Database Privacy

Balancing Confidentiality, Integrity and Availability

Martin S Olivier
Computer Science
University of Pretoria
Pretoria, South Africa
http://mo.co.za

ABSTRACT

The emphasis in database privacy should fall on a balance between confidentiality, integrity and availability of personal data, rather than on confidentiality alone. This balance should not necessarily be a trade-off, but should take into account the sensitive nature of the data being stored and attempt to increase all three dimensions to the highest level possible.

To achieve such a balance, technological means should be developed.

The paper illustrates some of the inherent problems in database privacy that should be addressed by technical solutions. It next demonstrates that the notion of privacy is complex; this complexity is likely to impede development of technical solutions.

Finally, the paper finally uses the notion of informed consent to illustrate how the privacy problem can be viewed from multiple angles to flesh out the underlying problems that may be addressed by technical solutions.

General Terms

Security

Keywords

Database privacy, personal data, confidentiality, integrity, availability, dataveillance

1. INTRODUCTION

For many people, when writing about privacy and computing, the temptation exists to focus only on the negative. It is simple to list many examples of cases where the use of computers were instrumental in some breach of an individual's privacy. From such examples and the sheer size of the global database collecting information about each of us, it is often concluded that the age of an Orwellian Big Brother has now arrived and personal privacy has forever been lost.

On the other side of the coin are the proponents of technology who enthuse about new technology, citing the many obvious advantages it has. However, in many of these cases they are either oblivious of the debate about privacy or see those who are concerned about privacy as uninformed scaremongers.

As is often the case when looking at such problems, the truth lies somewhere between the extremes. While it is not possible to pinpoint the appropriate spot between the extremes (since this will be influenced by time, location, culture and other factors) debate is required to establish some range of acceptable use of private information.

This is a well-known phenomenon in security: security is widely regarded as a balance between confidentiality, integrity and availability. Without the need for availability, the confidentiality problem is trivially solved by 'unplugging' the database. In the privacy sphere, the availability dimension has mostly been ignored when technical solutions to the privacy problem has been sought. In this mode of thought, anonymity presents itself as the natural solution: if private data cannot be collected, it cannot be misused. While this is clearly an important aspect of privacy, it is (also clearly) not applicable to a major class of privacy problems: For many (most?) transactions that have an effect in real life, details of the transaction and transacting parties need to be recorded — hence *availability* of personal information becomes an issue. And, once personal information has been recorded, the confidentiality problem changes from one of ensuring that private information is not disclosed for recording to one that ensures that private information is not improperly disclosed from where it has been recorded in a database.

The challenge of database privacy is therefore to enable the storage of personal information in databases in a manner that balances society's needs with those of the individual, with particular emphasis on the vulnerability of the individual.

Here 'society's needs' should be understood in an inclusive manner. It includes society's need for accountability (as exemplified by unique number or licence plates on motor cars), the need for information for civil interaction (as exemplified by 'sharing' one's credit card number with a merchant from whom one buys goods) and even the needs of the individual (as exemplified by the individual who wants to be unique rather than blend in with the crowd).

This paper argues that the concept of privacy is often more complex than realised. The next section argues that privacy mechanisms should extend security because the problems have fundamental differences. Section 3 mentions a couple of historical milestones to show that database privacy has received attention in the past — before communications privacy started to become the major focus in the area. I next use a personal example to highlight a number of relevant issues — in particular the need for availability. Section 5 shows that the concept of privacy is problematic, while section 6 attempts to cast some light on how the underlying concepts of privacy can be identified. Section 7 concludes the paper by summarising the major challenges to be addressed in the field.

2. DATABASE PRIVACY AND DATABASE SECURITY

This paper deals with database privacy. Database privacy concerns the protection of information about individuals that is stored in a database. Database privacy can (and should) use solutions developed for database security. Both database security and database privacy are based on a balance of confidentiality, integrity and availability. However, database privacy differs inherently from database security in some key respects. To mention just three of these differences, consider the questions

1. What is (was) the intended use of stored data?
2. Who carries the risk if data is disclosed to an unauthorised party?
3. Why does someone need to know a *specific* piece of information?

The first question is typically irrelevant in the case of database security. Information is considered an asset that gives an organisation a competitive advantage over its competitors. The fact that data is not only to be used for the intended reason why it was collected, is supported by the emergence of technologies such as data mining and OLAP *(on-line analytical processing)*. Security is typically concerned about keeping sensitive information out of the hands of competitors and and ensuring that data is not modified in some manner that will cause the organisation to lose money or suffer embarrassment. In contrast to this, the *intended use* of personal information is a cornerstone of privacy. In the ideal case, individuals are informed exactly what the information collected about them will be used for and the information is subsequently used solely for that purpose.

The relevance of the second question can be demonstrated with an example: Suppose an organisation accidentally discloses some sensitive medical diagnosis about someone in their database. The organisation may be liable and may be forced to pay damages to the individual. In many cases the organisation will be ensured against such mishaps and the only real effect may be some embarrassment to the organisation. In contrast, this may have a profound effect on the individual concerned, who may be ostracised from his/her community and for whom reparation in the form of damages or other remedy may be of little value. We have, after all, only one life.

The third question alludes to the standard need-to-know principle employed in security: tight security mandates that someone should only be given access to information he/she needs to know to do his/her job. The problem appears when an employee *may* need to access an individual's information. This may happen when the individual phones to discuss an item on his/her account or other record, when the employee happens to process an order, account or other item related to the individual and in many simple scenarios. Making it difficult for the employee to access personal information of, say, customers may hamper efficiency (and even the individual's perception of the organisation because they seem not have the information readily available to provide proper customer service). The problem with giving people access to all individuals' information is that it becomes relatively hard to distinguish between work and mere browsing of individual information.

What these questions have illustrated are the facts that database privacy needs to consider the purpose(s) for which data was collected, verification of protection mechanisms by those who would stand to lose most if private information is disclosed or modified and protection of each specific individual's information. Note that these and similar issues are not solved by simplistic solutions, such as banning the collection of personal information in databases: not only is the recording of some such information required by the society we live in at the beginning of the twenty-first century, but recording of such information is, in many cases, beneficial to the individual. Both of these aspects will be considered in more detail below. Note that many very useful approaches (such as P3P [21], onion-routing [12], Crowds [22], LPWA [9] and others) exist to prevent recording of information in a database in cases where such recording can be avoided. Preventing recording, however, is not our concern here, but rather dealing with cases where recording in unavoidable (or has already occurred).

3. HISTORICAL MILESTONES

Database privacy has received significant interest over the past three decades (even though the term *database privacy* was not used in most cases). It is only with the emergence of the World-Wide Web in the 1990s that preservation of privacy during communications has begun to overshadow privacy protection of stored data. In some cases — notably P3P — the intention of communications privacy is to only communicate data with a server that, according to the server's privacy policy will treat the private data in a manner one is comfortable with. In many other cases, the goal of communications privacy is to ensure anonymity and/or ensure that communications cannot be tapped.

Computer databases have forever changed the landscape of protection of private information. Where previously information may have been recorded in a manual file system that made it labour-intensive to locate records and costly to store and copy them, data is now stored in a manner that is cheap, easy to make perfect copies and that can be searched quickly using very sophisticated queries.

This section lists some of the major milestones in database privacy — ie in the protection of stored personal data. Much of what has happened in the field rests on legal and societal norms that address misuse of personal information. While such restrictions are important, they are insufficient on their own and need to be augmented by technical controls.

Already in 1973 the *Code of Fair Information Practices* established some of the fundamental principles of database privacy used to this day, including the requirements that the existence of no database containing personal information should be secret, that a person should be able to determine, correct and/or amend information stored about him/her, and that precautions against misuse of data should be taken by those who work with such data [10, p.7].

One of the next major milestones was the publication of privacy principles by the OECD *(Organisation for Economic Co-operation and Development)* with the intention to "harmonize national privacy legislation and, while upholding human rights, [to] at the same time prevent interruptions in international flows of data" [4, p.74].

The European Data Protection Directive limits the transfer of information across national boundaries to countries that will impose similar restrictions on use of the data as European countries do. In some countries, recording of specific data, such as ethnic origin, religious or life convictions and health information is prohibited (unless intended for a few specified cases) [24].

One of the major threats to consider is the aggregation of personal data from multiple sources. Bennett (as quoted by Whitaker [26, pp.125, 138]) uses the term *dataveillance* to refer to the extent that "the surveillance practices that the massive collection and storage of vast quantities of personal data have facilitated." The potential to invade privacy by combining different data sources is also recognised by the US Computer Matching and Privacy Protection Act of 1988 (5 U.S.C. 552a(o) *et seq*) that restricts federal agencies' ability to match data collected from different sources.

One of the best-known privacy cases, in fact, involved the collection and planned sale of information about the buying patterns of millions of Americans by Lotus Development Corporation and Equifax [3, p.17][10, p.9][4, p.57]. These plans were cancelled after about 30 000 letters were sent to Lotus and Equifax to protest against the sale of this information.

Many of the more recent developments in this area have not concentrated primarily on placing more restrictions on what data should be recorded, how it should (not) be used and when collection should be prohibited. To the contrary, such legislation has often placed a duty on communications and other providers to collect information about subscribers' use of their systems so that law enforcement agencies can get access to such data when required.

4. A PERSONAL EXAMPLE

While writing this paper, we were planning a trip to the Netherlands. Immediately after receiving the last of our required documentation we duly applied for the required visas, which we were told should take a day. With more than a week remaining, this seemed like ample time. However, the embassy phoned my wife that same afternoon and told her that her application would take several days to several weeks to complete because she is a medical doctor. Apparently, they are afraid that a visiting doctor (with a tourist visa) would work as a doctor while there and take work away from a registered doctor.[1] The fact that my wife would only be in the country for 3.5 days (Saturday afternoon to Wednesday morning) made no difference to the bureaucracy. Doctors need special permission. Since we did not want to wager the significant costs we have already incurred for travel and accommodation on the possibility that it would take several days *rather* than several weeks, she cancelled her visa application and we amended our travel plans to spend more time in a country that do not consider doctors as a 'threat.' Obviously the application fee (equivalent to about 19 Big Mac™ burgers, to express costs in terms of international purchasing power) was not refunded.

This example illustrates three major points[2]: Firstly, any personal attribute can be the basis for 'special' treatment: in this case, special skills held undesirable possibilities for the authorities.[3]

This is in stark contrast to traditional privacy wisdom that usually holds that information about an individual (or personal attributes) can be categorised into categories that are more or less sensitive, with one's own medical diagnoses often more sensitive than, say, one's surname.[4]

The second major point illustrated by this example is something that we will return to below: the notion of *informed* consent. On a visa application one is indeed warned that your personal information may be communicated to other countries. No indication is given that specific jobs will be treated different from others. If one argues that it is implied by the fact that your trade or profession is asked, it probably follows that marital status, age, the names of your spouse and parents, country of birth and similar information can all be used in non-obvious ways. In fact, when I went to the Embassy to try and get access to the directive or instructions that earmark doctors for nonstandard treatment, I was told that the document was confidential. When I tried to get a reference or other number or name for that document (so that I could try to get hold of it via the Dutch Freedom of Information Act[5], I was told that no

[1]For the record, she planned to accompany me solely to take a break from her hectic work schedule in South Africa.

[2]The visa example also raises a fourth point. Why is it necessary that a South African medical doctor's visit needs to be preapproved via a longer — and costly — process while this is, presumably, not required in the case of a doctor from, say, the United States who wishes to visit the Netherlands? One can only assume that the huge disparity between salaries earned in the first and third world makes it more lucrative for a doctor from a third world country to quickly earn a few Euros. While it is true that living costs are lower in third world countries, medical equipment, books and other necessities are actually much more expensive in third world countries. This inevitably makes one think of Chomsky's insight (expressed by Fox [8]) that "on the ill-balanced scales of global business, the favoured Euroamerican élites must inevitably grow richer, while the rest of the world could revert to the conditions of Blake's 'dark Satanic Mills'." While this has profound privacy implications — because being African is a personal attribute and because the greater the disparities between first and third world countries become, the greater the 'threat' that citizens from third world countries will hold for the first world will become — this point will not be discussed further in the current paper.

[3]This reminds me of an anecdote I heard about someone who, when the previous South African Government was still in control, decided to study Russian simply because he was interested in languages. He soon realised that, presumably because the African National Council had ties with Russia, learning Russian was an undesirable skill in South Africa then, that brought his activities under close scrutiny from government.

[4]However, a surname often has religious, nationality, ethnic and other connotations. My surname, for example, is associated with the French Huguenots who fled France in the late 1500s and early 1600s when Protestants were persecuted by the French state. A surname with a religious (or other) connotation become sensitive in a privacy sense when it is associated with a group from which — rightly or wrongly — a current national or other threat is perceived. The connotation holds whether one is a member of the specific 'threatening' group or not.

[5]Wet Openbaarheid van Bestuur, Wet van 31 oktober 1991

such identifying number existed.
The third — and perhaps most significant — point illustrated by this example, concerns controlled use of information one wants to share. If I used this example in years gone by, it would probably have been published in a paper-based form and distributed to technically inclined people. However, this paper is to be published on the Web, where it is accessible to the world. Conventional privacy wisdom holds that one should look at the privacy policy of whomever one gives information to and rather withhold it if one sees possible harm coming from the sharing. The 'privacy policy' for this paper is wide, public dissemination. The fact that I use an example that expresses an opinion that the procedure used by the Dutch government seems, at the least, silly, is a personal opinion that may be used against me later — I have no idea how willing national governments are to let foreigners who have criticised them in public visit their countries. Conventional privacy principles therefore dictate that I should rather not use this particular example. By implication, therefore traditional privacy limits freedom of speech. And this is a true challenge for database privacy: how can one enhance the goals of privacy with the least impact on freedom of speech? A concrete example may illustrate this point better. People routinely publish their e-mail addresses on websites with the intention that individuals should be able to contact them; however, those addresses are (usually) not intended for collection by spiders or bots to compile address lists that are subsequently used for spamming. Over the years, people have employed various mechanisms to be still able to publish e-mail addresses but withhold them from automated collectors. Initially, `mailto` links were dropped, because they unambiguously identify a piece of text as an e-mail address. Others are publishing their addresses with a portion that should be removed by a human before using it, eg `jsoap@somedomain.removethisbeforesending.edu`. Others are publishing the same address as `jsoap AT somedomain.edu` or, for lists of users, a webpage might say that the all e-mail addresses on the page should get an `@somedomain.edu` as a suffix and then list J Soap's e-mail address next to his name simply as `jsoap`. Yet others are using a (graphical) image to convey their e-mail addresses to humans in a manner that is hard to collect by bots. To take protection of e-mail addresses one step further, the standard format for articles in this publication includes an e-mail field; since I know the paper will be (widely) published I can opt-in or opt-out to include my e-mail address. In the first case, I lose control over how my personal information (e-mail address) is published; in the second case I make it hard for academics who would want to discuss any aspect of the topic with me (and the ease with which an author can be contacted probably has some impact — albeit small — on the number of times an author's work is cited, which again impacts on promotion, funding and other similar aspects of academic life). In the case of this particular article, I addressed the problem by giving a URL for a webpage from which my address can be obtained, and where I have control over the format in which the address is shared.
Note that the solutions given in the previous paragraph are limited because of the focus on one specific problem that is not indicative of the range of privacy problems to be addressed. Further, the solutions are all *ad hoc* meaning that do not generalise well to other privacy problems. Some more general approaches will be mentioned in section 7 below.
However, before considering technical solutions that will help balance confidentiality, integrity and availability of private data, the notion of *privacy* needs further attention. It will be clear that privacy as a concept is highly problematic and it is not possible to establish the required balance without fleshing the concept out somewhat.

5. PROBLEMS OF PRIVACY

The *Right to Privacy* is permeated with problems, such as the exact definition of *privacy*, whether it constitutes a fundamental right and whether people are and/or should be concerned about it.
Garfinkel [10, p.4] says privacy is "about self-possession, autonomy and integrity." According to Margalit [17, p.211] "self-respect and humiliation are based on a private space whose invasion is a symbolic act interpreted as humiliation, in the sense of lack of consideration for the victim's vital interests." Rosenberg [23, p.76] defines privacy as the "prevention of others from securing information about us that is immediately embarrassing (and so causes us pain) or of strategic value to others in their integration with us (and so imposes on us other material costs)." From these (and many other similar) statements it is clear that two concepts are central to the kind of privacy we are interested in in this article: *autonomy* and (the implied possibility of) *harm*: The right to privacy (if such a right exists) claims that individuals should have the greatest possible autonomy over information about themselves to avoid the harm that could be done if information about themselves were to become available to parties to whom they would not willingly give access to the information. Usually missing from definitions of privacy is a third aspect: the benefits that storage of personal data may hold for the individual, such as the possibility of improved customer service mentioned earlier.
However, beyond these two constants found in most definitions, confusion reigns. Rosenberg [23, p.76] argues that privacy may not be a right after all but a taste: "If privacy is in the end a matter of individual taste, then seeking a moral foundation for it — beyond its role in making social institutions possible that we happen to prize — will be no more fruitful than seeking a moral foundation for the taste for truffles." In 1758 Hume still thought it possible to establish a standard of taste [13]. Hume's argument is based on the existence of critics who are able to judge art and thereby establish this standard. Subsequent philosophers have denied the existence of such a standard, but have accepted the universal appeal that æsthetic judgment has: When one judges art as beautiful, Kant says, one judges as if others *ought* to also judge it as beautiful [14]. If privacy is indeed a matter of taste, rather than a right, it would explain society's difficulties to agree on exactly what privacy is. In that case much of what is commonly seen in the field of computing (including the last part of this paper) operates in the Humean fashion where 'experts' (critics) are trying to establish a standard that does not exist. And the attempts only seem realistic due to the universal nature of judgement of taste. Even though this is a core problem of privacy that should affect the manner in which it is dealt with in computing, it is not considered further in this paper.
If privacy is indeed a right, one should consider whether it is a fundamental human right (or a moral right), a property

right or some other right.

John Rawls's [20] theory of justice considers the *just* society and therefore can be used to think about fundamental human rights. The essence of his theory is that whatever people under a 'veil of ignorance' (in the so-called original position) would choose, would be just. In other words, if people, who do not know whether they will turn out to be the privileged or underprivileged in society, were to choose for or against privacy, what would they choose? That can then be considered just. Elsewhere [19] I have argued that his theory supports privacy. (This brief argument was presented to argue that a decrease in privacy levels for all to identify those who should be suspected of unlawful activities would be unjust.) In a similar vein, Garret [11] mentions some specific restrictions (or guidelines) that should be applied when collecting information about individuals and notes that such restrictions "would be endorsed by representative persons in the original position and included in the understanding of the Equal Liberties Principle that they would adopt to govern a just social order."

Rosenberg [23, pp.84–90], in apparent contrast, tacitly invokes Rawls to argue that a just society would *not* choose (absolute) privacy for medical information and credit reporting, since society has more to gain by forfeiting some privacy: By requiring medical information from applicants, medical insurance can be made affordable for people with average health; without it, medical insurance and hence medical treatment becomes prohibitively expensive for almost everyone. By collecting information about people who default on their loan repayments on blacklists, credit becomes possible in society; without such lists credit becomes prohibitively expensive with the implication that most people will not be able to buy expensive goods such as motorcars and houses.

This contrast makes is necessary to reconsider whether privacy is a fundamental human right. Note that privacy is indeed included in the bills of rights in a number of countries. The constitution of South Africa [1], as one example, entrenches this right. However, it is often difficult to relate such rights to database privacy. To illustrate, the clauses of the privacy right in the South African Constitution that may apply directly to database privacy are those that grant everyone not to have "their person ... searched" (§2(14)(a)), "their property searched" (§2(14)(b)) and "the privacy of their communications infringed" (§2(14)(d)). The first two will be considered again below, but are indeed problematic to apply in database privacy. The third clause does seem to be applicable: if one communicates private information with a second party (who stores it in a database) and the second party subsequently shares that information with a third party without one's permission, privacy has arguably been violated. The third clause does not, however, necessarily exclude violations of privacy by the second party itself (without involving any third parties).

The first two clauses quoted above tie in with the notion of privacy rights as property rights. One often sees claims that you *own* your name and other private information and therefore controls it. Branscomb [3], however, asks whether you own your name, address, telephone number, medical history and a list of related personal attributes.[6] In each case she comes to the conclusion that the information (for example, your name) is considered public knowledge, or that someone else (the post office, the doctor, etc) actually owns the information. You are only a stakeholder. And the fact that conferences can be held on the topic *"Who owns our genes?"* [18] clearly illustrates that it is far from clear that one can claim bodily integrity to protect personal attributes such as your name.

Moreover the fact that it is hard to treat privacy as an absolute right in any of the senses above, is painfully obvious. According to Rosenberg [23], if the right to privacy only becomes important once the relative value of private information compared to the costs to obtain it rises beyond some level, the right to privacy is a prudential right and not a moral right. Brin [4, p.14] supports the view that privacy is not an absolute right: "American judicial rulings tend to treat privacy as a highly subjective and contingent commodity, a matter of trade-offs and balanced interests, whereas freedom of speech and freedom of the press are defended with sweeping judgments of broad generality."

If privacy is not an absolute or fundamental right, one may ask whether it is a privileged right — one that takes precedence over many other rights. However, many authors are sceptical about the privileged status of a right to privacy.

Brin [4, pp.14–15] argues that too much privacy will actually undermine privacy. His argument is that, while most people prefer not to be stared at by strangers while eating, they nevertheless go to restaurants to eat amongst strangers. Now suppose a restauranteur improves the privacy of all guests by erecting thin screens around all the tables so that guests are protected from the gazes of strangers. Brin argues that the voyeur in their midsts will jump at the opportunity that his or her newly established privacy offers, to find a means to look at the other guests — by, for example, making a small hole in the screen through which other people can be watched, but through which the other guests cannot see the voyeur. In other words, while the visibility of the voyeur in the open system was checked by the possibility of being seen staring at others, this protective mechanism falls away in the more private setting.

Brin uses this example to consider more serious implications of too much privacy. Taking his cue from Popper's notion of an *open society* he argues that transparency in society is of the utmost importance. Where Popper's open society is a society that is open to criticism, Brin's transparent society is one where actions of the 'watchers' can be 'watched' so that criticism can effectively function. It is not that privacy should be totally abandoned as bad, but "transparency is *underrepresented* in today's fervid discussions about privacy and freedom in the information age" [4, p.18] .

Etzioni [6, p.5] uses a communitarian view to argue in favour of a "much needed social correction — [the] balancing of rights with a fresh emphasis on responsibilities — [that] has yet to be brought to bear on privacy issues."

I suggest that much of the confusion arises from an oversimplification of the concept of privacy (while acknowledging that I have not done justice to many of the more sophisticated views on privacy developed by some of the work cited above). One way to gain insight into the nature of privacy (or those aspects of it that can possibly be meaningfully protected) is to consider informed consent. Informed consent already plays a role in database privacy. However, informed consent also plays a significant role in medicine — in partic-

[6] While her legal arguments are from the US context, many of the arguments could also apply in other contexts.

ular in medical research. Like privacy, medicine also has to balance wellbeing with possible harm. And, based on a number of clear iniquities (such as the Tuskagee case) widespread debate has ensued over many years and these debates have culminated in ethical guidelines that address informed consent in medical research. While these guidelines explicitly consider informed consent, they implicitly guide the research process as such. Since informed consent can also be used in database privacy, guidelines for informed consent, in a similar manner, say something about the (practical) nature of privacy.

The next section first briefly considers some of the issues surrounding informed consent — including in the manner in which it is used in privacy. Next, some of the best-known guidelines for informed consent are reinterpreted to see what light they may shed on privacy.

6. INFORMED CONSENT

The first objection against informed consent is that — for privacy purposes — it is usually a binary decision: if one accepts the terms of the privacy policy, one can go ahead and use the offered service; if not, one has to look elsewhere. This is, amongst others, the way that the procedure has been automated for P3P [21]. The debate about whether allowing people to "opt-in" or "opt-out" of services is an example of this binary mode of thinking.

Often this (binary) choice that one purportedly has, is illusionary. As Etzioni [7] points out in many cases one has no alternative, but to use the service, and is therefore forced to give consent for using one's personal in manners one would have preferred not to.

Where one has a choice, consent is often not *informed* consent, due to the legalese used on consent forms, the time allowed for thinking about giving consent and the unpredictability of what records, that one is willing to share now, will contain in future. In fact, consent is often 'manufactured' throughout society: when the content of the media is driven by the needs of the state and big business, the public "will accept the meaningless and subordinate lives that are appropriate for them and they'll forget subversive ideas about taking control of their own lives" [27, p.85].

Finally, people "often do not read consent forms carefully because they assume that someone else has scrutinised the risks and benefits on their behalf" [28]. This will be of particular importance when considering my recommendations. Rather than using this state of affairs to argue that consent is of little use to protect privacy — "as a limited, secondary source for protection of privacy" [7] — I suggest that consent itself needs to be rethought.

Since informed consent forms a crucial part of medical research, it has seen much debate and current thinking is captured in ethics guidelines. I suggest that the following aspects hold the most potential for application in database privacy. Note that they are intended as points of dicussion both for informed consent when collecting information about individuals and as technical requirements of database privacy. To illustrate, the first point states that the individual should be given an explanation of the purpose *and* that purpose should form an inherent part of the materialised database.

1. An explanation of the purpose for which data is being collected [2, §46.116(a)(1)] [5, §5,3]. This is obviously not a new insight because, as mentioned in section 2 the *purpose* for which data is collected already forms part of the debate around privacy; however, I suggest that purpose should be specified with finer granularity. If a name, address and age of an individual are to be recorded in a database, a purpose can be associated with each of the fields.

2. Confirmation that individuals have the right to access their own data [5, §5.8].

3. A "description of any reasonable foreseeable risks or discomforts to the" individual [2, §46.116(a)(2)] [5, §5.9]. This places the burden on the owner of the database to disclose possible problems that may occur. In privacy terms this would include disclosing possible sharing of data with others, foreseeable problems if data ages (in other words, on whom is the burden placed to ensure the integrity of the data over time), etcetera.

4. Disclosure of alternatives that exist for the individual to having data recorded in this specific database [2, §46.116(a)(4)] [5, §5.13]. Obviously, the requirement cannot be that, say, an online bookseller should inform a customer of all other alternative booksellers where the customer's transaction can be recorded. Alternatives should be considered on the micro and macro levels. On the micro level it should be clear what aspects of storage one can opt-in or opt-out of. It should also be clear what information has to be supplied to use the service and whether anonymous or pseudonymous use is permitted On the macro level, it should be clear if the service is available from other parties. Alternatives to *storing* should also be considered, such as storing a reference [16] or hashed or encrypted value [15].

5. It should be stated what mechanisms are in place to protect the confidentiality and integrity of private information [2, §46.116(a)(5)] [5, §5.14,5.15].

6. Ways in which an individual will be compensated if privacy is violated should be given [2, §46.116.(a)(6)] [5, §5.15]. This is intended here to transfer some of the risk of privacy violations away from the individual to the owner of the database.

7. What secondary uses the data may be used for [5, §5.18].

8. Contact details for questions about the database and/or the privacy policy [2, §46.116(a)(7)].

9. It should be made clear when recording in a database is mandatory. This happens, for example, with storage of tax returns or recording a citizen in the population register [2, §46.116(a)(7)]. Such cases may need special security measures.

10. The approximate number of individuals whose information is recorded in the database should also be given [2, §46.116(b)(6)]. Since bigger databases are more at risk, the steps taken to protect it should be (see point 5 above) should be adequate given the size of the database.

11. How long the information will remain in the database [5, §5.3].

Having reconsidered informed consent from medical research in the light of privacy, the question should be asked: If privacy depended on an individual's informed consent for use of personal data, would it contribute to an increase in privacy? Clearly, informed consent (and its implications for database privacy) can become so complex that it would be of little use to the average individual. I contend, however, that technology can be used to help solve this problem. The next section makes some suggestions in this direction.

7. CHALLENGES

The picture of privacy that emerges from the preceding discussion is indeed a complex one. The question that this brings to mind is: Is it worthwhile to implement technical means to protect privacy, if it is so hard to capture privacy itself? Above I have tried to flesh out some of the details (that are obviously open to alternatives, criticism and debate). It is possible to use some such details to construct systems that enhance privacy, without the need of a precise definition of privacy.

The first challenge is to develop systems that can find a proper balance for confidentiality, integrity and availability of private information. Two approaches (at least) seem to have potential. The first is to encapsulate private information in some container or envelope. Whenever the private information is to be used the container verifies that the intended use is legitimate [25]. In this manner private information is made available, with the necessary confidentiality built-in. Implementation of such systems, however, still presents major obstacles.

Another approach is to distribute private information over a number of repositories and supply whoever has a legitimate need for such information with a pointer to the information [16]. Access to the information is controlled with the use of tickets or some other form of access control. Again the information is available to whoever has a legitimate need for the information. Moreover, when such data is properly maintained in such repositories rather than in the databases of those who need the data, integrity can be enhanced, since the repository can be updated once with all users immediately directed to the updated information, improving integrity.

The second challenge is to develop systems that can deal with the inherent complexities inherent in privacy. One possibility is the development of ontologies that express some of the underlying aspects of privacy, so that an individual's privacy policy may be better compared with that of another party (using automated means). Additionally, given the complexity for the average individual of such a view of privacy, mechanisms should be established to deal with the complexity. Whether these mechanisms should operate analogous to trade unions, activists, to some other existing societal structure or should use a new form of cooperation is not clear. However, to be effective, it is clear that this process should be implemented using technology — because only if technology is used can it be employed wide enough and can it be used in a proactive manner, rather than only once privacy has been violated.

This paper has intentionally steered clear of attempting to give specific technical solutions. Its intention was to highlight some of the issues that need to be considered to address database privacy. Issues will inspire solutions and solutions will suggest further issues. Such an iterative, balanced approach is required for a humane and prosperous future.

8. REFERENCES

[1] Constitution of the Republic of South Africa, 1996. Act 108 of 1996.

[2] Code of Federal Regulations, title 45, Public Welfare, part 46 Protection of human subjects, November 2001.

[3] A. W. Branscomb. *Who Owns Information? From Privacy to Public Access*. Basic Books, New York, NY, 1994.

[4] D. Brin. *The Transparent Society — Will Technology Force us to Choose between Privacy and Freedom?* Perseus Books, Reading, MA, 1998.

[5] Council for International Organizations of Medical Sciences (CIOMS). *International Ethical Guidelines for Biomedical Research Involving Human Subjects*. 2002.

[6] A. Etzioni. *The Limits of Privacy*. Basic Books, New York, NY, 1999.

[7] A. Etzioni. Medical records — enhancing privacy. preserving the common good. *Hastings Center Report*, 23(2):14–23, 1999.

[8] J. Fox. *Chomsky and Globalisation*. Icon Books, Cambridge, UK, 2001.

[9] E. Gabber, P. B. Gibbons, D. M. Kristol, Y. Matias, and A. Mayer. Consistent, yet anonymous, web access with LPWA. *Communications of the ACM*, 42(2):42–47, February 1999.

[10] S. Garfinkel. *Database Nation — The Death of Privacy in the 21st Century*. O'Reilly, Sebastopol, CA, 2001.

[11] J. Garret. John rawls on moral principles for individuals: With emphasis on implications for business ethics, February 2002. `http://www.wku.edu/~ jan.garrett/320jrpfi.htm`.

[12] D. Goldschlag, M. Reed, and P. Syverson. Onion routing. *Communications of the ACM*, 42(2):39–41, February 1999.

[13] D. Hume. Of the standard of taste. In I. Aalen, editor, *Philosophical Works, 3: Essays, moral, political and literary*, pages 266–284. Scientia Verlag, 1964.

[14] I. Kant. *Critique of Judgment*. Hackett, 1987. Originally published in German in 1790; translated by Werner S. Pluhar.

[15] F. A. Lategan and M. S. Olivier. Enforcing privacy by withholding private information. In S. Qing and J. H. P. Eloff, editors, *Information Security for Global Information Infrastructures*, pages 421–430. Kluwer, 2000.

[16] F. A. Lategan and M. S. Olivier. PrivGuard: A model to protect private information based on its usage. *South African Computer Journal*, 2002. In press.

[17] A. Margalit. *The Decent Society*. Harvard University Press, Cambridge, MA, 1996.

[18] Nordic Committee on Bioethics. *Who owns our genes?*, Tallin, Estonia, October 1999.

[19] M. S. Olivier. Position for panel on privacy. In *Sixteenth IFIP Working Conference on Database and Application Security*, August 2002. To be published in *Database and Application Security XVI* in 2003.

[20] J. Rawls. *A Theory of Justice*. Harvard University Press, 1971.

[21] J. Reagle and L. F. Cranor. The platform for privacy preferences. *Communications of the ACM*, 42(2):48–55, February 1999.

[22] M. K. Reiter and A. D. Rubin. Anonymous web transactions with Crowds. *Communications of the ACM*, 42(2):32–48, February 1999.

[23] A. Rosenberg. Privacy as a matter of taste and right. In E. F. Paul, F. D. Miller, and J. Paul, editors, *The Right to Privacy*, pages 68–90, Cambridge, 2000. Cambridge University Press.

[24] L. B. Sauerwein and J. J. Linnemann. *Handleiding voor Verwerkers van Persoonsgegevens — Wet Bescherming Persoonsgegevens*. Ministerie van Justitie, Den Haag, Nederland, April 2002.

[25] R.-C. Serban. *The Private Cyberspace: Modeling Electronic Environments inhabited by Privacy-concerned Agents*. PhD thesis, Vrije Universiteit, Amsterdam, The Netherlands, 2002.

[26] R. Whitaker. *The End of Privacy — How Total Surveillance is Becoming a Reality*. New Press, New York, NY, 1999.

[27] M. Winston. *On Chomsky*. Wadsworth, Bellmont, CA, 2002.

[28] J. Wise. Patients do not read consent forms. *BMJ*, 313:1421, December 1996.