Toward Standardization in Privacy-Preserving Data Mining

Stanley R. M. Oliveira[1,2] and Osmar R. Zaïane[1]
{oliveira | zaiane}@cs.ualberta.ca

[1]Department of Computing Science
University of Alberta, Edmonton, Canada, T6G 2E8

[2]Embrapa Informática Agropecuária
Av. André Tosello, 209 13083-886 - Campinas, SP, Brasil

Abstract. Issues about privacy-preserving data mining (PPDM) have emerged globally. The recent proliferation in PPDM techniques is evident. Motivated by the increasing number of successful techniques, the new generation in PPDM moves on toward standardization because it will certainly play an important role in the future of PPDM. In this paper, we lay out what needs to be done and take some steps toward proposing such standardization: First, we describe the problems we face in defining what information is private in data mining, and discuss how privacy can be violated in data mining. Then, we define privacy preservation in data mining based on users' personal information and information concerning their collective activity. Second, we analyze the implications of the Organization for Economic Cooperation and Development (OECD) data privacy principles in the context of data mining and suggest some policies for PPDM based on such principles. Finally, we propose some requirements to guide the development and deployment of technical solutions.

1. Introduction

The debate on PPDM has received special attention as data mining has been widely adopted by public and private organizations. We have witnessed three major landmarks that characterize the progress and success of this new research area: *the conceptive landmark*, *the deployment landmark*, and *the prospective landmark*. We describe these landmarks as follows:

- *The Conceptive landmark* characterizes the period in which central figures in the community, such as O'Leary [14, 15], Fayyad, Piatetsky-Shapiro and Smith [8, 16], and others [12, 5], investigated the success of knowledge discovery and some of the important areas where it can conflict with privacy concerns. The key finding was that knowledge discovery can open new threats to informational privacy and information security if not done or used properly. Since then, the debate on PPDM has gained momentum.
- *The Deployment landmark* is the current period in which an increasing number of PPDM techniques have been developed and have been published in refereed conferences. The information available today is spread over countless papers and conference proceedings[1]. The results achieved in the last years are promising and suggest that PPDM will achieve the goals that have been set for it.
- *The Prospective landmark* is a new period in which directed efforts toward standardization occur. At this stage, there is no consent about what privacy preservation means in data mining. In addition, there is no consensus on privacy principles, policies, and requirements as a foundation for the development and deployment of new PPDM techniques. The excessive number of techniques is leading to confusion among developers, practitioners, and others interested in this technology. One of the most important challenges in PPDM now is to establish the groundwork for further research and development in this area.

[1] The Privacy-Preserving Data Mining: http://www.cs.ualberta.ca/~oliveira/psdm/psdm index.html

Currently, one of the most important challenges in PPDM is to put forward standardization issues in PPDM because they will play a significant role in the future of this new area. In this paper, we lay out what needs to be done and take some steps toward proposing such standardization. Our contributions in this paper can be summarized as follows: a) we describe the problems we face in defining what information is private in data mining, and discuss how privacy can be violated in data mining; b) we define privacy preservation in data mining based on users' personal information and information concerning their collective activity; c) we describe the general parameters for characterizing scenarios in PPDM; d) we analyze the implications of the Organization for Economic Cooperation and Development (OECD) data privacy principles in knowledge discovery; e) we suggest some policies for PPDM based on instruments accepted world-wide; and f) we propose some requirements for the development of technical solutions and to guide the deployment of new technical solutions.

The effort described in this paper is by no means meant to be complete and comprehensive. Rather, our primary goal is to stir up the discussion on consensus about definition, requirements, principles and policies in PPDM. We argue that this line of work will eventually lead to standardization in PPDM.

This paper is organized as follows. In Section 2, we describe the problems we face in defining privacy for data mining. In Section 3, we describe some issues related to PPDM, such as privacy violation, and privacy definitions. In Section 4, we analyze the OECD principles in the context of data mining. We also suggest some policies for PPDM based on instruments accepted world-wide. In Section 5, we propose some privacy requirements for the development and deployment of technical solutions. Related work is reviewed in Section 6. Finally, Section 7 presents our conclusions.

2. Problems in Defining Privacy

Analyzing what right to privacy means is a fraut with problems, such as the exact definition of privacy, whether it constitutes a fundamental right, and whether people are and/or should be concerned with it. Several definitions of privacy have been given, and they vary according to context, culture, and environment. For instance, in an 1890 paper [22], Warren & Brandeis defined privacy as “the right to be alone.” Later, in a paper published in 1967 [23], Westin defined privacy as “the desire of people to choose freely under what circumstances and to what extent they will expose themselves, their attitude, and their behavior to others”. Schoeman [20] defined privacy as “the right to determine what (personal) information is communicated to others” or “the control an individual has over information about himself or herself.” More recently, Garfinkel [9] stated that “privacy is about self-possession, autonomy, and integrity.” On the other hand, Rosenberg argues that privacy may not be a right after all but a taste [18]: “If privacy is in the end a matter of individual taste, then seeking a moral foundation for it — beyond its role in making social institutions possible that we happen to prize — will be no more fruitful than seeking a moral foundation for the taste for truffles."

The above definitions suggest that, in general, privacy is viewed as a social and cultural concept. However, with the ubiquity of computers and the emergence of the Web, privacy has also become a digital problem [17]. With the Web revolution and the emergence of data mining, privacy concerns have posed technical challenges fundamentally different from those that occurred before the information era. In the information technology era, privacy refers to the right of users to conceal their personal information and have some degree of control over the use of any personal information disclosed to others [6, 1, 10].

Clearly, the concept of privacy is often more complex than realized. In particular, in data mining, the definition of privacy preservation is still unclear, and there is very little literature related to this topic. A notable exception is the work presented in [3], in which PPDM is defined as “getting valid data mining results without learning the underlying data values.” However, at this point, each existing PPDM technique has its own privacy definition. Our primary concern about PPDM is that mining

algorithms are analyzed for the side effects they incur in data privacy. Therefore, our definition for PPDM is close to those definitions in [20, 3] — ***PPDM encompasses the dual goal of meeting privacy requirements and providing valid data mining results***. Our definition emphasizes the dilemma of balancing privacy preservation and knowledge disclosure.

3. Privacy-Preserving Data Mining

3.1 Privacy Violation in Data Mining

Understanding privacy in data mining requires understanding how privacy can be violated and the possible means for preventing privacy violation. In general, one major factor contributes to privacy violation in data mining: *data misuse*.

Users' privacy can be violated in different ways and with different intentions. Although data mining can be extremely valuable in many applications (e.g., business, medical analysis, etc), it can also, in the absence of adequate safeguards, violate informational privacy. Privacy can be violated if personal data are used for other purposes subsequent to the original transaction between an individual and an organization when the information was collected.

One of the sources of privacy violation is called data magnets [17]. Data magnets are techniques and tools used to collect personal data. Examples of data magnets include explicitly collecting information through on-line registration, identifying users through IP addresses, software downloads that require registration, and indirectly collecting information for secondary usage. In many cases, users may or may not be aware that information is being collected or do not know how that information is collected [7, 13]. Worse is the privacy invasion occasioned by secondary usage of data when individuals are unaware of "behind the scenes" uses of data mining techniques [11]. In particular, personal data can be used for secondary usage largely beyond the users' control and privacy laws. This scenario has led to an uncontrollable privacy violation not because of data mining itself, but fundamentally because of the misuse of data.

3.2 Defining Privacy Preservation in Data Mining

In general, privacy preservation occurs in two major dimensions: users' personal information and information concerning their collective activity. We refer to the former as individual privacy preservation and the latter as collective privacy preservation, which is related to corporate privacy in [3].

- **Individual privacy preservation**: The primary goal of data privacy is the protection of personally identifiable information. In general, information is considered personally identifiable if it can be linked, directly or indirectly, to an individual person. Thus, when personal data are subjected to mining, the attribute values associated with individuals are private and must be protected from disclosure. Miners are then able to learn from global models rather than from the characteristics of a particular individual.
- **Collective privacy preservation**: Protecting personal data may not be enough. Sometimes, we may need to protect against learning sensitive knowledge representing the activities of a group. We refer to the protection of sensitive knowledge as collective privacy preservation. The goal here is quite similar to that one for statistical databases, in which security control mechanisms provide aggregate information about groups (population) and, at the same time, should prevent disclosure of confidential information about individuals. However, unlike as is the case for statistical databases, another objective of collective privacy preservation is to preserve strategic patterns that

are paramount for strategic decisions, rather than minimizing the distortion of all statistics (e.g., bias and precision). In other words, the goal here is not only to protect personally identifiable information but also some patterns and trends that are not supposed to be discovered.

In the case of collective privacy preservation, organizations have to cope with some interesting conflicts. For instance, when personal information undergoes analysis processes that produce new facts about users' shopping patterns, hobbies, or preferences, these facts could be used in recommender systems to predict or affect their future shopping patterns. In general, this scenario is beneficial to both users and organizations. However, when organizations share data in a collaborative project, the goal is not only to protect personally identifiable information but also some strategic patterns. In the business world, such patterns are described as the knowledge that can provide competitive advantages, and therefore must be protected [21]. More challenging is to protect the knowledge discovered from confidential information (e.g., medical, financial, and crime information). The absence of privacy safeguards can equally compromise individuals' privacy. While violation of individual privacy is clear, violation of collective privacy can lead to violation of individual's privacy.

3.3 Characterizing Scenarios in PPDM

Before describing the general parameters for characterizing scenarios in PPDM, let us consider two real-life examples where PPDM poses different constraints:

- **Scenario 1**: A hospital shares some data for research purposes (e.g., concerning a group of patients who have a similar disease). The hospital's security administrator may suppress some identifiers (e.g., name, address, phone number, etc) from patient records to meet privacy requirements. However, the released data may not be fully protected. A patient record may contain other information that can be linked with other datasets to re-identify individuals or entities [19]. How can we identify groups of patients with a similar disease without revealing the values of the attributes associated with them?

- **Scenario 2**: Two or more companies have a very large dataset of records on their customers' buying activities. These companies decide to cooperatively conduct association rule mining on their datasets for their mutual benefit since this collaboration brings them an advantage over other competitors. However, some of these companies may not want to share some strategic patterns hidden within their own data (also called restrictive association rules) with the other parties. They would like to transform their data in such a way that these restrictive association rules cannot be discovered but others can be. Is it possible for these companies to benefit from such collaboration by sharing their data while preserving some restrictive association rules?

Note that the above scenarios describe different privacy preservation problems. Each scenario poses a set of challenges. For instance, scenario 1 is a typical example of individual's privacy preservation, while scenario 2 refers to collective privacy preservation. How can we characterize scenarios in PPDM? One alternative is to describe them in terms of general parameters. In [4], some parameters are suggested:

- **Outcome:** Refers to the desired data mining results. For instance, someone may look for association rules identifying relationships among attributes, or relationships among customers' buying behaviors as in scenario 2, or may even want to cluster data as in scenario 1.

- **Data Distribution**: How are the data available for mining: are they centralized or distributed across many sites? In the case of data distributed throughout many sites, are the entities described with the same schema in all sites (horizontal partitions), or do different sites contain different attributes for one entity (vertical partitions)?

- **Privacy Preservation**: What are the privacy preservation requirements? If the concern is solely that values associated with an individual entity not be released (e.g., personal information), techniques must focus on protecting such information. In other cases, the notion of what constitutes "sensitive knowledge" may not be known in advance. This would lead to human evaluation of the intermediate results before making the data available for mining.

4. Principles and Policies for PPDM

4.1 The OECD Privacy Guidelines

World-wide, privacy legislation, policies, guidelines, and codes of conduct have been derived from the set of principles established in 1980 by the OECD[2]. They represent the primary components for the protection of privacy and personal data, comprising a commonly understood reference point. A number of countries have adopted these principles as statutory law, in whole or in part. The OECD Privacy Guidelines outline the following basic principles:

1. **Collection Limitation Principle**: There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject (consumer).
2. **Data Quality Principle**: Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and up-to-date.
3. **Purpose Specification Principle**: The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.
4. **Use Limitation Principle**: Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with [the Purpose Specification Principle] except: (a) with the consent of the data subject; or (b) by the authority of law.
5. **Security Safeguards Principle**: Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification, or disclosure of data.
6. **Openness Principle**: There should be a general policy of openness about developments, practices, and policies with respect to personal data. Means should be readily available for establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller (e.g., a public or a private organization).
7. **Individual Participation Principle**: An individual should have the right: a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him; b) to have communicated to him, data relating to him within a reasonable time; at a charge, if any, that is not excessive; in a reasonable manner; and in a form that is readily intelligible to him; c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and d) to challenge data relating to him and, if the challenge is successful to have the data erased, rectified, completed, or amended.
8. **Accountability Principle**: A data controller should be accountable for complying with measures which give effect to the principles stated above.

---

[2] Privacy Online - OECD Guidance on Policy and Practice. http://www.oecd.org/dataoecd/33/43/2096272.pdf

4.2 The implications of the OECD Privacy Guidelines in PPDM

We now analyze the implications of the OECD principles in PPDM. Then we suggest which principles should be considered absolute principles in PPDM.

1. **Collection Limitation Principle**: This principle states that some very sensitive data should not be held at all. Collection limitation is too general in the data mining context incurring in two grave consequences: a) the notion of "very sensitive" is sometimes unclear and may differ from country to country, leading to vague definitions; b) limiting the collection of data may make the data useless for knowledge discovery. Thus, this principle seems to be unenforceable in PPDM.

2. **Data Quality Principle**: This principle is related to the pre-processing stage in data mining in which data cleaning routines are applied to resolve inaccuracy and inconsistencies. Somehow, this principle is relevant in the preprocessing stage of knowledge discovery. However, most PPDM techniques assume that the data are already in an appropriate form to mine.

3. **Purpose Specification Principle**: This principle is the fundamental basis of privacy. Individuals should be informed of the purposes for which the information collected about them will be used, and the information must be used solely for that purpose. In other words, restraint should be exercised when personal data are collected. This principle is extremely relevant in PPDM.

4. **Use Limitation Principle**: This principle is closely related to the purpose specification principle. Use limitation is perhaps the most difficult principle to address in PPDM. This principle states that the purpose specified to the data subject (consumer) at the time of the collection restricts the use of the information collected, unless the data subject has provided consent for additional uses. This principle is also fundamental in PPDM.

5. **Security Safeguards Principle**: This principle is basically irrelevant in the case of data privacy, but relevant for database security. Security safeguards principle is typically concerned with keeping sensitive information (e.g., personal data) out of the hands of unauthorized users, which ensures that the data is not modified by users who do not have permission to do so. This principle is unenforceable in the context of PPDM.

6. **Openness Principle**: This principle, also called transparency, states that people have the right to know what data about them have been collected, who has access to the data, and how the data are being used. In other words, people must be aware of the conditions under which their information is being kept and used. However, data mining is not an open and transparent activity requiring analysts to inform individuals about particular derived knowledge, which may inhibit the use of data. This principle is equally important in PPDM.

7. **Individual Participation Principle**: This principle suggests that data subjects should be able to challenge the existence of information gained through data mining applications. Since knowledge discovery is not openly apparent to data subjects, the data subjects are not aware of knowledge discoveries related to them. While debatably collected individual information could belong to individuals, one can argue that collective information mined from databases belongs to organizations that hold such databases. In this case, the implications of this principle for PPDM should be carefully weighed; otherwise, it could be too rigid in PPDM applications.

8. **Accountability Principle**: This principle states that data controllers should inform data subjects of the use and findings from knowledge discovery. In addition, data controllers should inform individuals about the policies regarding knowledge discovery activities, including the consequences of inappropriate use. Some countries (e.g., the UK, Japan, Canada) that have adopted the OECD privacy principles do not consider this principle since it is not limited in scope, area, or application. Thus, the accountability principle is too general for PPDM.

Our analysis above suggests that the OECD privacy principles can be categorized into three groups according to their influence on the context of PPDM:

- **Group 1** is composed of those principles that should be considered as absolute principles in PPDM, such as Purpose Specification, Use Limitation, and Openness.
- **Group 2** consists of some principles that somehow impact PPDM applications, and their full implications should be understood and carefully weighed depending on the context. The principles that fall into this category are Data Quality and Individual Participation.
- **Group 3** encompasses some principles that are too general or unenforceable in PPDM. This group includes Collection Limitation, Security Safeguards, and Accountability. Clearly, the principles categorized in groups 1 and 2 are relevant in the context of PPDM and are fundamental for further research, development, and deployment of PPDM techniques.

4.3 Adopting PPDM Policies from the OECD Privacy Guidelines

One fundamental point to be considered when designing some privacy policies is that too many restrictions could seriously hinder the normal functioning of business and governmental organizations. The worst thing is that restrictions, if not carefully weighed, could make PPDM results useless.

Given these facts, we suggest some policies for PPDM based on the OECD privacy principles. We try to find a good compromise between privacy requirements and knowledge discovery. We describe the policies as follows:

1. **Awareness Policy**: When a data controller collects personally identifiable information, the data controller shall express why the data are collected and whether such data will be used for knowledge discovery.
2. **Limit Retention Policy**: A data controller shall take all reasonable steps to keep only personal information collected that is accurate, complete, and up to date. In the case of personal information that is no longer useful, it shall be removed and not subjected to analysis to avoid unnecessary risks, such as wrong decision making which may incur liability.
3. **Forthcoming Policy**: Policies regarding collecting, processing, and analyzing that produce new knowledge about individuals shall be communicated to those about whom the knowledge discovered pertains, in particular when the discovered knowledge is to be disclosed or shared.
4. **Disclosure Policy**: Data controllers shall only disclose discovered knowledge about an individual for purposes for which the individual consents and the knowledge discovered about individuals shall never be disclosed inadvertently or without consent.

5. Requirements for PPDM

5.1 Requirements for the development of technical solutions

Ideally, a technical solution for a PPDM scenario would enable us to enforce privacy safeguards and to control the sharing and use of personal data. However, such a solution raises some crucial questions:

— What levels of effectiveness are in fact technologically possible and what corresponding regulatory measures are needed to achieve these levels?

— What degrees of privacy and anonymity must be sacrificed to achieve valid data mining results?

These questions cannot have "yes-no" answers, but involve a range of technological possibilities and social choices. The worst response to such questions is to ignore them completely and not pursue the means by which we can eventually provide informed answers.

Technology alone cannot address all of the concerns surrounding PPDM scenarios [2]. The above questions can be to some extent addressed if we provide some key requirements to guide the development of technical solutions.

The following key words are used to specify the extent to which an item is a requirement for the development of technical solutions to address PPDM:

- **Must**: this word means that the item is an absolute requirement;
- **Should**: this word means that there may exist valid reasons not to treat this item as a requirement, but the full implications should be understood and the case carefully weighed before discarding this item.

1. **Independence**: A promising solution for the problem of PPDM, for any specific data mining task (e.g., association rules, clustering, classification), should be independent of the mining task algorithm.
2. **Accuracy**: When it is possible, an effective solution should do better than a trade-off between privacy and accuracy on the disclosure of data mining results. Sometimes a trade-off must be found as in scenario 2 in Section 3.3.
3. **Privacy Level**: This is also a fundamental requirement in PPDM. A technical solution must ensure that the mining process does not violate privacy up to a certain degree of security.
4. **Attribute Heterogeneity**: A technical solution for PPDM should handle heterogeneous attributes (e.g., categorical and numerical).
5. **Versatility**: A versatile solution to address the problem of PPDM should be applicable to different kinds of information repositories, i.e., the data could be centralized, or even distributed horizontally or vertically.
6. **Communication Cost**: When addressing data distributed across many sites, a technical solution should consider carefully issues of communication cost.

5.2 Requirements to guide the deployment of technical solutions

Information technology vendors in the near future will offer a variety of products which claim to help protect privacy in data mining. How can we evaluate and decide whether what is being offered is useful? The nonexistence of proper instruments to evaluate the usefulness and feasibility of a solution to address a PPDM scenario challenge us to identify the following requirements:

1. **Privacy Identification**: We should identify what information is private. Is the technical solution aiming at protecting individual privacy or collective privacy?
2. **Privacy Standards**: Does the technical solution comply with international instruments that state and enforce rules (e.g., principles and/or policies) for use of automated processing of private information?
3. **Privacy Safeguards**: Is it possible to record what has been done with private information and be transparent with individuals about whom the private information pertains?
4. **Disclosure Limitation**: Are there metrics to measure how much private information is disclosed? Since privacy has many meanings depending on the context, we may require a set of metrics to do so. What is most important is that we need to measure not only how much private information is

disclosed, but also measure the impact of a technical solution on the data and on valid mining results.

5. **Update Match**: When a new technical solution is launched, two aspects should be considered: a) the solution should comply with existing privacy principles and policies; b) in case of modifications to privacy principles and/or policies that guide the development of technical solutions, any release should consider these new modifications.

6. Related Work

Data mining from a fair information practices perspective was first discussed in [15]. O'Leary studied the impact of the OECD guidelines in knowledge discovery. The key finding of this study was that the OCDE guidelines could not anticipate or address many important issues regarding knowledge discovery, and thus several principles are too general or unenforceable. Our work here is orthogonal to that one in [15]. We investigate the influence of the OECD principles in the context of PPDM categorizing them in different groups of relevance. In particular, we show that the OECD guidelines are accepted world-wide and therefore they represent the primary components for standardization in PPDM. We discuss how the community in PPDM could derive some principles and policies from the OECD guidelines.

More recently, Clifton et al. discussed the meaning of PPDM as a foundation for further research in this field [3]. That work introduces some definitions for PPDM and discusses some metrics for information disclosure in data mining. The work in [3] is complementary to our work. The primary goal of our work is to put forward standardization issues in PPDM. Our effort encompasses the design of privacy principles and policies, and requirements for the development and deployment of technical solutions for PPDM.

7. Conclusions

In this paper, we make some effort to establish the groundwork for further research in the area of Privacy-Preserving Data Mining (PPDM). We put forward standardization issues in PPDM. Although our work described in this paper is preliminary and conceptual in nature, we argue that it is a vital prerequisite for standardization in PPDM.

Our primary goal in this work is to conceive a common framework for PPDM, notably in terms of definitions, principles, policies, and requirements. The advantages of a framework of that nature are: (a) a common framework will avoid confusing developers, practitioners, and many others interested in PPDM; (b) adoption of a common framework will inhibit inconsistent efforts in different ways, and will enable vendors and developers to make solid advances in the future in the PPDM area.

Our contributions in this paper can be summarized as follows: 1) we describe the problems we face in defining what information is private in data mining, and discuss how privacy can be violated in data mining; 2) we define privacy preservation in data mining based on users' personal information and information concerning their collective activity; 3) we describe the general parameters for characterizing scenarios in PPDM; 4) we analyze the implications of the Organization for Economic Cooperation and Development (OECD) data privacy principles in knowledge discovery; 5) we suggest some policies for PPDM based on instruments accepted world-wide; and 6) we propose some requirements for the development of technical solutions and to guide the deployment of new technical solutions.

Acknowledgments

Stanley Oliveira was partially supported by CNPq, Brazil, under grant No. 200077/00-7. Osmar Zaïane was partially supported by a research grant from NSERC, Canada.

References

1. M. Ackerman, L. Cranor, and J. Reagle. Privacy in E-Commerce: Examining User Scenarios and Privacy Preferences. In *Proc. of the ACM Conference on Electronic Commerce*, pages 1-8, Denver, Colorado, USA, November 1999.
2. R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Hippocratic Databases. In *Proc. of the 28th Conference on Very Large Data Bases*, Hong Kong, China, August 2002.
3. C. Clifton, M. Kantarcioglu, and J. Vaidya. Defining Privacy For Data Mining. In *Proc. of the National Science Foundation Workshop on Next Generation Data Mining*, pages 126-133, Baltimore, MD, USA, November 2002.
4. C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu. Tools For Privacy Preserving Distributed Data Mining. *SIGKDD Explorations*, 4(2):28-34, 2002.
5. C. Clifton and D. Marks. Security and Privacy Implications of Data Mining. In *Workshop on Data Mining and Knowledge Discovery*, pages 15-19, 1996.
6. S. Cockcroft and P. Clutterbuck. Attitudes Towards Information Privacy. In *Proc. of the 12th Australasian Conference on Information Systems*, Australia, 2001.
7. M. J. Culnan. How Did They Get My Name?: An Exploratory Investigation of Consumer Attitudes Toward Secondary Information. *MIS Quartely*, 17(3):341-363, September 1993.
8. U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From Data Mining to Knowledge Discovery: An Overview. In *Advances in Knowledge Discovery and Data Mining*. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smith, and R. Uthurusamy (eds.), pages 1-34, MIT Press, Cambridge, MA, 1996.
9. S. Garfinkel. *Database Nation: The Death of the Privacy in the 21st Century*. O'Reilly & Associates, Sebastopol, CA, USA, 2001.
10. P. Jefferies. Multimedia, Cyberspace & Ethics. In *Proc. of International Conference on Information Visualisation (IV2000)*, pages 99-104, London, England, July 2000.
11. G. H. John. Behind-the-Scenes Data Mining. *Newletter of ACM SIG on KDDM*, 1(1):9-11, June 1999.
12. W. Klösgen. KDD: Public and Private Concerns. *IEEE EXPERT*, 10(2):55-57, April 1995.
13. K. C. Laudon. Markets and Privacy. *Communication of the ACM*, 39(9):92-104, September 1996.
14. D. E. O'Leary. Knowledge Discovery as a Threat to Database Security. In G. Piatetsky-Shapiro and W. J. Frawley (editors): *Knowledge Discovery in Databases*. AAAI/MIT Press, pages 507-516, Menlo Park, CA, 1991.
15. D. E. O'Leary. Some Privacy Issues in Knowledge Discovery: The OECD Personal Privacy Guidelines. *IEEE EXPERT*, 10(2):48-52, April 1995.
16. G. Piatetsky-Shapiro. Knowledge Discovery in Personal Data vs. Privacy: A Mini-Symposium. *IEEE Expert*, 10(2):46-47, 1995.
17. A. Rezgur, A. Bouguettaya, and M. Y. Eltoweissy. Privacy on the Web: Facts, Challenges, and Solutions. *IEEE Security & Privacy*, 1(6):40-49, Nov-Dec 2003.
18. A. Rosenberg. Privacy as a Matter of Taste and Right. In E. F. Paul, F. D. Miller, and J. Paul, editors, *The Right to Privacy*, pages 68-90, Cambridge University Press, 2000.
19. P. Samarati. Protecting Respondents' Identities in Microdata Release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010-1027, 2001.
20. F. D. Schoeman. *Philosophical Dimensions of Privacy*, Cambridge Univ. Press, 1984.
21. E. Turban and J. E. Aronson. *Decision Support Systems and Intelligent Systems*. Prentice-Hall, New Jersey, USA, 2001.
22. S. D. Warren and L. D. Brandeis. The Right to Privacy. *Harvard Law Review*, 4(5):193-220, 1890.
23. A. F. Westin. *The Right to Privacy*, Atheneum, 1967.