29E CONFÉRENCE INTERNATIONALE DES COMMISSAIRES
À LA PROTECTION DES DONNÉES ET DE LA VIE PRIVÉE

LES HORIZONS DE LA PROTECTION DE LA VIE PRIVÉE

# TERRA INCOGNITA

PRIVACY HORIZONS

29TH INTERNATIONAL CONFERENCE OF
DATA PROTECTION AND PRIVACY COMMISSIONERS

## Séance d'information
## Dragon : *Quand la loi rencontre la technologie*
## Le forage des données

## Information Session
## "*Law Meets Technology*" Dragon
## Data Mining

28 septembre/September 28
9h00 – 10h00

Série Terra Incognita, cahier de travail # 15/Terra Incognita, workbook series # 15

Office of the Privacy Commissioner of Canada
Commissariat à la protection de la vie privée du Canada

# Table des matières / Table of contents


**Biographies — Conférenciers**

Mme Philippa Lawson — Présidente . . 2

M. Peter Fleischer . . . . . 2

M. Bradley Malin, Ph.D. . . . . 3

M. Richard Rosenberg, Ph.D. . . . . 3

**Résumé de la séance d'information** . . . . . 6

**Bibliographie** . . . . . 9

**Document de travail : « *Forage de données et sécurité nationale* »**

Introduction . . . . . 14

L'état en tant que consommateur d'information . . . . 15

Forage de données . . . . . 17

Risques associés au forage de données . . . . . 19

Forage de données et lois canadiennes en matière de protection de la vie privée . . . . . 25

Mesures de protection de la vie privée liées au forage de données . . . 28

Conclusion . . . . . 33

Notes de bas de page . . . . . 34

**Biographies — Speakers**

Ms. Philippa Lawson — Chair . . . . . 2

Mr. Peter Fleischer . . . . . 2

Dr. Bradley Malin . . . . . 3

Dr. Richard Rosenberg . . . . . 3

**Information Session Summary** . . . . . 6

**Bibliography** . . . . . 9

**Background Paper : "*Data Mining and National Security*"**

Introduction . . . . . 14

The state as an information consumer . . . . . 15

Data mining . . . . . 16

Data mining risks . . . . . 18

Data mining & Canada's privacy laws . . . . 23

Privacy measures for data mining . . . 25

Conclusion . . . . . 30

Endnotes . . . . . 30

## Biographies

### Présidente : Mme Philippa Lawson

Philippa Lawson est la directrice de la Clinique d'intérêt public et de politique d'Internet du Canada (CIPPIC), qui relève de la Faculté de droit de l'Université d'Ottawa. De 1991 à 2003, elle a pratiqué le droit administratif au Centre pour la défense de l'intérêt public d'Ottawa (Canada) et a représenté des groupes de consommateurs. M$^{me}$ Lawson est un défenseur de la vie privée de renommée nationale. Depuis le début des années 1990, elle apporte sa collaboration à des organismes de défense des consommateurs canadiens et internationaux dans de nombreux dossiers liés à la protection de la vie privée, notamment la réglementation de l'identification de l'appelant et le télémarketing, l'élaboration et la mise en œuvre d'une loi sur la protection des données s'appliquant au secteur privé et, de façon générale, la protection de la vie privée en ligne. Elle fait partie du comité qui a rédigé la norme nationale du Canada en matière de protection de la vie privée. Elle travaille comme cochercheuse/ collaboratrice dans le cadre du projet *On The Identity Trail* (voir www.idtrail.org), financé par le Conseil de recherches en sciences humaines, et elle est la chercheuse principale d'un projet de recherche pluriinstitutionnel ontarien sur le vol d'identité. Elle a dirigé la rédaction d'un rapport sur l'industrie canadienne du courtage de données et en est la coauteure. Ce rapport, publié en 2006, peut être consulté sur le site Web de la CIPPIC (www.cippic.ca).

### Conférenciers

#### M. Peter Fleischer

Peter Fleischer travaille à titre de conseiller en matière de protection internationale des renseignements personnels pour Google. Son travail consiste à faire en sorte que Google protège les renseignements personnels de ses usagers, respecte toutes ses obligations juridiques en matière de protection des renseignements personnels et aide à rehausser le niveau sur le plan de la protection de la vie privée dans Internet. Il est particulièrement déterminé à travailler en collaboration avec les intervenants, défenseurs et autorités de réglementation dans le domaine de la protection des renseignements personnels pour veiller à ce que Google réponde

## Biographies

### Chair : Ms. Philippa Lawson

Philippa Lawson is Director of the Canadian Internet Policy and Public Interest Clinic (CIPPIC) at the University of Ottawa, Faculty of Law. From 1991 to 2003, she practiced administrative law and consumer advocacy with the Public Interest Advocacy Centre in Ottawa, Canada. Ms. Lawson is a nationally recognized privacy advocate, and has worked with Canadian and international consumer organizations since the early 1990s on many privacy-related issues, including the regulation of Caller ID and telemarketing, the development and implementation of private sector data protection legislation, and online privacy generally. She is a member of the committee that drafted Canada's national privacy standard, is a co-investigator/collaborator on the SSHRC-funded "On The Identity Trail" project (see www.idtrail.org), and is a lead investigator for an Ontario-based multi-institution research project on identity theft. She directed and co-authored a 2006 report on the Canadian data-brokerage industry, accessible from the CIPPIC website at www.cippic.ca .

### Speakers

#### Mr. Peter Fleischer

Peter Fleischer works as Google's Global Privacy Counsel. His job is to ensure that Google protects its users' privacy, meets all privacy legal obligations, and helps to raise the bar in terms of privacy protection on the Internet. Mr. Fleischer is particularly committed to engaging with privacy stakeholders, advocates and regulators to ensure that Google is responsive to their privacy expectations. He works closely with public policy makers around the world to help update data protection concepts for the Information Age. Peter Fleischer has over 10 years' experience in the field of data protection, including his prior position as Microsoft's privacy lead for Europe and

bien aux attentes en matière de protection de la vie privée. Il travaille en étroite collaboration avec des décideurs publics du monde entier afin de les aider à mettre à jour les concepts de protection des données pour l'ère de l'information. Il possède plus de 10 ans d'expérience dans le domaine de la protection des données, ce qui comprend ses activités antérieures de responsable de la protection de la vie privée pour l'Europe et de directeur des affaires réglementaires à Microsoft. Il a fait ses études aux États-Unis (Harvard College et Harvard Law School) et en Allemagne (LMU- Munich) et travaille depuis dix ans à Paris. C'est, par ailleurs, un décrocheur qui a abandonné ses études dès l'école secondaire.

**M. Bradley Malin, Ph.D.**

Bradley Malin est professeur adjoint d'informatique biomédicale à la faculté de médecine de l'Université Vanderbilt et il a obtenu une nomination secondaire à la faculté de génie. Il est titulaire d'un baccalauréat en biologie moléculaire, d'une maîtrise en découverte des connaissances et en exploration de données, d'une maîtrise en gestion et politique publiques, de même que d'un doctorat en informatique, diplômes qu'il a tous obtenus à l'Université Carnegie Mellon. Il a publié de nombreux articles scientifiques sur l'informatique biomédicale, l'exploration des données et des liens informatiques et la protection des renseignements personnels. Ses recherches portant sur les bases de données génétiques et sur la protection de la vie privée lui ont valu plusieurs prix de l'American Medical Informatics Association et d'autres organismes internationaux. Il a dirigé plusieurs ateliers sur la protection de la vie privée et le forage des données pour le compte du IEEE et de l'ACM. De 2004 à 2006, il a été rédacteur en chef du *Journal of Privacy Technology* (JOPT) et il est le directeur scientifique invité d'un numéro spécial sur la protection des renseignements personnels et le forage des données, qui sera publié par *Data and Knowledge Engineering*.

**M. Richard Rosenberg, Ph.D.**

Richard S. Rosenberg est professeur émérite au Département d'informatique de l'Université de la Colombie-Britannique. Il s'intéresse à l'intelligence artificielle (IA) et aux répercussions sociales des ordinateurs. Ses recherches portent

Director of Regulatory Affairs. He was educated in the US (Harvard College and Harvard Law School) and in Germany (LMU- Munich), and has worked for the last decade in Paris.

**Dr. Bradley Malin**

Bradley Malin is an Assistant Professor of Biomedical Informatics in the School of Medicine at Vanderbilt University and holds a secondary appointment in the School of Engineering. He received a bachelor's in molecular biology, a master's in knowledge discovery and data mining, a master's in public policy and management, and a doctorate in computer science, all from Carnegie Mellon University. He is the author of numerous scientific articles on biomedical informatics, data and link mining, and data privacy. His research in genetic databases and privacy has received several awards from the American and International Medical Informatics Associations. He has chaired various workshops on privacy and data mining for the IEEE and ACM. From 2004 through 2006 he was the managing editor of the Journal of Privacy Technology (JOPT) and is the guest editor for an upcoming special issue on privacy and data mining for the journal Data and Knowledge Engineering.

**Dr. Richard Rosenberg**

Dr. Richard S. Rosenberg is a Professor Emeritus in the Department of Computer Science, at the University of British Columbia. His research interests include the social impact of computers and Artificial Intelligence (AI). His work on the

sur des sujets de préoccupation tels que la protection des renseignements personnels, la liberté d'expression, les droits de propriété intellectuelle, l'accès universel, le travail et l'éducation. Il a écrit de nombreux articles sur les questions liées à la protection de la vie privée, la liberté d'expression et l'éthique. Il a comparu devant des comités législatifs fédéraux et provinciaux et a présenté des exposés devant le National Research Council des États-Unis. Son livre le plus récent s'intitule *The Social Impact of Computers* (3e édition, San Diego, CA, Elsevier Academic Press, 2004). Il fait partie du conseil d'administration de la Civil Liberties Association de la Colombie-Britannique et préside la Freedom of Information and Privacy Association de cette province.

social impact of computers includes such areas of concern as privacy, freedom of expression, intellectual property rights, universal access, work and education. He has written many papers on privacy issues, free speech, and ethics. He has appeared before Federal parliamentary and provincial legislative committees, and has made presentations before the U.S. National Research Council. His most recent book is *The Social Impact of Computers*, 3rd Edition, San Diego, CA: Elsevier Academic Press, 2004. He is on the Board of the BC Civil Liberties Association, and the president of the BC Freedom of Information and Privacy Association.

29E CONFÉRENCE INTERNATIONALE DES COMMISSAIRES
À LA PROTECTION DES DONNÉES ET DE LA VIE PRIVÉE

LES HORIZONS DE LA PROTECTION DE LA VIE PRIVÉE

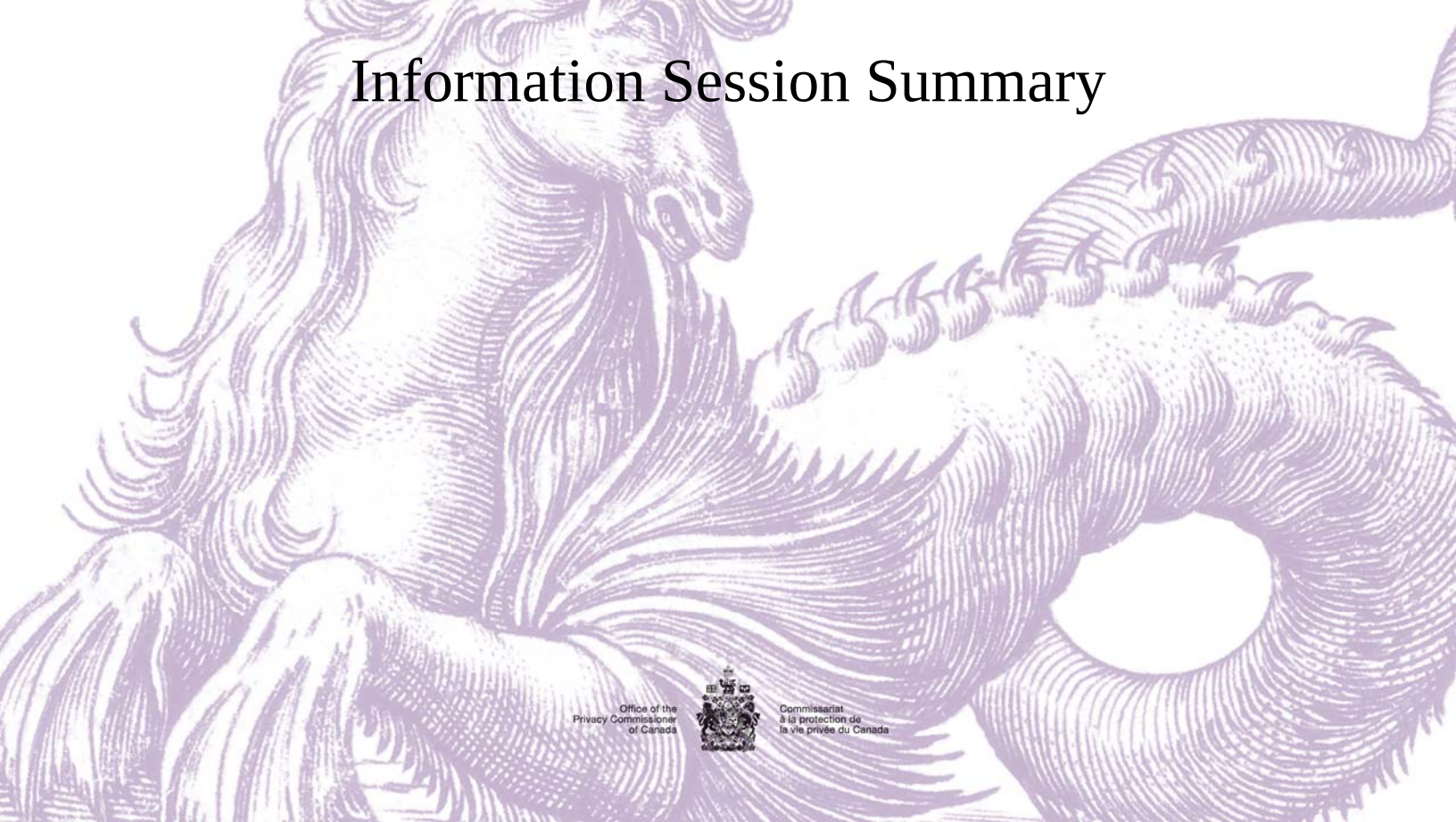
# TERRA INCOGNITA

PRIVACY HORIZONS

29TH INTERNATIONAL CONFERENCE OF
DATA PROTECTION AND PRIVACY COMMISSIONERS

## Résumé de la séance d'information

## Information Session Summary



Office of the Privacy Commissioner of Canada
Commissariat à la protection de la vie privée du Canada

## Séance d'information sur le forage des données – Résumé

Les technologies informatiques ont mené à la création d'une énorme industrie d'analyse, d'entreposage et de forage de données. Cette industrie repose sur le besoin toujours croissant des sociétés et des gouvernements de se procurer des renseignements détaillés sur des personnes, telles que des clients, des patients, des citoyens, et sur des menaces potentielles pour la sécurité. Les sociétés veulent en savoir le plus possible sur les consommateurs de façon à pouvoir mieux cibler leurs stratégies de marketing et augmenter ainsi leurs profits. Les gouvernements veulent repérer les menaces pour la sûreté publique et la sécurité nationale afin de prévenir des catastrophes publiques. Les chercheurs en soins de santé souhaitent améliorer leur capacité de prédire, de diagnostiquer, de prévenir et de traiter des maladies en recueillant des données cliniques et en surveillant de près des cas particuliers. Dans tous ces exemples, le forage des données est utilisé pour recueillir, trier et extraire des renseignements détaillés sur des personnes.

À l'aide des outils de forage de données, on peut explorer des bases de données pour y repérer des tendances cachées et de l'information prédictive qui serait autrement difficile à déceler. Ces outils servent à prédire de futurs comportements et tendances, permettant ainsi aux organisations de prendre des décisions en se fondant sur des analyses prospectives. En réponse aux préoccupations évidentes que suscite cette pratique sur le plan de la vie privée, les informaticiens tentent depuis quelques années de mettre au point des méthodes de forage de données moins « envahissantes », c'est-à-dire des méthodes qui protégeraient la vie privée des personnes tout en permettant de fournir aux chercheurs les renseignements dont ils ont besoin.

Au cours de cet atelier d'une heure, les participants examineront les techniques et pratiques de forage de données dans les secteurs public et privé, en vue de comprendre leurs répercussions sur la vie privée.

M. Bradley Malin, informaticien, professeur et expert en découverte des connaissances et en forage des données, offrira un aperçu des méthodes de collecte et de forage des données

## Data Mining Information Session - Summary

Computer technologies have spawned the creation of an enormous industry in data analytics, data warehousing and data mining. This industry is driven by ever-growing demand from corporations and governments for detailed information about individuals as consumers, patients, citizens, and potential security threats. Corporations want to know as much as possible about individual consumers so that they can more effectively target their marketing and thereby increase profits. Governments want to identify threats to national security and public safety so that they can prevent public disasters. Healthcare researchers want to improve their ability to predict, diagnose, prevent and treat diseases by gathering clinical data and tracking individual cases. In all cases, data mining is used to gather, sort, and extract detailed information about individuals.

Data mining tools scour databases for hidden patterns, finding predictive information that is otherwise not evident. They predict future trends and behaviours, allowing organizations to make decisions based on prospective analyses. In response to the obvious privacy concerns with this practice, computer scientists in recent years have work on "privacy preserving" methods of data mining – methods that would preserve individual privacy while still providing researchers with the information they want.

This one-hour workshop will examine data mining techniques and practices in both public and private sectors, with a view to understanding their privacy implications.

Dr. Bradley Malin, a computer scientist, professor and expert in knowledge discovery and data mining will provide an overview of data collection and data mining methods, with a focus on real world applications such as web mining, consumer personalization, social network analysis, and surveillance. He will then discuss specific data mining goals in population-based healthcare research, with recent examples drawn from clinical genomics and personalized medicine. Dr. Malin will also highlight emerging methods for facilitating data mining endeavors without violating personal privacy.

Peter Fleischer, Google's Global Privacy Counsel, will describe Google's data mining practices with

en mettant l'accent sur leurs applications concrètes, telles que l'exploration du Web, la personnalisation des clients, l'analyse des réseaux sociaux et la surveillance. Il passera ensuite en revue certains objectifs du forage de données dans le domaine de la recherche sur les soins de santé communautaires au moyen d'exemples liés à la génomique clinique et à la médecine personnalisée. M. Malin décrira également les nouvelles méthodes destinées à faciliter le forage des données sans porter atteinte au droit à la vie privée des personnes.

M. Peter Fleischer, conseiller en matière de protection internationale des renseignements personnels pour Google, décrira les pratiques de forage de données de Google en ce qui concerne l'exploration du Web et le Gmail, et expliquera les mesures que Google prend pour protéger la vie privée des personnes dans le cadre de ces activités.

M. Richard Rosenberg, informaticien et professeur émérite à l'Université de la Colombie-Britannique, se concentrera sur les risques pour la vie privée que représente le forage de données dans les secteurs public et privé. Il expliquera comment le gouvernement américain explore les données du secteur privé pour établir des profils de personnes à des fins antiterroristes, et décrira les répercussions sur la vie privée des pratiques de marketing affinées, mises au point grâce au forage des données.

plain the measures Google takes to protect individual privacy in the context of these activities.

Dr. Richard Rosenberg, computer scientist and professor emeritus at U.B.C., will focus on the risks of data mining for individual privacy in both the public and private sector contexts. He will discuss the U.S. government's use of private sector data mining to profile individuals for anti-terrorism purposes, as well as the privacy impacts of relentless and fine-tuned commercial marketing practices facilitated by data mining.

29E CONFÉRENCE INTERNATIONALE DES COMMISSAIRES
À LA PROTECTION DES DONNÉES ET DE LA VIE PRIVÉE

LES HORIZONS DE LA PROTECTION DE LA VIE PRIVÉE

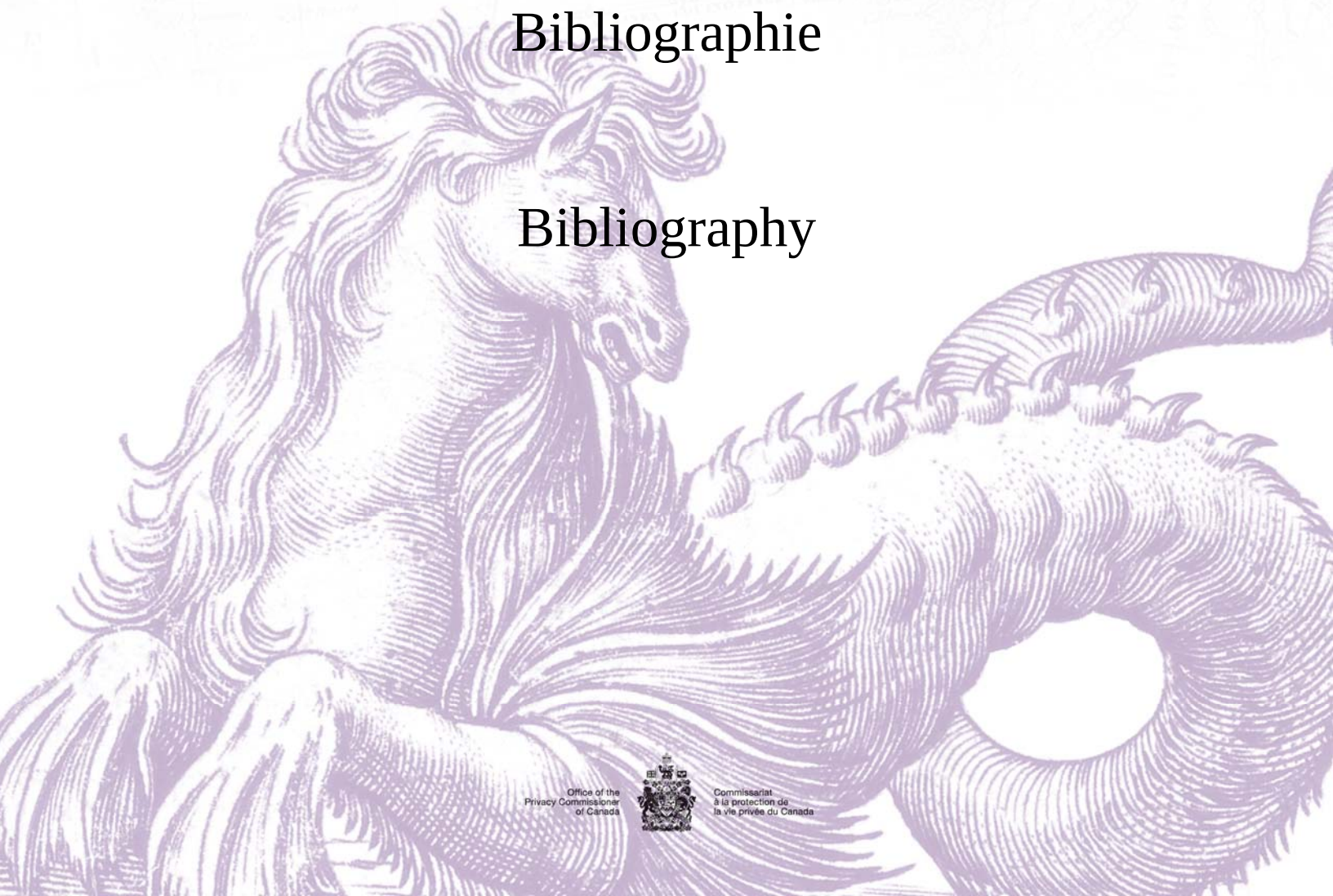
# TERRA INCOGNITA

PRIVACY HORIZONS

29TH INTERNATIONAL CONFERENCE OF
DATA PROTECTION AND PRIVACY COMMISSIONERS

# Bibliographie

# Bibliography



Office of the Privacy Commissioner of Canada
Commissariat à la protection de la vie privée du Canada

## Références choisies

### Forage des données : Méthodes et pratiques

Nemati, Hamid, et Christopher D. BARKO. *Organizational Data Mining: Leveraging Enterprise Data Resources for Optimal Performance*. Hershey: Idea Group, 2004.

Cet ouvrage passe en revue les applications liées à l'entreposage et au forage des données pour les organisations. Il décrit les aspects techniques et organisationnels des méthodes de forage de données en s'appuyant sur des études de cas d'applications commerciales réelles.

Berson, Alex, Stephen SMITH et Kurt THEARLING. *Building Data Mining Applications for CRM*. New York: McGraw-Hill, 2000.

Cet ouvrage, qui s'adresse à des professionnels de la technologie et du marketing, offre un aperçu des pratiques de forage de données dans le secteur privé et comporte des lignes directrices sur les méthodes de forage de données respectueuses de la vie privée.

### Méthodes de forage des données respectueuses de la vie privée

Privacy Preserving Data Mining: bibliographie en ligne (2006)
<http://www.csee.umbc.edu/~kunliu1/research/privacy_review.html>

Privacy Preserving Data Mining: site Web comportant des références (2004)
<http://www.cs.ualberta.ca/%7Eoliveira/psdm/psdm_index.html>

Vaidya, Jaideep, Chris CLIFTON et Michael ZHU. *Privacy Preserving Data Mining*. New York: Springer, 2006.

Cet ouvrage donne un aperçu détaillé des méthodes et techniques de forage de données respectueuses de la vie privée, ainsi que des problèmes en suspens. Il explique comment ces méthodes permettent le forage de données tout en

## Selected References

### Data Mining: Methods and Practices

Nemati, Hamid and Christopher D. Barko. Organizational Data Mining: Leveraging Enterprise Data Resources for Optimal Performance. Hershey: Idea Group, 2004.

This text provides an account of data warehousing and data mining applications for organizations. It explains technical and organizational aspects of data mining techniques, supplemented by case studies of real commercial applications.

Berson, Alex, Stephen Smith, and Kurt Thearling. Building Data Mining Applications for CRM. New York: McGraw-Hill, 2000

This book, written for technology and marketing professionals, provides an overview of data mining in the private sector, and includes guidelines for privacy-preserving methods of data mining.

### Privacy Preserving Data Mining

Privacy Preserving Data Mining: an online bibliography (2006)
<http://www.csee.umbc.edu/~kunliu1/research/privacy_review.html>

Privacy Preserving Data Mining: website with references (2004)
<http://www.cs.ualberta.ca/%7Eoliveira/psdm/psdm_index.html>

Vaidya, Jaideep, Chris Clifton and, Michael Zhu. Privacy Preserving Data Mining. New York: Springer, 2006.

This book provides a comprehensive overview of available approaches, techniques and open problems in privacy preserving data mining. It demonstrates how these approaches can achieve data mining, while operating within legal and commercial restrictions that forbid release of data. It is designed for practitioners and researchers in industry.

Sweeney, Latanya. "Privacy-Enhanced Linking" Special Interest Group on Knowledge

respectant les restrictions juridiques et commerciales qui interdisent la communication de données. Il s'adresse aux intervenants et aux chercheurs de l'industrie.

Sweeney, Latanya. *"Privacy-Enhanced Linking" Special Interest Group on Knowledge Discovery and Data Mining* 7.2 (2005): 72-75. 11 mai 2007. <http://www.acm.org/sigs/sigkdd/explorations/issues/7-2-2005-12/9-Sweeney.pdf>.

Sweeney examine les conséquences pour la vie privée de « l'analyse des liens », un domaine de l'informatique en plein essor qui consiste à recourir à des algorithmes pour glaner de l'information provenant de diverses sources. Elle propose que l'on tienne compte des pratiques équitables de traitement de l'information dans l'élaboration d'algorithmes, de sorte que la responsabilité de protéger le droit à la vie privée incomberait aux concepteurs de logiciels. Elle expose les avantages et les inconvénients d'une méthode d'établissement de liens davantage respectueuse de la vie privée et démontre que le droit à la vie privée peut être protégé au moyen d'une technologie destinée à recueillir des renseignements.

**Considérations liées à la vie privée associées au forage des données**

Jonas, Jeff, et Jim HARPER. *Effective Counterterrorism and the Limited Role of Predictive Data Mining. Policy Analysis*. 584 (2006) Cato Institute. 14 mai 2007. <http://www.cato.org/pubs/pas/pa584.pdf>.

Jonas et Harper font valoir que le forage de données de prédiction constitue un moyen inefficace et coûteux de contrer le terrorisme compte tenu de son faible taux d'exactitude. Pour être efficace, il faudrait être en mesure de comprendre le contexte et tirer des conclusions, ce que le forage de données de prédiction ne peut faire.

Loukidelis, David. *Information Technology, National Security and Privacy Protection*. Conférence sur l'Institut canadien

Discovery and Data Mining 7.2 (2005): 72-75. 11 May 2007 <http://www.acm.org/sigs/sigkdd/explorations/issues/7-2-2005-12/9-Sweeney.pdf>.

Sweeney considers the privacy implications of 'link-analysis', a growing area of computer science that constructs algorithms to glean information from different sources. She proposes that Fair Information Practices be addressed within the link-algorithm, thus placing responsibility on software developers to protect privacy interests. She identifies benefits and drawbacks of 'privacy-enhanced linking', and demonstrates that privacy interests can be protected through the construction of technology used to collect information.

**Privacy Perspectives on Data Mining**

Jonas, Jeff and Jim Harper. "Effective Counterterrorism and the Limited Role of Predictive Data Mining." Policy Analysis. 584 (2006) Cato Institute. 14 May 2007 <http://www.cato.org/pubs/pas/pa584.pdf>.

Jonas and Harper argue that predictive data mining is an ineffective and costly approach to counter-terrorism, given its low rate of accuracy. Effective counter-terrorism requires an understanding of context and the making of inferences which predictive data mining does not have and cannot do.

Loukidelis, David. "Information Technology, National Security and Privacy Protection."Conference on Canadian Institute for the Administration of Justice. Toronto. 29-30 Sep. 2005. <http://www.oipc.bc.ca/publications/speeches_presentations/CIAJSpeech(RevisedFinal)(Oct3-2005).pdf>.

Loukidelis, the British Columbia Information and Privacy Commissioner, focuses on data mining for national security purposes in this paper, written in 2005. He argues that, the neutrality of technology notwithstanding, Canadian privacy laws applicable to data mining for national security require substantial re-thinking if our rights to privacy are to re-

d'administration de la justice. Toronto. 29-30 septembre. 2005. <http://www.oipc.bc.ca/publications/speeches_presentations/CIAJSpeech(RevisedFinal)(Oct3-2005).pdf>.

Dans cet article rédigé en 2005, Loukidelis, le commissaire à l'information et à la protection de la vie privée de la Colombie-Britannique, se penche sur l'utilisation du forage de données à des fins de sécurité nationale. Il fait valoir que malgré la neutralité de la technologie, les lois canadiennes sur la protection des renseignements personnels applicables au forage des données à des fins de sécurité nationale doivent être repensées pour que notre droit à la vie privée puisse conserver tout son sens face aux rapides changements technologiques.

O'Harrow, Robert. *No Place to Hide*. New York: Free Press, 2005. <http://www.noplacetohide.net/>

O'Harrow décrit en détail le complexe sécurité-industriel créé à la suite des attentats du 11 septembre par le gouvernement des États-Unis, qui s'appuie de plus en plus sur les bases de données du secteur privé pour recueillir des renseignements destinés à identifier et à surveiller des suspects. O'Harrow met en évidence le rôle clé que le forage des données joue dans cette société de surveillance.

Solove, Daniel. *The Digital Person: Technology and Privacy in the Information Age*. New York: New York University Press, 2004. <http://docs.law.gwu.edu/facweb/dsolove/Solove-Digital-Person.htm>

Solove explique comment, dans le contexte américain, on utilise les bases de données électroniques pour créer des profils de personnes détaillés à des fins de marketing et de lutte contre le terrorisme. Il explique comment l'élaboration plus ou moins contrôlée de ces « dossiers numériques » crée une situation cauchemardesque digne de Kafka, où des personnes perdent à jamais le contrôle de leurs renseignements personnels.

main meaningful in the face of rapid technological changes.

O'Harrow, Robert. No Place to Hide. New York: Free Press, 2005 <http://www.noplacetohide.net/>

O'Harrow exposes in graphic detail the security-industrial complex created in the aftermath of 9/11 by the U.S. government as it increasingly relies upon private sector databases for information used to identify and track terrorism suspects. O'Harrow highlights the key role that data mining plays in this surveillance society.

Solove, Daniel. The Digital Person: Technology and Privacy in the Information Age. New York: New York University Press, 2004. <http://docs.law.gwu.edu/facweb/dsolove/Solove-Digital-Person.htm>

Solove explains, in the U.S. context, how electronic databases are being used to create detailed profiles about individuals for purposes ranging from marketing to counter-terrorism. He explains how the relatively unchecked development of these "digital dossiers" is creating a Kafkaesque privacy nightmare in which individuals lose control over their personal information and can never regain it.

Lyon, David. Surveillance Society: Monitoring Everyday Life. Buckingham: Open University Press, 2001. <http://www.mcgraw-hill.co.uk/html/0335205461.html>

Lyon, a sociologist and surveillance studies expert at Queen's University, Canada, examines the profound effects on social ordering created by data mining and other forms of technological surveillance now becoming an inescapable fact of daily life for ordinary citizens and consumers. Using examples from North America, Europe and Asia, he points out how data mining can be used to categorize and classify people in ways that reinforce stereotypes and discrimination.

Garfinkel, Simson. Database Nation: the Death of

Lyon, David. *Surveillance Society: Monitoring Everyday Life*. Buckingham: Open University Press, 2001. <http://www.mcgraw-hill.co.uk/html/0335205461.html>

Lyon, sociologue et expert en études sur la surveillance à l'Université Queen's (Canada), examine les profondes répercussions sur l'ordre collectif du forage des données et d'autres formes de surveillance technologique auxquels les citoyens et les consommateurs ne peuvent plus échapper. À l'aide d'exemples de cas observés en Amérique du Nord, en Europe et en Asie, il explique comment le forage de données peut servir à catégoriser et à classifier les gens d'une manière qui renforce les stéréotypes et la discrimination.

Garfinkel, Simson. *Database Nation: the Death of Privacy in the 21st Century*. Beijing; Cambridge: O'Reilly, 2000. <http://press.oreilly.com/pub/pr/593>

Garfinkel explique en détail comment les nouvelles technologies électroniques portent atteinte à la vie privée des personnes. Il décrit les nombreuses méthodes employées par les autorités et les sociétés américaines pour compiler des renseignements détaillés sur des personnes et ce, souvent à leur insu. L'ouvrage continue de servir de mise en garde contre la perte de vie privée associée au développement technologique.

Privacy in the 21st Century. Beijing; Cambridge: O'Reilly, 2000. <http://press.oreilly.com/pub/pr/593>

Garfinkel provides a detailed explanation of how new electronic technologies encroach upon individual privacy. The book catalogues the many ways in which U.S. governments and corporations compile detailed information about individuals, often without our knowledge. It continues to serve as a strong warning about the loss of privacy that is accompanying technological development.

29E CONFÉRENCE INTERNATIONALE DES COMMISSAIRES
À LA PROTECTION DES DONNÉES ET DE LA VIE PRIVÉE

LES HORIZONS DE LA PROTECTION DE LA VIE PRIVÉE

TERRA INCOGNITA

PRIVACY HORIZONS

29TH INTERNATIONAL CONFERENCE OF
DATA PROTECTION AND PRIVACY COMMISSIONERS

# Forage de données et sécurité nationale

# Data Mining and National Security

Par/by:

David Loukidelis

Office of the Privacy Commissioner of Canada

Commissariat à la protection de la vie privée du Canada

## Document de travail

### Forage de données et sécurité nationale

*Document d'information concernant l'atelier sur le forage de données*[1]

[TRADUCTION]

> Si la technologie de l'information continue de se développer sans entrave, l'utilisation de la persona numérique [la représentation numérique d'une personne] aura inévitablement des effets inéquitables et opprimants sur les personnes et des conséquences de l'ordre de la répression sur la société… C'est pourquoi la recherche doit s'intéresser spécifiquement à ces menaces et évaluer la mesure dans laquelle la réglementation suffira à les prévenir ou à y faire face. Si les risques s'avèrent gérables, il faudra alors exercer des pressions politiques efficaces sur les législatures pour qu'elles imposent des mesures et des mécanismes réglementaires appropriés. Dans le cas contraire, les technologues de l'information n'auront plus qu'à se débrouiller face à leur génie et à sa bouteille vide[2].

**INTRODUCTION**

Le présent document, rédigé en 2005 en vue de la conférence annuelle de l'Institut canadien d'administration de la justice, a pour but de montrer que, en dépit de la neutralité de la technologie, nous devons revoir en profondeur les lois canadiennes relatives à la protection de la vie privée applicables au forage de données effectué pour des raisons de sécurité nationale, pour continuer à faire valoir notre droit à la vie privée face à l'évolution rapide de la technologie. Cette question est particulièrement urgente étant donné l'expansion du pouvoir des États en ce qui concerne la collecte, l'utilisation et la communication des renseignements personnels.

Il est également urgent d'agir étant donné l'inévitable adoption, au Canada, de technologies de l'information destinées à exploiter nos

## Background Paper

### Data Mining and National Security

*Background Paper for Data Mining Workshop*[1]

> If information technology continues unfettered, then use of the digital persona [the digital construct of an individual] will inevitably result in impacts on individuals which are inequitable and oppressive, and in impacts on society which are repressive… Focussed research is needed to assess the extent to which regulation will be sufficient to prevent and/or cope with these threats. If the risks are manageable, then effective lobbying of legislatures will be necessary to ensure appropriate regulatory measures and mechanisms are imposed. If the risks are not manageable, then information technologists will be left contemplating a genie and an empty bottle.[2]

**INTRODUCTION**

The thrust of this paper, written in 2005 for the annual conference of the Canadian Institute for the Administration of Justice, is to suggest that, the neutrality of technology notwithstanding, Canadian privacy laws applicable to data mining for national security require substantial re-thinking if our rights to privacy are to remain meaningful in the face of rapid technological changes. This is particularly pressing given the expansion of state power to collect, use and disclose personal information.

It is also urgent given the inevitable adoption in Canada of information technologies to exploit our personal information in the interests of national security. The technology and practice of data mining in the interests of national security serve in this paper to illustrate the challenges to privacy, and privacy laws, raised by information technology.[3]

renseignements personnels dans l'intérêt de la sécurité nationale. Dans le présent document, la pratique du forage de données à des fins de sécurité nationale et les technologies afférentes servent à illustrer les défis que la technologie de l'information pose pour la protection de la vie privée et la législation dans ce domaine[3].

## L'ÉTAT EN TANT QUE CONSOMMATEUR D'INFORMATION

La meilleure façon de comprendre les conséquences, sur la protection de la vie privée, du forage de données est de les examiner à la lumière du regroupement et du forage des renseignements personnels par le secteur privé, et de la quasi-certitude que les gouvernements consommeront de plus en plus de données issues du secteur privé.

Depuis un certain nombre d'années, les entreprises du monde entier se tournent vers des techniques d'analyse de données de plus en plus sophistiquées pour, entre autres, évaluer le risque lié au crédit, mettre sur le marché des produits et des services, et gérer leurs relations avec leurs clients. Pendant la dernière décennie en particulier, les programmes de fidélisation se sont multipliés. Certains sont administrés par les entreprises elles-mêmes, tandis que d'autres le sont par des tiers. À l'aide de ces programmes, les entreprises recueillent de l'information très détaillée sur la vie, les habitudes, les finances, les attitudes, les préférences d'achat des consommateurs, et bien d'autres choses encore. L'ampleur et les détails des bases de données sont souvent, notamment aux États-Unis, accrus par l'information glanée dans les dossiers publics tenus par les gouvernements. Ces banques de données commerciales sont fréquemment mises en vente par de grandes sociétés, comme ChoicePoint et Axiom, qui offrent une gamme de plus en plus vaste de produits d'information concernant les consommateurs à quiconque est disposé à en payer le prix[4].

Depuis le 11 septembre, au Canada, la tendance a été d'accroître les pouvoirs de l'État en vue de forcer la production de renseignements personnels à des fins de sécurité nationale. De façon générale, les gouvernements semblent de plus en plus intéressés à acquérir les renseignements personnels des bases de données commerciales. Jusqu'à maintenant, c'est

## THE STATE AS AN INFORMATION CONSUMER

The privacy implications of data mining can best be understood against the backdrop of private sector aggregation and mining of personal information and the near certainty that governments will increasingly be consumers of data flowing from the private sector.

Businesses throughout the world have for some years now turned to increasingly sophisticated data analysis techniques to among other things assess credit risk, market goods and services and manage relationships to their customers. Over the past decade in particular loyalty programs have sprouted up everywhere. Some are operated by businesses on their own behalf while other programs are operated by third parties. Through these programs, businesses are collecting very detailed information about consumers' lives, habits, finances, attitudes, purchasing preferences and so on. The scope and detail of the databases is often, notably in the United States, enhanced by information gleaned from public records maintained by governments. These commercial data banks are very often for sale, with large corporations such as ChoicePoint and Axiom offering those willing to pay an increasingly wide array of information products about consumers.[4]

Since September 11, there has been a trend in Canada toward enhanced state powers to compel production of personal information for national security purposes. Governments also appear to have an increasing appetite for personal information acquired from commercial personal information databases. To date, evidence that governments are becoming consumers of personal information from commercial sources is found mostly in the United States.[5] US surveillance initiatives have started to use both public and private sector information to create powerful databases that can be mined for intelligence. These initiatives include the public-private national security and law enforcement surveillance partnership known as MATRIX and the Pentagon's (now defunct) Total Information Awareness research project[6]. They also include CAPPS I,[7] CAPPS II and Secure Flight. National security programs involving personal information will undoubtedly continue to roll out in the US and, as the announcement of Transport Canada's Passenger Protect no-fly list initiative shows, will soon arrive in Canada.[8]

surtout aux États-Unis que l'on trouve des preuves de cette tendance[5]. Les initiatives américaines de surveillance ont commencé à utiliser l'information provenant des secteurs public et privé afin de créer de puissantes bases de données qui peuvent être forées par les services de renseignements. Ces initiatives comprennent le partenariat de surveillance public-privé pour la sécurité nationale et l'application de la loi, connu sous le nom de MATRIX, et le projet de recherche (maintenant abandonné) *Total Information Awareness* du Pentagone[6]. Il y a également le CAPPS I[7] et le CAPPS II et *Secure Flight*. Les É.-U. continueront indubitablement de créer des programmes de sécurité nationale qui utilisent les renseignements personnels et le Canada leur emboîtera bientôt le pas, comme l'indique l'annonce de la liste d'interdiction de vol établie dans le cadre du Programme de protection des passagers de Transports Canada[8].

Il serait naïf de croire que les organismes canadiens chargés de la sécurité nationale et de l'application de la loi pourront résister longtemps à l'envie d'exploiter le trésor toujours plus riche de renseignements personnels numériques du secteur privé. Le Centre d'analyse des opérations et déclarations financières du Canada (CANAFE), par exemple, a reçu l'autorisation d'acquérir de l'information des bases de données commerciales. Le Programme de protection des passagers sera un autre exemple de l'utilisation par l'État de données provenant du secteur privé à des fins de transports et de sécurité nationale. Il est certain que la mise en œuvre de ce programme dépendra en grande partie des renseignements sur les passagers recueillis auprès du secteur privé.

Il est difficile de contester l'affirmation voulant que l'efficacité des services de renseignements est proportionnelle à l'abondance des données disponibles, aussi douteuse cette proposition générale soit-elle. Étant donné que les bases de données commerciales prolifèrent, qu'elles deviennent de plus en plus exhaustives et détaillées, et que les coûts de stockage des données diminuent continuellement (ce qui tend à allonger le cycle de vie des bases de données), l'État aura beaucoup de difficulté à lutter contre l'envie d'exploiter ces riches filons du secteur privé, sans compter ceux du secteur public[9].

It would be naïve to think that Canadian national security and law enforcement agencies will long be able to resist tapping into the ever-richer trove of digital personal information that exists in the private sector. FINTRAC, for example, has been given the authority to acquire information from commercial databases. The Passenger Protect initiative will be another Canadian example of state use of private sector data for transportation and national security purposes. Passenger Protect will surely depend in large part on passenger information collected from the private sector.

The assertion that more data means better intelligence is hard to resist, however doubtful it may be as a general proposition. Yet, as commercial databases continue to proliferate, as they become more and more comprehensive and detailed, and as data storage becomes cheaper and cheaper (tending to make databases life-long in scope), it will be very difficult for the state to resist exploiting the rich lodes of data found in the private sector, never mind in the public sector.[9]

## DATA MINING

Governments will want our personal data in order to use increasingly powerful computer technologies to create knowledge. Computers can be used in a variety of ways to derive knowledge from analysis of data using bespoke or off-the-shelf software. These techniques are generally referred to as 'data mining' and they are already in widespread commercial use in Canada and elsewhere.[10] The Congressional Research Service has defined data mining this way:

> Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods (algorithms that improve their performance automatically through experience, such as neural networks or decision trees). Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction.
>
> Data mining can be performed on

**FORAGE DE DONNÉES**

Les gouvernements voudront obtenir nos renseignements personnels et se servir de technologies informatiques de plus en plus puissantes pour créer des connaissances à partir de ces données. Les ordinateurs peuvent être utilisés de diverses façons pour tirer des connaissances de l'analyse de données, par exemple à l'aide de logiciels sur mesure ou de série. Ces techniques, généralement appelées « forage de données », sont déjà largement répandues dans le milieu des entreprises au Canada et ailleurs[10]. Voici comment le Congressional Research Service a défini le forage de données :

> [TRADUCTION]
> Le forage de données suppose l'utilisation d'outils d'analyse sophistiqués pour découvrir des tendances et des relations valides, antérieurement inconnues, à partir de vastes ensembles de données. Ces outils peuvent comprendre des modèles statistiques, des algorithmes mathématiques et des méthodes d'apprentissage automatique (algorithmes qui améliorent leur rendement automatiquement par l'expérience, comme des réseaux neuronaux ou des arbres de décision). Par conséquent, le forage de données comprend non seulement la collecte et la gestion de données, mais aussi l'analyse et les prévisions.
>
> Les données qui font l'objet d'un forage peuvent se présenter sous formes quantitative, textuelle ou multimédia. Des paramètres variés peuvent guider l'examen des données. Les applications peuvent comprendre l'association (la relation entre un fait et un autre, comme l'achat d'un stylo et celui de papier), le séquençage ou l'analyse des pistes causales (un fait qui en entraîne un autre, comme la naissance d'un enfant et l'achat de couches), la classification (la découverte d'un

> data represented in quantitative, textual, or multimedia forms. Data mining applications can use a variety of parameters to examine the data. They include association (patterns where one event is connected to another event, such as purchasing a pen and purchasing paper), sequence or path analysis (patterns where one event leads to another event, such as the birth of a child and purchasing diapers), classification (identification of new patterns, such as coincidences between duct tape purchases and plastic sheeting purchases), clustering (finding and visually documenting groups of previously unknown facts, such as geographic location and brand preferences), and forecasting (discovering patterns from which one can make reasonable predictions regarding future activities, such as the prediction that people who join an athletic club may take exercise classes).[11]

A key characteristic of data mining is that analysis of an individual's personal information can create new, secondary, information about that person. The hidden patterns and subtle relationships that data mining detects may be recorded and thus become personal information of the individual whose life is being scrutinized and analyzed. Information about an individual's credit history, credit card purchases, law enforcement record or interactions, travel habits and so on may be mined to derive evidence, or even a finding, that she or he is a possible terrorist who should be put on a terrorist watch list or be kept under surveillance. This new personal information becomes part of the swelling river of data whose channels are, in the private and public sectors, ever-changing and difficult to follow, much less control. The easier it becomes to accumulate and analyze personal information on a massive scale, the greater the potential for intentional or unintentional misuse and error. As data banks and data mining grow in sophistication and extent, each person's life will become more and more open to scrutiny, with further details becoming visible with each new advance in data analysis techniques.

On the other hand, data mining can yield benefits,

> nouveau lien, par exemple l'achat de ruban à conduits et de feuilles de plastique), le regroupement (la désignation et la documentation visuelle d'ensemble de faits antérieurement inconnus, comme un emplacement géographique et des préférences de marques) et les prévisions (la découverte de relations qui permettent d'effectuer des prévisions raisonnables concernant des activités futures, comme la participation à des cours de conditionnement physique des gens qui s'abonnent à un centre sportif)[11].

Une caractéristique clé du forage de données est le fait que l'analyse des renseignements personnels d'une personne peut donner de nouvelles informations secondaires à son sujet. Des caractéristiques cachées et des rapports subtils décelés par le forage de données peuvent être consignés et constituer d'autres renseignements sur la personne dont la vie est scrutée et analysée. Le forage de données peut porter sur les antécédents en matière de crédit d'une personne, ses achats effectués par cartes de crédit, des dossiers ou ses interactions liés à l'application des lois, ses habitudes de voyage et bien d'autres aspects de sa vie, pour trouver des indications, ou même conclure, que cette personne pourrait être un(e) terroriste dont le nom devrait faire partie d'une liste de surveillance ou dont les activités devraient être surveillées. Ces nouveaux renseignements personnels deviennent partie intégrante d'un flot croissant de données dont les voies de communication, dans les secteurs privé et public, changent constamment et sont difficiles à suivre et encore plus difficiles à contrôler. Plus il est facile d'accumuler et d'analyser des renseignements personnels à grande échelle, plus les risques d'utilisation inappropriée, intentionnellement ou pas, et d'erreurs sont grands. Plus les banques de données et le forage de données se complexifieront et prendront de l'ampleur, plus la vie des personnes sera exposée à l'analyse et plus chaque nouveau progrès technique d'analyse des données permettra de révéler des détails à leur sujet.

En revanche, le forage de données peut offrir des avantages; par exemple, des services améliorés

for example, in the form of improved services and greater efficiency. It may be that data mining will offer benefits for national security, although there should be no assumption as to the benefits—careful research is needed in each case to establish whether benefits can be realized. Certainly, the utility of data mining for national security purposes has been recognized by the US Congress, which has recommended its use by US agencies to combat terrorism.[12]

There are, however, a variety of concerns associated with data mining, which are heightened when data mining is used by the state for national security purposes. Experts have flagged the risks for a number of years and, more recently, US government studies of data mining initiatives have also noted the risks and recommended action.[13]

These risks are generally associated with use of data mining for surveillance of individuals, groups or populations. In the case of individuals or small groups, surveillance may be predicated on suspicion derived from other sources or it may be mass surveillance. Government should assess the risks associated with data mining for surveillance purposes—some of which are outlined below—before data mining expands in Canada, as it is likely to do, and then act to protect privacy.

## DATA MINING RISKS

The privacy risks of data mining are varied in nature and significance. While there is broad consensus in the literature about what the risks are, consensus on a hierarchy of risks is not evident. For this reason, and given the overview nature of this paper, the following outline is selective—not all of the risks are mentioned, they are not presented in any particular order, and they overlap in some respects. The goal is to establish that there are risks and then recommend action, since, when these risks are realized, they can entail real and possibly serious harm to individuals.[14]

An overall concern associated with data mining—and other information technologies—is the tendency to attribute reliability or even infallibility to the products of technology. It is therefore important that the following admonition be rigorously respected when creating and operating data mining projects:

> Although these techniques are

et une plus grande efficacité. Il peut procurer des avantages sur le plan de la sécurité nationale, même s'il ne faut présupposer aucun avantage particulier — seule une étude attentive peut établir pour chaque cas si le forage de données est profitable ou non. Une chose est sûre, le Congrès américain a reconnu l'utilité de cette technique à des fins de sécurité nationale et recommandé son utilisation par les organismes américains engagés dans la lutte contre le terrorisme[12].

Néanmoins, le forage de données suscite une gamme de préoccupations, lesquelles s'intensifient lorsque c'est l'État qui l'utilise à des fins de sécurité nationale. Depuis un certain nombre d'années, les experts attirent l'attention sur les risques et, plus récemment, le gouvernement américain a signalé ces risques et formulé des recommandations dans des études concernant certaines initiatives de forage de données[13].

Ces risques sont généralement associés à la surveillance de personnes, de groupes ou de populations. Dans le cas de personnes ou de petits groupes, la surveillance peut être fondée sur des soupçons issus d'une autre source ou résulter d'une surveillance à grande échelle. Avant que cette technique ne se répande au Canada, ce qui se produira probablement, le gouvernement devrait évaluer les risques associés au forage de données à des fins de surveillance — dont certains sont décrits ci-dessous — et prendre les mesures nécessaires pour protéger la vie privée.

## RISQUES ASSOCIÉS AU FORAGE DE DONNÉES

Les risques associés au forage de données sont de nature et d'importance variées. Bien qu'il se dégage de la documentation un vaste consensus sur les risques courus, on ne s'entend pas nécessairement sur l'ordre d'importance de ces risques. Comme le présent document ne vise qu'à donner une vue d'ensemble de la situation, son contenu est sélectif — tous les risques ne sont pas mentionnés, ils ne sont pas présentés dans un ordre particulier et se chevauchent à certains égards. L'objectif est d'établir qu'il existe des risques et de recommander des mesures pour les atténuer, étant donné que leur concrétisation peut entraîner des dommages réels et possiblement graves pour les personnes[14].

> powerful, it is a mistake to view data mining and automated data analysis as complete solutions to security problems. Their strength is as tools to assist analysts and investigators. They can automate some functions that analysts would otherwise have to perform manually, they can help prioritize attention and focus an inquiry, and they can even do some early analysis and sorting of masses of data. But in the complex world of counter-terrorism, they are not likely to be useful as the only source for a conclusion or decision. When these techniques are used as more than an analytical tool, the potential for harm to individuals is far more significant.[15]

### *Poor data quality*

The data quality problem can have a variety of causes. Missing data, fragmented data, outdated information and poorly authenticated or unauthenticated data can all contribute to error. Where data are acquired from commercial sources, data quality may suffer because the information was originally collected for purposes that do not require high accuracy.

Take an apparently trivial example from domestic life. Personal information collected through frequent shopper programs might find its way into databases exploited for national security data mining. Affinity programs do not require high assurances of identification upon enrolment and affinity cards may be shared among family, friends or mere acquaintances. Sharing of affinity cards could, for example, lead to false association of certain purchases or habits, and therefore religious beliefs, with the putative registered shopper. Moreover, the shopper's identity is likely not to have been robustly authenticated at the outset.[16] Use of such data for national security purposes, perhaps in conjunction with other flawed data, may paint an inaccurate portrait of an individual or, as the errors multiply across the class, skew more broadly-based analyses.

Une préoccupation générale associée au forage de données — et à d'autres technologies de l'information — est la tendance à accorder de la crédibilité aux produits technologiques, voire à les croire infaillibles. Il est donc important que les responsables de l'élaboration et de la mise en œuvre de projets de forage de données respectent rigoureusement les directives suivantes :

> [TRADUCTION]
> Même si le forage et l'analyse automatique de données sont des techniques puissantes, c'est une erreur de les considérer comme des solutions complètes aux problèmes de sécurité. La force de ces techniques réside en cela qu'elles sont des outils pour les analystes et les chercheurs/enquêteurs. Elles peuvent automatiser certaines fonctions que les analystes devraient, autrement, réaliser manuellement. Elles peuvent aider à déterminer les priorités et les points centraux d'une recherche. Elles peuvent également effectuer des analyses et des classements préliminaires de grandes quantités de données. Cependant, dans le monde complexe de la lutte contre le terrorisme, il n'est probablement pas possible de tirer des conclusions ou de prendre des décisions à partir de cette seule source d'information. Lorsque ces techniques ne sont pas uniquement utilisées comme outils d'analyse, les dommages qu'elles risquent de causer aux personnes sont beaucoup plus importants[15].

***Mauvaise qualité des données***

Les causes du problème de la qualité des données sont diverses. Les données manquantes, fragmentées, obsolètes, douteuses ou non authentifiées peuvent toutes être sources d'erreurs. Quand les données proviennent de

***Data leakage (intentional and accidental)***

Like water, information flows and it will find a way to escape. Data can and often will be spilled in a variety of ways. These can include loss or theft of poorly secured servers or storage media, the hacking of systems and retention of copies of data by contractors temporarily authorized to possess the data for service-related purposes. Data leakage magnifies the risk of misuse, including through inappropriate publication of damaging information.

***Data retention***

The information technology phenomena that are driving development of data mining techniques also enhance the likelihood that personal information fed into and derived from data mining projects will linger for longer and longer. Data storage is becoming cheaper every day and the technologies to find and exploit archived data are advancing all the time. These factors will be partly responsible for creation of the digital personality—the digital construct of each of us that will, in important ways, mediate between our true selves and the rest of the world, notably government.[17]

What makes the description of a person in today's global data world especially worrisome is that the portrait created is not a portrait of one's true self. Our digital selves, in other words, can hardly reflect our true selves. Analysis of data can create a caricature, but it does not create a person—and the essence of privacy is maintaining your personhood. This is of more than philosophical concern. The pooling of data streams and analysis of the data can have real and costly consequences for individuals. The longer these data linger, the harder it is to correct errors or to ensure currency, particularly where the information system is a secret national security system. Even where the data are accurate, their permanent retention will raise serious problems for those who might wish, and deserve, to be able to move on with their lives. It will become more and more difficult to obscure the folly, for example, of a youthful flirtation with radical politics. Aware of the power of our digital personae, we may withdraw or tend to the anodyne. This is hardly conducive to individual fulfillment or the wellbeing of society and government.

sources commerciales, leur qualité peut laisser à désirer, parce que l'information a été recueillie au départ à des fins qui n'exigeaient pas une grande exactitude.

Prenons un exemple apparemment banal de la vie courante. Les renseignements personnels recueillis dans le cadre de programmes d'encouragement à l'achat peuvent aboutir dans des bases de données exploitées à des fins de forage de données pour la sécurité nationale. Les programmes d'affinité ne demandent pas aux nouveaux inscrits des preuves d'identité solides et les cartes d'affinité peuvent être utilisées par des membres de la famille, des amis ou de simples connaissances. Par exemple, la multiplication du nombre de personnes qui utilisent une carte d'affinité peut donner lieu à des interprétations erronées des habitudes d'achat et, par conséquent, des croyances religieuses du consommateur dûment inscrit. En outre, l'identité de l'acheteur n'a probablement pas été authentifiée de façon rigoureuse au départ[16]. L'utilisation de ces données à des fins de sécurité nationale, parfois avec d'autres données aussi peu fiables, peut peindre un portrait inexact d'une personne ou, si les erreurs se multiplient, fausser les analyses plus globales.

***Fuites de données (intentionnelles et accidentelles)***

Comme l'eau, l'information « se répand » et trouve toujours moyen de passer par les interstices. Les données peuvent fuir de diverses façons, comme lors de la perte ou du vol de serveurs ou de supports de données mal sécurisés, du piratage de systèmes ou de la conservation de copies des données par des agents contractuels temporairement autorisés à utiliser les données. Les fuites de données accroissent les risques de mauvaise utilisation, y compris la publication inappropriée de renseignements dommageables.

***Conservation des données***

Les phénomènes de technologie de l'information qui sont à la base des techniques de forage de données accroissent également la probabilité que les renseignements personnels rassemblés dans le cadre de projets de forage de données soient conservés de plus en plus longtemps. Le stockage des données est de moins en moins coûteux et les technologies permettant de trouver

***False positives[18]***

US media reported last year the Senator Ted Kennedy was told he could not board more than one domestic flight because the name T. Kennedy was on the US no-fly list, CAPPS I. His name generated a hit when run against the list, so he was banned from flying. These were false positives—he was not the T. Kennedy on the list and should not have been flagged as a security risk, even if the 'real' T. Kennedy should have been. Senator Kennedy ultimately caught his flights because he was able to persuade managers on the scene that he was not a risk. Someone not as well known might not be so fortunate. A number of examples have been reported where individuals have been kept off flights in the US due to false positives.[19] This is more than a minor hassle for those unable to visit an ailing parent or attend a loved one's funeral. This is more than merely inconvenient for those who must fly on business. If they cannot travel when required, their jobs are in real jeopardy (and it will certainly not help an employee if the employer discovers that the employee cannot fly because she or he is on a terrorist no-fly list).

The problem of false positives is not unique to data mining, but our tendency to trust data and the scope for significant numbers of false positives promise, in combination, to make this a pressing issue. The risk of false positives is a system-design issue. If a data mining application cannot distinguish the 'noise' of ordinary behaviour from signs of possible terrorist activity, individuals will falsely be singled out for investigation or wrongly be put on watch lists. This is not to say that data mining should never be used for terrorism-related work. Rather, effective technological solutions must be found, and meaningful procedural and substantive protections must be implemented, to guard against the impact of false positives.

***Function creep***

It is an axiom of privacy that personal information gathered for one purpose will inevitably find other uses:

> Once the systems to access and use personal data are in place, there is an understandable interest in using those systems for other worthwhile purposes (e.g., preventing and prosecuting violent crimes). The consistent ex-

et d'exploiter des données archivées progressent sans cesse. Ces facteurs seront partiellement responsables de la création de personnalités numériques — c'est-à-dire des représentations numériques de chacun de nous qui serviront d'intermédiaires importants entre des personnes réelles et le reste du monde, notamment le gouvernement[17].

Ce qui est particulièrement inquiétant du cyberportrait que l'on crée de nous au moyen des bases de données, c'est qu'il ne nous ressemble pas. Autrement dit, nos doubles numériques reflètent mal notre véritable personnalité. L'analyse des données peut créer une caricature, mais elle ne crée pas la personne – et l'essence même de la vie privée est le maintien de l'identité individuelle. Cette question n'est pas purement philosophique. Le regroupement et l'analyse des données peut avoir des conséquences réelles et coûteuses pour les personnes. Plus la période de conservation des données est longue, plus il est difficile de corriger ou de mettre à jour ces données, particulièrement quand le système d'information est un système secret de sécurité nationale. Même quand les données sont exactes, leur conservation permanente peut entraîner des inconvénients graves pour les personnes qui ont le désir, et le droit, de laisser derrière elles certains épisodes de leur vie. Il sera de plus en plus difficile, par exemple, de cacher des folies de jeunesse comme l'adhésion à un groupe politique radical. Le pouvoir néfaste de notre persona numérique peut nous inciter à nous abstenir de toute prise de position et à devenir excessivement prudents. Cela serait contraire à l'épanouissement des gens, de la société et du gouvernement.

### *Résultats faussement positifs[18]*

L'an dernier, les médias américains ont rapporté que le sénateur Ted Kennedy avait été informé qu'il ne pouvait embarquer sur plus d'un vol intérieur parce que le nom T. Kennedy figurait sur la liste CAPPS I des personnes interdites de vol aux États-Unis. Son nom avait généré ce résultat et on lui avait interdit de monter dans l'avion. Ce résultat avait été faussement positif, car il n'était pas le T. Kennedy de la liste; il n'aurait pas dû être identifié comme personne présentant un risque pour la sécurité, même si « le vrai » T. Kennedy devait l'être. Le sénateur Kennedy a pu finalement monter à bord parce qu'il a réussi à convaincre les autorités qu'il ne présentait pas de menace à la sécurité publique. Une personne moins connue

> perience with data protection suggests that, over time, there is always pressure to use data collected for one purpose for other purposes. The expansive uses to which Social Security Numbers have been put are a practical example.[20]

In Canada, the two originally-intended uses for social insurance numbers have expanded to over two dozen federal government uses and the numbers are used for a myriad of other purposes in the private sector. This is not a merely trivial example, given the identifying, linking and organizing power of the social insurance number.

In the context of data mining for national security purposes, information generated for investigative purposes, or for use on a no-fly list, might be used for ordinary law enforcement purposes or to blacklist individuals.

### *Blacklisting*

Terrorist watch-lists are being used in the US and appear to exist in one form or another in Canada. Watch lists can have legitimate, even important, functions. Use of watch lists ought, however, to be confined to a limited scope of functions such as terrorism investigation, intelligence-gathering and security clearances. A watch list could turn into a blacklist—a list used as secret evidence, or effectively as a secret finding, to make decisions that directly affect individuals who have no knowledge of the evidence or any ability to challenge it.[21] Blacklists can, of course, be officially sanctioned or illicit. In either case, they are a concern, one that is magnified given the risks such as poor data quality that can be associated with data mining.

### *Wrongful misuse of data*

Concern about misuse of information derived from data mining activities is by no means unique to data mining. Examples abound from other areas, both in the public sector and the private sector.[22] Embarrassing or lucrative personal information tempts intentional misuse and the products of data mining will also be tempting.

### *Lack of due process*

Experience with national security data mining initiatives in the US suggests that authorities can be

aurait pu éprouver des difficultés beaucoup plus grandes. On a d'ailleurs signalé de nombreux cas où des personnes innocentes ont dû rester au sol par suite de résultats faussement positifs[19]. Ce genre de situation peut avoir des conséquences sérieuses pour les personnes qui se rendent au chevet d'un proche malade ou à des funérailles. Les personnes qui voyagent pour affaires risquent même de perdre leur emploi si elles ne peuvent se déplacer au besoin (et cela n'aide certainement pas la cause d'un employé quand son employeur découvre qu'il ne peut pas prendre l'avion parce que son nom figure sur une liste de terroristes interdits de vol).

Le problème des résultats faussement positifs n'est pas exclusif au forage de données, mais notre tendance à faire aveuglément confiance aux données et les nombreux risques d'erreurs concourent à donner à cette question une importance prépondérante. Le risque de résultats faussement positifs est lié à la conception même des systèmes. Si une application de forage de données ne peut faire la distinction entre le « bruit » de comportements courants et des signes d'activité terroriste, des innocents feront inutilement l'objet d'enquêtes ou seront inscrits à tort sur des listes de surveillance. Cela ne veut pas dire qu'on ne doit jamais utiliser le forage de données pour le dépistage des activités terroristes. Il faut plutôt trouver des solutions technologiques efficaces et adopter des mesures de protection adéquates afin d'éviter les conséquences fâcheuses des résultats faussement positifs.

### *Utilisation détournée*

C'est chose connue que les renseignements personnels recueillis pour une certaine fin seront inévitablement utilisés à d'autres fins :

> [TRADUCTION] Une fois que les systèmes de consultation et d'utilisation des renseignements personnels sont en place, on découvre naturellement des utilisations autres que celles pour lesquelles ils ont été conçus (p. ex. la prévention des crimes violents et la poursuite de ceux qui s'en rendent coupables). En matière de protection des données, l'expérience a montré qu'on finit toujours par céder aux

slow to recognize the need for due process and other protections. It appears, for example, that the Transportation Security Administration has been slow to devise due process protections for those who find themselves incorrectly placed on no-fly lists in the US.[23] Yet it is critically important that individuals mistakenly placed on no-fly lists or otherwise affected by errors or abuses of data mining systems be able to obtain redress through independent, fair, simple and as transparent as possible oversight processes.

## DATA MINING & CANADA'S PRIVACY LAWS

Canada's privacy laws are founded on a body of internationally-accepted fair information principles that are reflected in privacy laws throughout the world and in international instruments.[24] Our privacy laws aim to give individuals a degree of control over their own personal information throughout its life cycle. They give individuals the right to be told what information is being collected about them, who is collecting it, the uses to which it will be put, to whom it might be disclosed, and for what purposes it might be disclosed. In the private sector, the rules aim to give individuals a further degree of control by enabling them to generally choose which information to give up and for what purposes. Our privacy laws also give individuals the right to have access to their own information. They require organizations to take reasonable steps to ensure that personal information they hold and use is accurate and complete.

The following discussion illustrates how many of these rules are not fully equal to the task of meaningfully protecting privacy against risks associated with data aggregation, data sharing and data mining for national security purposes. The power of these information technologies, and the risks to individuals and society, are such that new approaches to privacy protection are required to supplement existing ones.[25]

### *Knowledge of collection*

An axiom of privacy protection is that, with limited exceptions, individuals must be given notice of collection of their personal information at the time it is collected.

As indicated earlier, data mining almost invariably depends on collection of personal information from a variety of sources and, certainly in the na-

> pressions qui s'exercent et utiliser les données recueillies à d'autres fins que celles prévues au départ. L'utilisation généralisée du numéro d'assurance sociale en est un bon exemple[20].

Au Canada, à l'origine, le numéro d'assurance sociale devait servir à deux fins. Aujourd'hui, l'administration fédérale s'en sert dans plus de vingt fonctions et le secteur privé encore davantage. Cet exemple n'est pas banal; n'oublions pas que le numéro d'assurance sociale donne accès à de nombreux outils d'identification, à diverses bases de données et à des systèmes d'organisation des données.

Dans le contexte du forage de données aux fins de la sécurité nationale, les renseignements obtenus à des fins d'enquête ou pour produire une liste des personnes interdites de vol peuvent servir à des fins courantes d'application de la loi ou pour inscrire une personne sur une liste noire.

***Inscription sur une liste noire***

Les autorités américaines utilisent des listes de surveillance des terroristes et le Canada en possède apparemment aussi sous une forme ou une autre. Ces listes ont souvent des fonctions légitimes, voire importantes. Toutefois, il faut en limiter l'utilisation, par exemple, aux enquêtes sur les présumés terroristes, à la collecte de renseignements et à l'habilitation de sécurité. Une liste de surveillance peut servir à produire une liste noire, elle peut constituer une preuve secrète ou motiver des décisions sur des personnes incapables de prendre connaissance des preuves présentées contre elles et, par conséquent, de les contester[21]. Bien entendu, les listes noires peuvent être légitimées ou illicites. Dans tous les cas, elles posent problème, compte tenu des risques d'erreurs que peut entraîner un forage de données basé sur des données dont la qualité laisse à désirer.

***Utilisation malveillante des données***

Les inquiétudes concernant l'utilisation malveillante de l'information ne se limitent pas aux seules activités de forage de données. On trouve de nombreux exemples préoccupants dans les secteurs public et privé[22]. Certains seront tentés d'utiliser des renseignements qui portent atteinte à

tional security context, this means affected individuals will usually not know of the collection of their personal information.

Some observers might suggest that this could be addressed by requiring information sellers or providers to notify affected individuals of the government's collection of information. This will be of questionable efficacy even where it is feasible at the time of collection. Further, such an indirect notice requirement is unlikely to work where personal information is collected for national security purposes, since notification will be dispensed with where national security is involved.

***Notice of the purposes for collection***

Another important privacy principle is that individuals are to be told the purpose for which their personal information is collected. The original collector of the information will not be collecting it for a national security purpose. This principle will therefore be honoured in the breach when the information is acquired later for national security uses. Requiring the person who originally collects the information to give notice of possible later national security use is unlikely to be acceptable to United States authorities. Nor is a notice given at the time of collection that it may be disclosed where 'required or authorized by law' sufficient.

***Direct collection***

Privacy laws stipulate that personal information can only be collected directly from the individual the information is about. There are exceptions to this, including for ordinary law enforcement needs—police can hardly be expected to ask a suspect for personal information needed to prosecute the suspect. The same will hold true for national security activities, meaning that indirect collection for national security data mining uses will be the norm, not the exception.

***Limited collection***

Although the precise standards vary somewhat, Canadian privacy laws permit organizations to collect only the personal information that is necessary for, or relevant to, the purpose for which it is collected. Where an individual's personal information that is initially collected for a commercial purpose is later used for national security data mining in conjunction with other information, the limited collection principle may have little meaning and

la réputation ou qui permettent d'extorquer des fonds, et les résultats des activités de forage de données auront le même attrait.

### *Absence de procédure régulière*

L'expérience des initiatives de forage de données pour la sécurité nationale aux États-Unis suggère que les autorités mettent du temps avant de reconnaître le besoin d'une procédure régulière et d'autres protections. Il semble, par exemple, que la Transportation Security Administration a tardé à élaborer des procédures régulières de protection pour les personnes qui se retrouvent à tort sur les listes d'interdiction de voler aux États-Unis[23]. Il est pourtant très important que les personnes inscrites sur ces listes par erreur ou qui sont victimes d'abus ou d'erreurs des systèmes de forage des données puissent obtenir réparation par la voie de mécanismes indépendants, justes, simples et aussi transparents que possible.

## FORAGE DE DONNÉES ET LOIS CANADIENNES EN MATIÈRE DE PROTECTION DE LA VIE PRIVÉE

Les lois qui protègent la vie privée au Canada s'inspirent d'un ensemble communément accepté de principes équitables sur la scène internationale pour régir l'information. Ces principes sont pris en considération par les lois en matière de protection de la vie privée du monde entier, de même que dans les instruments internationaux[24]. Nos lois sur la protection de la vie privée ont pour but d'accorder aux personnes un certain pouvoir sur leurs renseignements personnels au cours du cycle de vie de ces derniers. Elles leur reconnaissent le droit de savoir quels renseignements sont recueillis à leur sujet, qui possède ces renseignements et à quelles fins, à qui les renseignements peuvent être communiqués et à quelles fins. Dans le secteur privé, les règles visent à donner aux personnes le pouvoir de choisir les renseignements qu'ils veulent communiquer et l'utilisation qui peut en être faite. Grâce à nos lois sur la protection de la vie privée, les personnes ont le droit d'accéder aux renseignements personnels les concernant. Ces lois exigent des organisations qu'elles prennent des mesures raisonnables pour veiller à ce que les renseignements personnels qu'elles détiennent et utilisent soient exacts et complets.

L'analyse qui suit montre que bon nombre de ces

offer inadequate protection.

### *Individual access*

An important privacy right is the right to have access to one's own personal information. This enables individuals to find out what personal information an organization has about them, how it has been used and to whom it has been disclosed. It goes almost without saying that this right is illusory in the national security context.

### *Accuracy & completeness*

Most Canadian privacy laws require organizations to take reasonable measures to ensure that personal information they use to make a decision affecting an individual is accurate and complete. This is not a counsel of perfection, of course, but it does require positive, ongoing efforts to ensure data quality and completeness. Unlike the other traditional rules just mentioned, this duty is meaningful in the data mining context. It is necessarily imprecise and sensibly technology-neutral, but it can be particularized on an evergreen basis at a policy and operational level. A lingering concern, however, is whether meaningful independent oversight of the design of, and compliance with, this duty is available under the present privacy protection scheme.

### *Independent oversight*

As with any rights, rights to privacy mean little unless they can be vindicated through the rule of law. Independent oversight is a central tenet of internationally accepted privacy principles. Almost all of Canada's privacy laws provide for independent review and (to varying degrees) enforcement of privacy rights through commissioners or Ombudsmen with privacy oversight duties.[26]

## PRIVACY MEASURES FOR DATA MINING

As the preceding discussion shows, privacy risks associated with data mining present challenges that our existing privacy laws are in large measure ill-equipped to meet. This is not to say that our privacy laws are irrelevant in the context of data mining and other information technologies deployed for national security purposes. To be sure, the long-standing principles of limited (and proportional) collection of personal information, use of personal information for the purpose for which it

règles ne protègent pas toutes avec la même efficacité la vie privée contre les risques que posent l'agrégation, l'échange et le forage de données à des fins de sécurité nationale. La puissance de ces technologies de l'information, et les risques encourus par les personnes et la société, sont tels que les approches existantes pour protéger la vie privée ne suffisent plus. Il en faut de nouvelles[25].

### *Avis de collecte*

En règle générale, la protection de la vie privée exige, à quelques exceptions près, que les personnes soient informées de la collecte de renseignements personnels à leur sujet au moment de la collecte.

Comme on l'a indiqué ci-dessus, le forage de données repose presque toujours sur la collecte de renseignements personnels à partir d'un éventail de sources. Dans le contexte de la sécurité nationale, cela signifie qu'en règle générale les personnes touchées ne savent pas que leurs renseignements personnels font l'objet d'une telle collecte.

Certains observateurs suggèrent de régler cette question en exigeant des vendeurs ou des fournisseurs de renseignements d'informer les personnes concernées que le gouvernement a recueilli leurs renseignements personnels. On peut douter de l'efficacité d'une telle approche, même si elle est faisable au moment de la collecte. De plus, l'exigence d'un avis indirect de ce genre a peu de chance de fonctionner si les renseignements personnels sont recueillis à des fins de sécurité nationale, car l'avis ne sera pas obligatoire dans le contexte de la sécurité nationale.

### *Avis énonçant les fins de la collecte*

Un autre principe important veut que les personnes faisant l'objet d'une collecte de renseignements soient informées des fins de cette collecte. L'auteur de la collecte initiale ne recueille pas les renseignements pour des raisons de sécurité nationale. Ce principe devra donc être respecté par la suite au moment où l'information sera acquise pour des raisons de sécurité nationale. Il est fort improbable que les autorités américaines trouvent acceptable le fait d'exiger de l'auteur de la collecte de départ qu'il donne un avis à l'effet que les renseignements pourraient

was originally collected (or a very closely related purpose), information security and independent oversight remain relevant in the context of these new technologies.

While no single approach can adequately address all risks, solutions can and must be found. There is a pressing need for Canadian governments, notably the federal government, to study the available options and move quickly to implement effective and workable legal, policy and technological measures to protect privacy. Some, but not all, of the more significant measures worth considering are now outlined. Taken together, they can to some degree meet the pressing need for legislative and policy reform that provides for a comprehensive, one-stop approach to data mining approval and regulation for national security purposes.[27]

### *Data mining research*

Before federal government agencies engage in data mining—with the proposed Passenger Protect flight security initiative as an example—the federal government should undertake research into the effectiveness of data mining, with emphasis on technological and other tools for enhancing privacy protection. The research should also consider legal, social and ethical issues associated with data mining.

To be clear, a central focus of this research should be whether data mining for national security purposes offers meaningful benefits that are sufficiently important to override privacy and other civil rights concerns. It was acknowledged above that data mining can be useful for national security purposes, but before any data mining initiatives proceed in Canada, it is necessary to establish that any such benefits clearly and substantially outweigh the risks for privacy and other rights and liberties and that any such risks can be properly mitigated. This is not merely an exercise in assessing the constitutionality of proposals. It is a question of responsible and proportional policy making. Data mining should not be used for national security purposes in Canada unless stringent conditions are met.

### *Privacy impact assessments*

Many Canadian jurisdictions now have statutory or government policy requirements for a privacy impact assessment ("PIA") to be completed before a

éventuellement servir à la sécurité nationale, ou qu'ils pourraient être communiqués pour des motifs qui justifient ou autorisent cette communication aux termes de la loi.

***Collecte directe***

Les lois sur la protection des renseignements personnels stipulent que les renseignements personnels peuvent seulement être recueillis directement auprès de la personne concernée. Il y a cependant des exceptions, comme pour les besoins ordinaires d'application de la loi – les policiers peuvent difficilement demander à un suspect de fournir les renseignements personnels dont ils ont besoin pour le poursuivre. Il en va de même pour les activités de sécurité nationale; c'est-à-dire que la collecte indirecte de données dans le contexte du forage de données à des fins sécurité nationale sera la norme, et non l'exception.

***Collecte limitée***

Même si les normes diffèrent quelque peu, les lois canadiennes en matière de protection des renseignements personnels permettent aux organisations de recueillir seulement les renseignements personnels nécessaires, ou pertinents, aux fins pour lesquelles ils sont recueillis. Dans le cas des renseignements personnels recueillis pour des raisons commerciales et utilisés par la suite dans le forage de données à des fins de sécurité nationale avec d'autres renseignements, le principe de collecte limitée n'a plus guère de sens et n'offre plus vraiment de protection.

***Accès des personnes***

Le droit d'accès à ses propres renseignements personnels est un droit important en matière de vie privée. Il permet aux personnes de savoir quels renseignements personnels une organisation détient sur eux, comment ils ont été utilisés et à qui ils ont été communiqués. Il va presque sans dire que ce droit est illusoire dans le contexte de la sécurité nationale.

***Exactitude et intégralité***

La plupart des lois canadiennes sur la protection de la vie privée exigent des organisations qu'elles adoptent des mesures raisonnables pour veiller à ce que les renseignements personnels qu'elles

proposed program, policy or law is pursued.[28] A privacy impact assessment is a process—and an ongoing one at that—that requires an organization to assess the privacy risks of proposed programs, systems or laws and, to decide, whether they should proceed and to identify and implement mitigating measures where they do proceed. A PIA process enables privacy to be designed into new systems from the outset, thus promoting efficiency as well as good privacy practice and compliance. A mandatory PIA process, ideally with sign-off by the external oversight agency, should be a mandatory feature of any data mining governance framework.

***Chief Privacy Officers for national security agencies***

Large corporations now commonly have a chief privacy officer ("CPO") responsible for privacy compliance and oversight within the organization. These positions are often at the senior executive level, which recognizes the importance to a corporation's brand of good privacy practices and compliance.[29] A strong case can be made that Canada's federal, provincial and territorial governments should hire or designate CPOs in a similar fashion.

At the very least, federal agencies involved in national security and anti-terrorism activities should establish well-resourced, executive-level, CPO positions with responsibility for ensuring that information technologies such as data mining are designed and operated lawfully. These positions would not supplant, but would liaise with, external oversight agencies such as the Privacy Commissioner of Canada and the Security and Intelligence Review Committee. The US Department of Homeland Security established a CPO position over a year ago[30] and it is time such positions were created in Canada, with executive support and real internal authority.

***Prior judicial authorization for data mining activities***

There should be a strong rule that data mining can be performed only on anonymized data, with identification of individuals being possible only when specified quality and cogency criteria have been met and then only with prior judicial authorization. The technology exists to do this.[31] This rule would be relevant particularly in relation to data mining undertaken at a population or large group level.

utilisent pour prendre une décision touchant une personne soient exacts et complets. Ce principe n'exige pas la perfection, c'est évident, mais il exige des efforts positifs et continus pour veiller à la qualité et à l'intégralité des données. Contrairement aux autres règles traditionnelles mentionnées ci-dessus, cette responsabilité est significative dans le contexte du forage des données. Elle est imprécise et, heureusement, indépendante de la technologie, mais elle peut être élaborée plus en détail sur une base continue par des politiques et des mesures opérationnelles. Cependant, une préoccupation subsiste, celle de savoir si une surveillance indépendante et efficace de la prise en charge de cette responsabilité et de la conformité avec celle-ci est possible dans le régime actuel de protection des renseignements personnels.

### *Surveillance indépendante*

Comme tous les droits, le droit à la vie privée a peu de valeur s'il n'est pas protégé par des lois. La surveillance indépendante est au centre des principes de protection de la vie privée reconnus à l'échelle internationale. Presque toutes les lois canadiennes sur la protection des renseignements personnels prévoient l'examen et – dans une mesure variable – l'application indépendante du droit à la vie privée par des commissaires ou des ombudsmans chargés de surveiller le respect de la vie privée[26].

## MESURES DE PROTECTION DE LA VIE PRIVÉE LIÉES AU FORAGE DE DONNÉES

Comme le montre l'analyse précédente, les risques pour la protection de la vie privée qu'entraîne le forage de données constituent des défis que les lois actuelles de protection des renseignements personnels ne permettent pas vraiment de relever. Cela ne veut pas dire pour autant que ces lois sont pratiquement superflues dans le contexte du forage de données et d'autres technologies de l'information déployées pour des raisons de sécurité nationale. De toute évidence, les principes de longue date de collecte limitée (et proportionnelle) des renseignements personnels, d'utilisation des renseignements aux fins pour lesquelles ils ont été recueillis à l'origine (ou une fin connexe), de sécurité de l'information et de surveillance indépendante restent pertinents dans le contexte de ces nouvelles technologies.

Where data mining is proposed in relation to specified individuals, it should be permitted only with prior judicial authorization on the basis of particularized grounds that meet constitutional standards. These recommendations are commonly encountered in the US literature and official reports.[32]

### *Rules-based and other technological protections*

A number of technical approaches to data mining are available to enhance privacy in data mining, while more research is required to refine other techniques before they can credibly be deployed.

Rules-based processing techniques, it has been said, offer considerable promise for privacy protection in data mining. One technique would involve use of intelligent agents (or "proof-carrying code") to centrally query distributed databases by negotiating access and permitted uses on a database-by-database basis. Where data elements might move about, they could be labelled with meta-data stipulating how the element must be dealt with. This technique would allow rules specific to particular data elements to follow the data elements. A third approach involves software applications known as 'analytical filters', which are designed to filter and discard innocent noise and retain information of interest.[33]

### *Audit trails*

Information systems in health care and commercial applications are now commonly equipped with built-in audit systems. The best of these systems automatically log access to data files and create more or less immutable audit trails. At the most basic level, they can in real time identify when unauthorized access is attempted or succeeds. More sophisticated audit applications monitor authorized access for unusual patterns and can, either automatically or with human intervention, identify both inappropriate access and use by authorized users.

These systems enable administrators (and regulators) to ensure that rules are followed. In the context of sophisticated and powerful information technology like data mining, strong audit capabilities are of critical importance in preventing misuses of data, data spills and even function creep.

Même si on ne peut trouver une approche unique pour gérer tous les risques à la fois, on peut et on doit trouver des solutions. Les gouvernements canadiens, surtout le gouvernement fédéral, doivent sans tarder étudier les solutions possibles et rapidement mettre en œuvre des mesures législatives, stratégiques et technologiques efficaces pour protéger la vie privée. Certaines des mesures les plus significatives dont il vaut la peine de tenir compte sont présentées ci-dessous. Ensemble, elles peuvent répondre dans une certaine mesure au besoin pressant d'une réforme législative et stratégique en vue de se doter d'une approche intégrale unique d'approbation et de réglementation du forage de données à des fins de sécurité nationale[27].

### *Recherche sur le forage de données*

Avant que les organismes du gouvernement fédéral se lancent dans le forage de données – à titre d'exemple, mentionnons l'initiative de sécurité aérienne Protection des passagers –, le gouvernement fédéral devrait entreprendre des études sur l'efficacité du forage de données et mettre l'accent sur les outils technologiques et autres visant à rehausser la protection de la vie privée. Ces études devraient aussi se pencher sur les questions juridiques, sociales et éthiques que pose le forage de données.

Autrement dit, ces études devraient déterminer si les avantages du forage de données à des fins de sécurité nationale sont suffisamment importants et l'emportent sur les préoccupations en matière de protection de la vie privée et d'autres droits civils. On a conclu ci-dessus que le forage de données peut être utile à des fins de sécurité nationale. Toutefois, avant d'aller de l'avant avec le forage de données au Canada, il convient d'établir si les avantages de celui-ci sont clairement et considérablement plus importants que les risques pour la vie privée, les autres droits et les libertés individuelles et s'il y a moyen d'atténuer ces risques. Il ne s'agit pas simplement d'un exercice dont le but est d'évaluer la validité constitutionnelle des propositions, mais d'un exercice responsable et juste des pouvoirs publics. Le Canada ne devrait pas avoir recours au forage de données à des fins de sécurité nationale, à moins de répondre à des conditions rigoureuses.

### *Security of data mining systems*

Although a trite proposition, data mining systems must have strong security measures in order to prevent data leakages or corruption. As noted earlier, one generally-accepted privacy principle that applies to data mining in a meaningful way is the obligation to take reasonable security measures to protect personal information against unauthorized collection, use or disclosure. This is especially important in light of the risks that can be associated with data mining by the state. Data security must be a high priority in the design and operation of data mining systems.

### *Due process for affected individuals*

The fact that national security is involved cannot be allowed to oust due process for affected individuals. If someone is incorrectly placed on a watch list or no-fly list, or is investigated on false premises, they should have recourse to an effective process for redress. The process should, despite the national security nature of the enterprise, be as transparent as practicable in the circumstances,[34] should be inexpensive, and should be expeditious.

### *Ensuring effective external oversight*

Last, but by no means least, some way must be found of ensuring that there is effective, independent, oversight of data mining activities. As mentioned earlier, the Privacy Commissioner of Canada has authority to investigate privacy compliance by federal government agencies. The federal *Privacy Act*, however, is sorely in need of reform and the compliance powers of the Privacy Commissioner of Canada need to be enhanced to meet the challenges of information technology and national security. The Commissioner's powers need to be modernized and strengthened hand in hand with a substantially reworked set of legislated privacy rules that apply specifically to data mining. The Commissioner's role would complement that of the courts and Parliamentary oversight (if, in fact, enhanced Parliamentary oversight comes to pass, as it should). The Commissioner's role could also complement, in the case of CSIS, the oversight role of the Security Intelligence Review Committee.

### *Évaluations des facteurs relatifs à la vie privée*

Au Canada, de nombreuses juridictions prévoient désormais dans leurs lois ou politiques l'obligation de procéder à des évaluations des facteurs relatifs à la vie privée (ÉFVP) avant de lancer un programme, d'adhérer à une politique ou d'adopter une loi[28]. Une ÉFVP est un processus continu qui impose à une organisation qu'elle évalue les risques pour la protection de la vie privée que présentent des programmes, des systèmes ou des lois proposés afin de décider si elle peut aller de l'avant et si elle doit mettre en œuvre des mesures d'atténuation. L'ÉFVP permet d'intégrer les considérations relatives à la vie privée dans les nouveaux systèmes dès leur conception, ce qui encourage une plus grande efficacité, de bonnes pratiques en matière de vie privée et le respect des lois. L'ÉFVP, qui, idéalement, devrait s'accompagner d'une obligation d'approbation par une agence de surveillance externe, devrait être obligatoire dans tous les cadres de gouvernance de forage des données.

### *Responsable de la protection des renseignements personnels dans les organismes de sécurité nationale*

Il est commun pour les grandes sociétés de capitaux d'avoir un responsable de la protection des renseignements personnels chargé de la conformité aux lois en la matière et de la surveillance connexe. Ces postes sont habituellement à l'échelon administratif supérieur, ce qui montre que ces sociétés reconnaissent l'importance des pratiques et de la conformité en matière de protection de la vie privée pour leur image de marque[29]. On peut faire valoir que les gouvernements fédéral, provinciaux et territoriaux du Canada devraient faire de même et engager ou nommer des responsables de la protection des renseignements personnels.

À tout le moins, les organismes fédéraux qui contribuent à la sécurité nationale ou à des activités antiterroristes devraient créer des postes de responsables de la protection des renseignements personnels au niveau de la direction et les doter des ressources nécessaires pour veiller à ce que les technologies de l'information, comme le forage de données, soient conçues ou exploitées de manière légale. Les titulaires de ces postes ne remplaceraient pas les

## CONCLUSION

The rights and freedoms that we have come to expect will be upheld in Canada, including our privacy rights, are not absolute. Terrorism may necessitate new strategies to protect the security of all people. Although the risk of terrorist attacks on Canada is real, government must take great care not to overstep the line. Although it may be true that, the more freedom people have, the greater the potential risks, every increase in security almost inevitably curtails rights and freedoms that are at the heart of democratic societies. Rights and freedoms that we tend to take for granted because we have always been fortunate enough to have them can be easily eroded—in good faith or otherwise—and we must ensure that our elected officials maintain life and vibrancy in them.

No one can envy the difficult task lawmakers face in trying to strike the right balance between privacy and security, but it is critically important that they ask the hard questions and come up with appropriate answers. Meaningful reforms of Canada's privacy laws—particularly the federal *Privacy Act*—are urgently required in order to address the privacy challenges raised by data mining and other information technology applications. Those reforms are needed now.

## END NOTES

[1] This paper is a slightly revised reproduction of a paper prepared in 2005 for the Annual Conference of the Canadian Institute for the Administration of Justice, entitled "*Information Technology, National Security & Privacy Protection".* Portions of the paper were published in an article the author wrote that appeared in the August 19, 2005 issue of *The Lawyers Weekly*, published by LexisNexis Canada Inc., and appear with permission. The paper has not been updated to take into account developments since 2005.

[2] R. Clarke, "Computer Matching & Digital Identity" (paper delivered at Computers, Freedom & Privacy Conference, 1993) <http://www.anu.edu.au/people/Roger.Clarke/DV/CFP93.html> (hereinafter Clarke CFP Paper).

[3] For discussion of other privacy challenges presented by post-September 11 laws and policies, see *Privacy & the USA Patriot Act—Implications for British Columbia Public Sector Outsourcing* <http://www.oipc.bc.ca/sector_public/

agents de surveillance externes, comme la commissaire à la protection de la vie privée du Canada et le Comité de surveillance des activités de renseignement de sécurité, mais seraient chargés des relations avec ceux-ci. Le département de sécurité intérieure des États-Unis a nommé un responsable de la protection des renseignements personnels il y a plus d'un an[30] et le Canada devrait aussi créer de tels postes jouissant d'un soutien de la direction et d'une autorité interne réelle.

### *Autorisation juridique préalable pour les activités de forage des données*

Il faudrait une règle stricte établissant que le forage de données peut seulement viser des données rendues anonymes, que l'identification des personnes n'est possible que si des critères précis de bien-fondé et de qualité sont respectés et si on obtient une autorisation juridique. La technologie permet de faire cela[31]. Cette règle serait pertinente surtout pour le forage de données effectué à l'échelle d'une population ou d'un grand groupe. Le forage de données visant des personnes précises ne devrait être permis qu'avec une autorisation juridique préalable fondée sur des raisons précises qui répondent aux normes constitutionnelles. Ces recommandations sont courantes dans la documentation à ce sujet et les rapports officiels aux États-Unis[32].

### *Protections fondées sur les règles et les technologies*

Un certain nombre d'approches techniques du forage de données sont disponibles pour rehausser la protection de la vie privée, mais il faut davantage d'études afin d'améliorer les autres techniques pour qu'elles puissent être déployées efficacement.

On dit que les techniques de traitement fondées sur les règles sont très prometteuses pour ce qui est de protéger la vie privée dans le forage de données. Une de ces techniques exige l'apport d'agents intelligents (ou prothèses logicielles) pour balayer les bases de données distribuées à partir d'un centre en négociant l'accès et les utilisations autorisées pour chacune des bases de données. Aussi, si les éléments de données sont susceptibles d'être déplacés, on pourrait les marquer avec des métadonnées qui indiqueraient comment gérer l'élément en question. Grâce à

usa_patriot_act/pdfs/report/privacy-final.pdf> (Office of the Information and Privacy Commissioner for BC, October 2004).

[4] A recent news article about the consumer profiling database affiliate of Tesco, the large UK-based grocery chain, is an example of this kind of business activity. See Heather Tomlinson & rob Evans, "Tesco Stocks Up on Inside Knowledge of Shoppers' Lives", *The Guardian* (online), http://www.guardian.co.uk/business/story/0,3604,1573821,00.html (accessed September 21, 2005).

[5] The already extensive, and increasing, US federal government exploitation of commercial databases is well documented. See, for example, Daniel J. Solove, *The Digital Person* (New York University Press: New York, 2004), notably at 168-175.

[6] MATRIX stands for Multi-State Anti-Terrorism Information Exchange. To moviegoers, at least, this acronym has unfortunate echoes. Soon after its existence became public, the Total Information Awareness project was re-branded the Terrorism Information Awareness project, perhaps to avoid similarly unfortunate connotations.

[7] CAPPS stands for Computer Assisted Passenger Pre-Screening.

[8] In July of 2005, Jennifer Stoddart, Privacy Commissioner of Canada, wrote to Transport Canada officials and expressed concern about a no-fly list. On August 9, 2005, joined by other Canadian privacy commissioners, including the author, she again expressed concern about the implications of Passenger Protect.

[9] It is fair to say that commercial personal information databases in the US are richer in detail and on a much larger scale than those in Canada. This stems from at least two factors. First, there are a number of US federal privacy laws, but they cover specific sectors and are relatively generous, in part due to the long tradition of US public records laws, as regards compilation of personal information by database companies.

[10] Experts in the field, and some commentators, prefer the term 'knowledge discovery', with 'data mining' referring to a specific step in data analysis. Data mining is nonetheless the popularly used term adopted here.

[11] Congressional Research Service, *Data Mining: An Overview* (Library of Congress: Washington, 2004), at 1 (citations omitted) (hereinafter CRS Report). Data mining is fairly commonly encountered in the US federal government and promises to become more common. A May 2004 US General Accounting Office study of data min-

cette technique, les règles précises pourraient suivre les éléments auxquels elles se rattachent. Une troisième approche suppose des applications logicielles dites « filtres analytiques » conçues pour filtrer et supprimer les interférences inutiles pour ne conserver que l'information importante[33].

### *Pistes de vérification*

Les systèmes d'information des régimes de soins de santé et des applications commerciales sont désormais équipés de systèmes de vérification intégrés. Les meilleurs de ces systèmes enregistrent automatiquement l'accès aux fichiers de renseignements et créent des pistes de vérification relativement immuables. Les systèmes de base peuvent, en temps réel, signaler qu'un accès non autorisé est tenté ou réussi. Les applications de vérification plus sophistiquées surveillent les accès autorisés pour déceler les utilisations inhabituelles et peuvent contrôler, automatiquement ou manuellement, tant l'accès inapproprié que l'activité des utilisateurs autorisés.

Ces systèmes permettent aux administrateurs (et aux organismes de réglementation) de veiller au respect des règles. Dans le contexte des technologies de l'information sophistiquées et puissantes telles que le forage de données, il est essentiel d'avoir de solides capacités de vérification pour prévenir les utilisations indues des données, les fuites de données et le détournement d'usage.

### *Sécurité des systèmes de forage des données*

Même s'ils font l'objet d'une proposition bien établie, les systèmes de forage des données doivent être accompagnés de solides mesures de sécurité pour prévenir les fuites et la corruption de données. Comme on l'a indiqué ci-dessus, l'obligation d'adopter des mesures de sécurité raisonnables pour protéger les renseignements personnels contre la collecte, l'utilisation ou la communication non autorisées est un des principes généralement reconnus de protection de la vie privée qui s'applique de manière significative au forage de données. Cela est très important, surtout à la lumière des risques associés au forage de données par l'État. La protection des données doit être l'une des priorités de la conception et de l'exploitation des systèmes de forage des données.

ing revealed that, amongst 128 federal agencies, 52 were using or planning to use data mining. There were 131 operational and 68 planned data mining initiatives. Fourteen were related to detecting terrorist activities, 15 were aimed at detecting criminal activities or patterns and 23 at detecting fraud. See *Data Mining: Federal Efforts Cover a Wide Range of Uses*, <http://www.gao.gov/new.items/d04548.pdf> (accessed September 15, 2005) (hereinafter *GAO Data Mining General Report*).

[12] Joint Inquiry Into the Intelligence Community Activities Before and After the Terrorist Attacks of September 11, 2001 (House Permanent Select Committee on Intelligence & Senate Select Committee on Intelligence, H. Rep. No. 107-792, S. Rep. No. 107-351 (2002), at 4-6.

[13] For example, the *GAO Data Mining General Report* identified a number of commonly predicted deficiencies in US federal government data mining initiatives. In an August 2005 follow-up report, the newly re-named Government Accountability Office reported that selected agencies had taken steps to protect privacy, but privacy rights were still not being appropriately protected. See *Data Mining: Agencies Have Taken Key Steps* […] (Government Accountability Office: Washington, 2005) <http://www.gao.gov/new.items/d05866.pdf (accessed September 15, 2005). A review of data mining in and for the US Department of Defense also reported a number of deficiencies and risks. See *Safeguarding Privacy in the Fight Against Terrorism* (Report of the Technology & Privacy Advisory Committee) (US Department of Defense: Washington, 2004). <http://www.cdt.org/security/usapatriot/20040300tapac.pdf (accessed September 15, 2005) (hereinafter TAPAC Report). The Advisory Committee recommended a number of measures to address privacy risks, some of which are discussed above.

[14] For further reading on data mining and privacy risks, see the following selected publications: R. Clarke, "Information Technology & Dataveillance", 31 Commun. ACM 5 (1988) <http://www.anu.edu.au/people/Roger.Clarke/DV/CACM88.html#Dang> (accessed September 15, 2005); TAPAC Report; GAO *Data Mining General Report*; CRS Report; Clarke CFP Paper.

[15] M. De Rosa, *Data Mining and Data Analysis for Counter-Terrorism* (Center for Strategic and International Studies: Washington, 2004), at v.

[16] Although there is likely room to improve authentication for government-issued identification this is not to suggest that affinity card programs

### *Procédure régulière pour les personnes concernées*

Même si la sécurité nationale est en jeu, on ne peut pas ne pas prévoir de procédure régulière pour les personnes concernées. Si une personne est injustement inscrite sur une liste de surveillance ou d'interdiction de vol, ou encore si elle fait l'objet d'une enquête non fondée, elle doit pouvoir recourir à un processus de réparation efficace. Ce processus devrait être aussi transparent que possible dans les circonstances[34], être peu coûteux et être expéditif, et ce, même s'il est question de sécurité nationale.

### *Assurance d'une surveillance externe efficace*

Enfin, et c'est peut-être là le plus important, il faut trouver le moyen de veiller à la surveillance efficace et indépendante des activités de forage des données. Comme on l'a expliqué ci-dessus, la commissaire à la protection de la vie privée du Canada a l'autorité nécessaire pour enquêter sur la conformité des organismes fédéraux aux lois en matière de protection de la vie privée. Toutefois, la *Loi sur la protection des renseignements personnels* fédérale a grand besoin d'être modifiée et on doit accroître les pouvoirs de la commissaire à la protection de la vie privée pour qu'elle puisse faire face aux nouveaux défis que posent les technologies de l'information et la sécurité nationale. Les pouvoirs de la commissaire doivent être modernisés, renforcés et soutenus par des règles législatives de protection de la vie privée à jour qui s'appliquent précisément au forage de données. Le rôle de la commissaire viendrait compléter la surveillance exercée par les tribunaux et le Parlement (si, justement, celui-ci vote en faveur de pouvoirs de surveillance accrus, comme il devrait le faire). Le rôle de la commissaire compléterait aussi le rôle de surveillance du Comité de surveillance des activités de renseignement de sécurité, pour ce qui concerne le Service canadien du renseignement de sécurité.

## CONCLUSION

Les droits et les libertés que nous nous attendons à voir respectés au Canada, y compris le droit à la vie privée, ne sont pas absolus. Le terrorisme

need to do a better job of authenticating identity on enrolment. (A national identity card is not, for a number of privacy-related and efficiency reasons, the answer.)

[17] See Clarke, note 2.

[18] Of course, although it is not a privacy issue, the problem of false negatives is a serious one, since a system's failure to identify as a possible terrorist someone who actually is a terrorist can have drastic consequences.

[19] TAPAC Report, at 38-39.

[20] TAPAC Report, at 39-40. See also the classic, and still relevant, study by David H. Flaherty, *Protecting Privacy in Surveillance Societies* (Chapel Hill: University of North Carolina Press, 1989).

[21] In the 1970s, the US Senate investigation into FBI abuses, known as the Church Commission, discovered widespread use by the FBI of secret personal dossiers on American citizens whose allegiance or morals were considered suspect, with these dossiers being used to deny jobs and other opportunities (notably in the 1950s, during the McCarthy days). Timothy Garton Ash, the British historian and author, discovered long after his student days in Berlin that his visits to East German student friends had earned him a secret MI5 dossier which could have denied him public sector employment (even though MI6 had tried to recruit him in his student days). T. Garton Ash, *The File* (New York: Random House, 1997).

[22] According to the TAPAC Report, at 40, "thousands" of US Internal Revenue Service employees have been disciplined for inappropriately accessing and reviewing the tax files of well-known people. In the private sector, Bank of America employees were recently reported to have sold customer information for identity theft-related purposes. See J. Leyden, "US bank staff sold customer details", *The Register* (online) <http://www.theregister.co.uk/2005/05/24/us_banks_security_flap/ (accessed September 18, 2005).

[23] In a 2005 report, the GAO concluded that deficiencies remained in the oversight and due process aspects of Secure Flight, the latest version of the US no-fly list. See *Aviation Security: Secure Flight Development & Testing Under Way* […] (Government Accountability Office: Washington, 2005). A 2004 GAO report found related defects in CAPPS II and recommended changes. See *Aviation Security: Computer Assisted Passenger Pre-screening System Faces Significant Implementation Challenges* (Government Accounting Office: Washington, 2004). It remains to

pourrait bien forcer l'adoption de nouvelles stratégies pour assurer la sécurité de tous. Même si le risque d'attaques terroristes au Canada existe, le gouvernement doit veiller à ne pas dépasser les bornes. Même s'il est vrai que les risques possibles sont directement proportionnels à la liberté des personnes, tout rehaussement de la sécurité entraîne presque inévitablement une érosion des droits et des libertés qui sont au cœur des sociétés démocratiques. Les droits et les libertés que nous tenons pour acquis parce que nous n'en avons jamais été privés pourraient s'éroder facilement – malgré toute la bonne foi des dirigeants – et nous devons veiller à ce que nos élus les maintiennent.

Personne n'envie les législateurs à qui incombe la tâche de trouver un équilibre entre la sécurité et la protection de la vie privée. Il est essentiel pour eux de poser des questions difficiles, car c'est là le seul moyen d'obtenir les réponses appropriées. Il faut instamment apporter des réformes aux lois canadiennes sur la protection de la vie privée – en particulier à la *Loi fédérale sur la protection des renseignements personnels* – afin de faire face aux défis que posent le forage des données et les autres applications des technologies de l'information. Ces réformes s'imposent immédiatement.

## NOTES DE BAS DE PAGE

[1] Le présent document est une reproduction légèrement révisée d'un article rédigé en 2005 pour la Conférence annuelle de l'Institut canadien d'administration de la justice, intitulé *Information Technology, National Security & Privacy Protection.* Cet article a partiellement été intégré à un autre article de l'auteur paru, avec permission, dans le numéro du 19 août 2005 de *The Lawyers Weekly*, publié par LexisNexis Canada. La mise à jour du présent document ne tient pas compte des développements survenus depuis 2005.

[2] Roger Clarke, *Computer Matching & Digital Identity* (article rédigé pour la Conférence Informatique, liberté et protection de la vie privée, 1993)http://www.anu.edu.au/people/Roger.Clarke/DV/CFP93.html (cité plus loin sous le nom d'Article de Clarke pour la Conférence ILP).

[3] Pour consulter des articles sur d'autres problèmes de protection de la vie privée que posent les lois et les politiques depuis le 11 septembre, voir *Privacy & the USA Patriot Act*–

be seen whether Passenger Protect will implement due process to address errors and misuse in Canada.

[24] For example, the OECD's 1980 *Guidelines on the Protection of Privacy & Transborder Flows of Personal Data* (adopted September 23, 1980), to which Canada is a signatory.

[25] It may be objected that Canada's privacy laws already contain exceptions to the following principles. That is undoubtedly true, but it is no answer. As noted above, the nature and scale of the risks to individuals demand more.

[26] One exception to this is Nova Scotia. The Review Officer under Nova Scotia's public sector freedom of information and privacy legislation has no authority to investigate and enforce the law's privacy provisions. See the *Freedom of Information and Protection of Privacy Act*, S.N. 1993, c. 5.

[27] Whether data mining is appropriate for use for general law enforcement purposes is beyond the scope of this paper, although significant questions are raised by this prospect.

[28] For example, s. 69(5) of British Columbia's *Freedom of Information and Protection of Privacy Act*, R.S.B.C. 1996, c. 165, requires provincial government ministries to perform a PIA in accordance with standing ministerial orders respecting PIAs. An example of a PIA template, prepared by the British Columbia government (with input from the author's office), can be found through <http://www.oipc.bc.ca/sector_public/resources/pia.htm (accessed September 18, 2005).

[29] The proposition that good privacy is good for business is forcefully proved in A. Cavoukian & T. Hamilton, *The Privacy Payoff* (Toronto: McGraw-Hill Ryerson, 2002),

[30] The CPO's home page is found at <http://www.dhs.gov/dhspublic/interapp/editorial/editorial_0338.xml (accessed September 18, 2005).

[31] Taipale, note 10, at 79-80. Also see the TAPAC Report, recommendation 2.4. Also see De Rosa, note 21, at 17-18.

[32] See, for example, Taipale, note 10, TAPAC Report, Solove, note 11. In Canada, see Arthur J. Cockfield, "The State of Privacy Laws and Privacy- Encroaching Technologies after September 11: A Two-Year Report Card on the Canadian Government", (2003-2004), 1 U. of Ottawa Law & Technology Journal 325.

[33] Taipale, note 10, at 75-78.

[34] Because of the national security interests involved, it may be necessary, where the redress process could reasonably be expected to threaten

*Implications for British Columbia Public Sector Outsourcing* http://www.oipc.bc.ca/sector_public/usa_patriot_act/pdfs/report/privacy-final.pdf (Bureau du commissaire à l'information et à la protection de la vie privée de la Colombie-Britannique (C.-B.), octobre 2004).

[4] Un exemple de ce type d'activités commerciales est présenté dans un article récent qui traite des bases de données servant à établir des profils de consommateurs, affiliées à Tesco, une grande chaîne de magasins d'alimentation du Royaume-Uni,. Voir Heather Tomlinson et Rob Evans, « Tesco Stocks Up on Inside Knowledge of Shoppers' Lives », *The Guardian* (en direct), http://www.guardian.co.uk/business/story/0,3604,1573821,00.html (consulté le 21 septembre 2005).

[5] L'exploitation actuelle de grande envergure et croissante de bases de données commerciales par le gouvernement fédéral américain est bien documentée. Voir, par exemple, Daniel J. Solove, *The Digital Person*, New York, New York University Press, 2004, notamment aux pages 168-175.

[6] MATRIX est l'acronyme de Multi-State Anti-Terrorism Information Exchange (Échange d'information entre les États pour la lutte contre le terrorisme). Pour les cinéphiles du moins, cet acronyme rappelle des situations malheureuses. Peu après que le public a appris son existence, le nom *Total Information Awareness project* a été remplacé par *Terrorism Information Awareness project*, peut-être pour éviter des connotations négatives semblables.

[7] CAPPS est l'acronyme de *Computer Assisted Passenger Pre-Screening* (Contrôle préalable des passagers assisté par ordinateur).

[8] En juillet 2005, Jennifer Stoddart, commissaire à la protection de la vie privée du Canada, a écrit aux représentants de Transports Canada pour leur faire part de ses préoccupations à l'égard de la liste d'interdiction de vol. Le 9 août 2005, les commissaires canadiens à la protection de la vie privée, y compris l'auteur du présent document, ont uni leurs voix à la sienne pour exprimer leurs inquiétudes concernant les conséquences du Programme de protection des passagers.

[9] Il est juste de dire que, aux É.-U., les bases de données commerciales qui contiennent des renseignements personnels sont plus détaillées et de plus grande envergure que celles du Canada. Cela s'explique de deux façons; d'abord, il existe un certain nombre de lois

harm to national security, to permit independent representatives to examine classified information relevant to disposition of the matter.
These individuals would function in ways similar to *amicus curiae*.

fédérales américaines visant à protéger la vie privée, mais elles s'appliquent à des secteurs en particulier et sont relativement généreuses, en partie à cause de la longue tradition des lois traitant des dossiers publics aux É.-U., comme le montre la compilation des renseignements personnels par les entreprises de bases de données.

[10] Les experts du domaine et certains commentateurs préfèrent le terme « découverte de connaissances » à « forage de données » qui, selon eux, désigne une étape précise de l'analyse de données. Le terme « forage de données » est néanmoins le terme adopté ici en raison de sa popularité.

[11] Congressional Research Service, *Data Mining: An Overview*, Library of Congress, Washington, 2004, page 1 (citations omises) (cité plus loin sous le nom de Rapport du CRS). Le forage de données, assez souvent utilisé par le gouvernement fédéral américain, promet de devenir une pratique encore plus courante. Une étude sur le forage de données effectuée en mai 2004 par le General Accounting Office des É.-U. a révélé que, parmi 128 organismes fédéraux, 52 utilisaient cette technique ou prévoyaient le faire. Il y a eu 131 initiatives opérationnelles et 68 projets de forage de données. Quatorze organismes s'en servaient pour repérer des activités terroristes, 15 pour détecter des activités ou tendances criminelles et 23 pour détecter des fraudes. Voir *Data Mining: Federal Efforts Cover a Wide Range of Uses*, <http://www.gao.gov/new.items/d04548.pdf> (consulté le 15 septembre 2005) (cité plus loin sous le nom de Rapport général du GAO sur le forage de données).

[12] *Joint Inquiry Into the Intelligence Community Activities Before and After the Terrorist Attacks of September 11, 2001* (Enquête conjointe sur les activités collectives des services de renseignements avant et après les attaques terroristes du 11 septembre 2001), Comités spéciaux du Congrès sur les services de renseignements, Chambre des représentants des É.-U., no 107-792, Sénat, no 107-351, 2002, aux pages 4-6.

[13] Par exemple, le Rapport général du GAO sur le forage de données a souligné un certain nombre de défaillances courantes prévisibles dans les projets de forage de données du gouvernement fédéral américain. En août 2005, dans un rapport de suivi, le nouvellement rebaptisé Government Accountability Office a mentionné que certains organismes avaient pris

des mesures pour protéger la vie privée des personnes, mais que le droit à la vie privée n'était pas encore protégé adéquatement. Voir *Data Mining: Agencies Have Taken Key Steps* […] Washington, Government Accountability Office, 2005 <http://www.gao.gov/new.items/d05866.pdf (consulté le 15 septembre 2005). Une étude sur le forage de données au Département de la défense américaine, ou effectué pour cet organisme, a aussi révélé un certain nombre de défaillances et de risques. Voir *Safeguarding Privacy in the Fight Against Terrorism*, Rapport du comité consultatif sur la technologie et la protection de la vie privée, Département de la défense américaine, Washington, 2004).<http://www.cdt.org/security/usapatriot/20040300tapac.pdf (consulté le 15 septembre 2005) (cité plus loin sous le nom de Rapport TAPAC). Le comité consultatif a recommandé qu'un certain nombre de mesures soient prises pour atténuer les risques d'atteinte à la vie privée; le présent document traite de certaines de ces mesures.

[14] Pour consulter d'autres documents de référence sur le forage de données et les risques pour la vie privée qui y sont associés, voir les publications suivantes : Roger Clarke, « Information Technology & Dataveillance », 31 Commun. ACM 5, 1988 <http://www.anu.edu.au/people/Roger.Clarke/DV/CACM88.html#Dang> (consulté le 15 septembre 2005); Rapport TAPAC; Rapport général du GAO sur le forage de données; Rapport CRS; Article de Clarke pour la conférence ILP.

[15] M. De Rosa, *Data Mining and Data Analysis for Counter-Terrorism*, Center for Strategic and International Studies, Washington, 2004, à v.

[16] Bien qu'il soit certainement possible d'améliorer l'authentification des documents d'identité émis par l'administration publique, cela ne veut pas dire qu'il faille améliorer le processus d'authentification de l'identité de ceux qui s'inscrivent à un programme de cartes d'affinité. (Pour diverses raisons dont la protection de la vie privée et l'efficacité, la carte d'identité nationale n'est pas la solution.)

[17] Voir Clarke, note 2.

[18] Bien entendu, quoique ce ne soit pas un problème de protection de la vie privée, le problème des résultats faussement négatifs est sérieux, car l'incapacité d'un système à dépister un véritable terroriste peut avoir des conséquences tragiques.

[19] Rapport du TAPAC, pages 38 et 39.

[20] Rapport du TAPAC, pages 39 et 40. Voir aussi l'étude classique et toujours pertinente de

David H. Flaherty, *Protecting Privacy in Surveillance Societies,* (University of North Carolina Press, Chapel Hill, 1989).

[21] Dans les années 1970, l'enquête du Sénat des États-Unis sur les excès du FBI, connue sous le nom de Commission Church, a permis de découvrir que le FBI utilisait couramment des dossiers personnels secrets sur les citoyens américains dont les allégeances ou la moralité lui paraissaient suspectes. Ces dossiers ont servi à écarter des gens d'un emploi et à leur interdire certaines choses (tout particulièrement dans les années 1950 pendant le maccarthysme). Timothy Garton Ash, historien et auteur britannique, a découvert, longtemps après ses années d'études à Berlin, que ses visites à des étudiants d'Allemagne de l'Est lui avaient valu un dossier secret du MI5 qui aurait pu l'empêcher d'occuper un emploi dans la fonction publique (même si le MI6 avait essayé de le recruter pendant ses études). T. Garton Ash, *The File*, Random House, New York, 1997).

[22] Selon le rapport du TAPAC, à la page 40, des « milliers » d'employés de l'Internal Revenue Service des États-Unis ont subi des mesures disciplinaires pour avoir examiné sans raison valable les dossiers fiscaux de personnes connues. Dans le secteur privé, on a récemment annoncé que des employés de la Bank of America avaient vendu à des voleurs d'identité des renseignements sur certains clients. Voir J. Leyden, « US bank staff sold customer details », *The Register* (en ligne) <http://www.theregister.co.uk/2005/05/24/us_banks_security_flap/ (accédé le 18 septembre 2005).

[23] Dans un rapport de 2005, le GAO a conclu qu'il y avait toujours des lacunes relatives à la surveillance et aux procédures régulières en rapport avec la liste *Secure Flight*, la dernière version de la liste d'interdiction de vol des États-Unis. Voir *Aviation Security: Secure Flight Development and Testing Under Way* […], Government Accountability Office, Washington, 2005. Un rapport du GAO de 2004 a souligné l'existence de défauts semblables dans le CAPPS II et a recommandé des modifications. Voir *Aviation Security: Computer Assisted Passenger Prescreening System Faces Significant Implementation Challenges,* Government Accounting Office, Washington, 2004. Il reste à savoir si le programme Protection des passagers du Canada prévoira une procédure régulière pour gérer les erreurs et les utilisations malveillantes.

[24] Par exemple, les Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel publiées par l'OCDE en 1980 (adoptées le 23 septembre 1980) et dont le Canada est signataire.

[25] Certains voudront contester cette affirmation en rappelant que les lois sur la protection de la vie privée du Canada contiennent déjà des exceptions à ces principes. C'est vrai, certes, mais ce n'est pas suffisant. Comme on l'a indiqué ci-dessus, la nature et l'échelle des risques encourus par les particuliers exigent qu'on fasse davantage.

[26] La Nouvelle-Écosse fait exception. En vertu de la législation sur la liberté d'information et la vie privée dans le secteur public, l'agent d'examen n'a pas le pouvoir d'enquêter et de faire respecter les dispositions de la loi sur la protection de la vie privée. Voir la *Freedom of Information and Protection of Privacy Act*, S.N. 1993, ch. 5.

[27] Cet article ne traite pas du forage de données dans le contexte de l'application de la loi, mais cette idée est source de questionnements.

[28] Par exemple, le paragraphe 69(5) de la *Freedom of Information and Protection of Privacy Act* de la Colombie-Britannique, R.S.B.C. 1996, ch. 165, exige des ministères provinciaux qu'ils procèdent à une ÉFVP conformément aux règlements du ministre. Un modèle d'ÉFVP, préparé par le gouvernement de la Colombie-Britannique (avec les commentaires du bureau de l'auteur) est disponible à l'adresse <http://www.oipc.bc.ca/sector_public/resources/pia.htm> (consulté le 18 septembre 2005).

[29] Dans leur livre *The Privacy Payoff* (Toronto, McGraw-Hill Ryerson, 2002), A. Cavoukian et T. Hamilton prouvent que le fait de veiller à la protection de la vie privée est excellent pour l'entreprise.

[30] La page d'accueil du chef de la protection des renseignements personnels est accessible à l'adresse <http://www.dhs.gov/dhspublic/interapp/editorial/__editorial_0338.xml>(consulté le 18 septembre 2005).

[31] Taipale, note 10, p. 79-80. Voir aussi la recommandation 2.4 du *TAPAC Report*. Voir aussi De Rosa, note 21, p. 17-18.

[32] Voir, par exemple, Taipale, note 10, le *TAPAC Report*, Solove, note 11. Au Canada, voir Arthur J. Cockfield, « The State of Privacy Laws and Privacy-Encroaching Technologies after September 11 : A Two-Year Report Card on the Canadian Government », *Revue de droit et technolo-*

*gie de l'Université d'Ottawa,* 2003-2004, vol. 1, p. 325-375.

[33] Taipale, note 10, p. 75-78.

[34] Dans le cas des intérêts de la sécurité nationale, il pourrait s'avérer nécessaire de permettre à des représentants indépendants d'examiner les renseignements classifiés qui se rapportent au dossier en cause si le processus de recours (de réparation) peut raisonnablement constituer un risque pour la sécurité nationale. Ces représentants seraient appelés à titre d'*amicus curiae*.