A Positive-Sum Paradigm in Action in the Health Sector

Ann Cavoukian, Ph.D.

Information & Privacy Commissioner
Ontario, Canada

and

Khaled El Emam, Ph.D.

Canadian Research Chair in Electronic Health Information
CHEO Research Institute and University of Ottawa

A Zero-Sum versus Positive-Sum Paradigm

Individual rights are frequently pitted against societal rights or the public interest. When individual and societal rights collide, there is often an attempt to balance one against the other. The zero-sum paradigm dictates that the two goals (in this case, individual versus societal rights) are mutually exclusive and that each of the goals can only be attained at the expense of the other goal – the two goals can never be attained simultaneously.

Privacy is often viewed as an individual right that must be sacrificed in order to attain other socially desirable, but competing goals. For example, the right to privacy is often traded off to achieve national security goals. In the health sector, patient privacy may be sacrificed in the interests of health research and quality improvement. Over the years, the traditional zero-sum approach to managing competing goals has meant that privacy rights have been allowed to gradually deteriorate in favour of achieving other more urgent goals, such as minimizing a terrorist threat.

The Information and Privacy Commissioner of Ontario (IPC) is committed to bringing about a paradigm shift, by demonstrating how information technology, introduced to serve one function, can be designed and implemented in a manner such that privacy is maintained or enhanced, without derogating from the functionality of the technology. By building privacy into the design and implementation of information technology, the goal of protecting the individual's right to privacy and the original goal of the information technology can be attained simultaneously. This process, referred to as "Privacy by Design," shifts the traditional zero-sum paradigm to a positive-sum paradigm, in which both goals are maximized to the greatest extent possible.

A Zero-Sum Paradigm – Privacy versus Data Quality

Health care is an information-intensive industry. At the individual level, the efficient and effective delivery of health care depends on the ready availability of accurate and complete health information about individuals. At the societal level, maintaining and improving the health of populations requires extensive knowledge about the factors that contribute to good health, causes and treatments for medical conditions and diseases, emerging medical technologies, and policies and procedures for the efficient and effective delivery of health care. Such knowledge is typically generated through comprehensive research and the ongoing assessment of the care that is provided to patients. The predominant way in which such health research is conducted around the world is through access to health information that is accumulated during the course of providing health care to individuals.

Ontario's *Personal Health Information Protection Act* (PHIPA) permits the collection, use and disclosure of personal health information for secondary purposes, such as health research that is seen as benefiting society as a whole. Where the collection, use or disclosure is specifically permitted by PHIPA, health information custodians need not obtain consent from individuals. In some cases, certain conditions must be met. For example, in the context of health research, a Research Ethics Board (REB) must approve the use of personal health information, without consent. Where the collection, use or disclosure is not specifically permitted by PHIPA, health information custodians must either obtain direct consent from individuals or de-identify the health information. In practice, however, since it is often not practical to obtain consent, particularly with respect to previously collected data (i.e., retrospective data), health information custodians frequently rely on de-identification when using or disclosing health information for purposes that are not specifically permitted by PHIPA.

Under PHIPA, health information custodians have a general obligation *not* to collect, use or disclose personal health information if other information will serve the purpose, and *not* to collect, use or disclose any *more* personal health information than is reasonably necessary to meet the purpose. This means that health information custodians have a general obligation to collect, use and disclose *de-identified* health information rather than personal health information, if the de-identified information would be sufficient to serve the purpose. These general limiting principles apply whether or not the collection, use or disclosure is specifically permitted by PHIPA and whether or not individuals have consented to the collection, use and disclosure of their health information.

PHIPA defines identifying information 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." Health information that is de-identified in a manner such that an individual cannot be re-identified would fall outside of the scope of PHIPA. However, when traditional methods of de-identification are used, it is often possible to re-identify individuals. To the extent that it is reasonably foreseeable in the circumstances that it would be possible to re-identify individuals, the information would be considered to fall within the scope of the definition of personal health information and be subject to all of the limitations and restrictions imposed by PHIPA.

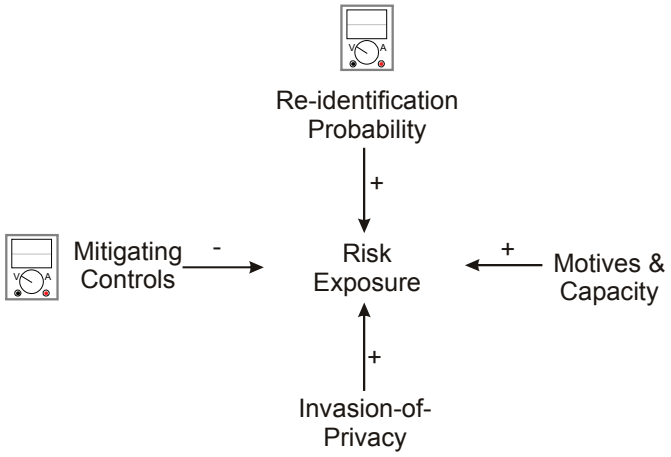
To reduce the re-identification risk to the level where re-identification is not reasonably foreseeable in the circumstances, health information custodians may alter and/or remove all direct and indirect identifiers prior to using or disclosing health information for secondary purposes. It is important to note, however, that the more variables that are altered and/or stripped from a database, the less useful the database will be for secondary purposes. Thus, individual privacy may be achieved through strict de-identification, but often at the expense of data quality. Alternatively, data quality may be preserved, but at the expense of patient privacy. This is the classic zero-sum paradigm, which we make every effort to avoid. In its place, we prefer to use a positive-sum paradigm, which maximizes the positive attributes of both interests.

Framework for Maximizing both Privacy *and* Data Quality

Dr. Khaled El Emam, a senior investigator at the Children's Hospital of Eastern Ontario Research Institute (CHEO), has resolved this dilemma through the development of a tool that de-identifies personal health information in a manner that simultaneously minimizes both the risk of re-identification and the degree of distortion to the original database. The application of this tool to any database of personal health information provides the highest degree of privacy protection, while ensuring a level of data quality that is appropriate for the secondary purpose. This privacy-enhancing technology provides an excellent example of what can be achieved using a doubly-enabling, positive-sum approach which maximizes both goals – in this case, individual privacy *and* data quality.

According to this framework, the overall re-identification risk exposure associated with a particular disclosure of personal health information is a function of four factors:

- The re-identification probability;
- The mitigating controls that are in place;
- The motives and capacity of the data recipient to re-identify the data; and
- The extent to which an inappropriate disclosure would be an invasion of privacy.



The last two factors are considered to be intrinsic to the data recipient and the personal health information that is disclosed and not subject to change by a health information custodian. In contrast, a health information custodian may change the re-identification probability (by increasing the amount of de-identification) and the mitigating controls. To reach an acceptable level of risk, the health information custodian may reduce the re-identification risk and/or add more mitigating controls. Since these two factors work in opposite directions, the health information custodian can manipulate them to balance one factor off against the other.

The re-identification probability can be controlled through the de-identification technique. More stringent de-identification techniques reduce the risk of re-identification. The other three factors are assessed using checklists.

Once a request for data has been received, the health information custodian can determine at the outset if the overall risk exposure is acceptable or not. If the risk exposure is not acceptable, the health information custodian may either de-identify the data further and/or put in place more mitigating controls. If the data recipient wants better quality data (i.e., less de-identified data), he or she must agree to additional mitigating controls which are included in a data sharing agreement. If the data recipient does not agree to additional mitigating controls, then the health information custodian must compensate by increasing the extent of de-identification and thereby reducing the exposure risk. The recipient and health information custodian must work together to achieve the level of data quality that is necessary for the recipient's purposes and the level of risk exposure that is acceptable to the health information custodian. A balance may be attained when the re-identification risk is low and the mitigating controls are low, or when the re-identification risk is high and the mitigating controls are also high.

A Positive-Sum Paradigm – Privacy *and* Data Quality

The value of the de-identification tool may be demonstrated through a real-life case scenario.

It is common for Canadian and U.S. hospitals to disclose prescription records to commercial companies. This data is then analyzed to provide research and market intelligence for the pharmaceutical industry, insurers, government agencies, and in some cases, to provide drug utilization benchmarking services back to the hospitals.

Prescription records which are provided to external organizations do not contain any information that directly identifies patients. For example, patient name and address are not included in these records. The assumption is made that because the prescription information is stripped of all direct identifiers, it falls outside of the scope of privacy legislation. However, this is an assumption that should not be taken for granted.

For example, if a record contains gender, date of birth, and postal code information about the patient, then the patient would be quite easy to re-identify by linking the record with other publicly available information (e.g., public registries about homeowners and borrowers). As another example, if a record contains the gender, age, some postal code information, as well as admission and discharge dates of a patient in a hospital, then these five pieces of information would likely make the patient unique among all admitted patients. Unique patients are much easier to re-identify. These re-identification risks pose a threat to patient privacy.

In 2008, a Canadian company, Brogan Inc., requested prescription records from CHEO, as part of a larger national effort to develop a hospital prescription records database. An analysis of the CHEO data indicated that the probability of re-identifying patients using the original variables requested by Brogan was unacceptably high to the hospital. The application of Dr. El Emam's framework provided a new de-identified record layout with an acceptably low level of risk of re-identification. Specifically, admission and discharge dates were replaced with quarter/year of admission and length of stay in days; patient age was provided in weeks; and the postal code was truncated to include only the first character. In addition, the data sharing agreement between Brogan and CHEO was modified to include additional mitigating controls (e.g., an audit requirement and a breach notification protocol). Thus, CHEO was able to achieve its goal of protecting patient privacy, while preserving the level of data quality that was deemed to be necessary for Brogan to include CHEO's prescription data in the national hospital prescription record database: a positive-sum, doubly-enabling solution, that satisfied the goals of both parties – win/win, not win/lose – a powerful reflection of Privacy by Design.

About the Authors

Dr. Ann Cavoukian, Information and Privacy Commissioner of Ontario, Canada

Dr. Ann Cavoukian is recognized as one of the leading privacy experts in the world. Noted for her seminal work on Privacy Enhancing Technologies (PETs) in 1995, her concept of *Privacy by Design* seeks to embed privacy into the design specifications of technology, thereby achieving the strongest protection. An avowed believer in the role that technology can play in protecting privacy, Dr. Cavoukian's leadership has seen her office develop a number of tools and procedures to ensure that privacy is strongly protected, not only in Canada, but around the world. She has been involved in a number of international committees focused on privacy, security, technology and business, and endeavours to focus on strengthening trust and confidence in emerging technological applications. Dr. Cavoukian also serves as the Chair of the Identity, Privacy and Security Institute at the University of Toronto, Canada and is a member of the Future of Privacy Advisory Board. Reappointed as Commissioner for an unprecedented third term, Dr. Cavoukian intends to grow *Privacy by Design* and hopes to make it go "viral."

Dr. Khaled El Emam, Canada Research Chair in Electronic Health Information CHEO Research Institute and University of Ottawa

Dr. Khaled El Emam is an Associate Professor at the University of Ottawa, Faculty of Medicine and the School of Information Technology and Engineering, a senior investigator at the Children's Hospital of Eastern Ontario Research Institute, and a Canada Research Chair in Electronic Health Information at the University of Ottawa. His main area of research is developing techniques for health data anonymization. Previously Khaled was a Senior Research Officer at the National Research Council of Canada, and prior to that he was head of the Quantitative Methods Group at the Fraunhofer Institute in Kaiserslautern, Germany. He has co-founded two companies to commercialize the results of his research work. In 2003 and 2004, he was ranked as the top systems and software engineering scholar worldwide by the Journal of Systems and Software based on his research on measurement and quality evaluation and improvement, and ranked second in 2002 and 2005. He holds a Ph.D. from the Department of Electrical and Electronics, King's College, at the University of London (UK).

Published: March 2010

Information and Privacy Commissioner of Ontario
2 Bloor Street East, Suite 1400
Toronto, Ontario • M4W 1A8 • Canada

Telephone: 416-326-3333 • 1-800-387-0073
Facsimile: 416-325-9195
TTY (Teletypewriter): 416-325-7539
Website: www.ipc.on.ca
Email: info@ipc.on.ca