1020 Brock Road South, Suite 2008
Pickering, ON L1W 3H2
Phone: (905) 837-0005
Fax: (905) 837-2199
www.boirefillergroup.com

By Richard Boire March 2005

A Practioner's Viewpoint on Data Mining & Privacy – Part 2

In the last article, I discussed how privacy has impacted the world of data mining. The discussion, without providing formal legal advice, conveyed some of the issues on how data miners have practically been impacted by the privacy legislation. Once again, as stated in the last article, these issues and thoughts are not meant to be construed as legal advice but rather as opinions from one data miner who has been lucky enough to have been a practitioner for a long period of time. Recognizing that this is one person's opinion, there will always be debate amongst data miners on almost any issue and debate certainly provides the forum for the optimal exchange of ideas. I suspect that this is also the case with privacy and its impact on data mining and I encourage you to express your viewpoint both within this newspaper as well as other media vehicles.

Although there are 10 guiding principles within the legislation, I focused on three which I felt most impacted the area of data mining:

- Identifying Purposes or Use of Information
- Consent
- Security or Safeguards

Much of the discussion in the last article dealt with the first principle of identifying purposes or use of information. I attempted to express the viewpoint that, in many ways, the use of data mining on customer data is not as privacy contentious as might be expressed by some of the leading privacy experts. The premise of my argument was that the use of mathematics and science essentially allows business users to make decisions on groups and not individuals. The reality of any data mining analysis is that business users will apply data mining results at an individual level, yet recognizing that the insights and learning regarding these decisions are based on the results and performance of a group of individuals.

The remainder of the article will deal with principles of consent and security or safeguards. Consent can appear to be relatively straightforward in terms of whether or not the customer gave permission to use his or her information. However, this principle is often more complex since it ties in very closely with how a given company will use the information. For instance, a company promoting products and services to its existing customers versus renting or selling its existing customer names to a third party represent very distinct business activities. One might argue that the former activity of continued marketing promotion to existing customers requires less restrictive means of

1020 Brock Road South, Suite 2008
Pickering, ON L1W 3H2
Phone: (905) 837-0005
Fax: (905) 837-2199
www.boirefillergroup.com

obtaining customer consent than the latter activity of renting or selling the customer's name to a third party. In fact, even within the activity of renting or selling names to a third party, there may be different levels of obtaining customer consent. Certainly, renting a name to a business publication would indeed be very different than renting a name to a charity. Yet, the reality in many cases for businesses regarding the renting and selling of names is that it is no longer an acceptable business practice. For those companies that do continue to conduct this practice, the consent requirement might be very stringent which usually means that the company requires an opt-in type consent as opposed to an opt-out type consent. Other restrictions might also indicate the specific organizations that would be allowed to receive the customer's name and address.

Opt-in vs. Opt-Out

The notion of opt-in vs. opt-out type consent represents a very important distinction in terms of gaining customer consent. In opt-in type consent, the customer or individual has to initiate some activity before customer consent is deemed to be obtained. This is often a check-box within some type of offer or communication piece which is being promoted by the company. The customer has to fill in the check box before customer consent is given. A blank check box is not deemed as customer consent. In other cases, the use of a questionnaire to obtain additional customer information might contain a clause which states that filling in the questionnaire is recognition of implied consent to use both questionnaire and behavioural information for future marketing purpose.

From a marketing standpoint, the opt-in type option for obtaining customer consent is very restrictive for most organizations and would severely impact their ability to conduct their business in a profitable manner. As we all are aware, many people will do nothing when it comes to receiving promotions. It does not mean that they do not want to receive additional products or services. It simply means that they don't care about filling in the check box. With an opt-in type clause, the net eligible universe available for marketing would be significantly smaller such that many marketers would not be able to take advantage of the economies of scale that are readily available with these larger universes. Many of these companies would cease activities that in the past have been very profitable and at the same time provided value and service to the consuming public. Because these activities are no longer profitable, jobs would be lost as well as additional consuming dollars which of course would certainly not help our economy.

Because of the deleterious effect of opt-in clauses, many organizations have attempted to deal with obtaining consent using the so-called opt-out provision. In the opt-out provision, the customer has to actually fill in a check box in order to state that consent is not given. A blank response to this check box is recognized as implied customer consent. The key for companies which choose to use opt-out consent is that this consent opt-out clause should be clearly visible and prominently displayed within the overall text. Any attempt to bury the clause in small print and within the middle or end of the text defeats the purpose of the overall privacy legislation. Furthermore, for those organizations that pride themselves as marketing experts, clearly defined opt-out type

1020 Brock Road South, Suite 2008
Pickering, ON L1W 3H2
Phone: (905) 837-0005
Fax: (905) 837-2199
www.boirefillergroup.com

clauses simply represent best business practices that will help to further cement the relationship of the customer with the organization.

Stating the Purpose

Within the actual consent clause, the text should be clear to the purpose of the consent. For example, if the purpose is to use the information to either rent or sell a given name to a third party for marketing purposes, then this activity should be clearly stated in the clause. Yet, if the purpose is to use the information to help sell other type of services and products to customers within the same organization, then that again should be clearly stated. However, some advocates might argue that more details should be provided to consumers regarding how this information should be used. For example, I have heard some opinions expressing the notion that the consumer be informed that mathematical and statistical techniques will be used to help analyze their information. My initial response to this opinion is "Do we really want to go there?" Do we really want to explain the intricacies of multiple regression and multivariate techniques to the public at large? Despite this being an arduous task, however, it may be something that we should do simply because it is the right thing to do? But how do we know what is right. Right in many cases with regards to the legislation attempts to look at reasonableness. That is the crux of the matter. Is it reasonable for consumers to know precisely how you are analyzing their data? What is reasonable is that consumers do want to know how you will use the information. They want to know how it might be used in other marketing activities. They also expect that future marketing activities which would be using this information are somehow related to the initial marketing activity in which they gave their consent. For example, it is reasonable to expect that a credit card customer having giving consent could be offered insurance to protect his overall credit balance in case of a job loss. Yet, it is unreasonable for a donor of a non profit company to expect to receive a promotion for a new credit card.

Reasonableness is a grey area and certainly the test of reasonableness is implicit throughout the legislation. It is why many pundits believe that this legislation is still work-in-progress which means that we will all gain greater knowledge and insights within this area as we acquire more practical experience in the application of the law.

Regarding this issue of using data mining tools and statistics to analyze data, does the consumer really care? If these tools are used to select consumers, then it might be argued that the consumer cares not about how he is being selected but rather that he might be selected for particular services and products. The use of these tools to aid in selecting a certain individual is of no interest to that consumer as long as he or she understands that they could potentially be selected for a particular offer or promotion. This might be the threshold of reasonableness today for the consumer. But, as stated above, as we gain more experience within the application of the law, the threshold of reasonableness might change to the point that consumers need to be informed about how this information will be analyzed. If that is the case, then I can see the book "Statistics for Dummies" becoming a best seller.

1020 Brock Road South, Suite 2008
Pickering, ON L1W 3H2
Phone: (905) 837-0005
Fax: (905) 837-2199
www.boirefillergroup.com

Channel Consideration

In customer consent, another key consideration is the type of vehicle or channel that is being used to promote products and services. For example, do direct mail, email, and outbound telemarketing all require the same level of customer consent? This is a question which is hotly debated today. Perhaps, different thresholds should be used which are dependant on the channel. Some channels are perceived to be more intrusive than others. For example, receiving a direct mail piece is certainly going to be less intrusive than receiving a phone call just after supper time.

Public Domain Information

Another issue in this debate is the use of what is commonly referred to as public domain information. In a nutshell, public domain information represents data that is available to the public anytime. For example, the use of StatsCan postal walk or Census data represents information that is readily available to the public. Yet, this information is only accessible at an aggregate level such as a postal walk or enumeration area. Within this context, the issue of customer consent is somewhat moot here since individual-level information is simply not available.

But there are sources of data that are at the individual level and which are still considered public domain. For example, mortgage registry information is public-domain information. Let us suppose that a bank were to obtain this list and market mortgage products to these names. Would this activity be in contravention of the law? Certainly, the information is publicly available. However, this is where the issue of reasonableness comes into play. Did the consumer upon filling out this registration form have a reasonable expectation that this information might be used by a third party to promote its services and products? In this particular scenario, the reasonableness test might lead us to a negative response to the previous question. In fact, no best practicing organization would use this data for the simple reason that the risk of damaging its reputation amongst consumers far outweighs the value of this information.

Data Security and Storage

As data represents the primary tool of the data miner, the security of this data or information as outlined in the legislation is of acute interest to the data miner. The protection of data that is entrusted to us is not an option but rather an obligation. Data miners are constantly exposed to the transfer of data either in receiving or sending data. During this transfer of data, the data miner should certainly be aware of the sensitivity of the information. This sensitivity will dictate the various protocols and procedures which are required in any data transfer. For example, files might be sent containing very rich demographic and behavioural information along with customer number as the key link between these files. Assessing the security risk here, we can determine that if information is lost or stolen, we essentially have a file of numbers along with demographic and behavioural information that is now available to the public. However, the reality is that no actions could be taken here which could potentially harm these consumers. Under this scenario, the appropriate protocol might be to use email with the actual information contained within a password-protected file.

1020 Brock Road South, Suite 2008
Pickering, ON L1W 3H2
Phone: (905) 837-0005
Fax: (905) 837-2199
www.boirefillergroup.com

But suppose that name and address are now being transferred instead of customer number and suppose we need to use name and address to create a match key in order to link the different files. Here, the use of an FTP site which allows companies to post data while giving the necessary instructions for the receiving company provides a much higher level of security. At the same time, the company posting this data would also want to make sure that its Internet firewalls provide superior protection than just the regular transmission of emails. The company receiving this data would also want to ensure that its firewalls are acceptable for this type of highly sensitive data. The data itself, which is transferred via FTP, would again be password-protected, thereby providing another layer of security. Another option besides the use of FTP in transferring this data is to simply burn the information onto a CD and transfer it via courier. Once again, the information on the CD would be password-protected.

In rare cases, organizations particularly credit card institutions will send data containing credit card information. The reason for this is that the credit card info represents the key information in linking all the files which may be sent to the data miner. Under this scenario, which is obviously the most sensitive from a security standpoint, organizations will use the FTP process as outlined above but also scramble or encrypt the actual credit card number. It is the actual encrypted card number that the data miner needs to use when linking files.

Another mechanism for ensuring that data is in a secure environment is to have proper storage procedures. Data miners, as a rule, would like to keep data as long as possible and have it readily accessible to them. But the reality is that there is always a certain degree of risk in keeping data for extended periods of time, especially data containing name and address. Because data miners deal with data that is both used for creating and applying a solution, they also need this data when validating the solution that has been applied to a campaign. For instance, the name and address information of a direct mail campaign using an applied data mining solution is required in order to link it to back to the responders of that campaign, which would also contain name and address information. The rule of thumb in most of these cases is usually six months which means that the expectation is that a backend or performance analysis should be done within six months of the campaign launch. If files are older than six months, procedures should be in place to either archive this data to physical media such as tapes or cartridges or perhaps transfer it to a server which is offline and not connected to the network.

As we all know, privacy is a hot button topic where people will have strong opinions on many of the principles and guidelines as outlined in PIPEDA. The various seminars that I have attended on this topic speak to both the interest as well as the varying opinions on some of the legislation. The legislation in many cases is very clear on what marketers as well as data miners can and cannot do. But in some cases, there are shades of grey when terms such as reasonableness are used. For the most part, marketers and data miners are highly respectful of the legislation since the legislation

1020 Brock Road South, Suite 2008
Pickering, ON L1W 3H2
Phone: (905) 837-0005
Fax: (905) 837-2199
www.boirefillergroup.com

itself has been a best practice by many organizations for many years. It is simply a good business practice to be respectful and attentive to the privacy needs of consumers if we are ever going to properly market the right services and products to them.

Richard Boire is a Partner with the Boire Filler Group, a database marketing consulting company that specializes in developing and implementing data mining strategies. He can be contacted at (905) 837-0005 or via e-mail at RichB@BoireFillerGroup.com.