**"Where Angels Fear to Tread"—Privacy in the Brave New World of Data Sharing**

David Loukidelis
Information and Privacy Commissioner for British Columbia

Speech delivered at

*Life in a Digital Fishbowl — A Struggle for Survival or a Sea of Opportunity?*

10th Annual Privacy and Security Conference
Victoria, British Columbia
February 4, 2009

Good afternoon everyone and a warm welcome to those of you who are visiting us here in Victoria.

It's always a pleasure to be a part of this conference and as always I'm very happy to see many friends and colleagues from out of town, including my colleagues Frank Work, Jennifer Stoddart and Ann Cavoukian, Bruce Phillips, former Privacy Commissioner of Canada, and so many others.

I'd like to congratulate Greg Spievak and his entire team at Reboot for organizing yet another successful conference this year. This is the 10th annual Privacy and Security Conference and, with thanks to the BC government for its ongoing support for this conference, I hope and expect to see many more of these conferences in the years to come.

This year's theme is life in a digital fishbowl, struggle for survival or sea of opportunity? I suppose the trite answer would be, 'Who can say?' But there's no doubt the topics being covered at this conference are serious and complex, fascinating and perplexing. Cloud computing, fusion centres, electronic health records, social networking, privacy enhancing technologies. All of these developments present opportunities but also some risks and they all raise a myriad of tough issues.

I'd like today to address aspects of one development that's not new, but is nowadays spreading rapidly, driven in material part by many of the information technologies we're examining at this conference. With apologies to Pope and Huxley for rudely bringing them together in the very bad trope that is my title today, my goal here is to ask whether data sharing will place citizens in a digital fishbowl in which they struggle for survival or swim happily in a sea of opportunity.

A necessary first point of discussion is the impact of modern information technologies. Information technologies have materially improved our lives in many, many ways. Drastic decreases in data storage costs, the ease with which data flows around the world, and the sophistication of data analysis techniques can make life easier and contribute to our individual and collective quality of life. At the same time, these technologies are enabling—in some senses driving—the creation of more and more personal information databases of increasing scope and sophistication. Whether collected by governments or corporations, vast

even egregious amounts of personal information—and, thinking about what we heard yesterday about video rental records, that personal information will sometimes be salacious—are being accumulated for a host of purposes.

More and more, our digital selves will be available, very often on a lifelong basis as various bits and bytes of ourselves accumulate and grow into a construct that may be distorted and only fleetingly resemble our true selves. Australian privacy expert Roger Clarke warned over 15 years ago of the prospect that our 'digital persona', constructs created through the collection, storage and analysis of data about us, will stand in for us in our dealings with governments or businesses. These are, Clarke argued, possibly threatening and dangerous phenomena, especially given the propensity for organizations in both sectors to use data surveillance—or 'dataveillance', as he called it—to control individual behaviour, perhaps, if not ordinarily, for nefarious purposes.

These digital profiles or constructs will be often used in new ways, for administrative or other government purposes unrelated to the original purpose for which the discrete data elements were collected, either to respond to new policy or legislative directives or in the name of law enforcement or national security.

One example of dataveillance is data analysis that involves "social sorting," which involves use of personal information to slot individuals, through their digital profiles, into categories of risk or desirability. What category you are in then affects your ability to get credit, insurance, and so on. Academics in the field of surveillance studies contend that social sorting affects individuals' quality of life, widens existing divisions in society—thus creating invidious distinctions—inverts the foundations of liberal democracy, and raises fundamental questions about what it means to be a citizen. An illustrative example reported in the *Ottawa Citizen* last week is that of a person living on an aboriginal reserve who was denied a loan application because of his postal code. His postal code was included in a category of postal codes that the bank had deemed unsuitable. His personal credit history was apparently of no consequence.

To better frame what I'd like to say later about data sharing and its privacy implications, it's worth my spending a moment now sketching out what privacy is and why it's so very important.

More than 40 years ago, Alan Westin famously conceived of privacy, information privacy that is, as a principle of informational self-determination—the claim of individuals to a reasonable but meaningful degree of control over collection, use and disclosure of their personal information. As Colin Bennett has written, we can't mature and flourish as thinking, responsible, caring persons and citizens without the self-autonomy that privacy protects.

Privacy's critical role in our individual and communal wellbeing is also reflected by privacy laws around the world. The fair information practices embodied in these laws—such as limited collection, notice, original purpose, openness, accountability, accuracy and correction—all offer some control over our own personal information and at the same time restrain the appetite of government to collect and use our personal information excessively.

Data sharing is the focus of these remarks. But what do we understand 'data sharing' to mean?

By 'data sharing' I mean the programmatic or planned disclosure of personal information by one government agency to another, by one government to another government, or by a government to a private sector organization. These disclosures might be one-way, two-way or multi-faceted. They may be one-off disclosures or regular, planned exchanges of data. Data sharing is likely to occur between or among networked or connected databases rather than depending on the creation of large uber-databases.

Let me pause for a minute to mention data mining and data matching. I know these mustn't be confused with each other or with data sharing, but it's worth noting that data sharing is often carried out to enable data matching and, increasingly, will be done to facilitate data mining.

Data mining is the application of various methods to extract information from large volumes of data and analyze that information using techniques, such as statistical analysis and modeling, to undercover hidden patterns or relationships.

Data matching, of course, is the comparison of personal information from various sources in order to find matches in information, very often for the purpose of making decisions about the individuals whose personal information is being matched. Data matching to identify fraudulent social benefits claims is a common example around the world. Programs of this kind often amount to surveillance of the many to find the guilty few.

Another reason for data sharing initiatives is research. Scientists may want to use personal information from the health system and other sources for epidemiological research into the health or disease of populations. Yet another purpose for sharing personal information will be to facilitate planning, program evaluation and resource allocation. Analysis of personal information at a system level can yield valuable insights into what programs are working or failing or where resources should be invested. As with research, much if not all of these activities can be carried out using appropriately de-identified information, and I note in passing that Professor Khaled El Emam, who is here with us, has recently done some very promising important work on de-identification of personal information for research.

As an aside, the Longitudinal Labour Force File story from 2000 offers a salutary warning to governments that are moving ahead with data sharing for these reasons. The then federal Privacy Commissioner, Bruce Phillips, who is also with us here today, expressed some concerns about the project's privacy features, but these concerns were not addressed. He later went public with his concerns and the resulting hue and cry ultimately forced the initiative to be abandoned, at least under that name. The unfortunate thing is that the project may have yielded a variety of benefits, but the government's failure to properly design it from a privacy perspective—and perhaps more important its failure to be transparent about what it was trying to do—contributed to a privacy disaster that need not have, by all accounts, ever have happened.

A third, and key, purpose for data sharing is so-called citizen-centred service delivery. For example, inter-disciplinary teams made up of workers from law, justice, social service and health agencies work to serve individuals involved with the Community Court on Vancouver's Downtown Eastside. The idea is to better meet the needs of these individuals, who often have serious health and poverty issues of various kinds. Data sharing arrangements of this kind are designed to facilitate the exchange of personal information between agencies who will use it to jointly make decisions directly affecting the individuals whose information is shared. And as service-delivery boundaries become more indistinct, we can expect more data sharing of this kind to occur.

So, data sharing can yield benefits, but what are some of the risks?

An obvious risk, of course, is misidentification of individuals. A poorly designed data matching program could easily confuse two different John Smiths and inaccurately identify one of them as having committed welfare fraud. Even if the innocent John Smith can prove his innocence—and, believe me, information technology in some ways can place a prima facie burden on affected individuals to prove innocence, not the reverse—who knows whether the truth about John Smith's innocence will be accurately reflected in all data holdings where the mis-identification might linger as part of his digital persona? Innocent people have found their names on no-fly lists in the past couple of years because surveillance of their personal information in data banks has turned up "false positives". This example is above all about data quality, where missing, fragmented, outdated or poorly authenticated data all can contribute to error. The frequency of mistakes that affects the lives of ordinary people reveals the pressing need for careful monitoring of data sharing systems and for better identity management solutions.

The John Smith example illustrates another risk, namely the ease with which inaccurate or incomplete information can, through data sharing programs, be replicated and widely distributed across various databases. Privacy laws give citizens the right to access their personal information and ask for it to be corrected, but in complex data sharing arrangements and complex systems, how will citizens even know where their information is or what's been done with it? This is a vexing question for which solutions must be found.

Another risk is what I believe is our tendency to often attribute excessive reliability, sometimes almost infallibility, to the products of technology, to confuse information for knowledge, information for proof. And when the product of the technology is a new picture of an individual, based on isolated bits of personal information that may or may not be an accurate reflection of that individual, citizens should be worried. Recognizing the potential for harm in these circumstances, in New Zealand, public servants are forbidden from making decisions about individuals simply based on the results of a data match. The match may only be considered as one piece of information, which must further be verified before it can be used in a decision that affects the individual.

There's little doubt that information sharing has already become a fixture of life in our technology-enabled society or that it will only increase in the future. As Richard Thomas, the

UK's Information Commissioner, wrote last year, "…the use and sharing of personal information are now permanent features of modern life, supported by mushrooming technological advances in the storage, analysis and use of large data sets. Public, private and voluntary sector organizations will continue to require access to personal information in order to provide goods and services, combat crime, maintain national security and to protect the public."[1] Major legislative changes to facilitate data sharing for just these kinds of purposes are now under debate in the UK Parliament, and similar initiatives are under consideration in New Zealand and elsewhere.

In BC, the drive to improve data sharing across government ministries comes from a variety of sources. The Premier's Technology Council—an advisory body made up of information technology experts drawn mostly from the private sector—has in at least two of its reports called for more investment in data sharing and the associated technologies and the resources needed to operate them. In its 9th report, published in January 2007, the Council reported that the "[r]esponsible sharing of information enhances the services that the BC Government can provide and generates significant benefits for citizens."[2]

The operative word here is the word "responsible". Who will determine what falls under the scope of "responsible" sharing of information? Will we find that public servants will see responsible information sharing as a wider church than some might find desirable? The law now says that public servants must collect only that information which is necessary for a given program or activity. Yet, given that data sharing is often seen as a critical and legitimate mechanism for protecting youth at risk, the vulnerable and the sick, will we see a push to liberate our personal data by overturning or relaxing the longstanding principles of necessity and proportionality to achieve these objectives? Even the sick and the vulnerable have privacy rights, remember.

Let me recap the discussion so far. We've seen how rapid changes in information technologies are transforming our world. We've remembered the fundamental importance of privacy in our ever-more networked world. We've noted how governments everywhere are looking to leverage their information holdings by increasing their sharing, and exploitation, of our personal information. And we've also touched on some of the risks that data sharing can pose for privacy. Against this backdrop, what's the way forward? How do we move ahead? One thing, at least, is certain, while no single approach can adequately address all risks, solutions can and must be found and I'd like now to touch on some but by no means all of the possible solutions.

First and foremost governments need to do research, research and more research. Research is needed to determine in each case whether data sharing offers meaningful benefits that are sufficiently important to override or reduce privacy. This isn't merely an exercise in assessing the constitutionality of data sharing proposals, although government would do well to remember our constitution. It is, rather, a question of responsible and proportional policy-making.

---

[1] Richard Thomas & Mark Walport, *Data Sharing Review: Data Sharing Review Report* (Information Commissioner, 2008) at 9-10.

[2] Premier's Technology Council Ninth Report (January 2007), at 11.

In conducting this research, government must examine real evidence, not just accept bald assertions about privacy barriers. Privacy is often blamed for a lot of things and time and again I'm struck by how often it's just a scapegoat, an excuse for inaction. My sense is that reluctance to share personal information because of professed concerns for privacy can be a cover for bureaucratic inertia (it's easier to say no), bureaucratic infighting (departmental rivalries do exist), an excuse for hoarding valuable assets (information after all is power and thus budget funding), or based on plain ignorance of the law and privacy principles (an ignorance that's cause for concern in other areas of privacy compliance).

We simply can't allow untested privacy claims to trigger unnecessary, and possibly even harmful, dilution of the balanced and reasonable privacy rights now found in the *Freedom of Information and Protection of Privacy Act*, which is more generous to government in the area of data sharing than other Canadian privacy laws. This is why I've repeatedly told the government that it mustn't declare privacy to be a barrier to be removed unless it's carefully studied each and every such assertion. And in return government has assured me that it will undertake this research, and that it won't even propose reducing your legislated privacy rights in the name of data sharing unless and until the law is found, after careful testing, to contain real barriers to data sharing.

Even if government does decide, based on clear evidence, that legislative changes are necessary—necessary, not nice or desirable—the implications of data sharing for privacy and other civil rights are sufficiently important that legislatives amendments should proceed only after full and meaningful public consultation. My office's forthcoming position paper is intended to be part of a public conversation and any government proposal to change the law must be subjected to a meaningful public consultation. Governments don't publish white papers in this province, but something like that should be done here, and in saying this I note that the next statutorily required all-party review of our law must begin this fall, thus presenting an opportunity for legislators to study the issues and make recommendations.

Regardless whether or not legislative change is contemplated, government needs to move now to implement effective and workable policy and technological measures to protect privacy where data sharing is going on. Taken together, the measures I'd like to touch on now will go some way to providing a meaningful framework for the governance and oversight of data sharing.

A key and much-needed tool in the privacy toolkit overall is a government-wide privacy management framework that includes a chief privacy officer for the provincial government. Large corporations now commonly have chief privacy officers, who are 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. It's time such a position is created in BC, with executive support and real internal authority across government, and I've been advocating this to government over the past year or so. This position should be at the pinnacle of a government-wide privacy management framework, a comprehensive framework of policies and practices to guide all aspects of government's collection, use and disclosure of our

personal information, including through data sharing programs. Personal information needs to be managed throughout its lifecycle and a chief privacy officer standing at the pinnacle of a privacy management framework is a necessary, indeed critical, part of governance for data sharing initiatives.

As you know, many Canadian jurisdictions—including BC—have statutory or policy requirements for privacy impact assessments to be completed before proposed programs, policies or laws are pursued. A PIA is not a panacea, but it's an important tool to assess and address privacy risks. It is above all a process, not an end product, requiring organizations to assess privacy risks in order to decide whether proposed programs, systems or laws 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 PIA process has to be a concomitant of any data sharing governance framework and must be seen as an evergreen, iterative process.

Another key feature of a responsible data sharing framework—certainly a framework that allows sharing of information whereby decisions directly affecting individual interests or rights—is to allow sharing only with prior external authorization. A model for this approach exists in New Zealand. Under the New Zealand *Privacy Act*, government can engage in data matching only with the prior authorization of the country's privacy commissioner. And the data sharing-related initiatives the UK government announced last week will allow data sharing only with the prior approval of a Secretary of State, with that approval requiring the further imprimatur of Parliament on positive affirmation.

Next, we need meaningful audit tools. 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 help system administrators and regulators to ensure that rules are followed. In the context of data sharing for service delivery, strong audit capabilities are of critical importance in preventing misuses of data, data spills and even function creep.

Although it's a trite proposition, data sharing systems also must have strong security measures in order to prevent data leakages or corruption. One traditional privacy principle that applies to data sharing in a meaningful way is the obligation to take reasonable security measures to protect personal information against unauthorized collection, use or disclosure. Data security must be a high priority in the design and operation of data sharing systems.

In closing, as I said at the beginning, information technologies are driving, or at least facilitating, great changes in how governments view their information holdings and how they exploit them. The sharing of personal information for a range of purposes can yields benefits for individuals and society, but there are risks that our government must take seriously and

that must be addressed meaningfully if we're to protect privacy, a fundamental cornerstone of our free and democratic society. We are, after all, talking about *our* personal information, not the government's, and government exists to serve us, not the other way around. Let's keep those two things in clear view as we walk down the road to a brave new world.

Thank you for your kind attention, and enjoy the rest of the conference.