International Journal of Law and Information Technology, Vol. 9 No. 1

A Data Protection Approach to Data Matching Operations Among Public Bodies

VAGELIS PAPAKONSTANTINOU[1]

Abstract

Data matching operations have been promoted as an administrative panacea that helps to reduce costs while maximising efficiency and combating fraud. Nevertheless, they may constitute at the same time a brutal invasion into individuals' privacy. After an extensive effort to define precisely data matching and distinguish such processing from similar filing operations, a data protection approach will be attempted through a given axis of analysis in order to demonstrate the strengths and weaknesses of this legislative approach.

1 Introduction

The term 'data matching' generates mixed feelings when used in different contexts. For individuals and civil rights groups it signifies a brutal invasion of personal privacy and a decisive step towards a 'Big Brother' state. Public administration greatly values data matching as an essential tool in the struggle to minimise expenditure and provide efficient service. Their enthusiasm is shared by private organisations that consider data matching an invaluable tool in planning marketing strategies and reducing costs. These contrasting views have created fierce disputes among the interested

[1] Drjur., LL.M, Athens Bar Association, Researcher, European Public Law Center.

parties since the first data matching operation, 'Project Match', was conducted in the United States in 1979.

Data matching is a form of surveillance. The attempt to match personal information collected in different contexts and process it automatically has the potential to create personal profiles on all data subjects involved. Such surveillance is not novel. Effective methods similar to data matching have been proposed long before the appearance of information technology. The English jurist and philosopher Jeremy Bentham developed the idea of a panopticon penitentiary, an innovation for the easy and efficient exercise of power over imprisoned criminals. The panopticon penitentiary was to be circular or polygonal with the cells around the circumference. At the core would be a central inspection area of galleries and lodge, from which authority could exercise a constant and absolute surveillance while itself remaining invisible.[2] These ideas on punishment are still influential. Even today proposals have been put forward for tracing criminals with the aid of signal-emitting bracelets, anklets, or implants. Foucault[3] indeed has proposed an expansion of the panopticon design, while Gandy[4] introduced the term 'the panoptic sort' in his political economy to describe the complex technology that involves the collection, processing and sharing of information about individuals and groups that is generated through their daily lives as citizens, employees and consumers and which is used to co-ordinate and control their access to goods and services.

Profiling is also not a novel technique. Although information technology strengthened the validity of its results, its fundamental concepts preceded our century. The criminologist Cesare Lombroso[5] used sociological observations to classify individuals and distinguish the 'born' criminal. By using the 'méthode pictographique' Lombroso aimed at demonstrating the external characteristics of criminals in an atlas. He also examined skulls of criminals to prove that crime was within their integral parts. His ultimate objective was to predict social groups that are destined to commit crimes. The same research axis has been employed 100 years later by the Health Council of Oslo, who purported to identify behavioural patterns of small children that might indicate psychological problems or later lead to 'anti-social' behaviour.[6] Similar projects have been conducted in France in 1973 and Germany with the therapeutic aim of treating and inhibiting 'dangerous behaviour' and thus guiding the child to adapt better to societal expectations.

Consequently, data matching or profiling or even population surveil-

[2] Semple J, Bentham's prison: A study of the panopticon penitentiary, Clarendon Press, Oxford 1993, p. 116.

[3] Foucault M, Surveiller et punir: Naissance de la prison Editions Gallimard, 1979, p. 39.

[4] Gandy O, The panoptic sort: a political economy of personal information, Westview Press, 1993 p. 132.

[5] Lombroso C, L'homme criminel: études anthropologique et médico-légale, Paris, 1887.

[6] Simitis S, Reviewing privacy in an information society, University of Pennsylvania Law Review 135, 1987 707–746.

lance are not recently emerged concepts. In fact, they build upon ideas that have been part of public administration since its appearance. Nevertheless, information technology has equipped bureaucracies with the means to perform tasks on a scale that previously seemed impossible. Human processing capability has increased exponentially during the last twenty years. Furthermore, our society has witnessed an enormous gathering of personal information. Even in 1985 the US Office of Technology Assessment identified 85 computerised record systems used for law enforcement, investigative and intelligence purposes, with collectively about 288 million records on 114 million persons. In the United Kingdom estimates showed a total of well over 113,141,000 individual files on citizens[7] Consequently, all conditions are met for successfully conducting previously impossible administrative operations.

Manual matching operations preceded the current fully automated data matching exercises. The US Office of Technology Assessment[8] reported that manual comparison of the contents of two record systems constituted a traditional audit technique. The US Senate Committee on Governmental Affairs[9] asserted that manual matches took place before the introduction of information technology in public administration. It is also suggested that data matching is far less of an invasion than manual searches since it searches only certain records instead of all fields within an entry.[10] Such examples further reinforce the assertion that data matching is not a novel technique. On the contrary, early labour-intensive implementations have preceded the fully automated processing afforded to modern bureaucracies by information technology.

This article purports to analyse the legal issues relating to data matching operations among public bodies. The terminology employed is that used by the UK Data Protection Commissioner.[11] In the US the term computer matching is used instead. Accordingly, the UK legal system constitutes the basis of this analysis; however, references are also attempted to the EU data protection Directive. While researching on matching operations, I have frequently encountered definitional difficulties, since the popular perception of data matching is misguided. This is why the first part of this article attempts to define data matching with regard to its essential components. In order to assess adequately the effectiveness of data protection legislation

[7] Lyon D, The Information Society: Issues and Illusions, Polity Press, 1991.

[8] US OTA, Federal Government Information Technology: Management, Security, and Congressional Oversight (OTA-CIT-297) 1986, next, US Office of Technology Assessment (1986).

[9] US Senate Committee on Governmental Affairs, Oversight of Computer Matching to Detect Fraud and Mismanagement in Government Programs Hearings before the subcommittee on Oversight of Government Management, 1982, next, US Senate Committee on Governmental Affairs; 1982.

[10] Kusserow R, The government needs computer matching to root out waste and fraud, Communications of the ACM Volume 27 1984.

[11] UK Data Protection Commissioner, A Guide to Developing Data Protection Codes of Practice on Data Matching, 1997, next, UK Data Protection Commissioner (1997).

with regard to data matching operations, a standard method of analysis is introduced based on the three stages forming the matching process. The data protection principles, as set in the UK Data Protection Act 1998, are then applied with further regard to the EU Directive. Finally, although they do not relate to data protection legislation *stricto sensu*, special reference will be made to the Audit Commission's National Fraud Initiative and the Social Security Administration (Fraud) Act 1997.

2 A definition of data matching operations

Filing operations can be of various types and, indeed, modern technologies have extended so much processing capabilities that conceptually different processes can easily be confused. This is perhaps the case with matching operations. On-line networks and connected databases have blurred the boundaries between data matching and similar types of processing. However, in order to proceed to concrete legal conclusions, all ambiguities require to be removed and the meaning of data matching clearly defined. Prior to a conceptual approach, an introductory technical description of matching technicalities will be attempted.

2.1 *The data matching process*

In computer dictionaries there is generally no definition of 'data matching'. Nor is there a standard process to be applied to each matching operation. The basic concept of correlating databases of personal data can, in fact, be realised through many technical variations depending on the actual circumstances. The data matching process described here is set broadly enough to cover different implementations and principles that have been adopted in recent matching programs.

Data matching participants

The participants in a matching operation can be classified into three groups: the organisation that actually runs the matching program, the organisations that provide the data intended to be matched and the organisations that receive the matched information. However, this classification is not absolute and in certain matching programs there are no distinct roles attributed to each of the agencies involved.[12] In most cases the organisations receiving the matching results are in fact the organisations that provided the personal information to be matched. Additionally, one of these organisations may undertake the task of running the matching operation on behalf of the other agencies. System operators conduct the

[12] US GAO, Computer Matching: Quality of Decisions and Supporting Analyses Little Affected by 1988 Act (GAO/PEMD-94-2) 1993.

matching exercise and system participants share their data through data matching systems and subsequently receive the matching results.[13]

We can distinguish three stages of matching operations. The first stage refers to the selection, the filtering and the appropriate transformation of data from separate databases in order to be processed during the matching exercise. The second stage involves the actual running of the matching algorithm. Finally, the third stage of a data matching exercise refers to the filtering of hits and the subsequent drawing of inferences based on the matching results. However, this classification is not absolute and certain elements of each stage may be excluded as a whole after the first match. The US General Accounting Office[14] has identified a – further – initial stage (the 'match definition and development') in which considerations of the viability of the match and planning are involved.

First data matching stage

The first stage mainly involves two tasks. First, each system participant must select which part of his records will be matched. This requires the detachment of the relevant section and its transmission to the system operator. The second task refers to the transformation of data into an appropriate form for the matching operation. Normally, it will precede the transmission of data to the system operator. During this stage it may be necessary that the organisation of data is altered according to instructions provided by the system operator, since the separate sets of records are linked by means of one or more common identifiers.[15]

Second data matching stage

The second stage refers to the matching operation itself. A matching algorithm is applied by the system operator to the files of personal data with the aim of identifying hits, that is discrepancies among records that seemingly relate to the same data subject. Alternatively, the algorithm may seek cases where an entry referring to a data subject does not exist where it should normally have been. An important distinction relates to the function of the matching algorithm. The matching algorithm can operate in either of two ways. The first method initially correlates information from several databases into a new one and subsequently the algorithm scans for discrepancies within the matched records. The second method does not create a new comprehensive database; instead, the matching algorithm

[13] UK Data Protection Commissioner (1997).

[14] US GAO, Computer Matching: Assessing Its Costs and Benefits (GAO/PEMD-87-2) 1986a.

[15] Lenk K, Automated information management in public administration OECD Informatics Studies No. 4, 1973.

scans simultaneously all participating databases and is programmed to store into a new database only the cases where it encounters discrepancies. The tool for this correlation of personal information is the usage of one or more common identifiers for each entry, representing a single data subject. Although common identifiers facilitate data matching, they are not essential, since databases can also be searched for combinations of selected factors.[16] The success of the algorithm is dependent on the transformation of data and the efficiency of the software and in certain cases the matching parameters may be changed and the matching rerun.[17]

Third data matching stage

Finally, the third stage of a data matching operation refers to inferences and decisions being drawn on the basis of the matching findings. The hits identified during the second stage form a new database to be held with the system operator. Subsequently, the database is passed to the system participants who will normally establish a procedure of filtering hits, in order to proceed to solid conclusions and actions. Additionally, the matched database may be used to update existing databases held within the system participants' information systems.

2.2 *Main characteristics of data matching operations*

Data matching constitutes a popular practice among government agencies and private organisations worldwide. A number of factors have facilitated its growth: the technological advances, the increased concern over fraud, abuse and waste, as well as the agencies' operational and technical capability to implement the match.[18] Therefore, it has been conducted since the middle 1970s in certain technologically advanced countries, largely unregulated and in secret.

The four data matching elements

Data matching ought to be distinguished from routine filing operations within automated record keeping systems. This can be accomplished by deriving the unique characteristics of data matching exercises, as opposed to other updating operations. Close examination of the few specialised statutes enacted as well as of the guidelines that several Data Protection

[16] US OTA, Federal Government Information Technology: Management, Security, and Congressional Oversight (OTA-CIT-297) 1986 p. 41.

[17] US GAO, Computer Matching: Assessing Its Costs and Benefits (GAO/PEMD-87-2) 1986.

[18] US GAO; Computer Matching: Factors Influencing the Agency Decision-making Process (GAO/PEMD-87-3BR) 1986.

Commissioners have already published[19] leads to the extraction of the common and distinct characteristics of data matching operations. First, they involve correlation of at least two databases. Second, they involve cross-examination of a significant number of records. Third, they depend on matching algorithms run on automated systems, mainly computers, and fourth, they result in administrative or marketing actions relating to data subjects whose personal information has been matched. In this context, it should be noted that the aforementioned elements must be applicable collectively for a filing operation to be characterised as matching exercise.

Data matching and closely related filing operations

At this point, a distinction needs to be made between data matching and other closely related techniques. Data linkage is the 'storage in an individual's record on one file of that person's identifier in one or more other files, to enable prompt and reliable inter-relationship of data in the future'.[20] Consequently, data linkage allows for permanent retrieval of personal information, whereas data matching is an *ad hoc* processing operation that aims at detecting discrepancies among different records relating to the same individual.[21] In certain data protection laws the issue of data linkage has been expressly elaborated. The Austrian Data Protection Bill 1974 laid down in section 8 a general prohibition: 'Personal data kept in storage in a data bank shall not be combined with personal data stored in another data bank unless explicit statutory provisions provide so, or unless it is compatible with the purpose for which the data were collected'. Furthermore it explained that this provision applied both to single instances and to permanent linkages.[22]

Data concentration is another concept closely related to data matching. It involves the merging of existing databases in order to create a 'national

[19] See especially the Hong Kong Privacy Commissioner, Matching Procedure: Some Common Questions, 1997. A different approach is adopted by the UK Parliamentary Office of Science and Technology, which adds to data matching the 'comparison of a given individual's details with one or more databases'. This definition seems to miss the fact that automated databases are not required for this kind of verification of personal data, which has always been an essential part of manual filing systems. The Parliamentary Office of Science and Technology note 93 (February 1997) seems also to be in contrast with the UK Data Protection Registrar's Guide (1997), where data matching is defined as 'the comparison of data collected by different data users . . . for the identification of anomalies and inconsistencies within single set of data or between two or more different sets'. A similar approach is adopted in the United States and by the Australian Department of Social Security.

[20] Clarke R, Computer matching by government agencies: The failure of cost/benefit analysis as a control mechanism <http://www.anu.edu.au/people/Roger.Clarke/DV/MatchCBA.html> 1994.

[21] A similar distinction is adopted by US GAO, where the term 'Front-end computer matching' refers to data linkage and is separated from computer matching. Kusserow (US Senate Committee on Governmental Affairs; 1982) considers front-end matching less intrusive as its role is to verify eligibility before payment is made. US OTA (1986) made use of the term 'Computer-Assisted Front-End Verification'.

[22] Hondius F, Emerging Data Protection in Europe North-Holland/American Elsevier 1975 p. 144.

database'.[23] Their difference is again related to the time factor: data matching is an *ad hoc* operation, whereas data concentration is rather a permanent correlation of information to be kept in a central database. Finally, a distinction within data matching operations refers to 'matching programs' and 'matches'. The US Office of Management and Budget[24] in its guidelines defines a matching program as a procedure in which a computer is used to compare two or more automated systems of records or a system of records with a set of non-Federal records to find individuals who are common to one system or set, adding that a single matching program may involve several matches among a number of participants. This distinction has not received publicity and wide acceptance and for the needs of this analysis it will not be taken into consideration.

2.2.1 Correlation of two or more databases

A fundamental element of data matching is the existence of two or more databases that incorporate personal information. Personal information stored within a database is kept temporarily in the form of raw data. When databases of personal information are assembled by public bodies, the collection of information is achieved either obligatorily, within their administrative powers, or voluntarily, when the individual wishes to subscribe to a service. Nevertheless, even in the case of a contract, a distinction needs to be made. Data may be provided without any 'opt-out' possibility or their collection may be considered essential for the provision of basic goods. On the other hand, data may be provided voluntarily by the individual who may consent to their being used for whichever causes the government regards appropriate. In any case, the individual has to make an informed choice about giving up personal data or subscribing to a service, based on information provided by the government.

An important definitional clarification relates to the different contexts under which each of the participating databases must be assembled. The UK Data Protection Commissioner[25] defines data matching as the comparison of data collected by different data users, or by the same user under different contexts. In case many system participants are involved, the relevant databases will obviously be collected for different purposes by each agency. However, when only one system participant is concerned, the databases must be assembled in different contexts, meaning that each of the databases intended to be matched must have been assembled for different purposes even within the same organisation. In the United States, however, the US Office of Management and Budget[26] supported that

[23] Clarke R, Computer matching by government agencies: The failure of cost/benefit analysis as a control mechanism <http://www.anu.edu.au/people/Roger.Clarke/DV/MatchCBA.html> 1994.

[24] US OMB, Privacy Act of 1974: Revised Supplemental Guidance for Conducting Matching Programs (Federal Register, vol. 47 No. 97) 1982, next, US Office of Management and Budget (1982).

[25] UK Data Protection Commissioner (1997).

[26] US Office of Management and Budget (1982).

matches done by an agency using its own records do not constitute data matching. Nevertheless, this opinion is not shared by the US Office of Technology Assessment.[27] Furthermore, the updating of a single database with factual changes, or even adding information in existent databases will normally not fall within the limits of data matching. Finally, it should be noted that distributed databases would be regarded as a single database for the purposes of data matching definition.

2.2.2 *Significant number of matched records*

The second fundamental element of data matching exercises is the fact that they intend to cross-examine a significant number of records. Although the term 'significant' cannot be specified arithmetically the number of matches must be sufficient to make manual searches impossible. In this context, the Hong Kong Privacy Commissioner[28] suggests more than ten while the US Office of Management and Budget[29] states that comparison of six individual student loan defaulters with the OPM file would not constitute data matching. Matching, in the sense of tracing information from different databases on specific individuals, is certainly not a recent phenomenon and is one of the basic functions of any filing system. Data matching however supposes operations that refer to a number of individuals, without any prior suspicion cast on them.

At this point a clear distinction needs to be made between data matching operations and case by case checking on an on-line system. Data matching requires a separate and *ad hoc* correlation of different databases of personal information. On the other hand, on-line systems, that may possibly be of an intranet or extranet architecture, enable constant checks on individuals on a case by case basis. The latter does not constitute data matching, but would rather fall within the area of data linkage, as defined above.

2.2.3 *Automated matching algorithm*

The data matching exercise must be based on automated processing, either on a computer or on any processing machine. Manual searches will not qualify as matching processes. The computer must apply an algorithm in order to find matches among the correlated databases and provide hits. The databases involved must be fully automated, meaning that the entire filing system must not be kept on paper. The algorithm should represent the questions that system participants address, although there is no need for it to be original, or built exclusively for this operation. Furthermore, the system operator or the provider of the software may belong either to the public or the private sector, regardless of the identity of the system

[27] US Office of Technology Assessment (1986).
[28] Hong Kong Privacy Commissioner, Matching Procedure: Some Common Questions, 1997.
[29] US Office of Management and Budget (1982).

participants. The computing process, however, complies with certain basic principles that may affect the validity of the results. The algorithm is constructed by software developers who follow the guidelines provided by the system participants. Therefore, its structure will reflect the system participants' perceptions and expectations. Socially prejudiced criteria and an accordingly developed matching algorithm may easily lead to manipulation of results.

2.2.4 *Resulting administrative actions*

The last criterion for a filing operation to be characterised as a data matching exercise suggests that its objectives must include the undertaking of administrative actions relating to individuals' whose data have been matched. An option that may vary from case to case relates to the procedure through which the aforementioned actions will be taken. The use of information systems for these decisions is a question of policies. The Privacy Commissioners of Canada and Holland[30] have identified three types of information systems: transaction-processing systems, programmed decision-making systems and decision-support systems. Programmed decision-making systems complete an entire task without any human intervention. Decision-support systems assist in the decision-making process either by generating potential solutions or by providing additional information on the basis of the initial processing. Administrative actions may be automated, based on decisions provided by expert systems and artificial intelligence, or may be effectuated by humans after examination of the matching results. In any case it should be noted that expert systems and artificial intelligence systems would only operate according to criteria fed to them by the very same humans that would otherwise take the appropriate actions.

A further issue that needs to be addressed is whether filing operations that only purport to lead to inferences on data subjects would qualify as data matching exercises. Inferences of this kind may belong in two broad categories, which, however, sometimes intersect: scientific inferences, and inferences of administrative nature. Scientific inferences are found at the basis of decisions that do not pose any threat to the individual's welfare, or in any way affect his social standing; consequently, the relevant filing operations will not meet the requirements of a data matching exercise.[31] Nevertheless, it should be noted that scientific inferences, although neutral in nature, may be used as the basis of designing administrative strategies, and therefore, may affect individuals indirectly; consequently the relevant filing operations must be considered as data matching

[30] Privacy Commissioners of Canada and Holland 'Privacy-Enhancing Technologies: The Path to Anonymity' 17th International Conference on Data Protection 1995 <http://www.ipc.on.ca/web—site.eng/matters/sum—pap/papers/anon-e.htm>.

[31] US Office of Management and Budget (1982).

exercises. In this context, Simitis[32] has stressed on the need to separate research results from any use outside of research, especially when they have taken advantage of the special status provided by data protection legislation. However, inferences of administrative nature are used as the justification of actions that directly influence individuals' social standing. Therefore, a filing operation that purports to make inferences of this nature would constitute a matching exercise when it leads to express administrative actions.

The administrative actions taken on the basis of data matching exercises could be beneficial or of an adverse nature for the individual. Acceptance of applications for social benefits or attribution of transfer payments to individuals that have not applied for them would constitute beneficial actions, whereas rejection of applications or cease of social benefits to those that had been receiving them fraudulently would be adverse administrative actions for the individual's interests. At this point a distinction needs to be made between transfer payments that are already paid or which individuals have applied for and legitimate expectations. Although there is no doubt that refusal of attributed rights affects adversely individuals' social standing, it is still unclear whether abolition of legitimate expectations bears the same effect. Valid actions, however, can only be based on correct data fed to the matching engine. Erroneous information will inevitably result in the algorithm providing false hits. Administrative actions based on unfounded inferences question the validity of data matching operations and are potentially harmful to data subjects.

3 A data protection approach to data matching among public bodies

Although data matching has been known and conducted for the last twenty years in a number of technologically advanced countries, only a few specialised statutes cover its operation and other relevant legal issues. The majority of international and national laws, rather than directly dealing with the issue of correlated databases, have introduced general principles and safeguards that would apply to any automated processing operation. Therefore, data matching is primarily regulated by indirect application of the data protection principles. Matching operations constitute automated processing of personal information held within databases and therefore fall within the scope of data protection legislation. Accordingly, the

[32] Simitis 'From the market to the polis: The EU Directive on the protection of personal data' Iowa Law Review 1995.

fundamental data protection principles address issues that are also critical in data matching exercises. Nevertheless, in certain cases data matching operations are explicitly referred to, either in specific articles of data protection statutes or in specialised legislation regulating individual categories of matching exercises. The UK Social Security (Administration) Fraud Act 1997 adopts a sectoral approach in those cases where data matching is to be conducted among known organisations for pre-defined purposes. In Sweden the Data Protection Commissioner must approve every data matching exercise. In Greece, Article 8 of the recently introduced Data Protection Act requires that a separate register on data matching operations be held with the Data Protection Commissioner.

The legal analysis of matching operations will essentially take into consideration the three stages of a typical matching exercise. The first stage refers to the selection, the filtering and the appropriate transformation of data from separate databases in order to be processed during the matching exercise. During this stage, the first legal issue raised relates to the scope of data protection legislation. Furthermore, the principles concerning the collection of personal information intended to be matched need to be analysed. In this context data quality issues have to be specifically addressed, relating to the information collected and used during the matching process. Special mention should be made of the processing of certain categories of sensitive information on individuals. The role of the monitoring authority on such automated operations is another issue of interest. Finally, the purposes of the data matching exercise would constitute the last legal issue that belongs to this stage.

The second stage involves the actual running of the matching algorithm. In this case, the principles that concern the processing of information will be analysed. Attention will be given to the actual programming of the algorithm, the criteria employed, as well as to the underlying principles of the processing in themselves.

Finally, the third stage of data matching exercises refers to the filtering of hits and the subsequent drawing of inferences based on the matching results. In this case the legal issues that will be addressed refer primarily to the logic of automated systems employed by system participants in order to reach their decisions and properly evaluate hits. A further legal issue concerns the security measures that should be taken with reference to the matched information so as to avoid unauthorised access. Apart from that, the right of individuals to access the matched data held on them, the right of subject access, has to be addressed. Finally, special attention will be drawn upon the categories of data matching operations whose purpose excludes them from a number of legislative provisions that protect individuals.

As far as the UK is concerned, the Data Protection Act 1998, perhaps in

line with the *laissez-faire* approach of its predecessor,[33] does not address explicitly the issue of data matching, a fact that has drawn criticism while the Act travelled through the parliamentary bodies.[34] Instead, data protection general principles are expected to be extended in order to cover data matching exercises.[35] Specifically, although concerns relating to data matching are considered legitimate, sections 49, referring to the development of codes of practice, and 21, concerning the preliminary assessment or prior checking of processing operations, are considered to be adequate safeguards against data matching and similar unforeseen types of processing.[36]

Furthermore, the Data Protection Commissioner issued a press release on data matching on July 16 1997 in response to the Social Security Administration (Fraud) Act. Titled 'Data matching: detection or intrusion?' the press release called for a statutory code of practice to be established on central government data matching activities. The Ministers agreed to a voluntary code but stopped short of putting it on a statutory footing. At the same time the Audit Commission, who facilitates data matching operations among local authority records to identify benefit and student award fraud, have decided to draw a code of practice for those involved in the National Fraud Initiative. The Office of the Data Protection Registrar has been asked to assist in its production. This background has led to the publication of a Guide[37] on drawing data matching codes of practice, in an effort to protect individual privacy and comply with the data protection principles.

3.1 *First stage of data matching operations*

Scope

The question of scope of data protection legislation is given a wide solution to cover processing conducted among public bodies. Recital 25 of the Directive explicitly refers to 'persons, public authorities, enterprises, agencies or other bodies' that could be processing personal information. Therefore, data matching operations conducted exclusively among public bodies are expected to fall within the provisions of data protection

[33] Lloyd I and Simpson M, Law on the Electronic Frontier (Hume Papers on Public Policy: Volume 2 No. 4) 1994.

[34] Columns 454, 467 House of Lords Second Reading.

[35] Column 478 in House of Lords Second Reading.

[36] Columns 130–131 House of Lords Committee meeting.

[37] UK Data Protection Commissioner (1997).

legislation. In cases of a sectoral approach, however, the answer would obviously be different, depending on the relevant provisions.

Collection of personal information (fair and lawful criteria)

The issue of collection of personal information intended to be matched is addressed through the principles of fair and lawful collection of data. The first three paragraphs of Part II of Schedule 1 of the Act, despite their placement, relate to the collection of information and not to the processing itself, expanding in fact the fair and lawful processing principle. The rules set in the Act are drawn in compliance to Articles 10 and 11 of the Directive. Specifically, paragraph 2 requires that data subjects be informed of the identity of the system participant either at the time of collection of their personal information, or during the first stage of data matching operations. Paragraph 3, however, limits this right, when data are collected indirectly and informing each data subject would involve a disproportionate effort or collection of this data is necessary for compliance with a legal obligation. Additionally, the Secretary of State can set further conditions to be met. The aforementioned limitations are expected to cover a large part of data matching among public bodies and may *de facto* abolish the requirement for fair and lawful collection of information intended to be matched.

Data quality

Data quality in the first stage of a matching operation refers to the already collected personal information that is intended to be matched. Each of the participating databases incorporates personal information that needs to be accurate and updated for the purposes of the matching. The requirement for data quality is also present while data are transformed into a recognisable format for the matching algorithm. The process of filtering of data, if not carefully planned, could deprive them of their accuracy by ignoring important elements of information. The Directive addresses the issue of data quality in article 6. Data should be accurate and, where necessary, kept up to date, while efforts must be taken to ensure that inaccurate or incomplete data are erased or rectified, having regard to the purpose of the processing. Since matching exercises are based on the accuracy of the information provided by the system participants, the aforementioned efforts towards accuracy should be of a highest level. The Act in Schedule 1 Part I requires that data should be 'accurate and, where necessary, kept up to date'. Since a relationship to the purposes of the processing is again attempted, in data matching the utmost care should be taken to this direction. Additionally, it should be noted that section II considers as compliant with the Act's provisions the mere indication of the data subject's views on the inaccuracy of his personal data. Nevertheless, in a fully automated operation such as data matching the incorporation of

such information in the final outcome will probably prove technically difficult and will thus be omitted.

Sensitive data

Special attention is normally given to sensitive data being processed during the matching operation. The Directive generally prohibits such processing; nevertheless, a number of exceptions is provided on grounds of consent, rights of third parties, protection of vital interests, medical purposes and public interest. Additionally, recital 34 authorises Member States to derogate from the data protection principles, if such derogations are justified by public interest in areas such as public health, scientific research, government statistics and social protection. Space for derogations is provided especially 'in order to ensure the quality and cost-effectiveness of the procedures used for settling claims for benefits and services in the health insurance system'. Section 2 of the Act adopts the definition of personal data proposed in the Directive. In case sensitive data are intended to be matched, Schedule 3 defines the conditions under which their processing is lawful. Nine subsections describe different circumstances when the processing would be acceptable, the 'management of healthcare services' case in section 9 being probably the one upon which a number of matching operations will be based.

Data Protection Commissioner

A further issue that conceptually belongs to the first data matching stage is the role of the Data Protection Commissioner. Articles 18 and 19 of the Directive provide a useful basis for efficient control on matching exercises, if such processing is customarily placed under prior notification and adequate control by the supervising authority. Article 20 allows for prior checking of operations by the supervisory authority, the government or the 'in-house' official in order to specify possible sources of risks for individual rights. Nevertheless, recital 54 diminishes the value of this provision by predicting that the 'amount posing such specific risks should be very limited'. This decision will be based on the nature, scope or purposes of the processing (recital 53). According to recital 54, the official may either give an opinion on the processing, or give an authorisation for the processing to go ahead. Data matching exercises, however, may 'present specific risks to the rights and freedoms of data subjects'. Additionally, the example provided in recital 53 ('exclusion of individuals from a right, benefit or a contract') matches the most common purpose of matching operations among public bodies. Prior checks, if combined with transparency of processing operations, through publication of their operational issues in a publicly available register, could prove a useful tool in safeguarding individual rights against matching processing.

In the United Kingdom the Commissioner is the supervisory authority

for the purposes of the Data Protection Directive (clause 51.1). As far as data matching is concerned, processing operations that would appear to cause damage or distress to data subjects are explicitly analysed in section 21. Before carrying such types of processing, which could include data matching, the data controller needs to submit a notification for preliminary assessment to the Commissioner, who needs to reply within twenty-eight days 'whether the proposed processing is likely or unlikely to comply with the provisions of the Act'. Additionally, section 22 adopts the Directive model of 'in-house' data protection supervisors 'responsible for monitoring in an independent manner the data controller's compliance with the provisions of the Act'. Finally, Part V of the Act describes the enforcement powers conferred upon the Commissioner. Specifically, the Commissioner may serve 'enforcement notices' (section 38) as well as 'information notices' in case he received a 'request for assessment' (sections 40 and 41) by a data subject. However, the Act introduces no specialised registry for matching operations. In such processing, however, the roles of the system participants and the system operator sometimes intersect and need further examination. The matching exercise may be conducted by a system operator that customarily conducts such exercises or by one of the system participants. The resulting database may be kept with the system operator or it may be shared with all system participants. Consequently, the Data Protection Commissioner needs to define which cases are covered by the obligation for registration and divide the legal responsibility among them accordingly.

Purposes of the matching

Finally, while still in the first data matching stage, the purposes of the matching operation need to be clearly set. The Directive adopts the finality principle, requiring that the purposes of the matching must be specified, explicit and legitimate at the time of the collection of personal information. Article 6.1, however, permits that they may differ from the purposes personal information was collected, but not substantially. The Act repeats the finality principle in Schedule 1 where 'personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes'. Therefore, unless an exemption is granted for the specific data matching operation, personal data cannot be used for matching exercises that lie outside the scope of their initial collection. Additionally, data should not be excessive for the purpose they are processed (Schedule 1 Part I); in this context, personal information to be included in data matching exercises should strictly relate and serve the declared purposes.

3.2 *Second data matching stage*

The second stage of data matching operations relates to the actual

processing, the running of the matching algorithm. The processing of information should be fair and lawful. The aforementioned principles, which also apply in the first stage of the exercise during the collection of personal information, need to be employed again at the stage of processing. Techniques and criteria need to be examined under this new perspective from a conceptual and a technical point of view. Since the system participants provide the guidelines to the matching software developers, further control needs to be exercised on these guidelines in themselves. Additionally, from a technical perspective, the principles of fair and lawful processing refer to the construction of such an algorithm that will avoid creating many mistakes. Attention should be given, on behalf of the system operator, to the criteria used, the techniques employed and the constraints of the system. Although appropriate filtering of hits in the third stage could reduce mistakes, the fair and lawful criteria require that quality standards of the matching algorithm be of a high level.

Article 6.1(a) of the Directive suggests that personal data must be processed 'fairly and lawfully'. Furthermore, Article 7(a) allows for matching of data only if the data subject has unambiguously given his consent. The 'data subject's consent' is defined in Article 2(h) as a freely given specific and informed indication of his wishes. However, the fact that article 8.2(a) refers to 'explicit' consent implies that in other circumstances implied consent is not ruled out. Alternatively, five other criteria still justify processing without requiring consent. Processing is allowed when related to contracts, legal obligations, vital interests of the data subject, the public interest and when legitimate interests of the controller do not override fundamental rights of the data subject.

Schedule 1 Part I of the Act requires that personal data 'shall be processed fairly and lawfully'. Part II further defines the term 'fairly' by taking into consideration the method by which personal data were collected. Furthermore, Schedule 2 sets the conditions under which processing of personal information is lawful. Probable legal grounds include, apart from individual's consent, entering or a performance of a contract, compliance with legal obligations, and vital interests of the data subject and the public administration. Undoubtedly, in data matching exercises among public bodies the individuals' consent is actually rarely given and therefore an *ad hoc* examination of each matching operation would seem appropriate. Section 6 considers lawful the processing for the purposes of legitimate interests pursued by the system participant, unless such processing conflicts the rights and legitimate interests of the data subjects. Section 6 is quite broad in scope and the Secretary of State may further specify it by order. Finally, it should be noted that the Data Protection Tribunal has taken the view that the fairness of processing will be judged by reference to the purpose of the processing, the nature

of the processing itself and the consequences for the individual affected by it.[38]

The Directive allows in Article 8.7 Member States to determine the conditions under which a national identification number or 'any identifier of general application' may be processed. Data matching exercises would improve their success rate if a unique numbering system were introduced. Nevertheless, a universal identifier would facilitate data surveillance techniques and would, therefore, constitute a serious threat to individuals' privacy. Finally, the Directive in article 14 requires that individuals be given the right to object to the processing of personal data. Section 9 of the Act confers individuals the same right when processing is likely to bring 'substantial damage or substantial distress to him or to another'. For the majority of data matching operations among public bodies such processing would seem to comply with the aforementioned requirements.

3.3 *Third data matching stage*

Automated decisions

The third stage of matching exercises refers to the filtering of information and the drawing of inferences and conclusions. In case this task is performed without any human intervention, data protection statutes explicitly provide that data subjects may object to such automated individual decisions being taken against them. In any case, the implementation of automated decision infrastructures is subject to the fulfilment of a number of conditions. Article 15 of the Directive confers the individual the right to choose not to be subject to decisions based solely on such automated processing of data. Nevertheless, the suggested exceptions on grounds of entering or performing a contract and on legal authorisation could diminish the value of article 15, given the imbalance of power that frequently exists in such situations.[39] The issue of automated decision-making is addressed in section 13 of the Act. Subsection 1 adopts the Directive provisions, by forbidding such decisions. Nevertheless, subsections 3 and 4 provide space for exemptions on grounds of entering or performing a contract, on authorisation by any enactment, as well as when the decision relates to granting a request or safeguards have been taken to protect data subjects' legitimate interests. Derogations are also justified by order of the Secretary of State, whereas data subjects can seek the assistance of a court. Although clause 13 seems to comply with the Directive, the

[38] UK Data Protection Commissioner (1997). The example further provided on 'unfair processing' may be applicable to data matching: it refers to a system that took adverse decisions about individuals without any human intervention.

[39] Lloyd I; Information Technology Law, Butterworths, 1997.

exemptions are described quite broadly and can justify data matching operations, which exclude human intervention in the stage of decision-making.

Security

The matched databases incorporate, undoubtedly, sensitive information on individuals. The Directive in article 17 requires that appropriate measures be taken to protect personal data from any form of unlawful processing. The same requirement is repeated in Schedule 1 Part I of the Act. Although implementation of such measures should relate to the nature of data to be protected, when data matching operations are conducted the sensitivity of data kept within the system calls for security standards of the highest technological and procedural level. The Directive continues in extending the above measures to agencies that process information on behalf of others. In data matching, the processing can be performed either by the system operator or by the system participants. In either case, Article 17 demands that adequate security measures are taken from all parties involved. Finally, the issue of the time period that data should be kept is addressed in Article 1(e): 'data must be kept for no longer than is necessary for the purposes for which they are collected or for which they are further processed'. Data matching exercises within the public sector purport to set a new database composed of hits derived from the initial correlated databases. The resulting database is then passed to the system participants in order to take further administrative actions. Consequently, it must be decided exactly how long is the 'necessary' period of time for the agencies to respond, after which the databases should be destroyed, or at least kept under stricter security measures. Finally, section 52 addresses the issue of unlawful obtaining or disclosure of data already held within the system participants' or system operator's infrastructures. Unless they consent, their use or disclosure to third parties constitutes an offence; space for exemptions is provided on grounds of the prevention or detection of a crime, public interest or reasonable belief.

The right of subject access

The right of subject access will normally refer to the matched database and will fall within the third data matching stage. The Directive confers individuals the right to access the matched information held on them in Article 12. System participants should also provide proof to the individual that rectification, erasure or retaining of inaccurate or unlawful data have actually been made. The relevant notification must also be addressed to all third parties. The Act refers to the right of subject access in Section 7.1. Nevertheless, it should be noted that the indexing and filing of personal information after a matching exercise will probably change, with records being updated or created in various public bodies. Therefore the policy of

demanding that data subjects address applications to each public body separately, may not always be effective for notifications or corrections of erroneous information. In this case a central registry of all matching operations run within the UK would prove invaluable.

Exceptions

Finally, it should be noted that a number of matching operations lie outside the scope of data protection laws. The Directive does not apply to areas of national sovereignty. Article 3 restricts its application to processing that falls outside the scope of community law. In this context, paragraph 2 specifies that 'the activities referred to in Titles V and VI of the Treaty on European Union regarding public safety, defence, state security or the activities of the State in the area of criminal laws fall outside the scope of Community Law'. Furthermore, recital 13 excludes processing that relates to the safeguard of the economic well being of States under the condition that it relates to State security matters. Nevertheless, this applies only to specific activities; public bodies whose activities belong usually outside the scope of community law may still be involved in processing operations that lie within its limits. This interpretation is further enforced by Article 13. The exempted provisions include the data protection principles, the notification of data subjects as well as the right of subject access. In cases of national security, defence, public security, criminal offences and important economic or financial interests, restrictions of the aforementioned rights and obligations are possible. Some of the above cases are explicitly excluded from the Directive's scope. Nevertheless, the Directive recognises that even processing that falls within community law may involve matters of this kind. On the other hand, the Act also provides space for a number of exemptions. Sections 26 to 36 refer to specific cases of matching, while section 37 authorises further exemptions to be set by order of the Secretary of the State. However, adequate procedures are put forward in a number of cases for the individual to dispute the nature of data matching and challenge its classification among the exempted categories. Crime and taxation are listed among the exempted categories, and it should be noted that the majority of matching operations among public bodies would fall within these categories. The Office of the Data Protection Commissioner has already stressed on the need of statutory codes of practice in order to protect the individuals' right to privacy, even in the above cases. A number of miscellaneous exemptions are also set in Schedule 7.

Processing for purposes of research, historical or statistical analysis

A final point needs to be made on matching operations serving purposes of research, historical or statistical analysis. The special requirements of such operations share public recognition and not only data protection statutes

but also all relevant legislative implementations involve special provisions in their favour. The Directive treats favourably such matching operations in a number of cases (Article 6(e), Article 11, Article 13.2, Recital 29). The US Office of Management and Budget[40] holds the opinion that matches done to produce aggregate statistical data without any personal identifiers do not constitute data matching operations. Such wide exemptions cannot be easily justified and involve risks of circumventing the data protection principles. Databases that are purely statistical or serve only research purposes are hard to find[41]. Additionally, the issue of safeguarding their integrity presents increasing difficulties. Integrity of a statistical database refers to whether it is safeguarded against wilful or accidental corruption of its contents, abuse of the privacy rights of data subjects, or aggregation of data so as to isolate a particular social group. Especially the latter can be achieved even without the assistance of modern technology and has always constituted a major concern[42]. Simitis[43] recognises that research presupposes a deliberate disregard of the purposes for which data were originally collected. Additionally, 'anonymisation' of data is not normally reconcilable with the needs of research projects. However, the privileged treatment for research could mean that anyone who wants to bypass the finality principle would need only to establish a research section. Consequently, it seems appropriate that further legal measures are taken to safeguard the research character of such matching operations.

4 The Audit Commission's National Fraud Initiative

The Audit Commission is a non-departmental self-financed public body, sponsored by the Department of the Environment, Transport and the Regions with the Department of Health and the Welsh Office. It was established in 1982, when the government brought local authority in England and Wales under its control. In 1990 the Commission's role was extended to include National Health Services authorities, trusts and other bodies. Finally, the Local Government Act 1992 gave the Commission additional responsibilities for the production of annual comparative indicators of local authority performance.

The Audit Commission[44] embarked on an effort to determine the extent

[40] US Office of Management and Budget (1982).

[41] Land F, The integrity of statistical data bases, OECD Informatics Studies Issue 101974, p. 114.

[42] Rodota S, Privacy and data surveillance: Growing public concern, OECD Informatics Studies Issue 10, 1974, p. 136.

[43] Simitis S, From the market to the polis: The EU Directive on the protection of personal data, Iowa Law Review 1995.

[44] UK Audit Commission, Protecting the Public Purse: Ensuring Probity in Local Government, 1993.

of fraud and corruption suffered by local government and to propose good practices for its prevention and detection. Data matching operations held a leading role among the solutions suggested. Such operations would combat efficiently benefit and student awards fraud. In this context, the Society of London Treasurers has established the London Committee for Action Against Fraud which in turn established a full-time team, the London Team Against Fraud. Among the tasks entrusted to the latter was the facilitation, with the help of the Audit Commission and the technical assistance of District Audit of data matching operations among local authorities in order to identify multiple claims filed in London Boroughs. The 1994 update of the report claims that some 90 per cent by value of 1993/94 student award frauds were discovered, as a result of such data matching operations. The update continues in noting that no other such bodies have been set up elsewhere in England or in Wales. Nevertheless, it encourages such initiatives while at the same time asking for their compliance with the Data Protection Act 1984.

The 1995 update analyses extensively the so-called 'inter-authority initiatives'. It praises the progress accomplished in the dissemination of matching operations and refers to a joint effort of local authorities with the Universities and Colleges Admissions Service (UCAS) to combat fraud in student awards. Furthermore, a National anti-Fraud Network has been established, within which the Societies of County, District and Metropolitan Treasurers have agreed to co-operate with the London initiative in an attempt to stem the potential spread of frauds of the types uncovered in the London area. The update also provides some numbers on data matching efficiency: The London initiative has resulted in the identification of student award fraud totalling £2 million. By matching housing benefits and student awards, the level of student award fraud detected nationally increased tenfold in 1993/94. The update concludes by providing a list of similar data matching initiatives taken in other parts of the country. Again, local authorities are asked to comply with the Data Protection Act 1984, although 'provided all aspects of this guidance are followed, problems should not arise in relation to data protection legislation'.

By 1996 data matching initiatives had been undertaken by several local authorities. In the Greater Manchester Area a system of data matching detected £1 million of fraudulent housing benefit and student award claims.[45] Consequently, the Audit Commission stated that they would like to see the use of data matching systems to identify people making multiple fraudulent claims implemented by all local authorities by 1997. Finally, in 1997 the Audit Commission asserted that councils were detecting more fraud than ever before and that they had major success in using computerised data matching techniques to focus their investigations.[46]

[45] UK Audit Commission, Protecting the Public Purse: Ensuring Probity in Local Government, 1996.
[46] UK Audit Commission, Protecting the Public Purse: Ensuring Probity in Local Government, 1997.

The update concludes that the Commission will continue with its data matching initiatives in partnership with councils.

Social Security Administration (Fraud) Act 1997

The Social Security Administration (Fraud) Act 1997 constitutes an example of sectoral legislative approach to data matching operations. Specifically, it refers to matching records among specific public bodies – the Inland Revenue, the Department of Social Security, the Duties and Excise Office and local authorities – in order to combat fraud in social benefits. The Act is justified by recent technical developments rather than the need for regulation, being publicly admitted that 'there was no legal obstacle to cross-checking those internal records' and that matching exercises have been conducted since 1985.[47] It should be noted that since these operations are conducted within public bodies whose activities are mostly exempted from the Data Protection Act, the Registrar has no way of knowing the criteria employed and no authority to monitor the process. Even an enforcement letter from her office would prove useless, according to section 38(2) of the Act. Furthermore, the issue of introducing smart card technology in public administration is also addressed.

According to sections 1, 2, 3 and 4 of the Act, data may be exchanged and matched among a number of public organisations and local authorities in order to combat fraud on transfer payments and social benefits. Nevertheless, an explicitly set aim is to 'maintain and improve the accuracy of social security information'. Furthermore, although data matching will be used in order to combat fraud, it will not be used in order to attribute benefits to those entitled but not receiving them.[48]

None of the parties represented in Parliament disputed the necessity for conducting data matching exercises. The problems created by fraud within social security policies were unanimously acknowledged and data matching was considered a powerful weapon in the public administration arsenal. Nevertheless, questions were addressed as to whether individuals' privacy would be adequately protected, whether the Act is in accordance with the Directive and whether its provisions do not constitute an overriding of the data protection principles. Finally, concerns were expressed as to whether such an exercise is plausible, considering contemporary technology level.

To the aforementioned concerns, the official response was that public interest in this case overrides possible privacy infringements and that appropriate safeguards would anyway be taken. The matching will be conducted by a limited number of skilled officials in a single site – although

[47] Mr Peter Lilley, Column 64, Parliamentary debates, House of Commons *Hansard* Debates of 25 November 1996.

[48] Nevertheless, the Parliamentary Office of Science and Technology (1997) warns that any system which flags cases for investigation needs to bear in mind that the majority may be legitimate claimants, or even receiving under payment.

not defined if they may belong to private organisations as well – and it will be an offence and 'gross misconduct within the Department' if information were disclosed to third parties. Finally, only relevant data will be transmitted by system participants and data will be provided on a 'need to know' basis. Explicit assurance was also made that 'the Department would not sell on information received under the data matching process'.

As far as legal conflicts to the Directive are concerned, the Act does not constitute a breach of its provisions, as it refers only to combating fraud, which is covered by the appropriate legislation, and, in any case, data matching will be conducted only after evidence of fraud. Finally, the issue whether the fundamental Data Protection Act's principles on data processing, namely the issues of fair and lawful processing and the finality principle, are overridden is not directly answered. Instead, it is admitted that an 'expansion of its rules' is attempted, and that, in any case, the system participants will be registered with the Registrar's office. Additionally, voluntary codes of practice – but not statutes – will be introduced in order to protect civil rights. It should also be noted that to the administration's mind 'we are not requiring any more information from individuals than is already taken'.

The Social Security (Administration) Act 1997 constitutes an example of sectoral data matching legislation, based on the space for exemptions provided by the general data protection legislation. Nevertheless, its implementation appears problematic. The Act's provisions conflict with the basic data protection principles and it also seems to ignore the Data Protection Commissioner's existence. Although its purpose appears to serve public interest in terms of financial saving, it does not involve a prior cost benefit analysis nor does it provide any assistance to individuals disputing or trying to rectify erroneous information. Additionally, the life span of the matched information is not defined nor is it defined who and under what safeguards will be responsible to keep it. Finally, the aim of the Act is defined as 'preventing, detecting, investigating and prosecuting offences relating to social security and for maintaining and improving the accuracy of social security information'. This description leaves space for matching exercises that do not purport to combat fraud but only to update government records, a policy that could be interpreted as opening the way for a Big Brother state.

5 Conclusion

A point that deserves special attention is the lack of definitional clarity on what exactly constitutes data matching operations. The public and even government bodies often confuse data linkage (popularly perceived as 'on-line' systems) to data matching. However, these terms refer to entirely different processing and they involve different risks and operational difficulties.

Data matching operations constitute a by-product of the information society. Public and bodies have collected enormous amounts of information in the course of their duties; their effective processing manually would require disproportionate efforts. Matching is a feasible method of processing for extracting quickly and efficiently the information required. As such, it is an essential part of modern administration and calls for its abolition are unfounded and irrelevant.[49] I believe that the objectives that matching operations efficiently serve and the fact that no alternatives exist have, in fact, imposed them in modern administration.

Problems have arisen, however, with data matching in the short term. It has resulted in unfounded hits that have caused distress and even damage to individuals. It has also brutally invaded their privacy, revealing information on their most intimate affairs: income, health, spending patterns and social life. However, data matching is a recent technique. It has only twenty years of history in the US and much less in other technologically advanced countries. Therefore, it can be disputed that all these problems refer to its 'infancy' stage. The issues of compatibility among diverse operating systems used in large organisations will hopefully be resolved in the near future.

Furthermore, mistakes would probably be abolished should a universal numbering system be introduced. Universal identifiers have attracted wide public criticism. They are accused of entailing civil rights risks and in some countries they are even constitutionally forbidden.[50] Nevertheless, such criticisms reflect a lack of social understanding. Information constitutes the basis of everyday life. We live, we work and we communicate on the exchange of information. Universal identifiers could save resources, making personal data correctly recorded by computer systems and automating transactions. I do not consider their global adoption threatening to privacy. Even countries that have opposed to their introduction (i.e. United Kingdom) use a number of other identifiers (registration number, driving licenses) for administrative purposes. A global universal system is in line with this concept and would sum up existing disparate schemes.

The individual is perhaps left legally and administratively unattended in his struggle against privacy infringements. Data matching indisputably invades his privacy in a most brutal manner. However, this is only a part of a general trend within the information society. Information is transmitted and recorded and processed and transmitted again and that is what information society is all about. Individuals live uneasily within this reality, since efficient data processing is only a recent phenomenon. Nevertheless,

---

[49] US Senate Commission on Governmental Affairs, Oversight of Computer Matching to Detect Fraud and Mismanagement in Government Programs Hearings before the subcommittee on Oversight of Government Management, 1982.

[50] Germany, Portugal, Austria and Hungary have constitutional limitations on the establishment of a single national number, Davies S, Big Brother, Pan Books, 1996, p. 79.

citizens should develop an understanding of the new information reality. The public should become aware that all information they transmit is now permanently stored and immediately processed. Older filing systems that made individual searches impossible have become obsolete. Consequently, individuals should control how much information they transmit in their everyday transactions. They should also be educated to take advantage of information technology tools. Systematic use of privacy enhancing technologies would constitute an obstacle to widespread matching operations.

From a legal aspect, it is not believed that additional legislation to data protection statutes is required. The safeguards set for all processing operations can be extended to cover data matching. Codes of practice could prove useful but only if they are set on a statutory basis. The only relevant proposal relates to a specialised registry of matching operations to be held with the Data Protection Office. That would facilitate exercise of the right of subject access and would enable the Data Protection Commissioner to monitor efficiently data matching operations. Additionally, specific provisions should make clear that each and every matching operation is assessed by the Data Protection Commissioner prior to its actual conducting. For the part of matching operations that lie outside data protection scope (social security, tax administration, state security) they should probably seek prior parliamentary approval and procedures for monitoring their results should be established within the appropriate departments.

As the Bangemann report notes, the information society will change the way people live and work together. Data matching and data linkage will become widely used by public and private administration in the very near future. This will promote efficient service and reduced costs. Individuals are caught in a transitional period, between the traditional manual and the modern automated administration. Nevertheless, it is only a matter of time until the latter dominates and individuals should start taking advantage of the new capabilities it offers and avoid the risks it involves.