A Study of Data Mining and Information Ethics in the Development of
Information Systems Professionals

James Lawler
jlawler@pace.edu
Seidenberg School of CSIS, Pace University
Information Systems
New York, NY 10038

John C. Molluzzo
jmolluzzo@pace.edu
Seidenberg School of CSIS, Pace University
Information Systems
New York, NY 10038

Abstract

Data and the information culled from that data is an extremely valuable organizational re-
source. Every data mining professional is aware of this, but few are well-educated on the im-
pact that data mining could have on privacy and the laws surrounding the privacy of personal
data. In a recent survey, van Wel and Royakkers (2004) showed that 20 data-mining profes-
sionals "prefer to focus on the advantages of web-data mining instead of discussing the possi-
ble dangers." These professionals argued that web-data mining does not threaten privacy.
Part of the reason some professionals are not concerned over the possible misuse of their
work and the possible harm it might cause might lie in the content of the data mining courses
they have taken and in the textbooks they used to learn their craft. This paper presents a
project that investigates the need for an expanded role of ethics in data mining education for
future information systems professionals. Our contention is that ethics and the social impact
of data mining must be an integral part of every data mining course. The methodology will
consist of a survey of data mining textbooks to determine the actual extent of coverage of
ethical issues (the focus of this paper), and in the future, a survey of schools on the coverage
of ethical issues in their data mining courses, development of ethical cases appropriate for use
in a data mining course, and the implementation of a data mining course that integrates ethi-
cal issues into its curriculum.

1. INTRODUCTION

Many companies today, like Wal-Mart, store
much of their business and customer data in
large databases called data warehouses.
Their customers are not told the extent of
the information that is accumulated on
them, how long it will be kept, nor the uses
to which the data will be put. (Hays, 2004)
This data is subsequently analyzed to pro-
duce new information to help the companies
evaluate business processes and customer
behavior. The technique usually used to do
the analysis is data mining. *Data mining* is a
"set of automated techniques used to extract
buried or previously unknown pieces of in-
formation from large databases." (Cavou-
kian, 1998) Thus, data mining seeks to find
new patterns and associations in databases.

Much of the data that is mined is public or
semi-public in nature - what we purchase at
the supermarket, where we surf the Web,

where we work, our salary. Fulda (1997) questioned whether it is possible that such data deserves legal or normative protection from those who would use this data for their own ends. Subsequently, Fulda (1999) answered his own question in the affirmative. He gave cases of how public data about an individual can be combined to produce new private knowledge about the individual.

The key ethical issues in mining personal data are that people are generally (1) not aware that their personal information is being gathered, (2) do not know to what use the data will be made, or (3) have not consented to such collecting or use. This data can be used to construct profiles and customer categories that can be used to target advertising. For example, Spangler et al (2003), describe how Personal Video Recorders can yield viewing data that can be mined to deliver through your television targeted custom advertising. In addition to data privacy issues, data mining raises other social concerns. For example, Danna and Gandy (2002) argue that data mining and the use of consumer profiles can actually exclude groups of customers from full participation in the marketplace and limit their access to information.

Thus, there are major ethical and social issues that arise from the practice of data mining. However, Nissenbaum (1998) observes that those in favor of using public information about individuals argue that as long as a person does not make an effort to conceal information about him or herself, then any restriction on what one does with that information is a limit on freedom. A recent study by van Wel and Royakkers (2004) seems to confirm Nissenbaum's observation. In a survey of twenty web-data mining professionals, they showed that the professionals "prefer to focus on the advantages of web-data mining instead of discussing the possible dangers." These professionals argued that web-data mining does not threaten privacy. Even if these professionals looked to the law and codes of ethics for direction, they might not receive it. Tavani (1999) points out that existing privacy laws and guidelines, such as those from the Organization for Economic Cooperation and Development (OECD) and the European Union Directive 95/46/EC of the European Council, do not adequately address privacy issues surrounding the use of public data in data mining.

One might wonder why professionals are not aware of or concerned over the possible misuse of their work, and the possible harm it might cause to individuals and society. Part of the reason might lie in the content of the data mining courses they have taken and in the textbooks they used to learn their craft. The purpose of this paper is to survey recent data mining textbooks to measure the extent to which they discuss the ethical, legal, and social issues surrounding data mining.

2. BACKGROUND

Our study analyzes the content of contemporary data mining textbooks to determine the extent to which they introduce and discuss issues relating to privacy of consumer data, laws that govern the use of personal consumer data, and professional guidelines for the collection and use of consumer data. See Table 1 in Appendix B.

Privacy

Privacy is not easily defined perhaps because the notion of privacy has evolved over time, and now means different things in different situations and in different cultures. Most scholars define three types of privacy (Tavani, 2004). *Accessibility privacy* is freedom from intrusion. Historically, this is the first notion of privacy to be codified into law. The Fourth Amendment to the U.S. Constitution protects citizens from unreasonable searches and seizures. *Decisional privacy* is freedom from interference in your personal choices – think Roe v. Wade and the right to have an abortion. *Informational privacy* is a person's ability to restrict access to and to control the flow of his or her private information.

What are the ethical bases of these notions? Tavani (1999) divides privacy theories into four types: non-intrusion, seclusion, limitation, and control theories. One of the first attempts to define privacy was the notion of "the right to be left alone" (Warren and Brandies, 1890) and has been identified with the right to *non-intrusion* in one's personal

"space" and interference in one's personal affairs. In the *seclusion* theory (Westin, 1967), privacy is equated with being alone. These two theories are primarily concerned with "psychological privacy" (Regan, 1995).

The non-intrusion and seclusion theories do not adequately deal with privacy issues surrounding personal information and access to that information – so-called informational privacy. Three theories of privacy address this problem. The *control theory* of privacy (Fried, 1970), maintains that you have privacy if and only if you have control over information about yourself. The *limitation* theory of privacy (Allen, 1988) defines privacy as being able to limit access to your personal information in certain contexts. In Moor's *control/restricted access* theory of privacy (Moor, 1997), a person has privacy in a situation if the person is "protected from intrusion, interference, and information access by others." He distinguishes between "naturally" private situations (i.e. protected by natural, physical means such as walls or distance) and "normatively" private situations (i.e. protected by norms or laws.) Thus, privacy needs to be understood in terms of situations in which access to individuals or their information is restricted.

Vedder introduced the concept of categorical privacy, which emphasizes attributing generalized properties to group members. The concept is particularly important with respect to the effects of data mining. *Categorical privacy*

> relates to data or information to which two conditions do apply: (1) the information was originally taken from the personal sphere of individuals, and – after aggregation and processing according to statistical methods - is no longer accompanied by identifiers of individual natural persons, but, instead, by identifiers of groups of persons; (2) when attached to identifiers of groups and when disclosed, the information is apt to carry with it the same kind of negative consequences for the members of those groups as it would for an individual natural person if the information were accompanied by identifiers of that individual. (Vedder, 1999)

Laws

In early March 2005, hackers stole the personal information of 32,000 people from the databases of LexisNexis. The stolen data included Social Security numbers and financial information. Although the CEO of LexisNexis claimed that the information they collect is governed by the U.S. Fair Credit Reporting Act, members of Congress disagreed. Rep. Joe Barton (R-Texas) said that "Under current law, anyone has a near-perfect right to package your personal information and do almost anything they want with it….They can change it, share it, rent it or sell it. The constraints are so flimsy there're laughable." (Gross, 2005) As a result of this and other large-scale identity thefts in recent years, Congress is considering new laws to govern what personal data a company (in particular data brokers like LexisNexis) can collect and share. For example, Congress is considering a law to prohibit almost all sales of Social Security numbers.

What laws are there governing the use of personal data? Although there is no explicit right to privacy in the Constitution, legislation and court decisions on privacy are usually based on parts of the First, Fourth, Fifth, and Fourteenth Amendments. Most of the laws in the United States govern what the federal government can do with personal data. Except for health care and financial organizations, and data collected from children, there is no law that governs the collection and use of personal data by commercial enterprises. Therefore, it is essentially up to each organization to decide how they will use the personal data they have accumulated on their customers.

Privacy Guidelines

Although, as noted above, there are few laws in the United States governing the use of personal data, many of the existing laws and businesses have used the Code of Fair Information practices of the Organization for Economic Cooperation and Development to guide them in setting informational privacy policy. The code is based on eight principles: Collection Limitation, Data Quality, Purpose Specification, Use Limitation, Security Safeguards, Openness, Individual Participation, and Accountability. (OECD, 2005)

European Privacy Policy

Given the global nature of today's economy, the European Union (EU) has realized that laws governing privacy must apply on an international scale. In 1995, the EU developed an informational privacy policy, known as the European Directive 95/46/EC, which applies to all member states. Following is the statement of their Principles Relating to Data Quality.

> 1. Member States shall provide that personal data must be:
> (a) processed fairly and lawfully;
> (b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Further processing of data for historical, statistical or scientific purposes shall not be considered as incompatible provided that Member States provide appropriate safeguards;
> (c) adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed;
> (d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified;
> (e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Member States shall lay down appropriate safeguards for personal data stored for longer periods for historical, statistical or scientific use. (CDT, 2005)

EU member states cannot export personal data to non-EU nations that do not meet the EU standard for privacy protection. Since there are no comprehensive privacy laws in the United States, how then can American companies exchange data with companies in the EU? To meet this problem, the U.S. Department of Commerce developed the "safe harbor" framework, which was approved by the EU in July of 2000. Being certified by Safe Harbor assures EU organizations that a company provides adequate privacy protection, as defined by the Directive. (USDC, 2005)

Data Mining and Privacy

Personal privacy is threatened by three computing practices. (1) Data gathering - the collection of personal information, often without the subject's knowledge and consent. (2) Data exchange – transferring personal data between databases, often without the subject's knowledge and consent. (3) Data mining – searching large databases of mostly public data to generate profiles based on peoples' personal data and behavior patterns. (Tavani, 2004)

Today's privacy laws and guidelines protect data that is explicit, confidential, and exchanged between databases. However, there is no legal or normative protection for data that is implicit, non-confidential, and not exchanged (in a data warehouse, for example.) Information gathered in data mining is usually implicit in patterns in the data. These patterns suggest new associations about persons, which place them into new categories. Data mining can also reveal sensitive information that is derived from non-sensitive data and meta-data – the so-called "inference problem." There is much research on the inference problem and techniques to control it. See Farkas and Jajodia (2002) for a survey of this research.

3. CONSTRUCTS AND FACTORS IN INFORMATION ETHICS

The descriptive constructs and factors of criticality in the education of information ethics are customized in this study of data mining and ethics from diverse expert literature sources, as defined in the categorical framework of Table 1 in Appendix B. A more complete description of each category is available from the authors.

These factors are imputed in this study to be important in introducing information ethics into data mining curricula. Few studies in the literature include this diversity of principles and practices in analyzing the adequacy of ethics in data mining education.

4. FOCUS OF THE STUDY

The focus of the study is to analyze the content of contemporary data mining texts for adequacy in introducing information ethics and privacy in information systems and computer science curricula. Closer examination of *business, consumer and ethical*, *governmental and organizational*, *managerial and methodological*, *pedagogical* and *technological content* enables fresh insight into the learning or non-learning of principles and practices of information ethics, and privacy as an issue in ethics, by higher education students. The study defines in essence a framework for improving the likelihood of a substantive information ethics, privacy, and mining model in the curricula of computer science and information systems schools.

5. RESEARCH METHODOLOGY

The research methodology of the study consists of *five iterative stages of analysis*.

In *stage 1* a sample of 29 data mining texts were chosen and analyzed by the authors of this study, in December 2004 – February 2005, as mostly representative of primary data mining texts in ABET certified schools of computer science and information systems. The authors were already knowledgeable instructors and researchers in data mining and information ethics. The authors identified the textbooks in a joint analysis that included diverse editors and marketing representatives of leading mining publishers and in an independent survey and analysis of scholarly and practitioner sources on the World Wide Web. The textbooks were identified from 2004 – 2000 publication dates of the publishers. The textbooks identified by the authors are detailed in Appendix A of this study.

In *stage 2* the 29 primary mining textbooks were analyzed by the authors and by a graduate student, in March – July 2005, for content inclusion of constructs of information ethics, from a checklist of 26 *business, consumer and ethical*, *governmental and organizational*, *managerial and methodological*, *pedagogical* and *technological factors*. The

factors are detailed in Table 1 in Appendix B. A more complete list of the factors is available from the authors on request. To the factors was applied a six-point rating scale of 5 – very high inclusion to 0 – no inclusion, based on author and aide perception of the scope of the principle and practice factors in the textbooks. The authors subsequently interpreted the data. The findings are summarized in Table 3 in Appendix B.

In *stage 3* the authors will expand the scope of the study to survey a sample of data mining instructors in ABET schools of information systems, for inclusion of primary mining texts and notably for secondary scholarly and practitioner sources in mining curricula. Instructors will also be surveyed in this stage as the ranking relevance of the constructs and factors in information ethics in Table 1 (Appendix B) in mining curricula and text. The instructors will furnish inherently an initial design for ideal mining texts and curricula. In *stage 4* the authors, where feasible, will survey a sample of students in mining courses of the schools in *stage 3*, for learning or non-learning of ethical information principles and practices. In *stage 5* the authors will finalize a design of a model that fully integrates information ethics in a computer science and information systems data mining curriculum. *Stages 3 and 4*, which will confirm or not confirm the findings and implications in *stages 1* and *2* of the current study, will be completed in August – November 2005, and *stage 5* will be completed in December 2005 – January 2006.

6. STATISTICAL ANALYSIS

Twenty-nine data mining texts were examined to determine the extent to which each covers the constructs listed in Table 1 of Appendix B. The rating scale of Table 2 was used for each of the 26 construct factors.

Table 3 (Appendix B) summarizes the results of the analysis. Each entry gives the number of books in the corresponding construct and rank. In the C1 categories, we see that only two books had chapters and eleven had sections devoted to a general discussion of privacy. Two books had chapters devoted to Federal privacy legislation (construct C2-3) and two more had sections on the topic. As might be expected, books devoted more space to managerial and methodological

constructs. Five books had chapters on either personalization techniques (construct C3-3) or the protection of systems (construct C3-5). Fifteen books had entire sections on the managerial and methodological constructs (all C3 constructs). Interestingly, only one book had a section discussing the role of the Chief Privacy Officer (construct C3-1). None of the books surveyed had a chapter on any of the pedagogical constructs, although eleven of them did at least mention one or more privacy publications (construct C4-2) or privacy advocacy groups (construct C4-5). Finally, almost no references were made to the various technological constructs. Only two books had sections in this area (construct C5-2) and one book had a few sentences.

Looking at the number of books that at least mention one of the constructs, see Table 4 in Appendix B, we can make a few interesting, but not unexpected, observations. Since most data mining texts aim at future data mining professionals, we expect that many of them would discuss some of the managerial issues surrounding the topic. As Table 4 shows, there are 56 references to these issues in the books surveyed. Business issues (C1 – 23 references) and governmental issues (C2 – 28 references) surrounding privacy are about equally referenced. The category with the least coverage is C5 – the technological constructs that enable the protection of personal privacy. As mentioned in the Introduction, there is a lack of awareness of the ethical issues surrounding data mining by data mining professionals. In light of the sparse coverage of the C1 and C5 constructs in the texts surveyed, this is not surprising.

We also can rank each text on the total number of references to the constructs. Table 5 (Appendix B) shows the total references to the constructs for each of the twenty-nine texts, which gives a means of rating the individual books as to their coverage of the 26 construct areas. Note that only three texts have 30 or more references. Table 6 (Appendix B) summarizes the results of Table 5 by aggregating books according to the total references into ranges 0-9, 10-9, 20-29, and 30 or greater.

These data lead us to conclude that unless individual instructors make an effort to discuss ethical issues in their courses, students will not be exposed to them through their textbooks. Data mining and its applications can have a profound effect on the privacy of individuals and groups. To make data mining professionals more aware of these issues, it is important that instructors and authors in this area fully discuss them. Instructors and authors should keep in mind the following observation made nearly ten years ago by Rotenberg (1996).

> Privacy will be to the information economy of the [the 21st] century what consumer protection and environmental concerns have been to the industrial society of the 20th century.

7. IMPLICATIONS OF INITIAL STUDY

"We are not just heading towards a world of Big Brother … but also toward a more mindless process … - a world that is beginning to resemble Kafka's vision in The Trail." (Solove, 2004)

The inadequacy of contemporary data mining texts, for instructors attempting to effectively introduce information ethics in computer science and information systems curricula, is an important implication of this current study. *Ethical principles for moral and philosophical theory (Grodzinsky, 1999) continue to be not included in depth in data mining texts. Ethical and non-ethical practices of citizen and consumer information mining that are current and immediate in government and industry are not often included in outdated texts (Schrage, 2005). Though instructors may enhance core foundational texts with further practitioner and scholarly resources, they are challenged in having comprehensively convenient and integrated texts (Richards, 2005). The text is a critical source for instructors in introducing students to the information ethics of data mining.*

The inherent limitation of data mining texts in introducing students to ethical issues is another implication of the study. *Though students learn fundamental dimensions and techniques of customer marketing and industry technology in a mining course, they may not learn complex ethical dilemmas and*

issues. Issues and pressures of customer privacy (Saporito, 2005) and legislation on privacy (Holmes, 2005) may not be clear to students if the mining text is limited to technology (Hackathorn, 2005). The limitation of the sampled texts of this initial study is an impediment in indicating ethical and professional responsibilities in society to students.

Implications include the importance of having an improved frame of reference for limited data mining texts. *Instructors in a mining course of a computer science and information systems school may not have sufficient creditability in introducing information ethics. Studies indicate improvement in the inclusion of philosophy, social science and information systems instructors (Brey, 2000) and in further interdisciplinary frames of marketing and business instructor teams (Gloeckler, 2005). Other improvements may include guest lecturers from firms in industry, government organizations, as the Federal Trade Commission, and professional societies, from which students might learn immediate issues and practices of ethics, privacy and security. Innovation in the introduction of information ethics in a mining course with limited texts is a needed step.*

The importance of collaborative liaison with customer relationship management (CRM) data mining and privacy software firms is an implication of the study*. To increase the* knowledge *of instructors in mining curricula and the learning of students that have limited data mining texts, schools of information systems may consider installing software of privacy enhancing technology (PET) firms, and of privacy invasive technology, in mining courses. Installation of early adopted beta marketing and mining software from relationship management firms, moreover, may be helpful in learning the potential and the risk of the technologies (Woods, 2004). (Pace University School of Computer Science and Information Systems is a recipient of Oracle CRM e-Business, SAS and SPSS technologies.) Such software and technology are often furnished to schools in low cost grants. Integration of current generation tools of mining technology in a mining course with an outdated text is a further needed step.*

The final implication of this study is in both the challenge and in the opportunity for schools of computer science and information

systems to improve curricula in data mining. Technology continues to evolve in power, contributing to increased *ethical issues that challenge information systems and computer science instructors and professionals (Turner, 2005). In a time of lower enrollment in technology programs, however, instructors have an opportunity to improve the immediacy of general technology curricula and specific mining curricula, by integrating information ethics. The subject is of increasing interest to potential information systems and computer science students and parents in United States Society. Instructors also have the opportunity to furnish an improved ethical foundation for the future of information systems and computer science professions. Information ethics is an issue that impacts what type of society schools will introduce in the 21st century (Solove, 2004).*

8. LIMITATIONS AND OPPORTUNITIES FOR RESEARCH

The findings of this study of data mining texts furnishes a foundation in *stages 1 and 2* for further researching mining and information ethics in the curricula of computer science and information systems schools. *Stages 3, 4 and 5* will expand the study of texts to instructors and ideally to students in schools, and *stage 5* will introduce an information ethics and mining course for a 21st century information systems curricula model. The future findings of the latter stages will be helpful and informative to instructors striving to integrate issues of ethics in mining pedagogy.

9. CONCLUSION

This study of data mining and information ethics in schools of computer science and information systems is insightful into the *business, consumer and ethical, governmental and organizational, managerial and methodological, pedagogical* and *technological factors* of ethical principles and practices in mining curricula. The inadequacy of current data mining texts for instructors and the limitation of the texts for student learning are issues indicated in the initial study.

The importance of further frames of reference in industry practitioner resources integrated into mining curricula is indicated in the study. Further research is needed in the factors of ethical practices and principles in information systems and computer science schools. The study in its preliminary state furnishes a framework however for improving the inclusion of information ethics in data mining curricula and is therefore timely.

10. REFERENCES

Allen, A. (1988) Uneasy Access: Privacy for Woman in a Free Society, Rowman and Littlefield, Totowan, NJ.

Brey, Philip (2000) "Disclosive Computer Ethics," Computers and Society, December, 10-16, 22-23.

Cannon, J.C. (2005) Privacy: What Developers and Information Technology Professionals Should Know, Addison-Wesley, New York, 17-52, 79-173, 289-299.

Cavoukian, A. (1998) Data Mining: Staking a Claim on Your Privacy. Information and Privacy Commissioner's Report, Ontario, Canada.

CDT, 2005. http://www.cdt.org/privacy/eudirective/EU_Directive_.html#HD_NM_2, accessed June 14, 2005.

Danna, A. and Gandy, O.H. Jr. (2002) "All The Glitters is Not Gold: Digging Beneath the Surface of Data Mining," Journal of Business Ethics, 40, 373-386.

Farkas, C. and Jajodia, S. (2002) "The Inference Problem: A Survey", SIGKDD Explorations, Vol. 4, Issue 2, 6-11.

Fried, C (1970) Privacy: A Rational Context. Chapter IX in Anatomy of Values, Cambridge University Press, New York.

Fulda, J. S. (1997) "From Data to Knowledge: Implications of Data Mining," Computers and Society, December, 1997, 28.

Fulda, J. S. (1999) "Solution to a Philosophical Problem concerning Data Mining," Computers and Society, December, 1999, 6-7.

Gloeckler, Geoff (2005) "This Is Not Your Father's MBA," Business Week, May 16, 74-75.

Grodzinsky, Frances (1999) "The Practitioner from Within: Revisiting the Virtues," Computers and Society, March, 9-10.

Gross, G. (2005) "U.S. lawmakers push for data privacy legislation," Computerworld, March 16, 2005. Accessed June 14, 2005 at http://www.computerworld.com/governmenttopics/government/legislation/story/0,10801,100405,00.html.

Hackathorn, Richard (2005) "The Ethics of Business Process Management," Business Integration Journal, March, 32.

Hays, C. (2004) "What Wal-Mart Knows About Customer's Habits," The New York Times, November 14, 2004. Accessed November 15, 2004 at http://www.nytimes.com.

Holmes, Allan (2005) "Riding the California Privacy Wave," CIO, January 15, 44-49.

Lawler, James (2003) "Customer Loyalty and Privacy on the Web," Journal of Internet Commerce, Summer, 1-10.

Moor, J. H. (1997) "Towards a Theory of Privacy in the Information Age," Computers and Society, 27, 3, 27-32.

Nissenbaum, H. (1998) "Protecting Privacy in the Information Age: The Problem of Privacy in Public," Law and Philosophy, 17, 5-6, 559-596.

OECD (2005), http://www.oecd.org/document/18/0,2340,en_2649_201185_1815186_1_1_1_1,00.html, accessed, June 14, 2005.

Preston, John, Preston, Sally, and Ferrett, Robert (2005), Computers in a Changing Society, Prentice Hall: Upper Saddle River, New Jersey), 156.

Regan, P. (1995) Legislating Privacy: Technology, Social Values, and Public Policy, U. of North Carolina Press, Chapel Hill, NC.

Richards, Clinton H. (2005) "Private and Public Sector Ethics," Proceedings of the Applied Business Research Conference, Puerto Vallarta, Mexico, March, 2.

Rotenberg, M. (1996) Quoted in James Gleick, "Big Brother is Us", New York Times Magazine, Sept. 29, 1996, p.130.

Saporito, William (2005) "Are Your Secrets Safe?," Time, March, 30.

Schrage, Michael (2005) "Ethics, Shmethics," CIO, March 15, 40.

Solove, Daniel J. (2004) The Digital Person: Technology and Privacy in the Information Age, New York University Press, New York, 55,149.

Spangler, W.E. et al (2003) "Using Data Mining to Profile TV Viewers," Communica-

tions of the ACM, December 2003, 46, 12, 67-72.

Tavani, H. (1999) "KDD, data mining, and the challenge for normative privacy," Ethics and Information Technology, 1, 4, 265-273.

Tavani, Herman T. (2004) Ethics and Technology: Ethical Issues in an Age of Information and Communication Technology, John Wiley and Sons, Hoboken, New Jersey, 92,121,124,140,144,146.

Turner, Freda (2005) "Anatomy of Unethical Leadership Crisis in Corporate America," Proceedings of the Applied Business Research Conference, Puerto Vallarta, Mexico, March, 1-4.

USDC, 2005,
http://www.export.gov/safeharbor/, accessed June 16, 2005.

Van Wel, L. and Royakkers, L. (2004) "Ethical issues in data mining," Ethics and Information Technology, 6, 129-140.

Vedder, A. (1999) "KDD: The challenge to individualism," Ethics and Information Technology, 1, 4, 275-281.

Warren, S. and Brandeis, L. (1890) "The Right to Privacy," Harvard Law Review, 14, 5, 1890.

Westin, A. F. (1967) Privacy and Freedom, Athenum Press, New York.

Woods, Deirdre (2004) "Powerful Partnerships," Computerworld, December 13, 40.

Appendix A

Data Mining Textbooks of Study (*Stages 1 and 2 of Study.*)

Note: The number in brackets refers to the number of the book in the study.

[6] Agosta, Lou (2000) The Essential Guide to Data Warehousing, Prentice Hall.

[21] Becker, Shirley (2002) Data Warehousing and Web Engineering, IRM Press.

[7] Berry, Michael J.A. and Linoff, Gordon S. (2000*) Mastering Data Mining: The Art and Science of Customer Relationship Management, John Wiley and Sons.

[16] Berry, Michael J.A. and Linoff, Gordon (2000*) Data Mining Techniques for Marketing, Sales, and Customer Support, John Wiley and Sons.

[9] Berson, Alex, Smith, Stephen and Thearling, Kurt (2000*) Building Data Mining Applications for CRM, McGraw-Hill Publishing.

[2] Biere, Mike (2003) Business Intelligence for the Enterprise, IBM Press / Prentice Hall.

[15] Delmater, Rhonda and Hancock, Monte (2001) Data Mining Explained: A Manager's Guide to Customer-Centric Business Intelligence, Digital Press.

[10] Dunham, Margaret H. (2003) Data Mining: Introductory and Advanced Topics, Prentice Hall.

[19] Hand, David, Mannila, Heikki and Smyth, Padhraic (2001) Principles of Data Mining, MIT Press.

[8] Hughes, Arthur M. (2000) Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable Customer-Based Marketing Program, McGraw-Hill Publishing.

[5] Humphries, Mark, Hawkins, Michael W. and Dy, Michelle C. (2000*) Data Warehousing: Architecture and Implementation, Prentice Hall.

[13] Inmon, W.H., Terdeman, R.H., Norris-Montanari, Joyce and Meers, Dan (2001) Data Warehousing for e-Business, John Wiley and Sons.

[11] Kantardzic, Mehmed M. and Zurada, Jozef (2005) Next Generation of Data-Mining Applications, IEEE Press.

[22] Kantardzic, Mehmed (2002) Data Mining: Concepts, Models, Methods, and Algorithms, John Wiley and Sons.

[12] Kimball, Ralph and Merz, Richard (2000) The Data Webhouse Toolkit: Building the Web-Enabled Data Warehouse, John Wiley & Sons.

[28] Kudyba, Stephan and Hoptroff, Richard (2001) Data Mining and Business Intelligence: A Guide to Productivity, Idea Group Publishing.

[24] Loshin, David (2003) Business Intelligence: The Savvy Manager's Guide: Getting Onboard with Emerging Information Technology, Morgan Kaufmann.

[1] Mallach, Efrem G. (2000) Decision Support and Data Warehouse Systems, McGraw-Hill Publishing.

[3] Marakas, George M. (2003) Modern Data Warehousing, Mining, and Visualization: Core Concepts, Prentice Hall.

[29] McGonagle, John J. and Vella, Carolyn M. (2003) The Manager's Guide to Competitive Intelligence. Praeger Publishers.

[20] Mena, Jesus (2000*) Data Mining Your Web Site, Digital Press.

[25] Moeller, R.A. (2000) Distributed Data Warehousing Using Web Technology, AMACOM.

[18] Mohammadian, Masoud (2004) Intelligent Agents for Data Mining and Information Retrieval, Idea Group Publishing.

[17] Sarker, Ruhul A., Abbass, Hussein A. and Newton, Charles S. (2002) Heuristics and Optimization for Knowledge Discovery, Idea Group Publishing.

[27] Simon, Alan and Shaffer, Steven (2001) Data Warehousing and Business Intelligence for e-Commerce, Morgan Kaufmann.

[14] Thuraisingham, Bhavani (2003) Web Data Mining and Applications in Business Intelligence and Counter-Terrorism. CRC Press.

[4] Turban, Efraim, Aronson, Jay E. and Liang, Ting-Peng (2005) Decision Support Systems and Intelligent Systems, Prentice Hall.

[26] Vitt, Elizabeth, Luckevich, Michael and Misner, Stacia (2002) Making Better Business Intelligence Decisions Faster, Microsoft Press.

[23] Wang, John (2003) Data Mining: Opportunities and Challenges, Idea Group Publishing.

Appendix B – Tables

Table 1 - Constructs, Factors, and Sources in Information Ethics

C1 Business, Consumer and Ethical Constructs	
Ethics Codes	**C1-1**
Definitions of Privacy	**C1-2**
Functions of Privacy	**C1-3**
Personal vs. Group Privacy	**C1-4**
Studies of Privacy	**C1-5**
C2 Governmental and Organizational Constructs	
Constitution	**C2-1**
Court cases	**C2-2**
Federal Legislation	**C2-3**
State Legislation	**C2-4**
Authorities	**C2-5**
Organizations	**C2-6**
C3 Managerial and Methodological Constructs	
Chief Privacy Officer	**C3-1**
Personal Privacy Policy Standards	**C3-2**
Personalization Techniques	**C3-3**
Privacy Systems	**C3-4**
Protection of Systems	**C3-5**
C4 Pedagogical Constructs	
Privacy Studies	**C4-1**
Privacy Publications	**C4-2**
Privacy Conferences	**C4-3**
Scholarly Journals	**C4-4**

Privacy Groups	**C4-5**
C5 Technological Constructs	
Digital Rights Management	**C5-1**
Platform for Privacy Preferences	**C5-2**
Privacy Aware Technology	**C5-3**
Privacy Invasive Technology	**C5-4**
Privacy Software Technology	**C5-5**

Table 2 - Rating Scale

Rating	Meaning
0	No mention of the construct.
1	A word or two about the construct
2	One or two sentences about the construct
3	A complete paragraph discussing the construct
4	A complete section about the construct
5	A complete chapter about the construct

Table 3 – Number of Books in Each Construct/Rank

		Rank					
		5	**4**	**3**	**2**	**1**	**0**
	Construct						
Ethics Codes	**C1-1**	2	6	1	1	1	18
Definitions of Privacy	**C1-2**	0	1	1	2	0	25
Functions of Privacy	**C1-3**	0	1	0	0	0	28
Personal vs Group Privacy	**C1-4**	0	2	1	1	0	25
Studies of Privacy	**C1-5**	0	1	1	2	0	25
Subtotal		2	11	4	6	1	121
Constitution	**C2-1**	0	0	0	0	0	29
Court cases	**C2-2**	0	0	0	1	0	28
Federal Legislation	**C2-3**	2	0	1	2	3	21
State Legislation	**C2-4**	0	1	1	2	3	22
Authorities	**C2-5**	0	0	3	4	0	22
Organizations	**C2-6**	0	1	2	1	1	24
Subtotal		2	2	7	10	7	146
Chief Privacy Officer	**C3-1**	0	1	2	2	4	20
Personal Privacy Policy Standards	**C3-2**	0	2	5	3	2	17
Personalization Techniques	**C3-3**	2	7	0	1	1	18
Privacy Systems	**C3-4**	0	3	0	1	1	24
Protection of Systems	**C3-5**	3	2	5	5	4	10

Subtotal			5	15	12	12	12	89
Privacy Studies	**C4-1**	0	0	0	0	0	29
Privacy Publications	**C4-2**	0	2	3	1	1	22
Privacy Conferences	**C4-3**	0	0	1	0	0	28
Scholarly Journals	**C4-4**	0	1	0	1	1	26
Privacy Groups	**C4-5**	0	0	0	0	4	25
Subtotal		0	3	4	2	6	130
Digital Rights Management	**C5-1**	0	0	0	0	0	29
Platform for Privacy Preferences	**C5-2**	0	2	0	1	0	26
Privacy Aware Technology	**C5-3**	0	0	0	0	0	29
Privacy Invasive Technology	**C5-4**	0	0	0	0	0	29
Privacy Software Technology	**C5-5**	0	0	0	0	0	29
Subtotal		0	2	0	1	0	142
Total		9	33	27	31	26	628

Table 4 — Any Mention

Construct Category	Any Mention at All (Rank > 0)
C1 – Business and Consumer Ethics	23
C2 – Government and Organizations	28
C3 – Managerial and Methodological	56
C4 – Pedagogical	15
C5 - Technological	3

Table 5 — Total References by Book

Book	Total References
1	10
2	0
3	14
4	27
5	3
6	4
7	24
8	0
9	23
10	6
11	0
12	18
13	0
14	30
15	2

16	1
17	8
18	10
19	0
20	28
21	7
22	4
23	37
24	13
25	8
26	2
27	34
28	16
29	17

Table 6 – Book Ranges

Range	Books in Range
0 to 9	15
10 to 19	7
20 to 29	4
30 or Greater	3