# Privacy Guidelines Workshop
# December 1, 2005

Prepared by: Mary Lysyk, University of Ottawa.
Khaled El Emam, PhD, University of Ottawa.
Carole Lucock, LLB, LLM, University of Ottawa.
Michael Power, LLB, Gowling Lafleur Henderson LLP.
Donald Willison, PhD, McMaster University.

## Document Information

| | |
|---|---|
| **Document Title:** | Privacy Guidelines Workshop Report |
| **Original Document Date:** | 20th December 2006 |
| **Current Version Date:** | 22nd March 2007 |
| **Document Version:** | 5 |
| **Copyright:** | CHEO Research Institute, 401 Smyth Road, Ottawa, ON K1H 8L1, Canada |
| **Contact:** | Khaled El Emam (kelemam@ehealthinformation.ca) |
| **More Information:** | http://www.ehealthinformation.ca/ |

# Table of Contents

# Introduction

In an era where information technology (IT) is increasingly taking a prominent role, there is a move by federal, provincial and territorial governments to accelerate the adoption of electronic health records (EHRs) [1-3]. The 2006 report of the Health Council of Canada recommended the rapid adoption of EHRs as a tool to improve access, quality and comprehensiveness of health care [2]. The Romanow Report on The Future of Health Care in Canada indicated that one of the major attributes of electronic health records is the accessibility of health information for research purposes [1]. IT facilitates clinical and population-based health research using personal health information (PHI), as in the development of disease registries or when conducting clinical and genetic database research [4;5].

While IT has the potential to enhance patient care, health system planning and research, it also brings with it the possible diminution of patient privacy [6]. Privacy has been identified as the issue that may slow and/or impede the progress of EHR implementation [7]. Information and privacy commissioners, as well as the two major reviews of Canada's health care system, have raised privacy as the key policy area needing to be addressed in relation to electronic health information (EHI) [1;3;7;8]. Protecting the privacy of personal information is becoming one of the most important ethical issues in IT [9]. Numerous surveys have shown that Canadians are concerned about their personal privacy eroding in an electronic world [7;10]. It is recognized that negative perceptions of PHI privacy by the public could stall the adoption of IT in health care, and reduce participation in clinical and population-based health research, especially when IT is used for data management [11]. There is also evidence that privacy considerations could hamper epidemiological research through strict consent requirements. Such consent requirements introduce problems like biases in recruitment due to non-consenting subjects, the inability to request consent, and increases in the costs associated with doing research [12-16]. Furthermore, Research Ethics Boards (REBs) typically do not have the necessary expertise to assess electronic health information privacy [17;18]. This lack of knowledge leads to inconsistent REB decisions across the country, making it difficult for researchers to follow standardized practices for protecting privacy [17;18].

Currently, privacy best practices specific to IT do not exist for the health research community.

## The First Steps to Developing EHI Privacy Best Practices

The Privacy Guidelines Workshop was held on December 1, 2005, in Ottawa, Ontario. The workshop was the first step in a broader research program focused on developing EHI privacy best practices and resource tools to aid researchers, clinicians and REBs. Privacy, policy and IT experts from across the country were brought together in Ottawa for the first workshop of its kind in Canada. The anticipated outcomes of developing these best practices and resource tools specific to IT use in the health research community would be:

- an increase in public confidence concerning the seriousness with which EHR privacy is handled;
- assisting members of the health research community in implementing best practices for EHI privacy;
- founding standardized guidelines to assist REBs in evaluating protocols that use IT and promote consistency across the country; and
- establishing a baseline that all custodians of EHI and IT vendors need to meet or exceed in ensuring privacy, confidentiality and security of personal health information within clinical research contexts.

## The Priority IT and Privacy Issues Currently Facing the Health Research Community

In order to identify the priority issues in electronic health information facing the health research community, interviews were conducted prior to the workshop with key privacy and electronic health information stakeholders from Alberta, Ontario and Quebec. Three areas emerged as priorities and formed the workshop focus:

1. The role of REBs in protecting privacy
2. Anonymization of EHI data

3. Outsourcing database development in research

## Scope and Structure of the Privacy Guidelines Workshop Report

This report will begin by outlining the background leading up to, and establishing the foundation for this first workshop. The specific methodologies used, and the workshop outcomes, findings, recommendations and next steps will be detailed. The report is divided into six parts. Following the Introduction, the Background section provides information and justification for why validated privacy guidelines are needed (and immediately so) for the use of electronic health information in health research. In the Overview, a brief summary of privacy legislation in Canada is presented, and includes specific privacy frameworks and ethical codes of conduct related to electronic health information. The current codes of conduct, which were identified as priorities in the interviews, that are here discussed include de-identified personal health information and IT outsourcing. This section concludes by discussing current best practices and challenges related to IT for the health research community. The Methodology section follows, and provides details regarding the interviews conducted. These interviews identified and prioritized the issues that Canadians felt most urgently needed to be addressed. This section also provides an overview of the workshop itself. The section on Workshop Outcomes provides details of the workshop proceedings, and is divided into the three focal areas identified from the interviews. The findings and outcomes of the workshop are supported in the text by direct quotations from participants. Lastly, the report concludes with a Discussion that includes a summary and key recommendations, along with next steps and priorities for action.

# Background

## The Increasing Role of IT in Health Care

The increasing role of IT in shaping the Canadian health care landscape is well recognized [2;3;19]. EHR is emerging as the key information and communications foundation for our health care system [2;3;19]. Due to the potential for interoperability and instantaneous access, EHRs have the capability to improve the quality of health care delivery, reduce costs and facilitate health services planning and research [2;3;19].

In Canada, EHR systems are beginning to emerge in a variety of stages, with a variety of components, within different jurisdictions across the country [6]. These include the development of clinical registries and networks for pharmaceutical, laboratory and diagnostic imaging information [6]. The Romanow Report on The Future of Health Care in Canada advocated EHRs for all Canadians as part of a pan-Canadian electronic health infrastructure [1] .

As Canadian health care moves towards the integrated use of EHRs in direct patient care, the health research community is seeing a proliferation of database research. This research incorporates both administrative and clinical database use, as well as the development of disease registries [4;5]. The common element for these trends is the increased utilization of IT and electronic health information. In the case of client registries, where research data come from clinical practice, EHR is the electronic data collection system [17]. While there are numerous clinical and research benefits to electronic health information, there also exists the potential for serious violations of privacy [4;6;20]. The Kirby Senate Report on health care states "…the implementation and full deployment of the pan-Canadian Health Infrastructure faces three major barriers: the protection of personal information, legal and ethical issues, and the interoperability of the various systems [3]". Elaine Gibson, from the Faculty of Law of Dalhousie University, reframed the concept of personal information

protection. She views this protection not as a barrier to EHR implementation, but rather as an essential component of an EHR infrastructure that will help to maintain the trust of Canadians and ensure that their PHI remains secure [21]. The results of numerous public opinion surveys of Canadians lends credence to Professor Gibson's position [7;10]. Canadians have serious concerns about the erosion of personal privacy and doubts about security [7;10].

To protect their privacy, some individuals have taken steps that may be detrimental to their well being [22], such as not being completely honest with their health care provider [23]. A survey in the US found that as many as 15% of adults have changed their behaviour to protect their privacy [24]. In order to protect their personal health information privacy, these people have taken steps that include: going to another doctor, paying out-of-pocket when insured to avoid disclosure, not seeking care to avoid disclosure to an employer, giving inaccurate or incomplete information on medical history, and asking a doctor not to write down the health problem or record a less serious or embarrassing condition [24].

Inaccurate data jeopardizes patient safety. Without complete information, clinicians could make treatment errors [25] and/or make errors ordering medications [26]. Furthermore, researchers may underestimate disease prevalence [27], and health system managers may underestimate compliance with standards of care, such as vaccination guidelines [28]. Health care organizations may also be in danger of receiving fines if they report inaccurate data to government agencies due to such misinformation by patients [29].

Some studies suggest that privacy considerations hamper epidemiological research as a result of the strict consent requirements they entail. These requirements may introduce biases in recruitment due to non-consenting subjects and the inability to request consent, and/or increase the cost of doing research [12-16]. Excessive restrictions on epidemiologists' access to identifiable PHI could be detrimental to society at large, as many beneficial studies could not be done [30]. Consequently, a pragmatic equilibrium still needs to be established whereby epidemiological research can progress, while PHI remains protected. Reaching this equilibrium, however, is more complex with the addition of electronic health information.

## Why Does Ensuring Privacy of Electronic Health Information in Health Research Need to be Addressed Now?

Negative perceptions of PHI privacy by the public could potentially stall the adoption of EHR in health care and reduce participation in health research, especially when IT is used for data management [17]. Survey results showed that addressing privacy and security issues are key to the success of EHRs [7], and that Canadians want clear assurances that privacy and security are being protected and respected [7;10;31].

Currently in Canada, there are no validated privacy guidelines specific to the use of electronic health information in health research that would both reduce privacy risk, and provide assurances that the necessary steps are being taken to protect PHI privacy, security and confidentiality.

The next section will review privacy statutes, frameworks and guidelines as they relate to electronic health information in clinical research. It will also address the issues of de-identified personal health information and IT outsourcing, and identify knowledge gaps that currently exist.

# Overview

## Privacy Legislation, Electronic Health Information and Current Best Practices for the Health Research Community

The Canadian privacy legislative landscape is often described as technology neutral and patchwork [6;32]. Technology neutrality is not surprising, as legislation applies to health information rather than the delivery or mode of delivery of the PHI [32]. Regarding the lack of uniformity, numerous organizations and institutions have expressed concerns about the challenges of complying with overlapping, or even conflicting legislations [6]. In terms of protecting personal health information, different Canadian jurisdictions often apply different rules. For example, some public sector legislation includes institutions such as hospitals, universities and regional health authorities, while others do not. The core federal legislative provisions aimed at protecting privacy, confidentiality and security of personal health information include the Canadian Charter of Rights and Freedoms [33] and the Privacy Act [34]. The 10 principles in the Canadian Standards Association (CSA) Model Code for the Protection of Personal Information were codified into law progressively between 2001 and 2004, through the Personal Information Protection and Electronic Documents Act (PIPEDA) [35]. PIPEDA governs how the private sector (including private clinical and physician practices) collect, use, disclose and dispose of personal information [35]. In the public sector, legislation governing health information privacy and management has been enacted by four provinces: Ontario [36] , Alberta [37], Saskatchewan [38] and Manitoba [39].

Recognizing the need to harmonize privacy practices, the Advisory Committee on Information and Emerging Technologies (comprised of Federal/Provincial/ Territorial (F/P/T) representatives) developed and implemented the Pan-Canadian Health Information Privacy and Confidentiality Framework [40]. The Framework, while not a legislative process, provides guidelines around the duties and obligations of custodians, or trustees, to protect personal health information [40]. For example, utilization of a risk management approach is emphasized. This approach serves to

protect data from loss, theft, corruption and unauthorized disclosure. To minimize these risks, the Framework recommends the use of a Privacy Impact Assessment (PIA) [40]. The PIA is a tool that evaluates whether a specific project, system, service or initiative will have an impact on the privacy rights of an individual. PIAs are essentially risk management processes, which were first developed and implemented by the Treasury Board of Canada in May, 2002. To date, only the Health Information Act in Alberta has legislated the use of PIAs for all new initiatives with privacy implications, including those specific to IT [37].

In addition to laws and guidelines, ethical codes of conduct that govern various health professionals (i.e., the Canadian Medical Associations' Code of Ethics [41]) need to be considered in the privacy landscape.

## De-identified (anonymized) Personal Health Information

All of the statutes, frameworks and ethical codes of conduct seek to protect personal information that is linked to an individual's identity. Information that is considered de-identified (anonymous) is typically not included in most Canadian statutes. However, most of the statutes do not define what is meant by de-identified information, leaving a gap concerning what is excluded from the definition of identified information. PIPEDA defines identified information, as "information about an identifiable individual, but does not include the name, title or business address or telephone number of an employee of an organization" [35]. Quebec legislation, on the other hand, defines identifiable information as "information concerning a natural person which allows the person to be identified" [42]. Other legislation includes information that can uniquely identify an individual, such as genetic materials and fingerprints, as well as sensitive information such as ethnicity.

The Saskatchewan [38], Alberta [37] and Ontario [43] health-sector privacy statutes expressly exclude de-identified information, and provide a description of what this means; however, this description varies between jurisdictions. In Saskatchewan, de-identified information refers to "personal health information from which any information that may reasonably be expected to identify an individual has been removed" [38]. Alberta's definition is worded somewhat differently and states, "that

the identity of the individual who is the subject of the information cannot be readily ascertained from the information", and prohibits the publication of "health information in a form that could reasonably enable the identity of an individual who is the subject of the information to be readily ascertained" [37]. Alberta's legislation also indicates that collection, use and disclosure will occur with the "highest degree of anonymity that is possible in the circumstances" [37]. Ontario's health sector legislation also considers the risk of combining information (e.g., such as data linkages) in its definition of de-identified information. It defines this as "information that identifies an individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or with other information, to identify an individual" [43].

In contrast, in the U.K. the data commissioner has taken a fairly strong position on the issue of what will count as anonymized data for the purposes of excluding it from the *Data Protection Act* (see Appendix B for more details):

> *The Commissioner considers anonymisation of personal data difficult to achieve because the data controller may retain the original data set from which the personal identifiers have been stripped to create the "anonymised" data. The fact that the data controller is in possession of this data set which, if linked to the data which have been stripped of all personal identifiers, will enable a living individual to be identified, means that all the data, including the data stripped of personal identifiers, remain personal data in the hands of the data controller and cannot be said to have been anonymised. The fact that the data controller may have no intention of linking these two data sets is immaterial.* [1]

In the United States, the 1996 Health Insurance Portability and Accountability Act (HIPAA) goes further than Canadian legislation by identifying 18 elements (Appendix A), from which, if at least one is contained in the data to be collected, then the information is considered identifiable. As a similar list does not exist in Canada, anecdotal evidence indicates that researchers use the 18 HIPAA elements as a guiding principle, despite the fact that research has not validated them in a Canadian context.

[1] U.K. Information Commissioner *Data Protection Act 1998: Legal Guidance* (London: Information Commissioner, 2002). A lengthier excerpt of this guidance is contained in Appendix B.

## IT Outsourcing

Some disclosures of PHI are intentional and also controversial. Unlike paper health records, electronic health information (e.g., EHRs) and their related management tasks can, and frequently are, processed and managed by third parties [45;46]. While access to the personal electronic health information still occurs within the organization, it is maintained and stored on servers usually located off site, possibly within another jurisdiction or even another country. In addition, IT makes it easier to outsource certain medical data management tasks, such as transcription and coding, to third parties that are in foreign countries. There are concerns about the extent to which the foreign companies can ensure the privacy of PHI in their possession, since they are governed by rules and traditions that may not be aligned with Canadian legislation and values [45;46]. An article in Rocky Mountain News describes a recent problematic outsourcing case as follows: “A women in Pakistan doing cut-rate medical transcription for the University of California at San Francisco medical centre threatened to post patients’ confidential files on the Internet unless she was paid more money” [45]. To prove her point, she attached several patient records to her email to UCSF. As it turns out, she is at the end of a long chain of command: The records she was handling had been subcontracted four times. The Pakistani woman withdrew her threat after being paid about $500 from a subcontractor indirectly involved in the scandal [45]. While this was an American story, it is unclear how much coding and transcription goes overseas from Canada, and what types of protection exist when such data management tasks are outsourced.

In British Columbia, proposed outsourcing of government services to the United States led to a study commissioned by the BC Privacy Commissioner [46]. The study identified the US Patriot Act, specifically s. 215, as a specific area of concern. This statute permits the US government access to personal information if threats to US national security are perceived. The concern over s.215, and its possible implications for Canadians, ultimately resulted in an amendment to the BC Freedom of Information and Protection Privacy Act (FIPPA), limiting use of US service providers. The province of Alberta is also currently reviewing the Patriot Act issue. These examples demonstrate the uncertainty and lack of policy direction that currently exist, particularly in trans-territorial (foreign) data outsourcing.

## Current IT Best Practice Guidelines in Health Research

Recognizing that protecting the privacy of all individuals whose information is used for research is one of the ethical challenges facing the clinical research community, the Canadian Institute of Health Research (CIHR) developed and published *CIHR Best Practices for Protecting Privacy in Health Research* in September 2005 [47]. These guidelines identify 10 privacy best practice elements, all of which are linked to the Tri-Council Policy Statement *Ethical Conduct for Research Involving Humans [48]*. A particular strength of this document is in its dynamic nature and intent that "these Privacy Best Practices must continue to evolve to reflect improved practices and innovative solutions, over time, and to reflect and influence ongoing legislative developments" [47, p.2]. It also identifies several areas that will require further study including foreign outsourcing and transfer of data.

The CIHR Best Practices for Protecting Privacy in Health Research, emphasize data limitation first and foremost and highlight the need for justification by the researcher of any data that are identifying or cannot be aggregated [47]. The document goes further by ranking data identifiability (i.e., direct identifiability, use of single coding and double coding strategies). As well, direct and indirect identification are distinguished with the latter recognizing that a combination of elements could result in indirect identification of an individual. Data security practices suggest use of PIAs in ensuring that data flow is secure and confidential.

While most of the 10 elements are relevant to IT, the CIHR Best Practices do not provide standardized guidelines specific to the use of electronic health information in health research. Canadian studies conducted by El Emam [17] and Willison [18] have shown that there are significant knowledge gaps in IT issues, such as data re-identification. The studies suggest that the health research community, including REBs, typically do not have the expertise to assess privacy in research protocols where IT figures prominently [17;18]. This leads to inconsistency across the country in dealing with privacy and electronic health information, making it difficult for researchers to follow standardized practices for protecting privacy [17;18]. Variation in how REBs handle EHI privacy is even more problematic in multi-center studies

[17;18]. The consequence is additional effort on the part of researchers to address differing requirements.

# Methodology

We conducted interviews with subject matter experts and key stakeholders to identify the core privacy issues and concerns related to the use of IT in clinical research. These issues would become the workshop priorities from which electronic health information best practices for the clinical research community would emerge. Subject matter experts and stakeholder groups who were invited to participate in both the interviews and workshop included: members of research ethics boards; federal and provincial health policy analysts; members of the Ontario privacy commissioner's office; privacy lawyers; security experts (industry and academic); clinical informaticians; individuals involved in IT implementation; and health services researchers and epidemiologists.

## Semi-Structured Interviews

Interview questions were determined based on a review of the literature in privacy and IT and from anecdotal information from privacy experts and REB members. The interview questions were open ended and evolved as the interviews progressed. Twenty-seven interviews took place with stakeholders from Alberta, Ontario and Quebec. Three areas emerged as priorities:

1. The role of REBs in protecting privacy
2. Anonymization of EHI data
3. Outsourcing database development in research

## Workshop- Privacy Guidelines

A total of 28 EHI and privacy expert stakeholders were invited to participate in the one-day workshop (see the actual attendee list in Appendix C excluding external facilitator and scribe). The workshop was divided into three sessions focused on the

three areas identified from the interviews: 1) REB policy alternatives for IT based clinical research study protocols; 2) De-identification/anonymization of electronic health information data; and 3) IT outsourcing. The three workshop sessions were led by recognized experts in each subject area, with a professional facilitator guiding the entire workshop. The session leads began each session with a twenty-minute background review. The background reviews will be provided in the next section in order to set the backdrop and context for the one-hour discussions that followed. The summary of key discussion points and outcomes will follow the background review.

# Workshop Outcomes

## The Role of REBs in Protecting Privacy: Four Policy Alternatives — Session Lead: Don Willison, McMaster University

### Background Review

Related to the role of REBs in protecting privacy, several approaches may be taken to ensure that privacy, confidentiality, and security issues are addressed. Findings from the semi-structured interviews outlined our non-mutually-exclusive policy alternatives for IT based clinical research protocols:

1. **Status Quo**
    - REBs are only one mechanism for privacy protection;
    - Health information custodians, sponsors, and researchers also take privacy protection into consideration and, together, provide adequate protection.
2. **Independent (prior) review of individual projects for privacy, confidentiality, and security issues.**
    - Many REBs do not have privacy , security, and IT expertise readily available;
    - A specialized committee would review the IT and the privacy elements of the protocol and then make a recommendation to the REB.
    - These recommendations would then be taken into consideration by the REB as part of the package of ethics issues to be addressed.
3. **Periodic audit of IT providers**
    - Currently, there is wide variation in the practices of IT vendors with respect to privacy and security protection. REBs do not really examine the IT vendor capability as part of the review;
    - If we follow an ISO 9000 model, there is an accepted standard for doing privacy impact assessments and there are accredited auditors;

- A registry could be maintained of PIAs that have been done for each IT system. The REB would then require the results of a recent PIA as a pre-requisite for approving a particular protocol;
- This may obviate the need for a PIA for each individual study;

4. **Train REBs and provide tools (e.g. standardized questions)**
    - The extent to which an REB is able to scrutinize a protocol that uses IT depends on the level of detail in the protocol. Lack of detail in the protocol as to the key issues around data access, de-identification (anonymization), safeguards, and disclosures may result in a research plan being passed by the REB with serious privacy and confidentiality flaws.
    - Standardized key questions would help to minimize the risk. This would need to be accompanied by adequate training of REB members in the relevant issues that each question addresses.

The advantages and disadvantages of each of these policy alternatives for REBs would be reviewed and discussed during the workshop.

The session began with a review of the study: *Academic REBs and Governance of Privacy, Confidentiality and Security in Database Research* [18]. The study's primary objectives were to document how REBs deal with privacy, confidentiality and security issues and to identify the concerns of these boards. Specific goals included documenting how REBs are presently dealing with issues of privacy, confidentiality and security in terms of common practices, variation in practices, innovative practice and common challenges. As well, the study aimed to identify common concerns of REB Chairs and REB Administrative Coordinators (e.g., educational needs), with a view toward increasing clarity and consistency in the approach used across REBs.

Two phases framed this study: Phase 1 was a pilot focused on identifying the scope of issues facing REBs; while Phase 2 focused on variation in REB practices and educational needs/challenges. Approximately thirty 'Generalist' REBs, affiliated with Faculties of Medicine in Canadian Universities, participated in the study. Methodologies included use of plausible, but hypothetical research case scenarios with mixed qualitative (e.g., interview) and qualitative approaches.

Two scenarios were highlighted. The first looked at a health record review study and asked if there is a need for individual consent in this case. The scenario research question aimed to identify the proportion of pregnant women tested for gestational diabetes from a retrospective chart review of 50 family practices (relevant records were identified through review of billing submissions). Patient data to be collected included postal code, ethnic origin, parity, mother's DOB and presence of glucose tolerance test. Approximately half of REBs that participated (n=14) indicated individual consent is required; while 12 did not feel individual consent was required. Of the 14 REBs who required consent, 10 cited identifiablity and access as a reason to obtain consent and were more inclined to regard abstracted data as potentially identifiable. All fourteen expressed concerns over access, use and disclosure. The 12 REBs not requiring consent were more inclined to regard data as de-identified/anonymous.

The second issue focused on patient/disease registries, biobanks and EHRs. The scenario was a multi-centre diabetes registry (planned to continue indefinitely), where patient accrual and data collection would take place through family physician practices. A research assistant (RA) would remove direct identifiers at the Principal Investigator's office and forward the de-identified data to the central registry. The RA would be the keeper of the ID key. Twenty-three of 30 REBs (approximately 77%) indicated that consent was required indicating that this is a prospective study and citing identifiably of data and intention of future research uses. Duration of consent was a concern as an almost equal number of REBs indicated the duration of the registry (n=8), or requiring periodic renewal (n=6), or indicated they were uncertain (n=7). There was a high degree of consensus on an ongoing reporting obligation (n=23).

Regarding concerns over electronic health information versus paper data, just over half (n=7) of the REBs sampled indicated that electronic data are different from paper data in that there are more security concerns and potential linkages with other databases. As well, additional concerns expressed were regarding the data transfer process and the need for additional security measures. Interestingly, an almost equal number of REBs did not share this concern. Data transfer concerns across national boarders once again showed a mixed response, with 60% saying concerns would be

dependent on numerous additional variables including the type of data, location of central registry and accountability structures.

This study concluded that there exists a substantial variation between REBs in terms of misinformation around core concepts of data identifiability, anonymity and personal information, and an incomplete conceptualization of what constitutes privacy-related risk. Variable standards and lack of training tools were also identified. Consequently, the four policy alternatives noted above were suggested:

1. maintaining status quo;
2. independent (prior) review of individual projects for privacy confidentiality and security issues;
3. periodic audit of IT providers; and
4. train REBs and provide tools (e.g. standardized questions).

These were to be reviewed and discussed by workshop participants in terms of advantages and disadvantages, and that consensus building take place regarding their role in determining best practices and consistency in both the development an REB review of clinical research protocols.

## The Four Policy Alternatives - Workshop Outcomes

The workshop participants began by clarifying several points from the review presentation and the 4 policy alternatives. The discussion asked how distinct these alternatives really are and, as they were not considered to be mutually exclusive, concluded that they might be combined. The distinction in the presentation slides between patient registries and electronic health records was also raised. It was felt that the issue of electronic health records overlays each of the scenarios.

> *"We talk about two types of studies; ones requiring data linkages and ones that don't. Whether it's simply a registry for collecting data over the Web or linked databases, both present the same risks."*

In addition, clarification was requested regarding the role of Privacy Commissioners in research protocols. While the role itself did not clearly emerge, the group agreed that the Privacy Commissioners' Offices should not approve the privacy components of research protocols.

The pros and cons of each alternative were discussed and the outcomes are summarized as follows:

**1) Maintaining the status-quo:**

All participants agreed that maintaining the status-quo was no longer an acceptable option particularly with the possibility of data linkages.

> *"Linked databases present increased risks to privacy of information; you can have four different databases that have been anonymized but when they are combined you can have identifiable information."*

**2) Independent (prior) review of individual projects for privacy confidentiality and security issues:**

The group decided that this approach, in theory at least, could work well, and one example was provided on the successful implementation of an independent prior review component.

> *"I sit on that kind of Independent Review Committee (IRC) at a hospital, which has a maternal child database that links information with a provincial health information database. The committee is composed of primary researchers involved in using databases for research and review applications for privacy. The REB was not actively involved with what was happening with the committee and confidentiality issues had arisen and we had to insist on a role for the REB. There is still an active role for REBs. The review committee is not comprised as a REB and there are other ethical issues that need to be dealt with by a REB. The committee responds fairly quickly and handles four to six applications (using linked databases) per year".*

Concerns were expressed regarding both the resources required for this approach, as well as the delays it would cause as the number of applications to REBs increase.

> *"This could present a major resource issue for REBs if they are to consider 20 applications per year. I'm in favour of using IRCs for reasons of expertise".*

Alternatively, privacy and security of data can be viewed as organizational infrastructure issues that can be ascertained through threat risk assessments at the local level (i.e., EHI custodian and IT vendor); and the importance of economies of scale were highlighted. One solution was to have a general infrastructure that can be used by many studies:

> *"In some institutions where they may have 20 applications, there may be economies of scale. Security is an infrastructure issue. If locally organized, it has benefits from a threat assessment perspective for IT and can [assist] the REB if the protocol is using standards such as what level of assurance is needed for the trial. We can develop a checklist with the level of security required for each application. If the infrastructure doesn't offer strong access control, we have to build it into each application."*

Within a health care organization, governance structures with privacy accountability policies were described as essential. As part of this framework, Privacy impact assessments were highlighted by the group as an increasingly utilized risk management approach that is currently legislated in Alberta for all new initiatives (including those involving electronic health information) with privacy implications.

> *"Research centers have incorporated privacy assessments into their processes and most of these things are very portable to hospitals where research is undertaken. Using the model of PIAs before a project is undertaken is a habit to get into. It doesn't cost money, its just part of the process."*

Furthermore, it was agreed that *"we need a standard for PIAs and for IT providers so they do it once and not for every project."*

**3) Periodic audit of IT providers:**

Workshop participants supported the position that audits are needed to ensure privacy, as well as security of electronic health information. However it was felt that audits are not sufficient on their own in terms achieving privacy and security.

> *"We can build security; but it doesn't mean we've achieved privacy. We need security in IT systems to reach privacy."*

Performing an audit for each individual research protocol was not considered reasonable. The ISO 9000 model was identified as a good approach to how an audit structure could be implemented in a cost effective and less onerous manner**:**

> *"ISO provides a good model and a useful approach to this with accepted standards, accredited activities, and periodic audits. Over the next two to five years, ISO is extending its process to certify the security of IT department."*

While costs were raised as a concern, the cost of *not* implementing a risk management system in terms of loss of public trust was emphasized:

> *"We haven't looked at the cost of not doing projects properly and the loss of trust internally and externally."*

**4) Train REBs and provide tools**

Standardized privacy guidelines and training are needed to improve consistency in reporting IT and privacy issues in research protocols, in order that researchers are clear about what is required and REBs can make informed decisions about research protocols. It was agreed by all that privacy training and privacy tools, such as standardized procedures with web based e-learning modules, were needed to assure consistent reporting of privacy practices. These strategies were identified as a priority for the entire clinical research community including REBs, researchers, clinicians and research coordinators; however, training and resources, alone were not considered sufficient:

> ***"****The issue for REBs is the creation of an institutional culture of privacy and security… it's not enough to have checklist, you have to ask general questions about risk, types of risk, personal specification information, individual identifiers etc."*

## Conclusions

> *"Judgment is required to attain privacy…the touchstone has to be the human beings being cared for."*

Workshop participants emphasized the need for policy alternatives 2 (Independent (prior) review of individual projects for privacy confidentiality and security issues), 3 (periodic audit of It providers), and 4 (training and tools for the clinical research community), to be integrated into the CIHR best practices framework. The importance of harmonizing common standards, processes and tools (e.g., standardized checklist), as well as maintaining practicality were highlighted.

Economies of scale were recommended due to concerns over costs, resource allocation and potential delays. Privacy and security of data were seen as infrastructure issues that can be ascertained through threat risk assessments at the local level (i.e., EHI custodian and IT vendor). Governance structures, with privacy accountability configurations, were described as essential within a health care organization. PIAs were viewed as a risk management approach that could

successfully mitigate risk, with a qualification that *"you have to get everyone involved."*

In addition to risk management mechanisms, IT audit mechanisms for ensuring both privacy and security are needed. Both mechanisms well help ensure and maintain a privacy culture. Training tools are needed, such as privacy workshops, e-learning modules and standardized privacy checklists that could be available to all members of the health research community.

Although the costs of implementing these practices were raised, the greater cost of not implementing them in terms of loss of public trust was stressed. Another discussion point raised by the group was that while there is a clear need in IT for privacy best practices to minimize risk, these should also be pragmatic and straightforward for researchers. Related to this point, concerns were also raised with database research, specifically, the need to determine data linkage privacy risk and the additional complexity that this adds for REBs, confirming again that REBs require more privacy strategies and guidelines beyond what currently exists.

It was also pointed out that REBs work as a part of a team including health care organizations, IT providers, research sponsors and researchers. All stakeholders who are involved in creating, maintaining or storing PHI must also share responsibility for maintaining privacy, confidentiality and security of personal health information**:**

> *"The single most important issue is to provide information to debate with REB members [in terms of] their role in linking the research community with the general community."*
>
> *"REB's can't be the sole guardian for all aspects of privacy and security; it needs to permeate the entire organization".*

Maintaining and facilitating this 'shared responsibility' approach helps build a privacy and security culture.

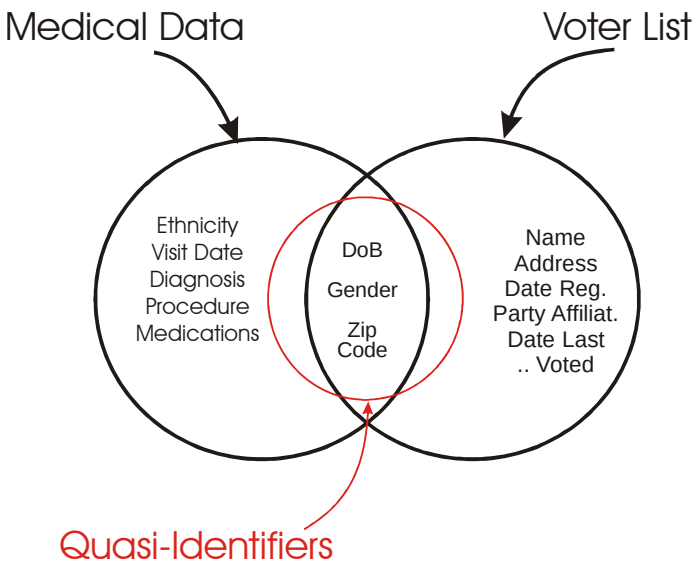
Finally, understanding e-commerce practices and how other countries evaluate the privacy and security of IT based research protocols was suggested. The need to protect the integrity of the health care relationship and create a dialogue between the research community and the community as a whole was a salient message.

## De-identification (Anonymization) of Electronic Health Information Data — Session Lead: Carole Lucock, University of Ottawa

### Background Review

A review of the work conducted by Sweeney [49], Carnegie Mellon University, Laboratory for International Data Privacy, regarding re-identification of anonymized medical data, set the context for this session. In the United States, Sweeney [49] demonstrated strategies for re-identifying data. Specifically, by linking three shared variables (date of birth, a portion of a zip code and gender) from two sets of data (voter lists and medical data), seemingly anonymized data could be re-identified (see Figure 1). These variables were termed quasi-identifiers, as alone they were anonymous, but combined with an external data source they could be used to re-identify data. Dr. Sweeney's work on k-anonymity found that data-holders who release de-identified data often do not know what external data sources, that can be used to potentially re-identify data, are available to the data recipient [49]. Consequently, data could be re-identified through the use of quasi-identifiers.

**Figure 1 [49]**




Important differences between Canadian and American systems, structures and available data sets mean, however, that the risks in the United States may not be

equally applicable in Canada. El Emam et al. [50] attempted to replicate the Sweeney study in Canada using similar variables: date of birth, postal code, gender and initials. Findings from this study showed that there is no comparable data set that is externally available to enable the same type of record linkage [50]. The study did find that de-identified data sets from well-defined professional groups of limited size (e.g., physicians, physical therapists, lawyers) could be re-identified with predictable success rates [50].

El Emam [17] also found that while REBs may require anonymization, there is no systematic or evidence-based approach concerning how this will be achieved. For example, although data limitation (data with variables eliminated) was the method used for anonymization, there was wide variation among practices and knowledge of which variables to remove or which variables were high risk. In general, decisions were made on the basis of tradition, rather than justified according to evidence. These findings were supported by work conducted by Willison [18] , which confirmed considerable variation in the ability to recognize the potential for re-identification through the combination of variables.

The knowledge gaps identified in this background review, and from the work of Sweeney [49], El Emam [50] and Willison [18], suggest the need for practical measures to raise general awareness in the health research community about the risks associated with variable re-identification.  Researchers in this community, once aware of the risks, then need to explore how Canadian practices could be improved.

## De-identification of Electronic Health Information - Workshop Outcomes

The group was asked to consider four questions: 1) The standard of de-identification: When data are anonymized, what standard ought to be aimed for or applied?; 2) How easy is it to re-identify data in Canada, and what can be done to fill the knowledge gap of those currently responsible for de-identification?; 3) The use of statistical and scientific methods: Are statistical and scientific methods available to assist in eliminating problematic variables? If so, what are the impediments to their use? Are there practical measures that can be taken to overcome identified impediments?; and 4) The use of other mechanisms to prevent re-identifying data linkage: What other

mechanisms are available to prevent re-identifying data linkage, and how can these mechanism be implemented?.

1) **The standard for de-identification (anonymization): When data are anonymized, what standard ought to be aimed for or applied?**
This question concerns the standard to be aimed for, or applied to determine whether or not data can be defined (or described) as de-identified or anonymous. The question assumes (based on the work of Sweeney [49], El Emam [50], El Emam [17] and Willison [18]) that uniform methods are not being used, which means that, in a given context, there is a risk that data could be unintentionally re-identified. However, it also assumes that there is not a common understanding of, or agreement about, what 'counts' as anonymous information.

   The CIHR document on Best Practices for protecting Privacy in Health Research incorporates all three definitions as part of the levels of data identifiability, which are ranked by the capacity to identify or re-identify individuals [47].

   As a result of this lack of consistency and clarity, additional questions emerge. These include:

   - Do we know the goal people who are de-identifying (anonymizing etc.) information are trying to accomplish?
   - Do we know what standard those who are using de-identified (anonymized, etc.) data are deploying (even if they are unable to meet the standard that they set)?
   - Do we know whether the standards (what is aimed for) that are being used correspond to legal and policy definitions, including those that would remove the information from the ambit of the legislative regime?

   These questions also engage more pragmatic considerations, which include whether data is sufficiently de-identified to exclude it from data protection legislation, and the risks of incorrectly assuming that data is

not re-linkable. General issues of public confidence are also prominent, based on the public's understanding of these terms.

Workshop participants strongly agreed that there is a need for consistent terminology when discussing de-identification, and that a standard is needed to determine what counts as de-identified data. This is particularly relevant as one participant remarked:

> *"Privacy interests are affected because data is being collected and when you combine data, it can be used for other purposes. When databases are created it attracts secondary interests and creates privacy issues".*

However, it was noted that these issues need to be considered within the context of current IT capabilities**.** Due to the complexity of this issue, it was also noted that other sectors would need to be involved. The private sector, for example:

> *"Regarding the standards of anonymization, we can't create fixed rules because we are dealing with a probability function. We have to look at the whole picture".*
>
> *"There are enormous incentives for government and private interests to use this information. For example, people being denied insurance".*
>
> *Involving other government sectors, such as Justice, was also advocated. Participants felt that this was also a judicial issue, as there exists a potential for fraud and identify theft which needs to be addressed within the discussion.*

**2) The ability to re-identify information and knowledge gaps concerning variables: How easy is it to re-identify data in Canada, and what can be done to fill the knowledge gap of those currently responsible for de-identification? How often have data been re-linked to an individual? Has this ever happened in Canada?**

These questions relate both to the sources of available data to enable re-identification, and to knowledge on the part of those who are anonymizing data as to the risks associated with variables contained in their data. There are currently insufficient grounds to assert that data-linkage of the type identified by Sweeney [49] is not an issue in Canada. As the work by El Emam et al. [50] is limited to externally available sources in Canadian provinces, options for further work in these areas include:

- Extending the study to include the possibility of re-linking across private data bases where data sharing is assumed to be on an anonymous basis;
- Investigating further sources of data, for example, what information is available commercially through data-brokers (in Canada and the U.S.);
- Extending the study to explore other variables that may pose equal problems to the ones found by Sweeney, using date of birth, gender and a partial postal code.

El Emam [17] and Willison [18] have also identified weaknesses in the practices of anonymizing data for a variety of purposes, including data-linkage. Workshop participants were asked to identify practical measures that could be taken to raise general awareness in the community about the risks associated with variables, and if these should include a list of variables that are particularly problematic. Although no strategies for raising general awareness were proposed, this area was acknowledged as needing further attention and discussion.

The group indicated that, while numerous examples of re-identification exist, we can not accurately know the extent of the problem as it has not been well documented in the scholarly literature or the Canadian press. Computer theft is one obvious example where data is vulnerable and privacy breaches can occur:

> *"Reports exist of laptops of federal government employees being stolen and thefts of computers from hospitals."*

It was felt that valuable lessons could be learned from the e-commerce literature and that, in the health sector, this area represents a significant knowledge gap. De-identification was described as a moving target; as IT knowledge and solutions advance, options that did not exist previously become possibilities. Identification of variables and their risk levels is currently lacking and is much needed.

**3 & 4) Use of statistical and scientific methods in IT applications: Are statistical and scientific methods, as well as other methods, and IT applications available to assist in eliminating problematic**

**variables? If so, what are the impediments to their use? Are there practical measures to overcome identified impediments?**

Workshop participants concluded that there is minimal use of statistical or other methods to assist in the identification and elimination of problematic variables. This deficiency is perhaps not surprising, since the use of these methods is complex. In addition, applications are available that can make variable elimination much easier (e.g. Datafly in the U.S.)[2]; however, these applications come with an associated cost. Moreover, these applications tend to be developed for the U.S. market, drawing Canadian users to the HIPAA standard (see Appendix A) which may or may not be suitable for the Canadian context. In Canada, mechanisms and applications are needed to properly identify problematic variables, so that clinical research does not continue to rely on intuition and tradition as a method of data de-identification [17;18;50].

Increasingly, a significant degree of reliance is placed on data-sharing agreements and REBs to act as gatekeepers. Legislation often requires the use of data-sharing agreements in the research context. When not required by law, such approaches are still usually recommended in the legislation, or through the offices of Privacy Commissioners or government agencies charged with administering privacy legislation.

It is interesting to note how institutions in the U.S. are approaching these issues, particularly the seriousness and sophistication of their approach. See the Human Investigation Committee of Yale University School of Medicine (http://www.med.yale.edu/hic/index.html) for an example. This committee provides resources for American researchers and others, as well as outlining procedural safeguards.

A concern was raised by participants that *"statistical techniques for blurring data are still a challenge*". Workshop participants felt that research was lacking in this area, and that it was needed on an ongoing basis in order to inform the health research community**:**

[2] Carnegie Mellon, Data Privacy Lab., online: < http://privacy.cs.cmu.edu/datafly/>

> *"We have to inform ourselves of the risks on a continuous basis and there has to be research to address these fundamental issues."*

Questions were raised about how REBs can possibly address this issue given how little is known:

> *"When REBs are asked about transferring data, all traceable identifiers are removed. What does the REB take as a reasonable standard?"*

As a result of this knowledge gap, it is difficult for REBs to determine a minimal risk standard. It was suggested that re-identifiability risk is not an issue unique to data security, but rather security within the broader health care system itself.

> *"We're talking about releasing data that could be re-linked. What are the benefits and do they outweigh the risks?"*

> *"The issue is not the research data security; it is security within the health care system."*

The group also felt that the matter is not black and white, and that variable identifiability depends very much on the context in which linkages are expected to occur:

> *"There are different cases where we need different levels of identifiably. When something is non-identifiable and a claim is made, what confidence can we have that the claim is valid?"*

It should be emphasized that there was a view by some participants that, to date, there have been no publicized incidences of data re-identification, that EHI is generally considered safe (particularly true in the context of observational research) in Canada, and that additional concerns over re-identification risks are unwarranted.

## Conclusions

> *"We have to inform ourselves of the risks [of data re-identification] on a continuous basis and there has to be research to address these fundamental issues."*

Fundamental to any discussion about data collection is the concern that the purpose of the collection be justified and of direct relevance to a particular research study. As a first step, a lexicon of common vocabulary needs to be developed and used so that

terms such as ‘anonymized’, ‘de-identified’, ‘non-identified’ and ‘quasi-identifier’ can be used in a consistent fashion. As well, it is important to recognize that de-identification, and re-identification, of personal health information data are evolving issues based on technological knowledge and capabilities. A standard for de-identification would be ideal, but not necessarily realistic, given the changing nature of IT. The group was concerned that there is much variability currently amongst researchers on defining minimal risk standards, within and between hospitals, research centers, and countries. While REBs may need to frame re-identification risk for different types of research, few, if any, empirical studies currently offer best practice guidelines. Workshop participants strongly identified the need for more research to guide REBs in this area, and specified that this research need always be linked to technological advancement. The group concluded that this was needed now, before incidences of re-identification cause harm and jeopardize public trust.

## IT Outsourcing — Session Lead: Michael Power, Gowling Lafleur Henderson LLP.

### Background Review

The session began by defining and describing outsourcing as either: 1) An entire business process being placed outside the organization to a third party service provider; or (2) The outsourcing of support functions to service providers. Several examples of outsourcing issues were reviewed, including the case in British Columbia (described in the Background section, under the IT Outsourcing heading) which led to changes in privacy legislation in that province. Central to this presentation was the message that the obligations for maintaining privacy and security ultimately rest with the Initial Collector (IC) of the personal health information, and that the IC needs to clearly address any potential issues through a contractual agreement with the service provider. This is particularly critical for non-Canadian service providers. Specific points for ICs that were highlighted included:

- Identify an individual who will ensure that all matters arising from the obligations are addressed in a timely manner
- Only collect and use personal information to fulfill obligations under the contractual agreement
- Only transfer or disclose personal information as permitted or required by law or judicial authority

- Maintain administrative, technological and physical safeguards to protect against theft and unauthorized copying, modification, use, disclosure or disposal
- Ensure employees' compliance through staff training, confidentiality agreements and sanctions. Ensure that staff who terminate employment follow termination procedures that ensure the return of all personal information, that the staff are not permitted access to secure information, sites and systems, and are not aware of ongoing confidentiality agreements.
- Promptly report events that may result in privacy, confidentiality and security breaches
- Ensure a coordinated strategy in the event of any requirement of a public response (e.g., press release/conference) with any alleged privacy breach

## IT Outsourcing - Workshop Outcomes

Workshop participants were asked to develop some key recommendations about IT outsourcing, or to suggest common themes for dealing with outsourcers. Before discussion began, two points regarding IT outsourcing were brought forth:

> *"Ensure that the contractual agreement with service providers (outsourcers) or their subcontractors clearly states what can be done with data."*

> *"The offices of financial institutions have guidelines for outsourcing but there are currently none for health care. Guidelines are needed for this sector as well."*

The following questions were addressed during this session's discussions:

1) **How much health research is outsourced?**
   Two-thirds of health research is sponsored by the pharmaceutical industry, and up to around 25% of that is IT related. The percentage of health research that is outsourced was not known. A question was raised as to whether outsourced data was less protected in foreign countries.
2) **Who is doing outsourcing? Are patient registries being outsourced?** Both commercial and academic agencies are outsourcing. It is not clear if patient registries are being outsourced.
3) **What is the responsibility of REBs for outsourcing and what are some concerns?**

   Genetic-based research in an oncology setting was identified as an example where getting sensitive information would be a concern. Within the scope of

a project, there are often protocols for handling information for each researcher involved. In outsourcing, workshop participants agreed that a contractual agreement is needed and should be included as part of the protocol.

4) **What are some of the topics/issues that should be addressed in outsourcing?**

Participants suggested that remedies for breach of contract be included in the contractual agreement as a precautionary measure. In some pharma-based projects, because data are moved to another country, there can't be a guarantee of confidentiality. It was noted that, while confidentiality and security cannot be guaranteed, risk can be managed with methods like contractual measures. It should be ensured that techniques to minimize risk are utilized in these situations. An additional concern raised was secondary outsourcing. Workshop participants emphasized the importance of knowing privacy legislation and standards, and that this should be a top priority:

> *"One issue is having data outside Canada and the other is secondary outsourcing. It is critical to know the privacy standard, regardless of where data is stored."*

The group also discussed the frequent occurrence of informal agreements between colleagues:

> *"Collegial agreements should never be used. Colleagues must also sign agreements."*

Of paramount importance, it was felt that data custodians must be held legally responsible. This will ensure that they enter any agreements related to personal data with prudence and caution. A *"set of defined vendor agreements"* was suggested to facilitate sound practices. Finally, the group recognized that institutions are frequently lacking internal policies and procedures to guide IT outsourcing decisions and practices.

The following list of possible recommendations for minimizing outsourcing risk was identified and discussed by the group:

- CIHR (in partnership with others) could develop a standard for dealing with IT outsourcing and only use companies registered to the standard.
- Provide education to the health research community about outsourcing and guidance to REBs as to what steps to take to ensure data is reasonably protected. For example:

    - responsibility, if any, of site investigators in multi-site studies
    - responsibility of REBs re: outsourcing e.g. in multi-site studies
- Identify responsibility of pharmaceutical manufacturers to ensure that “their” researchers follow proper safeguards for privacy.
- Explore the pros and cons of a regulatory requirement that data transfers can only go to certain countries (i.e. excluding those with less protection than in Canada).
- All outsourcing research agreements must be first reviewed by a privacy lawyer.
- Make available boiler plate contracts that include key elements to serve as starting points.
- Know what privacy standards will be applied in the outsourced country, and put in place common privacy standards.
- Develop minimum standards and selection criteria for IT vendors.
- Encourage institutions to make a policy statement regarding the obligations for data sharing within and between institutions with direct liability for all involved, including the right to third party audit.

From this list, the following key recommendations emerged:

1. As part of the informed consent process, disclose security and confidentiality measures including whether data is being outsourced.
2. Establish recognized mandatory standards and selection criteria for IT vendors.
3. Implement formal institutional policies detailing the obligations for data sharing within and between institutions (possible link to an accreditation system).
4. Provide training and best practices for the health research community related to IT outsourcing. Specifically, guidance is needed for REBs about what they need to know and ask.
5. Develop clear contractual agreements (including pro forma contracts), which outline what can be done with outsourced data.

## Conclusions

> *"You can't guarantee confidentiality. We are not guarantors; we are risk managers."*

Workshop participants raised the concern that researchers, REBs and health research custodians are not sufficiently informed about privacy issues with IT outsourcing in general, or foreign outsourcing specifically. When multi-centre studies take place, there may be no choice *"…but to [share data] with another site or country such as the US."* The group identified numerous issues related to IT outsourcing in health care research. Discussion took place regarding the concern that disclosing outsourcing to study participants and not guaranteeing confidentiality could be deterrents to potential research subjects. It was decided that as part of the informed consent process, disclosure about security and confidentiality measures, including whether data is being outsourced, is necessary. It was emphasized that while privacy, confidentiality and security cannot be "guaranteed", risk can be managed. Measures to minimize risk must be identified and implemented.

# Discussion

## Privacy Guidelines Workshop Summary, Recommendations and Next Steps

The Privacy Guidelines Workshop served to consolidate key recommendations for developing specific electronic health information best practices; practices that can be applied within the 2005 *CIHR Best Practices for Protecting Privacy in Health Research* guidelines. Like the CIHR Best Practices, and owing to the changing nature of IT, the recommendations should be reviewed at least every two years. The principle recommendations that emerged from the workshop are summarized below.

## Policy Alternatives for the Health Research Community:

1. A model is needed that will integrate the following three policy alternatives into electronic health information privacy best practices for health research. First, independent prior review of individual projects for privacy confidentiality and security issues; second, periodic audit of IT providers; and third, provide training and tools (e.g., standardized privacy checklist, e-learning module) for the clinical research community.

2. Privacy accountability frameworks, with clear privacy policies should be required for all electronic health information custodians.

3. Threat risk assessments, such as privacy impact assessments at the local level (i.e., electronic health information custodian and IT vendor), should be included as part of a total risk management approach.

4. Lessons learned from practices in e-commerce, as well as from other countries utilizing electronic health information systems should be compiled as part of a knowledge transfer mechanism.

## Next Steps

Future workshops will help guide the development of electronic health Information privacy models, while providing feedback and direction to the development of standardized privacy guidelines (e.g., EHI privacy checklist) and educational resources for the health research community. The need for standardized EHI best practices and educational resources have been identified as priorities for action.

## De-identification (anonymization) of Electronic Health Information:

1. Consistent vocabulary or a lexicon of commonly used terms (e.g., anonymization, de-identification, re-identification) should be developed.

2. More research is needed in the Canadian context to evaluate re-identification risk of potentially identifying variables, providing guidance in the use of health information data to the research community (i.e., similar to the HIPAA variable list). Due to the changing nature of IT, this research should continue on an ongoing basis. Specific areas of focus should extend current re-identification risk studies on:
   - Re-linking data across private databases where data sharing is assumed to be on an anonymous basis;
   - Identifying additional sources of data, for example, what information is available commercially through data-brokers (in Canada and the U.S.);
   - Identifying other variables that may pose equal problems to the ones found by Sweeney using date of birth, gender and a partial postal code.

3. Engage other sectors, such as justice and the private sector (e-commerce), in this dialogue.

### Next Steps

El Emam et al. [50] are currently expanding their study on data re-identification risk to include a larger sample size. They hope to also replicate this work in other jurisdictions across the country.

## IT Outsourcing

1. As part of the informed consent process, disclosure is required about security and confidentiality measures including whether data is being outsourced.
2. Establish recognized mandatory standards and selection criteria for IT vendors.
3. Implement formal institutional policies detailing the obligations for data sharing within and between institutions (possible link to an accreditation system).
4. Provide training and best practices for the health research community related to IT outsourcing. Specifically, guidance is needed for REBs about what they need to know and ask.
5. Develop clear contractual agreements (including pro forma contracts), which outline what can be done with outsourced data.

### Next Steps

IT outsourcing recommendations need to be incorporated into standardized best practice guidelines (e.g., EHI privacy checklist) that would be available to the health research community. A future workshop would serve to further develop, refine and discuss implementation these strategies.

**The outcomes of this workshop need to be disseminated broadly in order to promote awareness of these issues, and to generate further reflection and**

**discussion among the health research community, EHI and privacy experts and stakeholders, and the broader Canadian public.**

# Appendices

## APPENDIX A: HIPAA's 18 DATA ELEMENTS

1. Names

2. All geographic subdivisions smaller than a State, including:

   - street address
   - city
   - county
   - precinct
   - zip codes and their equivalent geocodes, except for the initial three digits of a zip code if, according to the current publicly-available data from the Bureau of the Census: (1) the geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people, and (2) the initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000.

3. Telephone numbers

4. Fax numbers

5. E-mail addresses

6. Social Security numbers

7. Medical record numbers

8. Health plan beneficiary numbers

9. Account numbers

10. All elements of dates (except year) for dates related to an individual, including:

   - birth date
   - admission date
   - discharge date
   - date of death
   - all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older

11. Certificate/license numbers

12 Vehicle identifiers and serial numbers, including license plate numbers

13. Device identifiers and serial numbers

14. Web Universal Resource Locators (URLs)

15. Internet Protocol (IP) address numbers

16. Biometric identifiers, including finger and voice prints

17. Full face photographic images and any comparable images

18. Any other unique identifying numbers, characteristics, or codes

## APPENDIX B: GUIDANCE FROM UK DATA COMMISSIONER

### EXTRACT FROM LEGAL GUIDANCE PROVIDED BY THE U.K. DATA COMMISSIONER

"The Commissioner recognises that the aim of anonymisation is to provide better data protection. However, true anonymisation may be difficult to achieve in practice. Nevertheless, the Commissioner would encourage that, where possible, information relating to a data subject, which is not necessary for the particular processing being undertaken, should be stripped from the personal data being processed. This may not amount to anonymisation but is in line with the requirements of the Data Protection Principles.

The Commissioner considers anonymisation of personal data difficult to achieve because the data controller may retain the original data set from which the personal identifiers have been stripped to create the "anonymised" data. The fact that the data controller is in possession of this data set which, if linked to the data which have been stripped of all personal identifiers, will enable a living individual to be identified, means that all the data, including the data stripped of personal identifiers, remain personal data in the hands of the data controller and cannot be said to have been anonymised. The fact that the data controller may have no intention of linking these two data sets is immaterial.

A data controller who destroys the original data set retaining only the information which has been stripped of all personal identifiers and who assesses that it is not likely that information will come into his possession to enable him to reconstitute the data, ceases to be a data controller in respect of the retained data.

Whether or not data which have been stripped of all personal identifiers are personal data in the hands of a person to whom they are disclosed, will depend upon that person being in possession of, or likely to come into the possession of, other information which would enable that person to identify a living individual.

It should be noted that the disclosure of personal data by a data controller amounts to processing under the Act.

For example:

The obtaining of clinical information linked to a National Health Service number by a person having access to the National Health Service Central Register will amount to processing of personal data by that person because that person will have access to information enabling him to identify the individuals concerned.

It will be incumbent upon anyone processing data to take such technical and organisational measures as are necessary to ensure that the data cannot be reconstituted to become personal data and to be prepared to justify any decision they make with regard to the processing of the data.

For example:

In the case of data collected by the Office of National Statistics, where there is a disclosure of samples of anonymised data, it is conceivable that a combination of information in a particular geographic area may be unique to an individual or family who could therefore be identifiable from that information. In recognition of this fact, disclosures of information are done in such a way that any obvious identifiers are removed and the data presented so as to avoid particular individuals being distinguished.

If data have been stripped of all personal identifiers such that the data controller is no longer able to single out an individual and treat that individual differently, the data cease to be personal data. Whether this has been achieved may be open to challenge. Data controllers may therefore be required to justify the grounds for their view that the data are no longer personal data. “

## APPENDIX C: PARTICIPANT LIST

| Participant | Organization |
|---|---|
| Khaled El Emam | University of Ottawa - Medicine |
| Mary Lysyk | CHEO RI |
| Sheila Chapman | CIHR |
| Jean-Francois Sauriol | Phirelight |
| Karim Keshavjee | Infoclin |
| Stan Matwin | University of Ottawa  Computer Science |
| Carol Lucock | University of Ottawa - Law |
| Philippa Lawson | University of Ottawa - Law |
| Alan Forster | OHRI |
| Debra Grant | Information and Privacy Commissioner-Ontario |
| Pamela Slaughter | ICES |
| Michael Power | Gowlings |
| Mary Marshall | Privacy Consultant |
| Wendy Robillard | Alberta Health & Wellness |
| Donald Willison | McMaster University |
| Susan Fox | Health Canada (Privacy Policy Division) |
| Elaine Gibson | Dalhousie - Law |
| Francis Rolleston | NRC O-REB |
| Valerie Steeves | University of Ottawa - Criminology |
| Jack Corman | IRB Services |
| Carlisle Adams | University of Ottawa – Computer Science |
| Jillian Oderkirk | Statistics Canada – Health Statistics Division |
| Leslie Paddock-Eliasziw | Alberta Research Tumor Bank |
| Kathryn Calder | Alberta Cancer Board |
| Sam Jabbouri | Carleton University – Computer Science |

# Acknowledgements

We wish to thank the Canadian Institutes of Health Research for providing funding to support this workshop. We also wish to thank the Ottawa Center for Research and Innovation for organizing the workshop and for providing financial support through the *Electronic Health Information and Privacy Conference* that was held in 2005.

# References


(1) Commission on the Future of Health Care in Canada, Romanow RJ. Building on values the future of health care in Canada : Final Report. Saskatoon: Commission on the Future of Health Care in Canada; 2002.

(2) Health Council of Canada (January 2005). Health Care: Renewal in Canada: Accelerating Change. 2005.

(3) The Standing Committee on Social Affairs SaT, Kirby MJL. The Health of Canadians-The Federal Role Final Report. 2002.

(4) Gostin LO, Hodge JG. Privacy and Security of Public Health Information. Model State Public Health Privacy Project 1999 February 24 [cited 2005 Dec 12];1-16. Available from: URL: http://www.critpath.org/msphpa/ncshdov.htm

(5) Willison D. Privacy and the secondary use of data for health research: experience in Canada and suggested directions forward. Journal of Health Services Research and Policy 2003;8(1):S1:17-S1:23.

(6) Electronic Health Records and the Personal Information Protection and Electronic Documents Act: University of Alberta, Health Law Institute and University of Victoria, School of Health Information Science. Office of the Privacy Commissioner of Canada; 2005.

(7) Office of Health and the Information Highway. Toward Electronic Health Records. Health Canada 2001 January [cited 2005 Aug 15];Available from: URL: http//www.hc-sc.gc.ca/ohih-bsi/

(8) Privacy Commissioner of Canada. Condition Critical: Health Privacy in Canada Today. 2006.

(9) Milberg SJ, Burke SJ, HJ Smith, Kallman EA. Values, personal information privacy, and regulatory approaches. Communications of the ACM 1997;38(12):65-74.

(10) EKOS Research Associates. Pan-Canadian Health Information Privacy and Confidentiality Framework Study. 2004.

(11) Electronic Health Information and Privacy Conference. Ottawa, Ontario: OCRI: IT in Healthcare; 2005.

(12) Armstrong D, Kline-Rogers E, Jani S, et al. Potential impact of the HIPAA privacy rule on data collection in a registry of patients with acute coronary syndrome. Archives of Internal Medicine 2005;165(1125):1129.

(13) McKinney P, Jones S, Parslow R, et al. A feasibility study of signed consent for the collection of patient identifiable information for a national pediatric clinic audit database. British Medical Journal 2005;doi:10.1136/bmj.38404.650208.AE.

(14) Melton III L. The threat to medical-records research. New England Journal of Medicine 2006;337(13):1466-70.

(15) Tu J, Willison D, Silver F, et al. Impracticability of informed consent in the registry of the Canadian Stoke Network. New England Journal of Medicine 2004;350(14):1414-21.

(16) Woolf S, Rothemich S, Marsland D. Selection bias from requiring patients to give consent to examine data for health services research. Archives of Family Medicine 2000;9:1111-8.

(17) El Emam K. Data anonymization practices in clinical research: A descriptive study. Report produced for the Access to Information and Privacy Division, Health Canada, 2006.

(18) Willison D. Academic REBs and governance of privacy, confidentiality and security in database research. Ottawa 2005.

(19) Irving R. 2002 Report on Information Technology in Canadian Hospitals: Canadian Healthcare Technology. 2003.

(20) Gostin LO. Public Health Law: Power, Duty, Restraint. Berkeley: University of California Press; 2000.

(21) Gibson E. Jewel in the Crown? The Romanow Commission Proposal to Develop a National Electronic Health Record System. Health Law Journal 2003;97-129.

(22) Mandl K, Szolovits P, Kohane I. Public standards and patients' controls: How to keep electronic medical records accessible but private. British Medical Journal 2001;322:283-6.

(23) Mitchell E, Sullivan FA. A descriptive feast but an evaluative famine: Systematic Review of published articles on primary care computing during 1980-97. British Medical Journal 2001;322:279-82.

(24) California Health Care Foundation. Medical Privacy and Confidentiality Survey. 1999.

(25) Leape L, Bates D, Cullen D, et al. Systems analysis of adverse drug events. Journal of the American Medical Association 1995;274(1):35-43.

(26) Ash J, Berg M, Coiera E. Some unintended consequences of information technology in health care: The nature of patient care information system-related errors. Journal of the American Medical Informatics Association 11, 104-112. 2004.
Ref Type: Magazine Article

(27) Johnson N, Mant D, Jones L, Randall T. Use of computerised general practice data for population surveillance: Comparative study of influenza data. British Medical Journal 1991;302:763-5.

(28) Wilton R, Pennisi A. Evaluating the accuracy of transcribed computer-stored immunization data. Pediatrics 1994;94:902-6.

(29) Davidson B, Lee Y, Wang R. Developing data production maps: Meeting patient discharge data submission requirements. International Journal of Healthcare Technology Management 2004;6(2):223-40.

(30) Gordis L, Gold E. Privacy, confidentiality, and the use of medical records in research. Science 1980;207(11):153-6.

(31) EKOS Research Associates. Healthcare and the Internet: Part of the Rethinking the Information Highway Study. Health Canada; 2003.

(32) Rozovsky LE, Inions NJ. Computerization and Information Linkages. Canadian Health Information. 3rd ed. Markham: Butterworths; 2002. p. 105-29.

(33) Government of Canada. Canadian Charter of Rights and Freedoms, Constitution Act, 1982, Part 1 of Schedule B to the Canada Act, 1982. 1982.
Ref Type: Statute

(34) Privacy Act, R.S. 1985, c.P-21, Privacy Act, R.S. 1985, c.P-21, (2006).

(35) Personal Information Protection and Electronic Documents Act, S.C.2000, c.5, Personal Information Protection and Electronic Documents Act, S.C.2000, c.5, (2006).

(36) An Act respecting Health Services and Social Services, R.S.Q., Quebec, (1993).

(37) Health Information Act, R.S.A. 2000, Alberta, (2001).

(38) Health Information Protection Act, Saskatchewan, (2003).

(39) Personal Health Information Protection Act, Manitoba, (1997).

(40) The Pan Canadian Health Information Privacy and Confidentiality Framework, The Advisory Committee on Information and Emerging Technologies, Health Canada, (2005).

(41) Code of Ethics, Canadian Medical Association, (2004).

(42) An Act respecting the protection of personal Information in the private sector, Quebec, (1994).

(43) Personal Health Information Protection Act, Ontario, (2004).

(44) Zoutman D, Ford B, Bassili A. The confidentiality of patient and physician information on pharmacy prescription records. Canadian Medical Association Journal 2004;170(5):815-6.

(45) Brand R. Overseas antidote: medical services are moving offshore, raising privacy issues. Rocky Mountain News 2005 May 21.

(46) Staff. Privacy and the USA Patriot Act: Implications for British Columbia Public Sector Outsourcing. Information and Privacy Commissioner for British Columbia; 2004.

(47) Canadian Institute of Health Research. CIHR Best Practices for Privacy in Health Research. Ottawa: Public Works and Government Services Canada; 2005.

(48) Tri-Council Policy Statement : Ethical Conduct of Research Involving Humans, Canadian Institute of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada, (1998).

(49) Sweeney L. Uniqueness of simple demographics in the US population. 2000. Carnegie Mellon University, Laboratory for International Data Privacy.
Ref Type: Serial (Book,Monograph)

(50) K. El Emam, S. Jabbouri, S. Sams, Y. Drouet, and M. Power, "Evaluating common de-identification heuristics for personal health information," *Journal of Medical Internet Research*, 8(4):e28, 2006.