed to be 'personal data'.

Concern on this issue is perhaps most acutely expressed in the medical research area. Medical researchers in pharmaceutical companies, clinical research organisations, universities and elsewhere use key coded data for various forms of clinical study. Such organisations also make extensive use of de-identified patient data from hospitals or general practice in pharmacovigilance, clinical epidemiology, retrospective medical research, clinical research planning, drug utilisation analysis, and health economics & outcomes research. Anonymised data have formed a valuable resource in understanding the delivery and dynamics of patient care, these being used by industry, governments, regulatory bodies, academic institutions and patient organisations. However, there is considerable uncertainty as to what constitutes an anonymised medical record. In situations where consent from patients is not readily obtainable, for example, with

historical data collections, there is a motivation to anonymise the data to facilitate important secondary use in research. However, if to anonymise the data much of the content has to be removed, for example environmental factors or diagnosis, due to disproportionate concerns about re-identification risk, then the research value of the information falls significantly.

We suggest that a significant step can be taken in solving this issue by the Working Party providing qualifying guidance on the limits of 'personal data'. This should be pragmatic and emphasise that identification must be subject to the reasonableness standard. For example, a definition such as that given in §3(6) of the German Federal Data Protection Act could be used as a basis for this interpretation:

"Depersonalisation means the modification of personal data so that the information concerning personal or material circumstances can no longer or only with a disproportionate amount of time, expense and labour be attributed to an identified or identifiable individual."

The Working Party could go further and add an important specific addition to their qualifying guidance to the effect that data are only deemed "identifiable" with respect to a particular controller only if that controller has information *in its possession, or which is likely to come into its possession*, that would enable it to identify the person to whom the data relate. The UK has adopted this position: data are deemed personal if the individual to whom they relate is identifiable "from those data and other information in the possession or likely to come into the possession of the data controller".³ As long as there is little or no chance of disclosure by the controller to a third party of information that could lead, in combination with data held by that person, to re-identification of individuals, then this approach seems more than reasonable.

Two practical examples will help to clarify the specific problem and demonstrate the practical value of the proposed overall solution⁴:

1 Key-Coded Data

Clinical trials with medicines are a key form of research sponsored by pharmaceutical companies. They are subject to the EC Clinical Trials Directive, existing national provisions, and international principles governing Good Clinical Practice. One of the principle aims of these controls is to ensure that adequate precautions are in place to guarantee that the privacy of the subject and the confidentiality of his other information are protected. The pharmaceutical company typically contracts with an independent investigator to conduct a clinical trial and the investigator is responsible for key-coding data related to trial subjects by replacing the individual's name and other identifiers with a unique numerical or alphanumeric code. The pharmaceutical company may receive only key-coded data from which all personal identifiers have been removed. Indeed, generally there are ethical and/or legal restrictions on providing non-coded data to the pharmaceutical company.

To protect the health of trial subjects, the investigator is also obliged to retain the keys that permit those subjects to be reverse-identified. Medicines can cause adverse effects that may manifest later during the trial or even only long after the clinical trial ends. Thus, it is

³ United Kingdom Data Protection Act 1998, § 1(1)

⁴ Examples originally articulated (but here developed) by Covington & Burling in their submission to the European Commission Review of the the Directive in 2002
http://ec.europa.eu/justice_home/fsj/privacy/docs/lawreport/paper/covington-burling_en.pdf

important to be able to go back to the individuals that participated in the trial to provide appropriate information, treatment, and/or compensation. However, the company never has access to the key. The company will merely provide the physician with the relevant numerical or alphanumeric code and the physician conducts the relevant follow-up.

Member State data protection authorities (DPAs) are split over whether key-coded data sent to the pharmaceutical company sponsoring a clinical trial are personal data regulated by the Directive and Member State law. Some DPAs take the position that such data are 'reversibly anonymised' and thus covered by national law as 'personal data'. According to these DPAs, such as the Belgium Privacy Commission, even though a pharmaceutical company does not have and should never obtain the key that would allow it to identify the person(s) to whom the data relate, the data are personal data because someone, somewhere – i.e., the investigator – can identify those persons. Therefore, Member State data protection laws apply. Other DPAs, such as that in the UK, take precisely the opposite view, concluding that the key-coded data transferred to the pharmaceutical company are anonymous and therefore not subject to the Directive and national implementing laws.

The Commission and Member States adopted the latter position in the EU-U.S. Safe Harbour Agreement. According to Frequently Asked Question Fourteen⁵, a transfer of key-coded data from an EU investigator in a clinical trial to a sponsoring U.S. pharmaceutical company is not a transfer of personal data triggering application of the Safe Harbour, if the U.S. company does not hold the key. If this way of distinguishing personal from anonymous data provides sufficient protection for sensitive health information in the context of an international transfer, it should suffice to protect such information when collected and further processed in the EU.

In effect, the Safe Harbour Agreement adopts the approach to defining 'personal data' that is recommended here. In the context of a clinical trial, the key-coded data that an investigator holds are personal data from the standpoint of the investigator because the investigator possesses information that enables it to identify the individuals to whom those data relate. However, as the sponsoring pharmaceutical company does not possess such identifying information, the key-coded data it holds should be considered anonymous data from the standpoint of the pharmaceutical company.

2 IP Addresses

The regulatory approach to IP addresses also illustrates the dilemma that the Directive's sweeping definition of 'personal data' can cause. According to the stated position of the Working Party, "IP addresses attributed to Internet users are personal data and are protected" by the Directive⁶. The Working Party reasoned that:

"data are qualified as personal data as soon as a link can be established with the identity of the data subject (in this case, the user of the IP address) by the controller or any person using reasonable means. In the case of IP addresses the ISP is always able to make a link between the user identity and the IP addresses and so may be other parties, for instance by making use of available registers of allocated IP addresses or by using other existing technical means".

⁵ FAQ 14—Pharmaceutical and Medical Products, at <http://www.export.gov/safeharbor/FAQ14PharmaFINAL.htm>

⁶ Article 29 Working Party, The Use of Unique Identifiers in Telecommunications Terminal Equipments: the Example of IPv6, Opinion 2/2002, WP 58, 10750/02/EN/Final, at 3.

The Working Party have assumed that if an IP address is identifiable by one company (e.g., an ISP) it is personal data as far as all other companies are concerned, even if they have no access to the information that permits an association to the individual. But this assumption is very questionable. ISPs typically do not divulge IP account names. Indeed, many Member States have interpreted Article 6 of the 2002 Electronic Communications Data Protection Directive as prohibiting ISPs from divulging user information connected to IP addresses. If a third party cannot receive assistance from an ISP in associating an IP address with a particular user, the IP address is not personal data as far as the third party is concerned. From the third party's perspective, the IP address is anonymous.

Of course, if that third party intends to subsequently match an IP address with the identity of an individual, perhaps obtained through an online data collection form, then there is grounds to consider the IP address (and associated data) as personal data 'before the event' and process it accordingly. Online companies may seek to make policy commitments not to undertake such a data matching exercise. Such matching is particularly difficult with IP addresses due to the facts that a. most ISPs assign dynamic IP addresses, b. most companies use proxy servers, and c. IP addresses could relate to anyone using a particular device. It is again a matter of context, and the conclusion likely turns on whether the company processes the data with the intent of identifying an individual.

The Working Party's approach to IP addresses contrasts sharply with the approach adopted in the Safe Harbour Agreement and some Member States to key-coded clinical trial data. As noted above, the Safe Harbour Agreement treats such data transferred to the United States as anonymous because the U.S. recipient does not receive the relevant identifying information and is generally barred from receiving that information by ethical and/or legal principles. With respect to IP addresses, regulators have reached precisely the opposite conclusion on almost identical facts. A third party that holds an IP address without identifying information is deemed to hold 'personal data', even though in some countries it may be illegal for the third party to obtain the identifying information from the entity that holds it, i.e., the ISP. In our view, the Safe Harbour Agreement adopts the correct approach, and this approach should be applied to IP addresses held by parties other than ISPs.

It is of note that this more pragmatic position is supported by jurisdictions with data protection legislation outside Europe, for example, Hong Kong. In May 2006, in a written reply to a member of the Legislative Council, the Secretary for Home Affairs (Dr Patrick Ho), outlines a policy position on IP addresses similar to that advocated above:

"An Internet Protocol (IP) address is a specific machine address assigned by the web surfer's Internet Service Provider (ISP) to a user's computer and is therefore unique to a specific computer. An IP address alone can neither reveal the exact location of the computer concerned nor the identity of the computer user. As such, the Privacy Commissioner for Personal Data (PC) considers that an IP address does not appear to be caught within the definition of "personal data" under the PDPO..."⁷

While exact location and/or the particular user identity may not be required to qualify the IP address as personal data, Mr. Ho's point that the IP address only identifies a machine is important. In fact, this raises a slightly different, but associated, aspect of the concept of

⁷ <http://www.info.gov.hk/gia/general/200605/03/P200605030211.htm>

identifiability. In determining whether an IP address can be considered an item of personal data in itself, consideration should be given to the fact that the number is not allocated to a natural person but rather to an item of networked equipment. Data generated through the use of such equipment may be the result of intervention by a number of individuals, perhaps the members of an extended family each making use of a home pc, a whole student body utilising a library computer terminal, or potentially thousands of people purchasing from a networked vending machine. We should note that the number of internet-connected devices is set to explode in the coming years. To illustrate the point, it is envisaged that in the future every light bulb will have an IP address, to turn it on and off, and to send a signal when it needs to be replaced. In fact, the logic of this argument could be applied to a variety of unique identifiers that are not necessarily associated with a particular natural person, for example, RFID numbers. Clearly the more divorced the use of such a number is from the identity of a single natural person, the less strong the argument for considering such 'identifiers' as an aspect of personal data. Whether or not these identifiers are personal data will turn on the context in which they are collected and how they are stored and processed. The Working Party should look to this context and determine whether the organization is intending to use the unique identifier to relate to an individual or to a machine.

Conclusion

EPOF members represent a group of multi-national companies that have taken the lead in positively seeking to comply with data protection laws and regulations on a pan-European basis, in doing so having practical hands-on experience in applying the rules on an everyday basis. We contend that significant issues are raised as a consequence of how 'personal data' is variously defined across the Community, in particular in respect to what makes an individual 'identifiable'. EPOF recommends that the Article 29 Working Party consider the issues raised by this paper in the context of their review of the definition of personal data, with a view to providing guidance that seeks to harmonise interpretation across the Community on a pragmatic basis.