

Data mining: Privacy and ethical issues associated with consumer information

Brian D. Otte

Capella University

Title page .....	1
Abstract .....	3
Introduction.....	4
Discussion .....	5
Data mining definition. ....	5
Enabling technology. ....	8
Faster more powerful CPUs.....	8
Massive storage for cheap.....	8
Fast communication channels linking data source, databases, and data warehouses. ....	9
Organizational perspectives on data mining.....	10
The organization and management of markets. ....	12
The organization and management of risk.....	12
The organization and management of fraud. ....	13
Individual perspectives on data mining .....	14
The consumer perspective.....	15
The consumer and privacy.....	19
The consumer and the ethical ramifications. ....	22
Future trends in data mining. ....	23
Conclusion .....	24
References.....	26

### Abstract

Data mining consumer information such as purchases and activities furnish organizations with information about consumers' habits. An investigation into the subject of data mining concerning privacy and ethical issues associated with consumer information is the focus of this literature review. The sources for this literature review include published academic research, recent books, and corporate web sites. A view of data mining consumer information and the related privacy and ethical issues through the lens of the organization, the individual, and the data, frames the context for this study. The conclusion details a distillation of perspectives presented from authors with specific concerns about current data mining practices and future issues, considering current trends with data mining.

## Introduction

Data mining extracts knowledge from volumes of data, using mathematical models to find new, patterns in the data that are not apparent without using these methods (Cooper & Schindler, 2008, p. 702). The focus for this literature review is on data directly collected from consumer purchases, activities, or choices. These inferences inform the data miner about the lives and actions of the people from which the data is collected. Indications exist that data mining details personal characteristics from data that contains no personally identifiable information through profiling (Hildebrandt, 2006). Profiling indicates personal traits about individuals even though the source data has removed all personal information. The original data does not indicate who the person is; however, the data reveals likely personal attributes by way of membership in a given profile as derived from their actions.

Data mining can include the study of association, sequence, classification, clustering, and forecasting (Seifert, 2008). This literature review focuses on the general concept of data mining from three specific viewpoints. Business issues an organization may face, consumer exposure to privacy issues, and data management issues associated with the data, frame the context of this literature review. An investigation of these three issues through an ethical and privacy lens, while relating each to the consumer and the organization interested in the results.

An investigation into what organizations are looking to get from the data, the mechanisms they get the data, and what they do with the data is conducted through this literature review. This requires a focus on the business strategy and the alignment of data mining techniques an organization uses to harness the power of Information Technology (IT). The actions taken by organizations should reflect their business strategies (Ward & Peppard, 2002).

An organization is compelled to gather information about their customers developing methods resulting from following their strategies. Fueling the need for consumer information is the need for organizations to gain a competitive advantage. The methods organizations employ, and the interaction with the consumer about the data collection is a reflection of their organizational strategy. There is little to hold organizations accountable for the data collected once the data is anonymized (Hildebrandt, 2006). Viewing organizations through an ethical lens enables a vision as to how they should handle the data. The effect of profiling on the individual with respect to privacy is the approach for this literature review.

While addressing each of these issues in their respective sections, the issues also continually linked to their ethical implications. Specifically, this literature review presents an investigation of data mining with respect to the organization, privacy issues, and data management issues. The ethical dimensions that each of these domains present through data mining reveals information to those who are investigating the secrets about the consumer, while consumers remain largely unaware of data mining, or its effects.

## Discussion

Data mining definition.

Landry, Debreceeny & Gray (2004) offer a clear picture of data mining stating “Data mining incorporates a variety of tools and processes that can work independently or together to analyze and discover relationships in collections of data” (p. 26). Vedder (1999) presents a view where data mining is a subcomponent of Knowledge Discovery in Databases (KDD). Vedder, discussing KDD, states “The KDD process is usually divided into three phases: the data warehousing phase, the data mining phase, and the interpretation phase” (p. 275). Firestone

(2005) extends the definition of data mining stating, “Data mining is the activity of automatically extracting hidden information (patterns and relationships) from large databases without benefit of human intervention or initiative in the knowledge discovery process” (p. 50). Other definitions detail the process utilized to extract hidden information entailing the use of algorithms, applying a “...non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data...” (Firestone, 2005, p. 50)

Landry, Debreceeny and Gray (2004) posit three phases of data mining as 1) data warehouse, data screening, aggregates, and 2) data cleaning, and 3) Artificial Intelligence and statistical analysis tools (p. 26). They determine that there are two separate styles of data mining, called directed and undirected data mining. The type of data mining employed reflects the business strategy that an organization employs to seek its data. According to Landry, Debreceeny and Gray “...directed data mining, is designed to test or measure expected patterns of business behavior” (p. 26). Landry, Debreceeny and Gray, defining undirected data mining state that it “...seeks patterns or relationships without any preconceived expectations of hypothesis” (p. 27). Therefore, an organization, which has some vague notion of consumer activity, might employ directed data mining, whereas an organization attempting to formulate new ideas might employ undirected data mining.

Cunningham, Song and Chen (2006) detail data warehouse design for a Customer Relationship Management (CRM) system. The purpose of a CRM system is to aid an organization in managing the information about its customers. These systems might include information on customer profitability or customer orders and trending histories. According to Cunningham, Song and Chen who state, “[t]he purpose of a data warehouse is not just to store

data but rather to facilitate decision making. As such, the first step to designing the schema for the CRM data warehouse is to identify the different types of analyses that are relevant to CRM” (p. 66). The takeaway for the organization implementing data warehousing is that before implementing a system, the organization needs to understand what it is attempting to retrieve from the system.

Data mining applications range from predicting consumer demand to predicting consumer preferences. Data mining offers the ability to look at past consumers preferences and predict the actions of consumers. As an example, Hildebrandt (2006) suggests that Ambient Intelligence will anticipate events concerning consumer activities beyond hotel room temperature preferences. Specifically, Hildebrandt believes that organizations employing a data mining toolset may seek to gain competitive advantage, suggesting that this form of data mining may “...cater to your specific health needs, travel plans or your preferred professional infrastructure” (p. 550). In this case, the use of data mining enables an organization to anticipate travel plans; it might aid the business traveler and indeed give a competitive advantage to the organization. Similarly, by catering to specific health needs, data mining adds value to both the organization and society through extending or creating a better quality of life by using these tools.

Organizations interested in marketing may seek information indicating what leads consumers who buy product X to buy product Y, which yields a higher margins. Grocery stores may use this information for product placement, and marketers are interested in this information to take advantage of complementary products. According to Seifert (2008), “Although data mining can help reveal patterns and relationships, it does not tell the user the value or significance of these patterns.” (p. 3) Organizations that correctly interpret and strategically act

on the information, provided by data mining, gain a competitive advantage. To enable a clear conversation about data mining, a discussion of the enabling technology naturally leads into the recent emergence of data mining in many organizations.

Enabling technology.

Many factors enable data mining. Cabena, Hadjinian, Stadler, Verhees, and Zanasi (1998) detail three enabling technologies for data mining. They suggest that some enablers are enormous amounts of data stored on computer systems, the growth a data warehousing, new IT solutions, and research into machine learning. Three enabling technologies, seen as a subset of the list provided by Cabena et al., include the ability to process more data, the ability to store more data, and as for new IT solutions, communication channels linking the data sources with the data base that stores the information.

*Faster more powerful CPUs.*

According to Morris and Truskowski (2003), since 1983, processor speeds have improved by a factor of multiple thousands. Moore's law, a commonly quoted law concerning CPU power, states, "...the number of transistors per unit area of silicon doubles every 1.5 years" (Morris & Truskowski, p. 206) (Seifert, 2008) (Moore, 1965). Contrast this growth with storage, where the media is doubling the bits stored per unit of HDD media every year (Morris & Truskowski, p. 206).

*Massive storage for cheap.*

Morris and Truskowski (2003) present the trends in storage capacities indicating that hard drive capacities nearly double each year (p. 215). The increased capacities cause the prices per unit of storage to drop. A crucial aspect of successful data mining is the existence of large

database storage (Cooper & Schindler, 2008, p. 111) (Seifert, 2008, p. 2). The cost per unit of storage continues to fall (Morris & Truskowski).

*Fast communication channels linking data source, databases, and data warehouses.*

The ability to communicate information that originates from consumer activity to a data repository suitable for data mining is possible through data networks facilitated by the Internet. Consumer activities on the Internet are themselves sources of information that data mining may find useful. The ability for personal information to flow freely throughout the world may also be a mechanism that enables data mining. Agre and Rotenberg (1997) cite Bennett who suggests, "...personal information now routinely flows across jurisdictional boundaries" (p. 3). The effect on privacy is that the individual has great difficulty in identifying where their data exits, and any legal rights associated the data they create are obscured by the free flow of information.

Morris and Truskowski (2003) posit, "...more important than improvements in device density or cost have been new applications that have been enabled by these advances" (p. 206). Morris and Truskowski, looking at future challenges, indicate that machine-generated data applications will become increasingly prevalent (p. 214). Some other machine-generated technologies include digital computers, cell phones, or automated tollbooth-scanning devices and have the potential for providing an uninterrupted stream of consumer information. Agre and Rotenberg (1997), discussing machine generated technology currently in use, discuss a photo radar system that is "...monitoring all cars and photographing the license plates of speeders" (p. 170). These and other non-interactive data collection venues bring to life the concept of Ubiquitous computing espoused by Weiser (Weiser & Gold, 1999). Inexpensive storage, powerful CPUs, and automatic collection and communication of data enable data mining.

These sources of data naturally lead to supplying massive amounts of data. Even with the enabling technology of massive storage for cheap, organizations needed a location to store its quality data that it selected from the massive amounts of collected data (Cabena et al., 1998, p. 19). At this point in the KDD process, as presented by Vedder (1999) the data is destined for a data warehouse.

Consolidating information typically occurs in a system called a data warehouse, where algorithms find statistical inferences or correlations, revealing the secrets that the data contains. According to Cooper and Schindler (2008), the evolution of data mining stepped through data collection in the 1960's, data access in the 1980s, data navigation in the 1990s, leading to data mining in the 2000s (p. 111). Many interests exist for finding out what the data reveals about consumers through data mining, such as economic, marketing, or even national security concerns. Data mining reveals this interesting data through various mathematical techniques applied to the data (Cooper & Schindler, 2008).

A data warehouse enables an organization to make decisions based off the data, whereas data mining analyzes the data and presents new knowledge from the data (Cabena et al., 1998). Enabled by the ability to process more data, store more data, the proliferation of data collection sources, and connecting all of this data to a data warehouse, offers data miners a rich source of information to gather and represent reality through digital means. The organizational perspective on data is the next area of focus is this literature review.

#### Organizational perspectives on data mining

Cabena et al. (1998) suggest that an organization has three general areas with which to apply data mining techniques in the management of markets, risk, and fraud. Seifert (2008) adds

other areas to use data mining such as detection of waste and medical research. Detection of waste is a method of managing risk or fraud. An exploration of medical research through data mining market information for a pharmaceutical organization is a method of managing sales markets for medical research.

The organizational use of data mining extends the knowledge about consumers' actions to the organization. Through data mining, an organization can discern information about consumers, which their competition cannot. According to Cooper and Schindler (2008) "Data-mining technology provides two unique capabilities to the researcher or manager: pattern discovery and prediction" (p. 111). Pattern discovery such as purchasing habits or correlations between purchases and predicting how consumers react to marketing allows an organization to discover and exploit consumer trends (Cooper & Schindler).

These capabilities present an organization a competitive intelligence (CI). According to Cabena et al. (1998) CI is "...the process of collecting, analyzing, and disseminating information about industry developments or market trends to enhance a company's competitiveness" (p. 31). CI offers the organization an advantage over their competition giving the organization a strategic advantage.

Hoffer, George, and Valacich (2001) present work by Porter, detailing two activities related to the value chain, which are primary activity and support activity. Hoffer, George, and Valacich list the primary activities as inbound logistics, operations, outbound logistics, marketing and sales, and service. However, data mining presents the ability to experience a competitive advantage in any category should an organization discern a method to capitalize on the information. Ward and Peppard (2002) stressing strategy state, "[a]ll organizations have

some form of strategy, whether implicit or explicit, and the essence of business strategy lies in creating future competitive advantages faster than competitors” (p. 65). This is one method where data mining enables a competitive advantage.

*The organization and management of markets.*

While the discussion of the technical merits of data mining is beyond the scope of this literature review, the policies that an organization embraces associated with data mining is within the context of this review. Hidden secrets within massive amounts of data have their contents revealed through data mining, only discernable after removing the noisy data from the useful data (Hildebrandt, 2006). Organizations that do not have access to the data or those who do not mine the data to reveal its secrets cannot take advantage of the information it contains. This means that those who do have access to the information, can make decisions based grounded on that information.

However, these secrets also have the potential to reveal secrets of the individual, impinging on their privacy. Rygielski, Wang, and Yen (2002) believe that “...privacy is more of a policy issue than a technology one” (p. 495). This indicates that the organization has choices to make concerning personal privacy. To mitigate the concerns of privacy Rygielski, Wang, and Yen suggest that organizations that collect data inform consumers of what they are doing with the data, and what data that they are seeking (p. 495).

*The organization and management of risk.*

According to Cabena et al. (1998) organizational risk include issues “...arising from competitive threat, poor product quality and customer attrition” (p. 30). Organizations may use data mining to address these risks. Predicting the actions of consumers is vital to understand

capacities, demands, and materials needs. Cabena states, "...data mining is used in the finance, retail and telecommunications industries to predict likely customer losses" (p. 30). The strategy to employ data mining aids an organization with another source of valuable data to manage its risk.

*The organization and management of fraud.*

Cabena et al. (1998) state, "Human nature dictates that some level of fraud is inevitable in all industries" (p. 34). Organizations adopt techniques to provide a competitive advantage. Landry, Debreceeny, and Gray (2004) state, "Data mining can help you better understand your business by providing new insights into areas such as customer buying and return patterns, product profitability, instances of general selling and administrative (GSA) expenses, and fraud detection" (p. 25). Unfortunately, for organizations, identifying fraud is a necessary reality that they must confront.

Fraud exists and a fact of doing business is that property is stolen. Furthermore, internal problems are bound to happen. Data mining presents the ability to comprehend the issues by removing all of the extraneous noise present with the massive amounts of data (Hildebrandt, 2006). Business sectors that involve many transactions are especially susceptible to fraud as more data obscures details, which data mining addresses by offering the tools to remove obscurity through its use.

Meeting a strategic need for knowledge management requires the strategy employed to be in alignment with the espoused strategy. Smith (2004) discussing Knowledge Management (KM) strategies, suggests that "Creating customer value, turning a profit, and managing people, are some of the major areas that companies must inspect in detail when determining what KM

strategy is going to be most effective” (p. 7). Smith (2004) states, “The goal of every organization is to manage their knowledge successfully” (p. 8). Once an organization gets consumer information, consumers have little recourse as to where the information travels, or what they do with the data (Da Costa, Schulte, & Singer, 2006). Moreover, if the organization mismanages the data, the consumer is unlikely to be aware that it happened, or what the ramifications of the mismanagement mean to the consumer.

Published research indicates many applications for data mining, but this literature review focuses on the organization, the individual, and the data management perspective. Investigating through a privacy and ethical lens the current focus is on group profiling and targeted marketing. These activities directly affect the consumer, either directly through targeted marketing, or indirectly through profiling. Presentations of the issues associated with data mining from the view of the consumer are the next.

#### Individual perspectives on data mining

The focus of this section is on the data generated by the consumer. In some cases, consumers willingly link their demographic information to their actions, incentivized through money saving and convenient data collection methods such as preferred shopping cards. Fresh Brands (2006) entices users to use the preferred card indicating, “Using the Preferred Club Card entitles you to valuable discounts and savings throughout the store” (para. 1). Consumers get discounts on purchases, and Fresh Brands get consumer purchasing information.

Preferred Club Cards do offer limited savings, the card enables the linking of an individual’s purchases to specific demographic information of specific individuals. The store creates marketing programs based on the results obtained. If purchasing habits remain private

and the use of information benefits the consumer, these programs are advantageous. On the other hand, what does it mean if purchasing habits become public, or consumers become disadvantaged in some way by the marketing program?

Another example details a method that organization's use to identify consumers through their shopping habits. Etzioni (1999) presented an issue concerning a consumer who, while using a credit card, frequently purchased liquor at a supermarket. During one visit, the consumer fell resulting in knee surgery. The consumer sued the supermarket, but the supermarket attempted to argue, based on past purchases linked to the consumer through credit card transactions, that the consumer was at fault (Etzioni, p. 163). The consumer did not opt-in to data collected with use of a credit card, which the supermarket utilized to argue its case. This case brings into view the ethical issues that organizations and consumers have concerning data collection. Consumer preferred shopping cards are a method where consumers explicitly opt-in to an organization monitoring their purchasing habits. However, organizations have shown their willingness to use other methods, not obtaining an explicit opt-in from the consumer. The next issues involve the perspective of the consumer.

*The consumer perspective.*

Organizations gather consumer data by a variety of means, some already detailed. Once an organization collects consumer data, and successfully removes personally identifiable linkages between an individual and their data, the anonymized data can still determine specific traits associated with like members of a group through profiling (Hildebrandt, 2006). Hildebrandt discusses profiling from the perspective of the data and reveals a concept of Ambient Intelligence data mining. According to Hildebrandt, Ambient Intelligence seeks

statistical correlations without any preconceived notions of what it is looking at, or in a sense, it is autonomous data mining.

Hildebrandt (2006) suggests that profiling, through data mining, has advanced to a point where once an individual is assigned to a profile, likely characteristics of that individual are discernable. This raises an interesting perspective concerning data mining and profiling in that “...in the case of profiling we are not dealing with data, but with inferred knowledge” (Hildebrandt, p. 550). According to Hildebrandt, while discussing profiling states, “...profiles may reveal sophisticated knowledge about a person that is more intimate than sensitive personal data” (p. 550). Even though personally identifiable information is removed from the data, through the process of data mining, Hildebrandt suggests that personal data may still be derived from the data.

The research presented by Vedder (1999) suggests that data, which has all linkages to an individual person removed, still has the potential to harm that person. This can occur either through misapplying profiles or through accurately applying a profile and limiting products or services which the excluded member of a profile values. Da Costa, Schulte, and Singer (2006) are concerned precisely about this as they state, “...the important aspect is not the extra offers made to a specific group of people, but rather the limited choices for people outside the target group” (p. 82). Vedder concurs that categories of people can be identified and actions taken to the detriment of the group from which that person resides.

Vedder admits, “[t]he data used and the generalizations and profiles created do not always qualify as personal data” (p. 277). Nonetheless, Vedder presents a view that just because personal information is scrubbed from a data repository the ability to glean information from

massive databases and accurately apply to a class of people might need some sort of privacy protection.

Agre and Rotenberg (1997) discuss some uses that organizations have for their profiling activities. Some of these uses are:

Companies consult private databases containing consumer profiles to decide if individuals should be employed or should be allowed to purchase insurance, have credit, rent an apartment, open a bank or brokerage account, see a doctor, or engage in other common and basic human activities (Agre & Rotenberg, p. 210).

Profiling has the ability to touch many aspects of individuals' lives with decisions being made based on membership of one profile, or non-membership in another profile. The effect of referencing the resulting information allows businesses to select ideal candidates who they conduct business, limit choices available to specific individuals, or through attempts at focusing their marketing programs to specific individuals.

One common approach by privacy advocates is to limit the exposure of private information, essentially not releasing any private information wherever possible. Hildebrandt (2006) suggests that "Even though the protection of personal data can limit profiling by limiting the input of data, anonymisation [*sic*] will not limit but rather facilitate large scale group profiling" (p. 552). Therefore, consumer limiting information does not appear to be effective against profiling tactics.

Etzioni (1999) offers the view of triangulation where information is derived from multiple databases, comprising a picture more complete than any one database of information. Etzioni offers a suggestion that "...information that seems innocent at one point in time may turn

out to be highly sensitive at another, and hence all information should be treated as subject to the same rules” (p. 163). The recourse an individual has concerning membership in a profile is difficult if not impossible to change. No direct linkage to one’s personal information exists in the data source. Yet even though the linkage is missing to the person, through data mining, the person can be affected from the process of KDD. This makes sense from an organization’s perspective as organizations continue to collect and data mine the information.

Rygielski, Wang, and Yen (2002) suggest “Customer privacy can be better protected when customers do not have to reveal their identities and can remain anonymous even after the data mining probing” (p. 495). As an example, some services require the divulgement of some of the most personal information such as a social security number or personal medical information for entry into the military service. In this instance, personal medical, and social security information are required.

Consumers need the ability to understand what data is collected, what information is being derived from the data mining techniques, and to what purpose the data gets used (Vedder, 1999). Rygielski, Wang, and Yen (2002) discuss a Privacy Bill of Rights, which includes:

1. Fair access by individuals to their personal information
2. Responsible linkage of online and off-line information
3. Suitable criteria for opt-in and opt-out privacy options
4. Standardizing the disclosure to consumers of any existing privacy policy
5. Independent verification of implementation and execution of privacy and security policies

6. Fair mechanisms for resolving disputes by a trusted third party (Rygielski, Wang, & Yen, p. 495)

Perhaps if consumers were made aware of the where and when data is collected, the use of the data, and the life of the data, privacy concern could be minimize. Vedder (1999) believes that

...the data subject has some specific rights with regard to ‘his or her’ personal data. Among these rights are the right of access (knowing what data is being stored and whether the data relating to the data subject are being processed), the right of rectification, the right to know to who the data has been disclosed, and the right to object to the processing of the data relating to the subject (p. 276)

Hildebrandt (2006) informs that legally once the data is made anonymous, by removing any linkages to an individual, “...citizens have no legal right to even access the knowledge that is inferred from these anonymised [*sic*] data and may be used in ways that impact their lives” (p. 550). However, when a profile is once again linked to an individual, as in credit scoring, the individual has some legal protection (Hildebrandt, 2006, p. 550). This leaves the consumer in a relatively powerless situation. Even though personal information is not included in the data, personally identifiable characteristics are derived from the data through profiling activities (Hildebrandt, 2006; Vedder, 1999)

*The consumer and privacy.*

Whitman and Mattord (2005) detail that when it comes to protecting personal data, organizations “...should ensure that this data receives at least the same level of protection that the other important data in the organization—such as intellectual property, strategic planning,

and other business-critical data—does” (p. 482). While privacy advocates suggest not divulging any information, how can the individual know the security placed on the systems that house their data? Agre and Rotenberg (1997) specifically believe that data-protection laws and officials are valuable for privacy (p. 171). Laws and privacy officials may be the only effective method to ensure that private data, once collected, is kept secure. In most cases, the consumer will supply the information, whether through opt-in methods like a consumer preference card, or using a credit card. There is little recourse for the consumer but to link their purchasing habits through their purchases, unless their purchases are made with cash at random locations purchasing random items. This seems extreme and impractical.

Firestone (2005) details some concerns about data mining in cases where the data “donor” does not intend the information for use in a data mining fashion. Specifically, when a consumer agrees to give information to an organization what limitations are imposed on the data? Firestone states, “Privacy is a concern within the commercial and governmental sectors. For example, will data used for one purpose be reused for purposes to which the donor never agreed?” (pp. 50-51).

Privacy conversations, in an attempt to present the range of possibilities, usually present a vision of Big Brother, as delivered in Orwell’s 1949 work titled *1984* (Lindberg, 2002, p. 964). This view of Big Brother, an extremist centralized grip of monitoring and collection of personal information, persevered through the 1990’s fueling privacy advocates case. Unlike the centralized view of Orwell’s Big Brother, from the 1960’s through the 1970’s the data collected on individuals remained unconnected in central repositories, for the most part available only to the collectors of information (Agre & Rotenberg, 1997, p. 103). With the availability of

interconnection facilitated by the Internet, Agre and Rotenberg state, “No longer can personal information be said to be located anywhere” (p. 103).

The locus of the threat to privacy has inverted from an Orwellian “Big Brother”, which is a centralized threat to privacy, to the ubiquitous and non-centralized threat to privacy through the explosion of technology and the usefulness of data collected by multiple and diverse organizations. Agre and Rotenberg (1997) highlight this fact by succinctly stating:

One’s privacy is now less threatened by the omniscient gaze of a centralized ‘Big Brother’ than by the unknown and unseen collection, matching, and profiling of transactional data, a trail of which is left by every one of us as we purchase goods, apply for services, make entertainment choices, and so on” (p. 103).

While the deterioration of privacy continues to concern privacy advocates, the focus of the concern has changed. This requires a different stance with respect to privacy arguments, which focused on one centralized entity, such as Orwell’s “Big Brother”. Da Costa, Schulte, and Singer (2006) agree stating, “Today, at least in the case of the United States, this metaphor is less useful and even misleading in the description of contemporary surveillance societies” (p. 79).

Privacy advocates need a different character representative of current privacy concerns.

Currently information is gathered and used within the same organization that gathers the information. In some cases, brokers gather information and organizations purchase the information from the brokers. ACNielsen fills this void by collecting data on consumers who may purchase items through a variety of retail locations. ACNielsen completes the picture of consumer purchasing habits left incomplete through lack of preferred consumer cards, credit cards, or other methods used to identify consumers and their habits. ACNielsen (n.d.) enables

organizations by keeping track of consumers purchases, stating that "...consumers can find the same product in many different outlets, [and] complete retail coverage is essential to help manufacturers develop the right marketing and sales programs, and to help retailers understand their competitive positions" (para. 2).

Little research indicates organizational sharing information between similarly positioned organizations beyond purchasing it from an organization like ACNielsen. In some instances research has indicated that the different methods organizations use to store their data inhibits data sharing. As an example, Firestone (2005) cites Seifert who states, "Interoperability among platforms and databases used by different organizations and agencies remains a concern..." (p. 52). However, this trend is addressable with technology, and is only a matter of time.

*The consumer and the ethical ramifications.*

Vedder (1999) suggests an ethical perspective that "Once the data has become anonymous, or has been processed and generalized, an individual cannot exert any influence on the processing of the data at all" (p. 277). While the data can identify likely individual traits of a consumer, the individual cannot influence what is done with the data. Vedder realizes this issue stating, "The rights and requirements make no sense regarding anonymous data and group profiles" (p. 277).

Vedder (1999) presents a view where data warehousing, data mining and interpretation comprise the three phases in Knowledge Discovery in Databases (KDD). An organization must employ a strategy throughout each of these phases, which generally direct its actions. Legal requirements discern the collection of data, but once individually identifiable information is removed from the data, the organization is free to handle the data, as it deems appropriate.

Once the information about an individual is within an organization, according to Agre and Rotenberg (1997), "...personal information is used inside organizations in ways that are not visible to the outside world or to the subject of the information" (p. 211). This gives consumers no way to know what information is collected, what organizations are doing with the information, and no way for consumers to react to incorrect or erroneous information, which may negatively affect the individual.

According to Etzioni (1999), the ethical aspects of data mining will require a compromise between organizations and consumers. Consumers simply cannot keep everything secret, nor should they have to be concerned about organizations profiling their purchasing habits. Organizations need to self regulate their data mining activities. Etzioni suggests that this may not be enough stating "...additional mechanisms that draw largely on institutional changes, as well as some new legislation, are needed" (p. 163). This area needs further investigation in future research.

Future trends in data mining.

The main perspective in this literature review focused on organizational data mining techniques concerning a consumer habits for business interests. Although business interests are important, other organizations exist which are driven by different sources of motivation, and apply different strategies for data mining. Cabena et al. (1998) suggest that future application areas for data mining might apply the technology to text mining and web analytics. Examples in the business and consumer context include analyzing customer complaint letters through text mining to identify trends, or analyzing web server logs in an attempt to identify what a users' intent was concerning their initial visit to the site (Cabena et al., p. 36).

Firestone (2005) presents a view of the future of data mining stating “This wave, of course, is that of distributed multi-agent-based data mining” (p. 52). Potentially combating the future vision of Firestone, Hildebrandt (2006), looking for consumer transparency as to where their information goes, calls for Privacy Enhancing Technologies (PETs). PETs restrict the flow of information about people’s online activities and provide a mechanism for consumers to understand where their data went.

The issue of data mining, deriving detailed information from consumer habits and trends is not a technological issue, as it has proven successful and applicable for businesses to achieve a competitive intelligence. This competitive intelligence offers an advantage over their competition. Rygielski, Wang, and Yen (2002) state, “It is businesses and managers who determine how to exploit collected data, in other words, more of a policy issue than a technology issue” (p. 500). The methods used to exploit data mining determine how consumers are affected. However, mismanagement of consumer data will also affect the organizational strategy.

### Conclusion

Data mining has many useful applications such as minimizing fraud or targeting specific consumers (Cabena et al., 1998). Data mining is useful for organizations to fulfill the needs of the customer, through targeting high value customers, presenting organizations the ability to deliver products people want at a premium. On the other hand, technology enables organizations to collect data, derive previously hidden information about consumers, and apply this information to gain a competitive advantage. Most of this occurs without consumer knowledge.

What recourse do consumers have? Automated agents are collecting information in a distributed manner and individuals have no control over what data is collected, what happens to

the data, or who has access to the data (Firestone, 2005). Consumers also have limited ability to correct any errors or limit access to this data (Agre & Rotenberg, 1997). Consumers can limit the availability of data by the best available option concerning data collection, which is opting-out of voluntary data collection programs. Da Costa, Schulte, Singer (2006) suggest subverting data collection methods to taint the data store. Specifically they state, “Some common strategies are paying with cash ... or using another customer’s loyalty card to add noise to the store database” (p. 85). The effectiveness of these strategies are not clear, but these strategies make obtaining consumer data more difficult for use in data mining, and potentially more expensive for organizations conducting this type of surveillance. Unfortunately, some of these strategies also remove the incentives available for the consumer through participating in these programs. There is a price for privacy.

This literature review has investigated the concept of data mining, from the organizational perspective, the individual perspective, and from the perspective of the data in light of the privacy and ethical issues that each is presented through the collection of data. Automated data collection devices ensure that a massive amount of data is becoming easily available. The future of data mining continues to become more affordable to smaller organizations as the cost of obtaining, storing, and processing the data is decreasing in price. Organizations utilizing data mining techniques find the results more attractive for their organization. The consequences of data mining are shaping the choices available to consumers with little effective recourse available to consumers except to opt-out when they can.

## References

- ACNielsen. (n.d.). Retail measurement services. Retrieved November 27, 2008, from <http://www2.acnielsen.com/products/rms.shtml>
- Agre, P. E., & Rotenberg, M. (1997). *Technology and privacy: The new landscape*. Cambridge: The MIT Press.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering data mining: From concept to implementation*. Upper Saddle River, NJ: Prentice Hall PTR.
- Cooper, D. R., & Schindler, P. S. (2008). *Business research methods* (10th ed.): Irwin/McGraw-Hill New York.
- Cunningham, C., Song, I.-Y., & Chen, P. P. (2006). Data warehouse design to support customer relationship management analyses. *Journal of Database Management*, 17(2), 62.
- Da Costa, B., Schulte, J., & Singer, B. (2006). Surveillance creep! New manifestations of data surveillance at the beginning of the twenty-first century. *Radical History Review*(95), 70-88.
- Etzioni, A. (1999). *The limits of privacy*. New York: Basic Books.
- Firestone, J. M. (2005). Mining for information gold. *Information Management Journal*, 39(5), 47-52.
- Fresh Brands Inc. (2006). Piggly Wiggly Preferred Club Card. Retrieved November 19, 2008, from <http://www.shopthepig.com/>
- Hildebrandt, M. (2006). Profiling: From data to knowledge. *Datenschutz und Datensicherheit-DuD*, 30(9), 548-552.

- Hoffer, J. A., George, J. F., & Valacich, J. S. (2008). *Modern systems analysis and design* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Landry, R. J., Debreceeny, R., & Gray, G. L. (2004). Grab your picks and shovels! There's gold in your data. *Strategic Finance*, 85(7), 24.
- Lindberg, C. (Ed.) (2002) *The Oxford American College Dictionary*. New York: G. P. Putnam's Sons.
- Moore, G. (1965). Cramming more components onto integrated circuits. *Electronics* 38(8). Retrieved January 19, 2008 from <http://download.intel.com/research/silicon/moorespaper.pdf>
- Morris, R. J. T., & Truskowski, B. J. (2003). The evolution of storage systems. *IBM Systems Journal*, 42(2), 205.
- Rygielski, C., Wang, J.-C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in Society*, 24(4), 483-502.
- Seifert, J. W. (2008). *Data mining and homeland security: An overview* (No. RL31798).
- Smith, A. D. (2004). Knowledge management strategies: A multi-case study. *Journal of Knowledge Management*, 8(3), 6.
- Vedder, A. (1999). KDD: The challenge to individualism. *Ethics and Information Technology*, 1(4), 275.
- Ward, J., & Peppard, J. (2002). *Strategic planning for information systems* (3rd ed.). New York: John Wiley & Sons, Ltd.

Weiser, M., & Gold, R. (1999). The origins of ubiquitous computing research at PARC in the late 1980s. *IBM Systems Journal*, 38(4), 693. Retrieved February 17, 2008, from Academic Search Premier database.

Whitman, M. E., & Mattord, H. J. (2005). *Principles of information security* (2nd ed.). Boston: Thompson Course Technology Press.