12-31-2002

Privacy and Customer Data Quality: exploring the Issues

Martin Gibbs
University of Melbourne

Graeme Shanks
University of Melbourne

Reeva Lederman
University of Melbourne

Roselle De Silva
University of Melbourne

Privacy and Customer Data Quality: exploring the issues

Martin R. Gibbs

Graeme Shanks

Reeva Lederman

Roselle De Silva

Department of Information Systems
The University of Melbourne
Parkville, Australia
m.gibbs@dis.unimelb.edu.au

Abstract

New privacy protection legislation has recently been enacted in Australia. The Privacy Amendment (Private Sector) Act 2000 regulates the way private sector organisations can collect, use, keep secure and disclose personal information. As such, it has required many organisations to change the ways in which they handle personal information. The ability of organisations to respond to the requirements of this new legislation is affected by the quality of their customer data. This paper explores the issues created by poor customer data quality for organisations as they adjust their business practices to meet the provisions of the new privacy legislation. A number of key issues emerge including managing large amounts of fragmented customer data, understanding what information is required for organisational activities, controlling use and disclosure across the organisation, keeping track of pre- and post- December 2001 data, and allowing anonymity when interacting with customers.

Keywords

Privacy, Customer Data Quality, Information Systems Management

INTRODUCTION

There is no general legal right to privacy in Australia. Australian common law does not recognise the right to privacy (*Victoria Park Racing and Recreation Grounds Co. Ltd. V Taylor* 1937, 58 CLR 479) and the Australian constitution is silent on the issue. Currently, an individual's privacy is protected through the operation of a patchwork of Federal, State and Territory legislation, as well as the operation of some aspects of the common law of contract, torts and confidential information. Until recently there has been a notable and glaring gap in this patchwork of protection. There has been no unified and overarching legislation in Australia to protect the personal information of individuals held by private sector organisations (AGD, 1999). While the Commonwealth Privacy Act 1988 (Cth) (Privacy Act) regulated how government agencies handled personal information there was no similar regulation covering the private sector. The new Commonwealth privacy law in the form of the Privacy Amendment (Private Sector) Act 2000 (Cth) (Private Sector Amendment) came into effect on the 21st of December 2001 in order to remedy this situation. The Private Sector Amendment extended the Privacy Act to the private and health care sectors and has given individuals the legal right to access and correct information held about them by private sector organisations for the first time (OFPC, 2001a). These new private sector provisions have extended the Privacy Act to regulate the ways in which large, private sector organisations can collect, store, use and disclose personal information. As such, the Act presents a number of challenges to organisations that collect, use and distribute personal information. Meeting these challenges has required significant changes to the ways in which organisations handle personal information and has created new responsibilities for organisations to maintain the data quality of the personal information they hold.

The Private Sector Amendment applies to the vast majority of private sector organisations with annual turnovers of $3 million or more and to organisations that provide health services or hold health related information. It also applies to organisations with smaller annual

turnovers that trade in personal information. The main repository of personal information in organisations is customer databases. Previous studies have shown that maintaining consistently high levels of customer data quality is a significant challenge and considerable expense for organisations (Redman, 1998; Strong *et al*., 1997; Wang, 1998). In this paper we argue that the ability of organisations to comply with the provisions of the Privacy Amendment Act will be significantly influenced by the data quality of the personal information they hold. In particular, poor customer data quality will create a number of serious problems for organisations in this regard. The connection between poor customer data quality and privacy is one that has not been explored in any detail previously, yet is clearly of great importance to Australian organisations.

We use an exploratory research study to identify the key issues created by poor customer data quality that face organisations as they adjust their business practices to meet the provisions of the new privacy legislation. These issues have great significance for organisations and should serve as a 'wake-up call' for those private sector organisations that are somewhat remiss in maintaining the quality of their customer data quality. In addition, our findings have significance for organisations that extend beyond our national borders and beyond the reach of the Australian legislation we are considering in this paper. As the Australian privacy legislation is based on the OECD's 1980 *Guidelines on the Protection of Privacy and the Transboarder Flows of Personal Information,* the issues we identify should not be thought to be limited to the Australian context but are generalisable to all organisations that must comply with privacy laws derived from these OECD principles.

This paper first outlines the history, purpose and provisions of the Private Sector Amendment and provides a brief description of the relevant privacy principles. The following section describes a semiotic framework for understanding data quality (Shanks and Darke, 1998), and relates relevant privacy principles to data quality dimensions. We follow this with a discussion of the research approach used in this exploratory study. The next section includes a detailed discussion of five key issues that emerged from the study concerning how poor customer data quality impacts the ability of organisations to comply with the provisions of the Private Sector Amendments. We observe that although the explicitly stated intention of the Private Sector Amendment is to 'give people some control over the way information about them is handled' (OFPC, 2001b), poor customer data quality degrades an organisation's ability to effectively manage and use the information it holds. This degradation in the control over the personal information an organisation possesses, undermines its ability to cede some of that control back to the individuals concerned with serious implications for the protection of their information privacy. A number of important implications for practice and directions for future research conclude the paper.

BACKGROUND

The Privacy Amendment (Private Sector) Act 2000

Definitions of privacy abound. Most definitions of privacy invoke one or more of the following rights: the rights to act anonymously, the right to live free of unwanted harassment, and the right for individuals to chose how they present themselves to the wider community. When discussing privacy and information technologies, the last of these listed rights is often restated as the right for individuals to control the access others have to their personal information. Given that a number of possible definitions of privacy have common currency it is interesting to note that neither the original Privacy Act, nor its more recent amendments, explicitly define privacy. However, they do provide sets of 'privacy' principles for the protection of personal information held by government and private sector organisations respectively and legally oblige non-exempt organisation to abide by these principles. In addition, it has been explicitly stated on numerous occasions that the primary focus and goal of this legislation is to give individuals some control over the personal information held about them by government and private sector organisations (FPC, 2000; OFPC, 2001a; 2001b). As such, it is useful to understand the definition of privacy implicit within the Privacy Act and its later amendment as pertaining solely to 'information privac'. Roger Clarke has usefully defined information privacy in the following way:

> *Information privacy refers to the claims of individuals that data about themselves should generally not be available to other individuals and organisations, and that, where data is possessed by another party, the individual must be able to exercise a substantial degree of control over that data and its use.*
>
> (Clarke, 1999)

The Privacy Act, including the new private sector provisions, regulates the way personal information is collected, stored and used. Personal information is defined within the Privacy Act as:

> *Information or an opinion (including information or an opinion forming part of a database), whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion.*
>
> (The Privacy Act, 1988 (Cth), Sect 6)

The new laws cover non-government organisations – including limited and listed companies, partnerships, trusts and individuals operating a business – with an annual turnover of more that $3 million. In addition, organisations with a smaller turnover that provide a health service or hold health records are also obliged to comply with the provisions of the new legislation, as are organisations that trade in personal information. There are, however, some notable and controversial exceptions to the amendments for organisations such as small businesses, political parties, media organisations, and in relation to employee records.

The Private Sector Amendments established ten National Privacy Principles (NPPs) as the minimum standard for information privacy in the private sector. The legislation was designed to encourage organisations to develop their own code of conduct regarding information privacy, which, once approved by the Federal Privacy Commissioner, would set the standards for the handling of personal information by the organisation. In the absence of an approved code, all non-exempt organisations are obliged to comply with the NPPs.

The NPPs govern how an organisation should handle personal information. They cover: collection (NPP1); use and disclosure (NPP2); data quality (NPP3); data security (NPP4); openness (NPP5); access and correction (NPP6); use of government identifiers (NPP7); anonymity (NPP8); transborder data flows (NPP9); and sensitive information (NPP10). In relation to the current study concerning the impact of poor customer data quality, five of these NPPs are relevant: they are briefly elaborated below.

NPP1 – Collection An organisation must only collect personal information that is necessary for its activities and collection must be fair, lawful and not intrusive.

NPP2 – Use and disclosure An organisation must only use or disclose personal information for the primary purposes it was collected. Use or disclosure of personal information for secondary purposes is only permissible when consent has been given or the secondary purpose is related to the primary purpose and it is reasonable to expect such a secondary use or disclosure to occur. However, personal information may be used for the secondary purpose of direct marketing without the individuals prior consent provided the individual is given the opportunity to 'opt-out' of future marketing campaigns.

NPP3 – Data quality Reasonable steps must be taken to ensure that the personal information an organisation collects, uses or discloses is accurate, complete and up-to-date.

NPP6 – Access and correction An organisation must give individuals access to their personal information if requested and they must correct that information if it is inaccurate, incomplete or out-of-date. In addition, an organisation may charge for providing access to personal information if those charges are not excessive.

NPP8 – Anonymity Where possible, organisations must give individuals the option to remain anonymous during any contact or transaction with the organisation.

The Private Sector Amendment is one of several recent measures introduced by the Australian Government to facilitate Australia's transition to an information economy (DCITA, 2000). The new provisions have been implemented with the aim of balancing individual's

rights for information privacy against the 'right of government and business to achieve their objectives in an efficient way' (FPC, 2000:2). In the spirit of promoting a 'culture that respects privacy' (FPC, 2000:2), privacy has been promoted as being 'good business' (OFPC, 2001c) as well as being good for individuals. In particular, the legislation has been implemented with the recognition that consumers 'lack of trust' in the way commercial organisations handle their personal information is a major barrier to the growth of e-Commerce (OFPC, 2001c). The legislation is also an attempt to bring Australia into line with international privacy regimes especially those of the European Union (EU) given the possibility that the EU will impose trade restrictions on nations that do not adequately protect the personal information of EU citizens.

In developing the provisions of the Private Sector Amendment the government has deliberately opted for a 'light-touch' co-regulatory approach to the regulation of privacy with the aim of encouraging compliance through facilitation rather than through the threat of punitive actions for non-compliance (OFPC, 2001d; 2001e). This approach has been designed to minimise the burden of compliance for businesses. It is also an approach that has attracted strong criticism and has led to the amendments being dubbed 'anti-privacy laws' (Roger Clarke quoted in Haslam and Mitchell, 2001) and described as 'reducing existing privacy protection' (Clarke, 2000) due to the large number of exceptions and qualifications built into the legislation and because it seemingly 'legitimises many unreasonable uses of personal data' (Clarke, 2000). The legislation has also be criticised for lacking in 'grunt' and being 'toothless' due to the Federal Privacy Commissioner not being granted with significant investigative powers or an ability to impose significant punitive penalties for breaches of the Privacy Act (McClelland in Australia, House of Representatives, 2000:2223-7). While these criticisms are significant and have a significant bearing on how private sector organisations have responded to the new privacy provisions, they are not the major focus of this paper. Rather, we wish to discuss the unexplored issues associated with how poor data quality in personal information will affect an organisation's ability to comply with the provisions set out in the Privacy Act.

Customer Data Quality Management

Customer information is increasingly viewed by organisations as an important asset that can be used to deliver competitive advantage and support business initiatives that focus on the customer. Accordingly, it is crucial that the collection, storage and use of customer data is properly managed within organisations. A key aspect of customer data management is to ensure that the data is of high quality (Shanks and Tay, 2001).

A customer may be defined as a person or organisation that is engaged in legal contract with an organisation. There are three main categories of customer information: fundamental, demographic and behavioural. Fundamental customer data includes properties such as the customer's name, address, postcode and telephone number. Demographic data includes properties such as the socio-economic, location and age profile of customers, especially in relation to census data. Behavioural data includes properties related to the customer's interaction with organisations including products and services previously purchased and records of interaction. Customer data may be used to support the operational activities of an organisation and also for strategic marketing and sales activities. An important objective of customer data management is to achieve a 'whole of customer view'. This involves being able to consolidate customer data from numerous legacy databases. Consolidating this customer data is a very difficult task mainly due to data quality problems.

We define quality as 'fitness for purpose'. Much of the existing work on data quality focuses on the intrinsic quality of data in databases and consists of lists of desirable information quality dimensions (Wand and Wang, 1996). These lists typically include dimensions such as completeness, accuracy, reliability, consistency, timeliness, precision and conciseness. Several frameworks have been developed that organise and structure important concepts in information quality (see for example Wand and Wang, 1996; Wang and Strong, 1996; Kahn *et al.*, 2002). In this paper, we use the framework of Shanks and Darke (1998). This framework is soundly based in semiotic theory and includes both product-oriented and service-oriented aspects of data quality. *Semiotics* is the study of the use of symbols to

convey knowledge and suggests four discrete levels of data quality: syntactic, semantic, pragmatic and social (Stamper, 1992).

Syntactic data quality is concerned with the structure of data and is focused on form rather than content. The goal of syntactic data quality is consistency of representation in one or more databases. For customer data, this includes consistent representation of such things as names, addresses and telephone numbers as well as consistent representation of codes that are used to identify and categorise customers. Customer data that is inconsistent can lead to serious difficulties in consolidating customer data from various legacy systems into a data warehouse for customer relationship management purposes.

Semantic data quality is concerned with the meaning of customer data as assigned by users of the information. The goals of semantic data quality are that data is complete and accurate and up-to-date. For customer data, completeness means that data is recorded for each customer of the organisation. Data can be incomplete if there is either no data at all about a customer or parts of the data about a particular customer is missing. Accuracy of customer data means that data recorded about a specific customer is correct and matches the customer in the business world at a particular point in time. Up-to-date is a measure of how recently data values have been updated in a database. Customer data that is incomplete, inaccurate or not up-to-date can lead to many problems including extra costs, lost business opportunities, offended customers and frustrated employees.

Pragmatic data quality is concerned with the use of customer data, and varies with the person involved, the task at hand and the organisational context. The goals of pragmatic data quality are usefulness and usability. For customer data, usefulness is concerned with the degree to which the data supports a person in accomplishing a task. Usability refers to the degree to which the customer data may be accessed and used effectively. For customer data to be useful and usable, it should be easy to access, timely, easily understood, and concise.

Social data quality concerns the shared understanding of data by various social groups within organisations or societies (Shanks and Corbitt, 1999). It is particularly relevant for large, multinational organisations and organisations developing global inter-organisational electronic commerce systems where cultural differences become important. It is also relevant for organisations that have multiple points of contact with customers. The goals of social data quality are shared understanding of meaning and awareness of bias among different users. These goals are crucial in understanding the viewpoints of different users and the consolidation of data from different legacy systems. Using customer data without a shared understanding of its meaning and awareness of bias may lead to problems in correctly interpreting reports based on the data and problems with combining data from multiple sources.

Managing customer data quality involves understanding and measuring data quality problems and designing improvement strategies for both existing data stocks and incoming data flows (English, 1999). Managers are increasingly asking for clear business benefits to be realised from expenditure on fixing data quality problems. Improving customer data quality leads to greater customer dissatisfaction (correct names and addresses, accurate billing, filled orders and receipt of appropriate marketing materials), decreased operational costs (less time and other resources spent detecting and correcting errors), more effective decision making (accessing and using relevant and accurate information) and increased employee satisfaction (greater trust of information in databases) (English, 1999; Redman, 2001; Wand and Wang, 1996). Increased data quality should also lead to greater compliance with the Privacy Amendment Act.

Research Question

NPP3 of the Private Sector Amendments sets out expectations for the maintenance of data quality and requires organisations to ensure that the personal information it collects, uses or discloses is accurate, complete and up-to-date. That is, data quality is defined solely in terms of its semantic properties. However, if we adopt a broader, semiotic view that understands quality as ‘fitness for purpose’, it is possible to see that data quality extends beyond the semantic dimension and has implications for many of the other requirements set

out in the new legislation. That is, each relevant NPP may be related to one or more data quality dimensions. NPP1 (Collection) constrains an organisation to only collect personal information that is necessary for its activities. This is related to completeness at the semantic data quality level. Poor completeness indicates either missing data or excess data that is not required. This will violate NPP1. NPP2 (use and disclosure) constrains an organisation to only use or disclose personal information for the primary purposes it was collected unless the individual concerned has given their consent. This is related to usefulness at the pragmatic data quality level. Data that is useful will support the activities it was intended to support. NPP3 (data quality) ensures that the personal information an organisation collects, uses or discloses is accurate, complete and up-to-date. This is the complete set of data quality dimensions at the semantic data quality level. NPP6 (access and correction) concerns giving individuals access to their personal information if requested and allowing them to correct it. This is related to the accessibility data quality dimension at the pragmatic data quality level. NPP8 (anonymity) concerns giving individuals the option to remain anonymous during any contact or transaction with the organisation. This is related to the syntactic data quality level and concerns the representation of codes and identifiers.

Clearly, data quality issues are strongly related to the principles found in the Privacy Act. This research project will explore that relationship by identifying the data quality issues that occur in practice as organisations come to grips with the recently introduced Private Sector Amendments. The research question addressed in this study is:

> *What difficulties associated with data quality are organisations experiencing in their attempts to fulfil their obligations to maintain the information privacy of individuals as set out in the Privacy Act?*

RESEARCH APPROACH

This research study was exploratory in nature and involved two main phases: a conceptual study and then an empirical study. The conceptual study phase of the research included an extensive review and synthesis of the Privacy Act legislation, associated submissions to parliament, press commentary and other relevant literature from both academic and practitioner sources. This was then synthesised with concepts from the semiotic framework for understanding data quality in order to develop an initial understanding of how poor customer data quality may prevent organisations from fulfilling their obligations to maintain the information privacy of individuals, and to develop an interview protocol for data collection in the empirical phase of the research.

The empirical phase of the study involved in-depth interviews with eight experienced practitioners. Interviewees were identified opportunistically and selection for interview was based on the criteria that they had extensive experience with privacy and data quality issues. Five interviewees had data management roles in organisations that handled large amounts of customer data in different industry sectors. The other three were consultants specialising in the areas of privacy and/ or data management. Empirical data was conducted through semi-structured interviews and review of documents contributed by interviewees. Interview duration ranged from 60 to 90 minutes. All interviews were recorded on audiotape and fully transcribed. Qualitative data analysis techniques (Miles and Huberman, 1994) were used to identify key issues associated with data quality faced by organisations as they responded to the provisions of the Act. Interview transcripts were initially analysed independently by each author. These were then compared and any inconsistencies reconciled to produce the final list of issues.

KEY ISSUES EMERGING FROM THE STUDY

A number of key issues emerge from the data analysis. These include:

Managing and Permitting Access to Large Amounts of Fragmented Customer Data

Many organisations have a history of separate business units developing and maintaining independent customer databases. Typically these legacy systems have been developed autonomously and use different data structures and identifiers to record personal information. In addition, these databases are often ‘owned and operated’ by separate

functional units within the organisation. Consequently, the personal information an organisation holds about individuals is fragmented across multiple and heterogenous databases. This makes accessing and collating personal information difficult and time-consuming. Rarely in these cases is there a unified and consolidated view of the personal information an organisation holds about an individual (Shanks, 1997).

One of our interviewees, the information systems manager for a large metropolitan teaching hospital, said that locating and identifying all the databases within the organisation that contained identifiable personal information was a major problem for her organisation's ability to comply with the new privacy legislation. While the paper-based patient record recorded all treatment that patients received within the hospital, various units within the hospital also maintained their own, separate records for a variety of purposes associated with research, treatment and service evaluation as well as for the purposes of providing a health service to the patient. This organisation had approximately 30 different function units that collected and used personal information; many of them using more than one information system to so. While a portion of these information systems were modest in scale – spreadsheet applications and small databases – the difficulties faced by the organisation as a whole in compiling a view of the totality of personal information held about any one individual, are obvious. This degree of fragmentation creates serious pragmatic data quality problems and has severe implications for this organisation's ability to respond in a timely and efficient manner to an individual's request to access their personal information as required by NPP6 (Access and Correction).

Two other interviewees who worked for large retail organisations also reported problems of a similar nature. Although the problems were on a smaller scale, the organisations they worked for were grappling with similar issues associated with multiple and fragmented databases and the difficulties this situation presented in forming a unified whole of customer view. They anticipated that this situation would create some difficulties with providing customers with access to all the personal information held about them and would impede their organisation's ability to strictly comply with NPP6 (Access and Correction).

It is also worth noting that the legislation allows organisations to charge individuals for providing access to their personal information, although 'excessive' charges are prohibited by the legislation. However, the definition of 'excessive' in this context has yet to be established (Sinclair, 2002a). Recent press reports indicate that some large organisations will use an hourly-rate to calculate these charges on a cost-recovery basis. For example, the four major Australian banks are charging between $25 and $70 an hour for accessing personal information (Sinclair, 2002a). Poor data quality at the pragmatic level created by the fragmentation of personal information across different databases will impact on the time it takes to recover personal information and hence the charges made by an organisation on a cost-recovery basis. Charges made on this basis are unlikely to be deemed 'excessive' unless fragrantly extravagant. This creates the possibility that customers exercising their right of access and correction will end up paying a high premium for this right as a result of an organisation's poor customer data quality.

Understanding what Personal Information is Required for Organisational Activities

Under NPP1 (Collection), organisations may only collect personal information if it is required for a specific function or activity. Collecting personal information 'just in case' it is needed at some future point in time is no longer permissible. Limiting the collection of information in this way has been regarded as good information management practice for some time and is strongly associated with an organisation's ability to maintain its customer data quality. It is well known that data that is collected but not used, or not connected to a functional area of the organisation will tend to be of poor quality and/ or will degrade in quality very quickly (Orr, 1996).

However, many organisations have difficulty with identifying precisely what personal information is and is not necessary for their ongoing functions and activities. In addition, there is a prevalent tendency to strive for data quality at the semantic level of completeness by collecting a broad range of personal information about customers. One interviewee with many years experience consulting in the areas of information systems management and

privacy observed that many firms she'd worked with had trouble identifying the personal information they actually needed to fulfil their business activities:

> *It's interesting when you go and work in these organisations because you think they'll be logical and they'll be structured and they'll collect things sensibly but when it comes to information, they don't have any clear way to pin down exactly what information they need… The approach has tended to be ' Let's collect a whole lot, just in case, and then we're pretty well covered if, you know…*

Striving for completeness in customer data can easily and quickly lead to poor data quality across its other dimensions. Remembering that quality is defined as 'fitness for purpose' it is easy to see why this would be the case. Customer data that has no purpose is, by definition, of poor quality. That is to say, when looked at pragmatically, data that has no purpose is not useful and hence represents an unnecessary burden and cost to organisations. In addition, personal information that is not connected to a ongoing purpose or activity is like to be of poor quality at the semantic levels of being accurate and up-to-date for two reasons. Firstly, data collection that is unconnected to a current organisational function tends to be of low quality because there is no motivation to maintain strong quality control at the point of collection. Secondly, data quality problems of these kinds with personal information are typically identified and corrected when that data is actually used such as when a transaction is being completed with the customer.

The challenges faced by organisations with undisciplined collection practices in complying with the provisions of the new legislation will be to determine the precise purposes that motivate the collection of personal information. Having made this determination, they will need to change their data collection practices and only collect personal information that is required for specific activities or functions. Thus we can observe that complying with these requirements of the Private Sector Amendments will have a positive impact on data quality by encouraging good information management practices.

Controlling Use and Disclosure Across the Organisation

One interviewee, a consultant from a large accountancy firm with a number of years experience in performing privacy audits as well as providing consulting services to public and private sector organisations, made the following observation:

> *I think some organisations get quite a surprise when they actually look at how they use information and who they disclose it to. Whereas they might think it's all relatively under control, suddenly – you know, you might have a department collecting [information] in a structurally diverse organisation with different people sharing across different quarters – it becomes a big issue because, generally, no one knows what everyone else is doing with the information.*

NPP2 (Use and Disclosure) requires organisations to only use or disclose personal information for the primary or related secondary purposes for which it was collected and/ or for which consent has been obtained. In this regard the legislation is very clear; use of personal information must be restricted to those purposes the individual has been informed about and to which they have consented. However, large organisations will commonly use customer data for a variety of purposes and sharing of this personal information across an organisation's functional units or business lines is often necessary. If different units in an organisation are to use a common customer data set, they must maintain a shared understanding of what is, and is not, an appropriate use or disclosure of the personal information it contains. Data quality at this social level of shared understanding and common interpretation must also be maintained for the total duration that each piece of personal information is stored and used by the organisation; a duration that could easily extend for many years or decades.

Therefore, in order to comply with NPP2 (Uses and Disclosures), it will be necessary for organisations to tag personal information held in customer data sets with its approved and allowable uses. This is particularly important when customers are giving consent for certain

uses and not for others. It also important when different functional units are using customer data for unrelated purposes. The cost of reprogramming information systems to allow for this tagging is estimated by some industry analysts as costing large retail and financial service firms millions of dollars (Sinclair, 2002b).

An important exemption built into NPP2 (Uses and Disclosures) allows organisations to use non-sensitive personal information for the secondary purpose of direct marketing without an individual's prior consent provided the organisation gives the individual an opportunity to 'opt-out' of future marketing campaigns and it is impractical for the organisation to seek consent from an individual beforehand. Maintaining a record of customers that have and have not, opted-out of direct marketing campaigns can pose a number of difficulties for large organisations. As the chief compliance officer at Westpac, Susan Brooks, has commented:

> *We have around 1,000 locations around the country, so we have to establish a flag on any account if that customer comes in and says they don't want to receive any marketing material. That might sound easy but it has required a significant change so that our systems speak to each other across the group, including AGC, Westpac, and Westpac Financial services.*
>
> (Quoted in Gluyas, 2001)

Establishing a flag for this purpose in the organisation's customer databases is one viable option provided the technical capacity for an expansion of this nature exists. However, some organisations do not possess this capacity. Also, some large organisations wish to keep this kind of marketing information separate from their customer databases for a variety of reasons. One of our interviewees, the CIO for a large retail company, reported that his organisation maintained opt-out lists on a PC-based system separate to their mainframe-based customer database and used the opt-out lists to cleanse their customer lists before mounting a direct marketing campaign. Another interviewee who consulted to large companies on privacy issues also suggested that this kind of practice was quite common in industry. These kinds of practices raise a number of issues for the maintenance of customer data quality particularly if this opt-out list is maintained within a single department, such as marketing, and general access is not possible to staff who have direct customer contact such as front of house and call centre staff. The real risk is that unless this opt-out information can be updated in an easy and timely manner, it will get lost or improperly recorded, with implications for the organisation if individuals who have opted-out of direct marketing campaigns continue to receive this type of material.

Keeping Track of Pre- and Post-21 December 2001 Data

A number of NPPs do not apply to personal information collected before the date the legislation came into effect (21 December, 2001). In particular NPP2 (Use and Disclosure) does not apply and organisations may continue to use this personal information for a wide variety of primary and secondary purposes without the consent of the individual (OFPC, 2001f). However, all aspects of the Private Sector Amendment, including NPP2, will apply to new personal information that is added to a customer record as well as existing personal information that is updated. This creates a situation in which the personal information within a single database, and even within a single customer record, is subject to different levels of protection depending on whether or not the information was entered into the database before or after the date the legislation came into effect. The main reason for allowing pre-existing databases to be treated less stringently under the new legislation was because it was thought that requiring all NPPs to apply to existing data would '…impose unjustifiably high compliance costs on business, and these costs may well be passed on to the consumer' (AGD, 2000). For an organisation to comply with the requirements of the legislation but still not pay an 'unjustifiably high compliance cost' they must be able to differentiate pre- and post-21 December personal information. That is, they must be able to store the dates of additions and changes to personal information in their customer records, or at minimum, flag those changes that occurred post-21 December. Ideally, they must be able to do this at the level of the individual data field. As one of our interviewees pointed out, for a small proportion of organisations this kind of data tagging is simply not possible given the inherent constraints and capacities of their customer record systems. Where it is

possible, many organisations will need to reprogram their information systems to achieve this level of data tagging at considerable expense.

Allowing Anonymity when Interacting with Customers

Where possible, organisations must give individuals the option to remain anonymous during any contact or transaction they have with the organisation. For example, it should be possible for an individual to telephone an insurance company to obtain an estimate for house and contents insurance without having to provide any identifying information such as a name, telephone number or street address although non-identifying information such as postcode and value of goods to be insured might need to be provided.

However, for many organisations, conducting anonymous transactions with members of the public is difficult due to constraints built into their transactional information systems. According to one of our interviewees involved in privacy consultancy work:

> *Some organisations have been saying,' we can' t because of our systems constraints. We can't actually [have anonymous transactions], we have to go through certain identification fields before we can provide information.'*

Many information systems require personal information to be entered into mandatory data fields before a transaction can be processed and information or a similar service can be provided. Thus, in the example above, in order to obtain a quote for insurance, it might be necessary for a call-centre operator to complete a number of data fields with identifying information before the system will provide an insurance estimate. Of course, one common workaround often used to circumvent these kinds of information systems constraints – constraints that have often been built in deliberately in the first place in an attempt to improve data quality by ensuring completeness – is for operators to enter dummy data in order to move through the system to gain the required information. These kinds of practice can serious degrade customer data quality. While existing customer data quality does not, in itself, directly affect an organisation's ability to comply with NPP8 (Anonymity), information systems constraints designed to preserve semantic data quality can impede an organisation's ability and willingness to comply with the anonymity requirements of the Private Sector Amendments.

IMPLICATIONS FOR PRACTICE AND FURTHER RESEARCH

Implications for Practice

A number of important implications for practitioners emerge from this study. First, the Private Sector Amendments have been designed to improve the information privacy of individuals by 'giving them some control' over how their personal information is used by private sector organisations. However, in order to 'give control' organisations must have control over this information in the first place. As we have shown in this paper, poor customer data quality severely undermines the ability of organisations to control the personal information they hold about individuals and inhibits their ability to comply with the new legislation.

Second, it would seem that most legislation aimed at protecting information privacy is based on the assumption that organisations have an integrated customer data set and that it is relatively easy to access, collect and collate all the personal information they hold about an individual. The reality is quite different for most organisations. These organisations cannot readily achieve the whole of customer view necessary for strict compliance with the provisions of the Privacy Act due to problems with their customer data quality.

Third, this problem is particularly pernicious for organisations with multiple points of customer contact. These organisations are often characterised by semi-autonomous functional units that have been in the habit of amassing their own customer databases without reference to a centrally coordinated information management strategy. As a result, the sum total of personal information held about any individual is fragmented across multiple and incompatible databases creating significant data quality problems that severely hinder the formation of a unified and integrated whole of customer view. This inability to develop an integrated whole of customer view compromises an organisation's ability to effectively

manage its customer data and hence compromises its ability to it meet its obligations under the Privacy Act. These organisations will need to exert strong control over the ways in which their function units collect and manage personal information if they are to improve their customer data quality sufficiently to achieve a whole of customer view and comply with the Privacy Act.

Fourth, the ability to develop a unified and integrated, whole of customer view about an individual enables organisations to unproblematically comply with the provisions of the Privacy Act. Establishing and maintaining high levels of customer data quality across all four data quality levels is an important part of developing this kind of view of the personal information held about a particular individual. It is ironic to note that it is precisely these kinds of information systems that use good quality, highly integrated databases of personal information that have raised the hackles, suspicions and fears of privacy advocates and political commentators for several decades due to their ability to data mine and match data from multiple sources (see for example Clarke, 1988; Davies, 1997). Yet, it is precisely those organisations with highly integrated and carefully managed customer data that are in the best position to comply with the provisions of the new legislation.

Fifth, complying with new legislation always has the potential to be an expensive activity for business. Reports and commentaries to date suggest that while many organisations have put some effort into developing privacy policies and statements, they have expended far less effort than expected in terms of updating their information systems to ensure they are technically capable of compliance with the provisions of the Act. As we have pointed out, the new legislation will require many organisations to make significant and expensive changes to their information management systems including the introduction of new tags to track authorised uses and disclosures, customer consent, and dates of information accession as well as other initiative for the maintenance of data quality. Establishing good levels of customer data quality is a necessary and potential expensive step that any organisation must take in order to be privacy compliant.

Further Research

A number of suggestions for further research emerge from this study. First, our results indicate that the Privacy Act should have a positive impact on data quality. That is, privacy legislation can be seen as promoting good data quality practices, and the benefits organisation can derive from these practices could extend well beyond merely being compliant with legislative provisions. This suggests a line of inquiry to investigate how the new privacy legislation is encouraging organisations to improve their data quality and the subsequent 'spin-off' benefits they are deriving from these activities. An investigation of how the Privacy Act is changing the information management practices of organisations and whether additional benefits are accrued from these changes warrants further investigation. This form of investigation would best be pursued through detailed case studies.

Second, as a result of this study, we are able to suggest that further research is need in the policy arena. Legislation to protect personal information tends to be based on the assumption that organisations have an integrated set of customer data they can access, collate and modify in a relatively unproblematic fashion. However, this kind of unified whole of customer view is by no means common. The implications this has for the formulation of effective policy and legislation for the protection of information privacy needs to be more fully explored. In addition, within the Privacy Act data quality is defined solely in terms of its semantic properties. By adopting a broader, semiotic understanding of data quality we have been able to show that serious data quality issues exceed this narrow definition. Our study suggests that more research is required to support the development of policy and legislation that takes into account a more sophisticated understanding of data quality.

Third, as we have we argued in this paper, the ability of an organisation to comply with the Privacy Act crucially depends on its ability to maintain fitness for purpose of the personal information it holds about its customers. It should be pointed out that this so-called 'fitness' is itself transformed by the provisions of the legislation. That is, poor customer data quality not only affects an organisation's ability to comply with the provisions of the legislation but the legislation itself actually changes what counts as good or poor data quality. Fitness for purpose is transformed by the requirements in the amendments relating to collection, use

and disclosure, access and correction, and anonymity. This insight suggests that our semiotically based framework for understanding data quality needs to be revisited and revised to include a new data quality dimension that acknowledges the deep and intimate link between information privacy and data quality. Information privacy could thus be put forward as a new dimension of pragmatic data quality related to the usefulness and useability of customer data.

CONCLUSION

It is clear that there is still a great deal of work to be done in terms of privacy. Prior to the legislation coming into effect in December of last year, many industry analysts and commentators were predicting that it would be 'next big thing' for the Australian IT industry following the massive amounts of work generated by Y2K and the introduction of the GST. While the Privacy Act has been a boon for legal firms and accountancy consultancies it not had had the expected impact on the Australian IT industry. Indeed, it would seem that while many organisations have spent some time and effort on the window dressing of privacy policies and disclaimers they have done little to change their underlying processes and to redesign their data infrastructure to deal with the required changes to they way in which they handle personal information. Given the issues identified in this study, we believe this signals a significant problem. In this context it is interesting to note that Alan Fels and the Australian Competition and Consumer Commission (ACCC) have recently stated their willingness to prosecute organisations under the Trade Practices Act for making deceptive statements and misleading advertising if they do not meet the standards for privacy declared in their own privacy policies. This may well act as a wake-up call for organisations that have adopted privacy policies and issued privacy statements to their customers that they are unable to comply with due to poor levels of customer data quality. We may yet see more organisations take their data quality issues seriously in the future.

The Private Sector Amendments have been criticised by a number of commentators for the range of exemptions they contain and for 'light-touch' approach to enforcement adopted by government. Nevertheless, it cannot be denied that the Privacy Act is has advanced information privacy protection in Australia by requiring large organisations formulate privacy policies and to take reasonable steps to protect the information privacy of individuals. While the legislation may be weak on enforcement, it is still in an organisation's best interests to comply with the legislation. Privacy issues consistently rank highly in consumers' list of concerns (Clarke, 1997) and have been attributed as one of the factors limiting the growth of e-Commerce (OFPC, 2001g). Complying with the principles not only prevents possible damage to reputation resulting from non-compliance but can also help develop much needed trust between private sector organisations and their customers. In addition, the new legislation has succeeded in raising awareness of information privacy issues amongst IS practitioners as well as in the broader community and has quite visibly started a cultural shift in the private sector towards a culture that respects and values information privacy.

REFERENCES

Attorney General's Department (AGD) (1999) The Government's Proposed Legislation for the Protection of Privacy in the Private Sector, AGPS, Canberra.

Attorney General's Department (AGD) (2000) Fact Sheets from Attorney General's Department Privacy Law and Policy Reporter 7, URL http://www.austlii.edu.au/au/journals/PLPR/2000/7.html Accessed 12 April 2002.

Australia, House of Representatives (2000) Parliamentary Debates, 6 November 2000.Clarke, R. (2000) Privacy Bill Needs Much More Work, The Australian, 15 February, URL http://www.acs.org.au/news/oz150200.html Accessed 22 Jan 2001.

Clarke, R. (1988) Information Technology and Dataveillance, Communications of the ACM, 31, 498-512.

Clarke, R. (1997) What Do People Really Think? MasterCard's Survey of the Australian Public's Attitudes to Privacy, URL http://www.anu.edu.au/people/Roger.Clarke/DV/MCardSurvey.html Accessed 21 Jan 2002.

Clarke, R. (1999) Internet Privacy Concerns Confirm the Case for Intervention, Communications of the ACM, 42, 60-67.

Davies, S. (1997) Re-Engineering the Right to Privacy: How Privacy Has Been Transformed from a Right to a Commodity in P. Agre and M. Rotenberg (eds) Technology and Privacy: The New Landscape, MIT Press, Cambridge Mass.

Department of Communications Information Technology and the Arts (DCITA) (2000) Submission from the Department of Communications Information Technology and the Art to the Senate Select Committee on Information Technologies Inquiry into e-Privacy – July 2000.

English, L. (1999) Improving Data Warehouse and Business Information Quality, Wiley Computer Publishing, New York.

Federal Privacy Commissioner (FPC) (2000) Submission from the Federal Privacy Commissioner to the House of Representatives Standing Committee on Legal and Constitutional Affairs Inquiry Into the Privacy Amendment (Private Sector) Bill 2000, URL http//:www.privacy.gov.au Accessed 10 April 2002.

Gluyas, R. (2001) Rush is on for Privacy Policies. Australian IT, 10 December, URL http://austalianit.news.com.au Accessed 2 Jan 2002.

Haslem, B. and Mitchell, S. (2001) New Laws 'Complex and Full of Holes' Australian IT, 21 December, URL http://austalianit.news.com.au Accessed 2 Jan 2002.

Kahn, B., Strong, D.M. and Wang, R.Y. (2002) Information Quality Benchmarks: Product and Service Performance, Communications of the ACM, 45(4): 184-192.

Miles, M.B. and A.M. Huberman (1994) Qualitative Data Analysis: An Expanded Source Book, 2nd edn, Sage Publications, Thousand Oaks.

Office of the Federal Privacy Commissioner (OFPC) (2001a) Information Sheet 1-2001: Overview of the Private Sector Provisions, URL http//:www.privacy.gov.au Accessed 11 Dec 2001.

Office of the Federal Privacy Commissioner (OFPC) (2001b) Information Sheet 2-2001: Preparing for 21 December 2001, URL http//:www.privacy.gov.au Accessed 11 Dec 2001.

Office of the Federal Privacy Commissioner (OFPC) (2001c) Good Privacy, Good Business: Privacy in Australia, AGPS, Canberra, URL http//:www.privacy.gov.au Accessed 11 Dec 2001.

Office of the Federal Privacy Commissioner (OFPC) (2001d) Information Sheet 13-2001: The Privacy Commissioner's Approach to Promoting Compliance with the Privacy Act, URL http//:www.privacy.gov.au Accessed 11 Dec 2001.

Office of the Federal Privacy Commissioner (OFPC) (2001e) 'Implementation of the Privacy Amendment (Private Sector) Act 200', Privacy Law and Policy Reporter, 8 URL http://www.austlii.edu.au/au/journals/PLPR/2001/3.html Accessed12 April 2002.

Office of the Federal Privacy Commissioner (OFPC) (2001f) Information Sheet 10-2001: Application of the Privacy Act to Information Already Held, URL http//:www.privacy.gov.au Accessed 11 Dec 2001.

Office of the Federal Privacy Commissioner (OFPC) (2001g) Privacy and the Community, AGPS, Canberra, URL http//:www.privacy.gov.au Accessed 9 March 2002.

Orr, K. (1996) Data Quality and Systems Theory, Proceedings of the International Conference on Information Quality, R. Wang (ed) MIT, Boston, November.

Redman, T. (1998) The Impact of Poor Data Quality on the Typical Enterprise, Communications of the ACM, 41(2), 79-82.

Redman, T. (2001) Data Quality: The Field Guide, Digital Press, New Jersey.

Shanks, G. (1997) The Challenges of Strategic Data Planning in Practice: An Interpretive Case Study, Journal of Strategic Information Systems, 69-90.

Shanks, G. and Corbitt, B. (1999) Understanding Data Quality: Social and Cultural Aspects, Proceedings of the 10th Australasian Conference on Information Systems, Wellington, December.

Shanks, G. and Darke, P. (1998) Understanding Data Quality in Data Warehousing: A Semiotic Approach, Proceedings of the International Conference on Information Quality, I.Chengilar-Smith and L. Pipino (eds) MIT, Boston, November.

Shanks, G. and Tay, E. (2001) The Role of Knowledge Management in Moving to a Customer-focused Organisation, Proc. European Conference on Information Systems, Bled, June

Sinclair, J. (2002a) The Charge to See, The Age, 18 February, Money Manager Suppl. 3.

Sinclair, J. (2002b) Eyes Wide Open, The Age, 9 April, Next Suppl. 7.

Stamper, R. (1992) Signs, Organisations, Norms and Information Systems, Proceedings of the 3rd Australian Conference on Information Systems, Wollongong, December.

Strong, D.M., Lee, Y.W. and Wang, R.Y. (1997) Data Quality in Context, Communications of the. ACM, 40(5), 103-110.

Wand, Y. and Wang, R. (1996) Anchoring Data Quality Dimensions in Ontological Foundations, Communications of the ACM, 39(11), 86-95.

Wang, R.Y. (1998) A Product Perspective on Total Data Quality Management, Communications of the ACM, 41(2), 58-65.

Wang, Y. and Strong. D.M. (1996) Beyond Accuracy: What Data Quality Means to Data Consumers, Journal of Management Information Systems, 12:4, 5-34

COPYRIGHT