Available online at www.sciencedirect.com

SCIENCE DIRECT®

PERGAMON International Journal of Information Management 23 (2003) 185–199

International Journal of
Information
Management

www.elsevier.com/locate/ijinfomgt

Data Base Management Systems (DBMSs): Meeting the requirements of the EU data protection legislation

Anastasios Gounaris[a,*], Babis Theodoulidis[b]

[a] *Department of Computer Science, University of Manchester, Oxford Road, Manchester M13 9PL, UK*
[b] *CRIM, Department of Computation, UMIST, P.O. Box 88, Sackville Street, Manchester M60 1QD, UK*

Abstract

As the size of the databases containing personal data is expanding very fast worldwide, the mass collection and processing of personal data has raised a lot of concerns about the manner in which the personal data of an individual are processed. In an effort to address privacy concerns, the European Parliament adopted the Data Protection Directive, which enforces organisations to take steps to ensure their compliance. Current database technology fails to allow organisations to comply with the requirements of the new data protection legislation. In this paper, a complete set of the DBMS operability requirements is presented, in order to support the EU Directive. These requirements affect the database facilities to identify individuals and for audit trail, the security and processing mechanisms of the DBMSs, and the kind of data that needs to be stored. An implementation model is also proposed.

Keywords: Data protection; Databases; Database management systems; Privacy

1. Introduction

Innovations and enhancements in computer processing power, disk storage, memory and networks have been close to explosive. Databases with information about every aspect of life are now measured in gigabytes and terabytes. Organisations are constantly collecting data about their present or potential customers through the Internet, by buying marketing lists and so on. Thus, the size of the databases containing personal data is expanding very fast worldwide. At the same time, this mass collection and processing of personal data has raised a lot of concerns about the manner in which the personal data of an individual are processed (Cranor, 1999; Reden, 1999). Privacy and security are already the two critical ingredients in building customer confidence. In an

*Corresponding author.
E-mail address: gounaris@cs.man.ac.uk (A. Gounaris).

0268-4012/03/$ - see front matter
doi:10.1016/S0268-4012(03)00023-9

effort to address privacy concerns, the European Parliament adopted the European Union (EU) Data Protection Directive 95/46/EU in 1995 (EU, 1995). Each country member of the EU had the option to create its own implementation of that directive. In the present paper the focus will be on the British implementation, without this meaning that there are essential differences with the statutes in other EU countries. On the contrary, as the British Data Protection Act is an extension of the EU directive, this work covers the legislation of any country member. Furthermore, it fully implements the OECD[1] Privacy Guidelines on how best to balance privacy protection with the free flow of personal data. Member countries of the OECD include Australia, Canada, Hungary, Japan, Korea, Mexico, New Zealand, Switzerland, Turkey, the US etc. So, the problem addressed is a global one.

In the United Kingdom, the EU Directive is implemented by the Data Protection Act (DPA, 1998, Chap. 29), which came into force on March 1, 2000 (DPA, 1998), and is supervised by an independent authority, called the Information Commissioner (IC,[2] former Data Protection Commissioner). The DPA regulates the way personal data, and especially sensitive ones, are collected and manipulated 1998 (DPA, 1998; IC, 1998; BCS, 1998; BSI, 1999). According to it anyone who processes personal data must comply with the eight enforceable principles of good practice. These principles state that personal data must be processed for limited purposes in a fair and lawful manner and in accordance to the individual's rights. Moreover, personal data should be adequate, relevant and not excessive in relation to the process purpose, accurate and secure. Additionally, personal data must not be kept longer than necessary or transferred to countries without adequate protection measures in place (DPA, 1998). Data controllers have to ensure that information about individuals is gathered, processed, used, disclosed, and disposed of, in a way that ensures confidentiality, data accuracy and legitimate behaviour. The challenge is to apply the information management techniques (like data mining, text mining, pattern recognition, machine learning, statistical data analysis, neural networks, visualisation) without breaching any of the Data Protection Act principles. Since DPA (1998) came into force, data protection has ceased to form just a matter of good practice, resting with organisation's good will. Nowadays, safeguarding individuals' data is a Law requirement as well. Directors can also be criminally liable if they do not take steps to ensure their organisation's compliance (McKilligan, 2000).

The most important elements of good practice, which the Commissioner has sought to promote, are "the transparency, the fairness, the purpose limitation and the security". By transparency it is meant that except in cases where it would be prejudicial to the prevention and detection of crime or the collection of taxes, data subjects should be fully informed of the uses to which their personal data may be put. Fairness is the general principle that personal data should be processed in a way, which is fair to data subjects. Although fairness may be difficult to define in precise terms, clearly it is a concept, which embraces the notion of equity and is opposed to any action, which is discriminatory. Purpose limitation defines that, having collected information for specified purposes, that information should not be used or disclosed for other purposes. Security implies ensuring that data are stored in a manner appropriate to the sensitivity of those data and are not disclosed to others. Finally, data controllers should notify the Commissioner of the

[1] Organisation for Economic Co-operation, Development. http://www.oecd.org/.
[2] Guidance & other publications. http://www.dataprotection.gov.uk/dpr/dpdoc.nsf.

personal data kept, how and why the data are processed, who they are disclosed to, and where they might be transferred.

Database Management Systems (DBMSs) are the main 'tool' used to store and manipulate personal data. As a result, companies should pay particular attention to their database design and manipulation and the way these can conform to the Data Protection Act. Current database technology is not able to comply with the requirements of the new data protection legislation. In addition, database systems cannot ensure full data confidentiality and most of them disclose to users, either directly (Denning, 1999) or indirectly (Brodsky, Farkas, & Jajodia, 2000; Chor, Kushilevitz, Goldreich, & Sudan, 1998) more data than they need to perform their official duties, thus violating one of the eight principles of the Data Protection Act. Furthermore, addressing a data subject access request is not included explicitly in the system's functional requirements. Also, out-of-date back-up data can rarely be deleted and is difficult to provide a response to a subject access request within the requirements of the Data Protection Act. New systems are still being designed without the capability to maintain audit trails (McKilligan, 2000; BSI, 2000). No information about data sources, process purposes, data subject's consents, retention time, access rights, recipients and relevant filing systems containing personal data is held. And this is just an incomplete list of potential problems that can arise given the current support that database technology provides to address the data protection legislation (IC, 2001; PISA[3]).

The main problem lies in the fact that although organisations put adequate procedures in place to safeguard the way the personal data they hold they can be used, problems are generated by DBMSs which do not support the relevant procedures. In this paper, a complete set of the DBMS operability requirements is presented, in order to support the EU Directive. The requirements refer to front-end database applications and their implementation is a big step towards the full compliance with the DPA (1998). The full description, analysis and explanation of these requirements has been presented in Gounaris and Theodoulidis (2001).

2. Related work

At the best of our knowledge, there is no prior work that examined the implications of the Data Protection Legislation for the DBMSs at such a level. The Office of the Information Commissioner has recently published a series of drafts of consultation and codes of practice towards the compliance with the DPA 98 (IC[2]). All these guides are general and not technical. They provide a clear description of the data protection context and the data protection issues that need to be considered. In this way, they provide the basis for further investigation in data protection issues and for the specification of data protection enhancing database requirements.

The British Standards Institute has also published guides (BSI, 1999, 2000, 2001) with the assistance of experts from industry and advice and support of the Office of the Information Commissioner. All these guides, apart from BSI (2000) are general, too and give advice to organisations about how to ensure compliance with the Data Protection Act 1998 when managing their information processing operations. The BSI (2000) takes one step further and proposes some IT operability requirements for systems that deal with personal data. It focuses on the

[3] Building a privacy guardian for the electronic age. http://www.tno.nl/instit/fel/pisa/.

implications of the legislation for database management, including maintaining data accuracy, marketing issues, managing data subject access, and obtaining and recording consent. The importance of these guides lies in the fact that they provide the background for the database requirements to be derived, along with the guides published by the Information Commissioner. Pounder and McLean (1998) offers another approach for compliance with the DPA in practice.

The DBMS industry has taken some actions recently in order to solve the privacy problems that organisations face with their legacy DBMSs. Examples of such actions are the solutions presented by NCR,[4] IBM[5] and Oracle.[6] NCR has enhanced its data warehousing products with certain services that help organisations to comply with the Data Protection Law. These services can recommend an approach for a privacy implementation within a data warehouse environment, assist data controllers in determining their business need for data privacy and review and give feedback on a privacy implementation (NCR[4]). IBM Privacy Research Institute has been working on the development of novel data protection technology. Although the main focus is not on DBMSs, they have presented a technique for data protection-enhanced data mining (IBM[5]). Complementary to that, Oracle has concentrated its efforts on meeting the data security aspects of the DPA (1998) (Oracle[6]).

3. Building DBMSs that are complied with the Data Protection Legislation

The recent Data Protection Legislation affects many aspects of the processing of personal data held in databases and manipulated by DBMSs. Such aspects are the infrastructure that needs to be provided by the DBMS in order to identify a data subject, the additional data and metadata that are required, the audit trail facilities, the security of the data and the mechanisms to process personal data. Each of these aspects is examined separately in the following subsections.

3.1. Data subject identification

The first step towards the enforcement of a data privacy policy in databases is to provide the capability to identify uniquely a person whose details are kept in that database. In many cases, companies and organisations use the name of the individual combined with other attributes, like his address, in order to infer his identity. In practice, this approach is inadequate. E.g., if a person is recognised by the combination of his name and his address, there is no means to distinguish between two persons with the same name living at the same property. The solution is to use unique identifiers for every data subject. All the personal identifiers that are used in local data stores or in organisational procedures should be linked to that unique personal identifier, in order to enable a complete view of someone's personal data. That measure does not necessarily implies that the system is capable of locating and printing out all the personal data held in the organisation's databases concerning a single individual. The reason for this is that even if unique identifiers are in place, they are of little help unless the underlying schema of the database is

[4] NCR Worldwide Services—Services for Data Warehousing Solutions. http://www.ncr.com/services/sol_dw.htm.
[5] IBM Privacy Research Institute. http://www.research.ibm.com/privacy/.
[6] Oracle Corporation. http://www.oracle.com/.

designed in such a way that all pieces of someone's personal data are either in the same record with or related to the individual's identifier.

Moreover, a key right under the recent Data Protection Legislation is the right for each individual to access his personal data. At this point, adequate attention should be paid not to disclose this data to persons that pretend to be the relevant data subject. Apart from other policies within the organisation, it is desirable for the DBMS to have built-in facilities for that functionality.

Two other issues that are closely related to the requirements described above, are the indirect identification of a person and the erasure of someone's personal data. It is very common the context of the information to be what makes the information personal or not. E.g., in a company with many thousands of customers, a customer's initials are not personal data. But in a customer list with only a hundred entries, the initials either identify uniquely or help to identify the relevant person, and thus it becomes personal data and is protected by the Law. Complementary to this is how the total erasure of someone's personal data from a database can be achieved, as all the links and combinations of data that help identifying him need to be detected and deleted.

To summarise, the requirements for a DBMS with regard to the identification of a data subject and his personal data is that it should have the capability to:

- identify uniquely every data subject;
- keep details about all the data subject identifiers used in the organisation systems;
- locate all the personal data held in organisation's database concerning a single individual;
- print out all data concerning a specific individual (including data in back-up and other files and providing a translation of any codes used);
- deploy authentication procedures to ensure that the person who makes a request to access his personal data is in fact the relevant data subject;
- keep details about what data identify the data subject in which context;
- delete the information and the links that enable individuals to be identified.

3.2. Additional data and metadata

The Data Protection Law has made obligatory for everybody processing personal data, in any way, to keep additional information that had negligible importance so far. In most of the real-world cases, and in all the cases where personal data is stored in databases for commercial reasons (e.g., databases with customers' data), personal data entries should be accompanied by entries about the consent of the data subjects for the companies to process and use their personal data. Further, the consent of the data subjects for transferring their data in countries outside the European Economic Area, where the local Data Protection Legislation may be more relaxed, and for publishing their data on the web and thus making them publicly available, should be kept in the databases.

The set of information that data controllers need in order to enforce the data protection policies within their organisation is even more extended, as they are obliged firstly to support certain activities related to the DPA (1998), and secondly, to allow governmental bodies to check whether the organisation has complied with the data protection rules. Such an activity is the notification, according to which the data controller has to notify of the process purposes for the personal data

of an individual to help ensuring that this data is processed in a fair and controlled way. Also, the data subjects should be allowed to make use of their right to define for themselves the manners they can be contacted for marketing purposes (e.g. via telephone, mail, fax, e-mail, SMS, etc.). If they declare that they do not prefer to be contacted, the company is obliged to respect their decision, and to proceed to specific actions like removing their entries from their lists and ensuring that their details will not be added again. In case that the data subjects request to access their personal data, the data controller should include in the report, apart from the personal data held in databases, the personal data held in back-up files, in flat electronic files (e.g., emails), and in manual files as well. Moreover, the report should include information about the automated mechanisms that are used to process his personal data and about the persons to whom his personal data have been or might be disclosed. The recipients of the personal data do not have to be outside the organisation, as they include the employees of the data controller who are authorised to process the data. Keeping information about the data source is another aspect of the personal data that should exist in these reports. This information is also important for achieving high quality of data, as it can help distinguishing data deriving from unreliable sources. Minimising the amount of excessive information stored is one of the most challenging tasks of the data controller. A promising solution to that is to restrict the usage of "free text" fields within the databases. Thus it becomes more difficult to enter information that is not really needed and consequently forbidden by the Data Protection Law. Another activity for which the data controllers are responsible for is the application of a retention policy, as they are not allowed to keep personal data for an arbitrarily long time. For each category of personal data, a specific retention policy exists and should be followed.

Applying a data protection policy is a problem with many dimensions. Data controllers may want to employ different policies for different pieces of personal data. Also, data may be collected for different purposes and from different sources. Consequently, the DBMSs should be capable of treating the items of information individually or in logical groups with regard to data protection metadata (e.g., it would be unrealistic to have to assign one value for the consents of the individual for all his or her personal data). Moreover, the system should allow for a clear distinction between sensitive and ordinary personal data and between personal data that processed for research purposes and data processed for other legitimate purposes, as there are specific provisions of the Data Protection Legislation for these kinds of personal data and processes.

In general, the DBMS should have the capability to:

- treat items of information individually or in logical groups;
- keep details about the data subjects' consents
  - for processing their personal data,
  - for transferring their personal data outside the EEA,
  - for publishing their personal data on the Web,
- keep details about
  - the process purposes of data subjects' data,
  - the data subjects' marketing preferences,
  - all back-up, flat electronic and manual files that contain data subjects' data,
  - the possible recipients of the personal data,

  - the persons authorised to manipulate the personal data,
  - the roles of the authorised users within the organisation,
  - the sources of the personal data,
  - the data subject's access requests,
  - the retention time of a piece of personal data,
  - the logic of any automated decision mechanism that is used to process a data subject's personal data,
- distinguish personal data between sensitive and non-sensitive;
- distinguish which data is processed for research purposes;
- use extra optional information fields and free text fields only for justified purposes.

3.3. Audit trail

Audit trail, in relation to databases, is a means of tracing all activities affecting a piece of information, such as a data record, from the time it enters the database to the time it leaves. An audit trail documents the path from the input to output and should provide enough information to reconstruct or verify the entire sequence. It is deemed as a key functionality of a database system in order to be in compliance with the Data Protection Directive of EU. Each DBMS should have the capability for audit trail in order to keep details about who, when and for what purpose accessed, amended, deleted or disseminated which data. It is critical for ensuring the fair processing of personal data and the processing of such data in line with specified and lawful purposes. It also safeguards personal data against unauthorised use and is needed in order to address a data subject access request.

The audit trail should also be used to check whether every process performed on personal data is related to a notified process purpose, and thus ensuring the adequacy of the notification.

Compared with traditional audit trail mechanisms, a data-protection enabled audit trail differs in that it does not only detect misuse of data, but tries to prevent it as well.

To summarise this aspect of the DBMS requirements, the system should have the capability:

- for audit trail and logging;
- to audit the processes of personal data to ensure that the notification adequately covers the processing activities;
- to deploy mechanisms to ensure that misuse of personal data within the organisation can be identified and remedied;
- to restrict the processing of personal data unconnected with the organisation activities using the organisation resources.

3.4. Data security

The high-level objectives of data security are the secrecy, the integrity and the availability of the data (Jajodia, 1996). The secrecy, or confidentiality, is concerned with the unauthorised disclosure of information, the integrity is concerned with the unauthorised modification of information or processes, while the availability is concerned with improper denial of access to information. The means to achieve these goals and the foundation for database security is provided by the

authentication, access control and audit together. Auditing was examined separately and this subsection deals with the first two dimensions.

A common mistake in the database security policy of an organisation is that the access control policy is deployed, usually by using passwords, at a very coarse granularity. A single password is often enough to gain access to all the data held in the database, thus relaxing significantly the security constraints. But even if robust access control mechanisms are in place, covert channels and inference channels may bypass them, and sensitive personal data may be disclosed to unauthorised persons. Moreover, it is quite challenging for a database administrator to manage to give access rights to the system users, exactly up to the extent it is required to perform their duties. Another common cause of unlawful data disclosure is that personal data of many persons exist in the same file or record, and so, when one of these persons requests to access his data, he is also revealed personal data belonging to others. Finally, very few systems can claim that they have taken efficient security measures to safeguard the data from electronic attacks or system failures.

So, as far as the data security is concerned, the DBMS should have the capability to:

- deploy appropriate access control mechanisms;
- ensure that the database does not allow unlawful disclosures of personal data;
- ensure that when someone requests personal data about another person, such data should be released only, and to the extend that, he requires the data in order to perform his official duties;
- ensure that no personal data of another data subject are disclosed, when a data subject requests to access his personal records;
- deploy security techniques to safeguard data from internal and external electronic attacks;
- back-up data;
- provide a secure method of transmission, when personal data are collected or processed on-line.

3.5. Data manipulation

Until now, the capabilities of the DBMSs were exclusively motivated by the necessity of new functionalities that facilitate the information processing. The DPA (1998) asks for new functionalities, which may complicate the design of such systems but are crucial in meeting the requirements imposed by the Act.

E.g., for many reasons, a data controller may choose to create copies of a database. Some of these copies are to be kept online and in use, and some serve as back-ups. In general it is desirable to have back-up copies, as it was stated in the previous paragraph. On the other hand, when it is required to delete or amend some data, the DBMS should have the capability to do that not only in the main database, but also in any place where this data is copied, including all the replicas and backup files. Apparently, this procedure should be automated and form and integral component within the DBMS. Imagine a situation where the retention time of thousands of records expires simultaneously. It would be unrealistic to update the database manually.

Another important aspect is how the quality of the data held in a database can be preserved. The major threat comes from new data that are inserted without having previously been checked. Situations like storing new lists with potential customers in the system without checking whether the persons on that list have already stated their marketing preferences, or integrating databases by simple matches on the name attribute, should be avoided. For the former case, ''suppression''

files with the preferences of persons whose data are not currently in the database seems to be a good approach. Similar solutions need to be found for all the other kinds of processing, to which a data subject has the right to object. Furthermore, the user of the system should be provided the set of the legitimate process purposes for each personal data item he manipulates.

Loss in data accuracy and integrity can be incurred also by changes in the organisation's IT infrastructure and this should be kept in mind during system upgrades. For avoiding the storage of excessive data, a promising direction can be to develop tools that taking as an input the processes that run within the organisation, produce the complete set of personal data that need to be collected and stored in the organisation's DBMSs. In most of the cases, in order to permit audit trailing and more processing options, the valid time and the transaction for the personal data should be kept. However, the most challenging task is to develop such mechanisms for all the kinds of databases, and not only for the traditional ones that hold structured data. A big amount of personal data exists in paper files, flat electronic files, emails and so on.

Consequently, and complementary to the rest of the DBMS requirements explained previously, the system should have the:

- capability to
  - amend and permanently delete the data,
  - automatically delete the personal data, when its retention time expires,
  - build and use suppression files,
  - block or suppress the processing of data subjects' personal data when the data subject objects to the processing,
  - merge databases and integrate information sources using complete matches of all criteria,
  - hold scanned paper files and other computer files apart from databases, that contain personal data (e.g. multimedia databases),
  - change the organisation's hardware and software systems avoiding loss in data accuracy and integrity,
  - map directly the process diagrams of an organisation to the class diagrams of the organisational database,
  - support bi-temporal semantics,
  - manipulate back-up files,
- capability to provide those who are dealing with personal data information about the purposes for which this data has been collected.

4. An implementation model

Each organisation has many sources from where it acquires the data that is necessary to perform its operations. These sources vary from paper to structured electronic files. The relevant data populate the organisation's databases. In general, these databases are independent and heterogeneous and they are part of a distributed database management system if they are connected together. In order to address the data protection problems (and many other problems related to efficient information management), firstly all the elements of these databases that contain personal data have to be integrated, taking into consideration the Law requirements. So,

a new database is created. The data controller operates on this integrated database. The persons to whom the personal data is disclosed are called recipients. The recipients can be either the system's authorised users that access the data in order to do their job or people outside the organisation, who can acquire this data either directly (e.g. via the organisation's website) or through the system users. The integration procedure depends on whether the integrated database is conceptual or physical. Also, the whole implementation procedure of this architecture will be different if it is constructed from the beginning or it is applied to an existing database system.

The architecture of the DBMS we envisage is shown in Fig. 1. It refers to the integrated view of all the databases that exist within an organisation and contain personal data. The operability requirements presented in the previous section can be reclassified into three new categories according to their implementation procedure (this classification is presented in the appendix). Some of requirements require extensions to the set of metadata held in the database. Such data protection-related metadata include information about the retention time of personal data, their source, process purposes and so on. This category matches more or less the category "additional data and metadata" of the initial classification, which has been discussed in Section 3.2. The implementation of the extensions result in modifications to the database schema. As a result, the database derived contains both the integrated view of personal data scattered across the organisation's databases and their metadata. In other words, the procedure mentioned above results in a new enhanced database, which is capable to hold additional information that is necessary for the compliance with the Law.

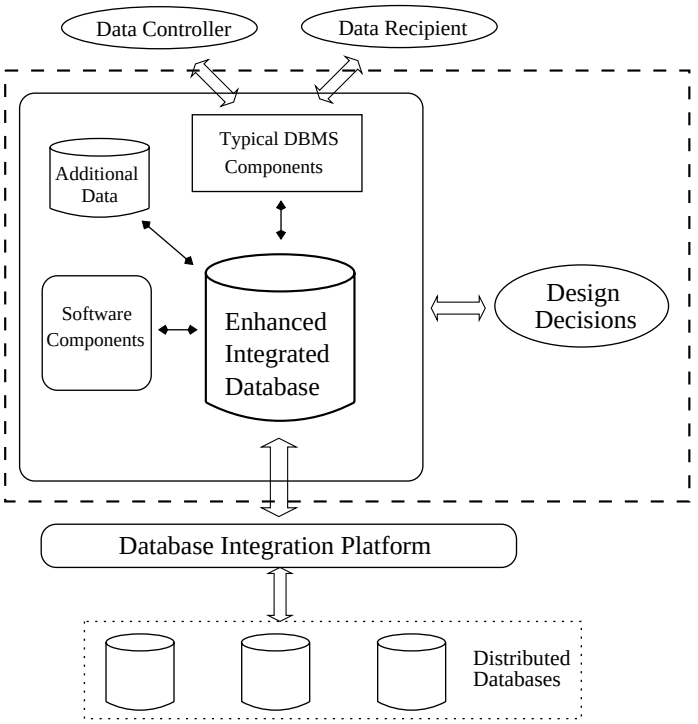

Fig. 1. The architecture of a data protection-enabled DBMS.

Another part of the requirements implies the enhancement of the DBMS with new capabilities. The data controllers may choose to develop the relevant mechanisms within the database applications. A second approach is to implement the new functionalities within the main DBMS. Apparently, this method is more efficient as the same components will not have to be inserted to each database application separately. So, a software module is incorporated into the DBMS for each new data protection-related capability. Examples of such modules are the mechanisms responsible for audit trail, logging, back-up and data restoration during disaster recovery procedures.

There are also some requirements that cannot be applied in a straightforward manner. Certain design decisions should be made and specific business and technical issues should be researched. These decisions and the outcome of the relevant research will define the level at which the system will support the Data Protection Act and also affect the design of the system and the software components. Solutions may vary significantly among different organisations. For example, each organisation has its own security policy and an investigation of all possible ways of electronic attacks is a prerequisite for taking decisions on it.

The typical components module in the figure consists of all the sub-modules that are common in a DBMS, such as the Query, Data Definition Language and Data Manipulation Language Compilers, the Data Dictionary and the System Catalogues, the Run-time database Processor, the Concurrency Control and Backup/Recovery Subsystems and so on. As the focus in this paper is on the data protection aspects of a DBMS, these components are not presented in detail.

The model of a DBMS proposed provides to the data controller of an organisation the necessary means to perform his or her duty. Any operations performed on such DBMSs on behalf of data recipients are ensured to be lawful at a very large extent. The data controllers themselves of the organisation can act on the data in three modes. Firstly, when they address a data subject access request. Secondly, when they notify to the Commissioner or check whether the notification covers the organisation's activities. Finally, when they ensure the compliance with the other parts of the DPA (1998) preparing the system for a possible assessment by the Commissioner. The data controllers need to be provided with all the relevant information in order to perform their duties. This necessity is what defines the set of the system requirements and forms an evaluation method of the requirements' validity.

5. Conclusions

At the best of our knowledge, the work described in this paper is the first attempt to investigate in depth the implications of the EU and British data protection legislation for databases and then to present technical solutions for the arising problems and proposals for systems complied with the Law. One of the main reasons that this has not happened yet is that there are some general misunderstandings about the notions of data protection and privacy. Most people focus only on the security aspects of the data protection and they do not take into consideration all the other factors for legitimate processing of personal data. Moreover, privacy is very commonly linked to the anonymity on the web instead of embracing the notions and principles of proper acquisition and retention, integrity, aggregation and derivation of data, information sharing and proper access (Jajodia, 1996). A widely accepted definition of privacy is "the claim of individuals, groups,

or institutions to determine for themselves when, how, and to what extent information about them is communicated to others (Westin, 1967; Walker, 2000).'' According to this definition, the work described hereby helps preserving and protecting privacy in databases.

After analysing the codes of practice published by the Office of the Information Commissioner, who is responsible for enforcing and supervising the DPA (1998), and the ''Guide to managing your databases'' of the British Standards Institute, a complete list of database operability requirements was created. Following, an implementation model was presented. The main concept of the architecture is first, to extend the underlying database schema of the database that keeps the personal data with appropriate meta-data. The second step was to extend the DBMS with built-in components that ensure compliance with the Law. The advantage of this approach is that the data protection features do not need to be added to the database application after the design and creation of the database, but they exist within the DBMS along with many other operational capabilities.

A limitation of the current work is that the database model is applied to the centralised database of an organisation. This centralised database derives from the integration of all the databases and data sources that exist within the organisation. In practice, few organisations have implemented an integrated view of the data they hold and no robust technical solutions have been developed to this end. The requirements of the databases scattered across the organisation may be different from those described in this thesis. For example, if the users retrieve personal data from remote databases and not from the central one, then the remote databases should also have full audit trail facilities and the capability to share information about a user's activities on someone's personal data, with other databases containing such data. Also, there may be changes to the implementation model in order this to be adapted to a distributed database system.

Appendix A

A.1. Schema requirements

The DBMS should have the:

- capability to keep details about;
  - all the data subject identifiers used within the organisation,
  - the data subjects' consents for processing their personal data for transferring their personal data outside the EEA and for publishing their personal data on the Web,
  - the process purposes for each data subjects' data,
  - the data subjects' marketing preference,
  - all back-up, flat electronic and manual files that contain data subjects' data,
  - the possible recipients of the personal data,
  - the persons authorised to manipulate the personal data,
  - the sources of the personal data,
  - the retention time of a piece of personal data,
  - the automated decision mechanisms that are used to process a data subject's personal data,
  - the data subject's access requests,

- capability to distinguish personal data between sensitive and non-sensitive;
- capability to distinguish which data is processed for research purposes;
- capability to use extra information fields and free text fields for justified purposes;

A.2. Software components requirements

The DBMS should have the capability:

- to identify uniquely every data subject;
- to locate all the personal data held in an organisation's databases concerning a single individual;
- to ensure that the person who makes a request to access his personal data is in fact the relevant data subject, through the deployment of authentication procedures;
- for audit trail and logging;
- to audit the processes of personal data in order to ensure that the notification adequately covers the processing activities;
- to ensure that misuse of personal data within the organisation can be identified and remedied, through the deployment of mechanisms;
- to restrict the processing of personal data for purposes unconnected with the organisation activities using the organisation resources;
- to permanently amend the data;
- to permanently delete data;
- to automatically delete the personal data, when its retention time expires;
- to ensure proper access and implement the authorisation policy, through the deployment of access control mechanisms;
- to back-up data and restore them in a possible disaster recovery procedure;
- to provide a secure method of transmission, when personal data are collected or processed on-line;
- to print out all the data concerning only a specific data subject providing a translation of any codes used;
- to use suppression files for marketing lists;
- to block or suppress the processing of data subjects' personal data when the data subjects object to the processing;
- to provide information to those who are dealing with personal data about the purposes for which this data has been collected.

A.3. Requirements related to design decision and research issues

The DBMS should have the capability to:

- ensure that no personal data of another data subject are disclosed, when a data subject requests to access his personal records;
- keep details about what data identify the data subject in which context;
- delete the information and the links that enable individuals to be identified;

- ensure that the database does not allow unlawful disclosure of personal data. When someone requests personal data about another person, such data should be released only, and to the extend that, he requires the data in order to perform his official duties;
- safeguard data from external electronic attacks, through the deployment of security techniques;
- safeguard data from internal electronic attacks, through the deployment of security techniques;
- hold scanned paper files and other computer files apart from databases that contain personal data (e.g. multimedia databases), through the deployment of appropriate mechanisms;
- merge databases and integrate information sources using complete matches of all criteria;
- treat items of information individually or in logical groups;
- change the organisation's hardware and software systems avoiding loss in data accuracy and integrity;
- map directly the process diagrams of an organisation to the class diagrams of the organisational database;
- support bi-temporal semantics;
- manipulate of back-up files.

References

BCS (1998). Data Protection—everybody's business: A practical guide for professionals and Business Managers. The British Computer Society, ISBN 1-902505-04-2.

Brodsky, A., Farkas, C., & Jajodia, S. (2000). Secure Databases: Constraints, inference channels, and monitoring disclosures. *IEEE Transactions on Knowledge and Data Engineering*, *12*(6), 900–919.

BSI (1999). Guide to the Practical Implementation of the Data Protection Act 1998. DISC PD 0012:1999. British Standards Institute, ISBN-0-580-33029-X.

BSI (2000). Guide to Managing your Database, DISC PD 0012-4:2000, British Standards Institute, ISBN 0-580-33254-4.

BSI (2001). Data Protection—pre-audit workbook by A. Shipman, DISC PD 0012-5:2001, British Standards Institute, ISBN 0-580-33264-0.

Chor, B., Kushilevitz, E., Goldreich, O., & Sudan, M. (1998). Private information retrieval. *Journal of the ACM*, *45*(6), 965–981.

Cranor, L. F. (1999). Agents of choice: Tools that facilitate notice and choice about web site data practices. *Proceedings of 21st International Conference on Privacy and Personal Data Protection*, Hong Kong, September 13–14, 1999.

Denning, D. (1999). *Information warfare and security.* Reading, MA: Addison-Wesley, ISBN: 0201433036.

DPA (1998). Data Protection Act 1998, http://www.hmso.gov.uk/acts/acts1998/19980029.htm#aofs.

EU (1995). Directive 95/46/EC of the European Parliament. http://europa.eu.int/.

Gounaris, A., & Theodoulidis, B. (2001). Building a DBMS for Data Protection, Centre for Research in Information Management. UMIST. TR-01-1 June 2001.

IC (1998). The Data Protection Act 1998: An introduction. The Data Protection Registrar, ISBN 1-870-466-21-7.

IC (2001). Information Commissioner Annual report and accounts for the year ending 31 March 2001. Data Protection Commissioner Office. http://www.dataprotection.gov.uk/dpr/dpdoc.nsf.

Jajodia, S. (1996). Managing Security and Privacy of Information. Computing Surveys 28(4es): article 79.

McKilligan, N. (2000): Data Protection and Customer Privacy: Getting your Act Together. http://www.ecominfo.net/arts/090_bsi-disc.htm.

Pounder, C., & McLean, K. (1998). IDA Programme: A Guide to Data Protection Compliance. http://europa.eu.int/comm/internal_market/en/dataprot/studies/idaintro.htm.

Reden, A. (1999). Data Protection Activities in the Private Sector. *Proceedings of the 21st International Conference on Privacy and Personal Data Protection.*

Walker, K. (2000). Where everybody knows your name: A pragmatic look at the costs of privacy and the benefits of information exchange. Stanford Technology and Law Review.

Westin, A. (1967). *Privacy and freedom*. New York: Atheneum.