

Data Data Everywhere; But Researchers Dare to Drink

Reviewed by Neil Seeman

Data Data Everywhere: Access and Accountability?

Edited by Colleen M. Flood

Queen's Policy Studies Series

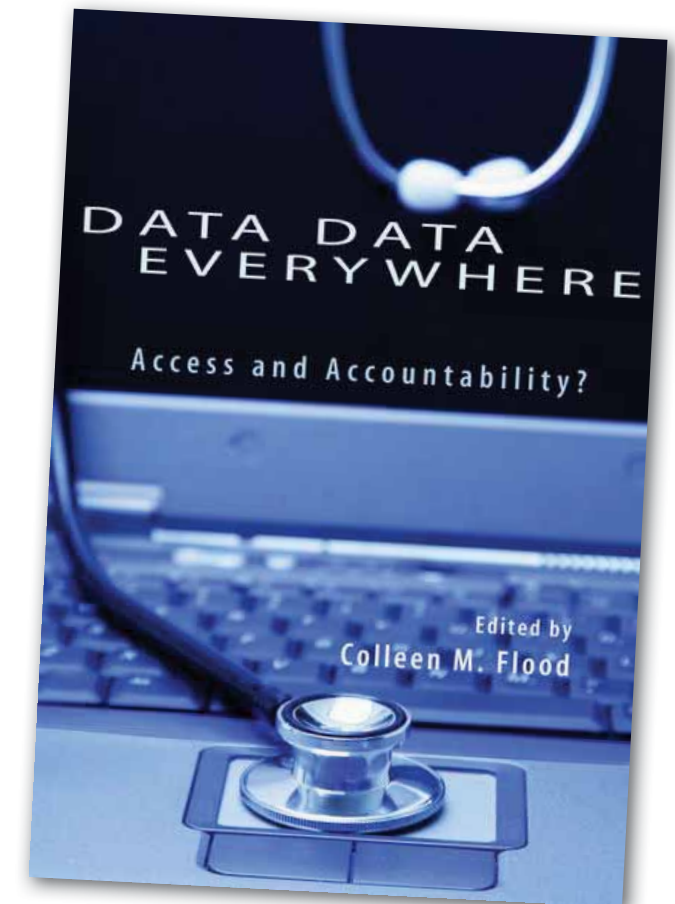
Montreal and Kingston, ON: McGill-Queen's Press; 2011

Data, data everywhere, yet nothing but a gripe. Gripe – characterized by academic chauvinism and pre-Internet era thinking – has been the chorus of some outspoken lawyers and ethicists, government officials and researchers who are terrified of the Big Brother of “Big Data” (Seeman 2012). To be fair, there are legitimate concerns about data mining, in particular the harnessing of data for secondary purposes to improve Canada’s healthcare system. It is to this end that Colleen Flood, Canada research chair in health law and policy and editor of *Data Data Everywhere*, has assembled a sterling group of thought leaders to debate whether folks closer to my way of thinking (“data-access absolutists”) or, on the other side (the camp I call the “data-access restrictionists”) need to be more nuanced.

Many of the concerns of the Big Data revolution will be familiar to readers. Beware: Online snooping of open-access blogs and social networks! Browser cookies (see Wikipedia 2012b) that track what websites I’ve visited so that marketers can sell me stuff I don’t want! iPads left on airplanes! USBs gone missing in parking lots with personal data stored on an unsecured cloud platform! (see Wikipedia 2012a).

While some are immersed in angst over these issues, the vendor community has a vested interest in solving these problems quickly – because consumers care a great deal about this. Increasing consumer attention to such issues are *protecting* Web users, with regulatory and industry-initiated changes trending toward greater and greater personal control over one’s online data. Following an investigation by the privacy commissioner of Canada, Facebook took action in June 2010 to stop applications (“apps”) from accessing private parts of a user’s profile, unless the user grants formal permission. Facebook then created a control panel that allows users to see which apps are accessing the various categories of their personal information.

Emerging Web practices support user autonomy and data control. Meanwhile, a growing number of Internet tracking companies (e.g., BlueKai Inc., Lotame Solutions Inc. and



eXelate Inc.) are joining the Open Data Partnership. This initiative – designed to allow consumers to edit personal information collected about them from online sources – builds upon the self-regulatory principles for online behavioural advertising announced by the Digital Advertising Alliance in July 2009.

We should not forget that black-letter negligence law applies to hospital employees who leave patient data in the wrong places. Yet fear builds upon fear, and this escalation is even less rational when it comes to new technology. Remember the “digital divide”? Today, Web-enabled cell phone technology is saving lives in the poorest parts of the world (Seeman 2010). Remember, too, that when clinicians talk openly about patients’ medical conditions in elevators or in shared bedroom hospital

accommodations (with a light curtain divide), this is about as private as an electronic medical record (EMR) channel built on Twitter. We *never* had real privacy in healthcare interactions, and data passed sloppily via word of mouth can be just as intrusive to patient autonomy as data passed sloppily via computer.

It seems to me that the sea change in antipathy to all things digital and data-related is a function of anti-corporate animus. (If someone's making money off it, then it must be scary!) So-called "stealth marketing," which concerns Wendy Armstrong, one of the contributors to *Data Data Everywhere*, has, in fact, been around since the 1940s. Today's "digital footprint" on someone

We never had real privacy in healthcare interactions.

stumbling upon a "new mom" website is similar to yesterday's subscription to a print magazine about raising children. Magazines kept tabs on who you were, and they sold the lists. Today, it's our e-mail inbox that gets flooded; 10 years ago, it was our physical mailbox. And keep in mind that Groupon has a multi-billion dollar valuation for a reason: people love its offers ... and it's among the easiest services to deactivate if you're frustrated by it. (Google Plus, Take note.) But, to be sure, we need strong advocates like Wendy Armstrong to ensure that consumers continue to stamp out the few, but influential, vendors that play in the "grey area" of Internet ethics.

Today, most human resources managers Google potential recruits and advisors; in the past, people phoned their friends and family to find someone to hire. The Internet is therefore a platform for *democracy* in hiring, recruiting and all aspects of commerce (with notable exceptions of the guild industries such as law or real estate) – giving consumers and citizens what they want (think Arab Spring and #occupy). This is why companies that try to game Google through fancy search-engine optimization tools often get punished and reduced in their ranking.

If you want to know about the power of the Internet for consumers, ask patients with stigmatized conditions (such as human immunodeficiency virus/acquired immune deficiency syndrome [HIV/AIDS] or a psychiatric illness). They want to share their stories, and they are among the highest users of the Internet. *The Internet is about the very opposite of social exclusion.*

Flood and her colleagues have convinced me, someone who believes in unlocking health data archives, that I need to think more objectively on these matters. And so too do the data-access restrictionists. For this accomplishment alone, this book is a smashing success. And, not to emphasize too obvious a point, educating *me* is not that important in the grand scheme of policy dynamism; what's more important, as Patricia Kosseim underscores, is educating policy makers about the importance of different perspectives.

Many of the contributors, particularly Andrea S. Gershon and Jack V. Tu, also note that educating research ethics boards (REBs), data custodians and the public about new, expanding methods distinct from "opt-in" informed consent procedures can offer new-age alternatives that can protect patients while at the same time serving to advance the quality agenda. Among the most interesting chapters to me was that served up by Dale McMurchy and colleagues, who showed that engaging HIV/AIDS stakeholders has been essential to the success of a major HIV cohort study, since this very engagement won affected individuals' trust and ongoing support and led, ultimately, to one of the richest data sets on HIV. Such engagement helps to ensure that research areas reflect the real needs of real patients.

Here's my thesis: we value diversity of opinion in this country. If we don't engage people thoughtfully in this debate over the role and reach of Big Data, if we devolve the responsibility for this debate to bureaucrats, academics, lawyers or pollsters, we will all lose, and Canadian healthcare quality will suffer for it. One of the best contributions researchers can make in this regard, Kosseim notes, is to describe real-life case studies where hard choices need to be made regarding the potential knock-on effects of opening up databases in ways that were once closed.

Let my bias be clear: the sky is not falling thanks to more data in our midst. If anything, liberals and conservatives (the latter group being especially confused on this matter, forgetting the policy needs of, say, capturing, long-form census data) have sidelined the importance of opening up administrative and other closed databanks now owned by government, with restricted access given to anointed researchers, in the name of liberty or autonomy, the ethical pillars to which all lawyers and ethicists pay obeisance.

Then comes that samurai word, *privacy*, which lawyers and information technology (IT) security consultants and other newly created professionals wield like a Herculean sword whenever anyone mentions the idea that, well, maybe people who code for a living (but don't take care of patients) can map the geo-located socio-economic status of every street in the country and link it to clinical and financial outcomes, food bank usage data, frequency of hospital visits, chronic disease states – and, maybe insights into reallocation of scarce preventive care resources might reveal themselves.

Hey, maybe the hackers who know more about linking data sets in the era of semantic data scraping off the Web (see Wikipedia 2012c) than do traditional policy researchers can reduce the number of Canadians (currently 16,500) who die each year in a hospital as a result of preventable medical errors. I offer a case example: post-market drug safety surveillance is difficult. Finding self-reports, from the Web, biased as they may be toward tech-savvy patients, can prevent the next Vioxx incident, especially if the early warning signs from these self-reported patient and caregiver stories can be rapidly captured

(in 24 hours) and weighed against administrative databases (Rizo et al. 2011). As contributor Robyn Tamblyn explains, current information on adverse events is acquired via voluntary reports from providers, with 99% of adverse events failing to be detected. This is both tragic and absurd in 2012.

With two taps on Google, and a little bit more time if I code the right algorithms in multiple languages, I could find out the rate, per region, of people who die in the shower, and I could find out their identities. Yes, the example is an absurd one. (Usually, I'm led to understand, dying in the shower is secondary to something else, such as a drug overdose.) But I chose it because some enterprising 17-year-old could, theoretically, create a call to action on Facebook for such information, post it on Factual (<http://www.factual.com/>) and then invite the world to manipulate, correct and edit the data. If administrative databases don't engage the public, they (and the researchers and funders behind them) will become less relevant and less influential in policy and allocation decisions.

I have identified just a few of the advantages of diversifying the club of people who deserve access to linked administrative and clinical data sets, provided, of course, that these data are encrypted and de-identified (a challenge vastly overestimated by some in the IT security business who make a living convincing people how hard this is). Robert Ouellet, in his contribution, makes a solid case that it's ethically slippery to delineate a patient's "circle of care" with sharp lines: "Arguably, housekeepers/cleaners, dieticians, clergy, social workers, nurses, aides, and doctors all play a part in an individual's health care." The bias to action should be to let more people in.

Missing, until now – and thank heavens for the timely arrival of this important book by Colleen Flood and colleagues – is a sober wake-up call to action, an eloquent and balanced suite of brainy, open-minded perspectives of the policy implications and challenges associated with the emergent world of Big Data. *Big Data* is a catch-all term (disclosure: I run a for-profit company in this sector) for large data sets – for profit or government, outside healthcare or inside – that house data whose manipulation, or linkage to other open data sets, can potentially raise issues of ethical concern while at the same time, should they be given less-fettered restrictions, could open up a world of insights into the quality-driven agenda in healthcare, notably, moving Canada a tiny bit closer toward addressing individual patient needs, wants and expectations.

One issue of potential concern is the new post-Big Data era of informed consent, as William W. Lowrance notes. In the context of electronic health records (EHRs), one can imagine well-intentioned secondary use of data (think of patient-based values encoded into the EHR) that could be linked to other "habits" data or patient-linked clinical data, and thereby render dated the notion of highly-specific informed consent rules. "The distinction between research and other secondary uses of data

has blurred over time," Don Willison, Elaine Gibson and Kim McGrail write. A flood of confusion and uncertainty has ensued.

On these issues, a huge thank you is due to Andrea S. Gershon and Jack V. Tu, who carefully describe a taxonomy of issues and related legislation that applies to consent-related matters in this evolving landscape. The example provided by Patricia Martens of how the Manitoba Centre for Health Policy addresses these issues, ensuring optimal use of its data repository while ensuring data security, is invaluable to any research unit or

For this accomplishment alone, this book is a smashing success.

company that wishes to emulate best practices. As Willison and colleagues explain, it is important to move to consensus across multiple affected parties on the acceptable circumstances for the use of health information, including the place for specific individual consent; the required architecture for secure data management; and the role of third-party commercial vendors and how it may affect consent issues.

Although balanced, the subtext of the chapters by many of the contributors leads me as a reviewer toward a conclusion, informed by my own bias toward open data, that says that the status quo is too restrictive, researchers need more access to data, delays in access are absurdist and the possibilities afforded by making it easier to link administrative and clinical databases in healthcare are profound. This conclusion is perhaps most strenuously articulated by the always-eloquent essayist Steven Lewis. Lewis, in putting paid to "the belief that the open and highly diffused use of health information for research and other non-clinical purposes is inherently sensitive and dangerous" has, in my view, two crossbow arrows in his quiver: the very spirit of natural law and, not least, the spirit of medicare itself.

Natural law and how it has been interpreted by eminent jurists would suggest, in my view and as Lewis states, that the onus should clearly fall on those who restrict data access as opposed to those (no matter what their academic pedigree) who insist upon ring-fencing access. Further, medicare is based on many principles, but overarching is the idea that we are all equal with equal capabilities and that we share solidarity of interest in lifting the whole boat to make the tide of quality healthcare rise. If this is the case, then patients and caregivers are equal stakeholders in this debate, and, although poll numbers fluctuate, most patients (and certainly most with chronic disease) are content with wider access, with much more free flow of information and with the easy linkage of databases – without the current ecosystem of approvals of REBs from dozens of academic clubs that set their own variant rules over who may apply, how and why. States Simon B. Sutcliffe: "Within a publicly funded system with expectations of transparency, accountability, and

sustainability of evidence-based, effective care, these data are relevant to many stakeholders.”

Yes, dear readers, I am biased. I am continually flummoxed (as is every entrepreneur I know, $N > 100$) by the reticence of bodies we are all well familiar with to unbolt the Fort Knox protocols that define the fiefdoms of their data warehouses. Let us please rid ourselves of the very word *privacy*. It is no longer apt in the world of Web 3.0, where I can capture the names and locations of folks around the world with various chronic illness, such as amyotrophic lateral sclerosis (ALS) or cancer, by joining open-access communities. I can then map their real names on to their

Let us please rid ourselves of the very word *privacy*. It is no longer apt.

real Facebook identities and find out what music they like and who their friends are. And, if I'm particularly nefarious (for the record, there are a litany of strong codas forbidding this, including my own ethics), I can sell these data to corporate or political interests. What matters today – ask Mark Zuckerberg – is *user-defined control over one's own data*, not *privacy*, the latter being a term that predated the ascendance of social media in 2004.

I am reminded here of a session of the privacy section of the Canadian Bar Association in 2009 at which I asked an audience of 500 lawyers, “Have any of you heard of the website PatientsLikeMe?” Nobody raised a hand. This is interesting coming from a group of folks who are paid to be experts on the changing notions of privacy, their education having been subsidized by taxpayers. People on PatientsLikeMe (PLM [patientslikeme.com]; now over 140,000 people and 1,000+ conditions) share their intimate personal information with the community and identify off-label uses of medications they're taking – and Canadians (as of 2008) are among the highest proportion of PLM users (Seeman 2008).

It concerns me that lawyers and ethicists are unaware of the speed of the Web and the rise of technology's power to capture and mine data. There is no time for luxury in setting ethical standards and regulations years after Facebook or Groupon has already set the very standard that the consumer loves and to which the customer has become accustomed. But I digress.

Colleen Flood and Bryan Thomas, in their introductory chapter, suitably called “Searching for a Sweet Spot,” correctly note that the goal is not to generate more and more data (that is happening and will continue to happen inexorably) but, rather, to transform those data into reliable evidence. This will allow Canada to cross the quality chasm. It is now clearer that quality is less about finding a nirvana technology solution to reducing adverse events or improving transitions of care than it is about mining data intelligently to direct providers at the point of care, and directing planners who distribute dollars to the data that

matter. Dorothy Pringle puts a backdrop on this that we should all remember: “Research that would answer most health services-related questions relevant to the roles and effectiveness of professional and non-regulated health-care practitioners – including what influences quality health care – has yet to be done.”

Canada can be an engine of innovation in healthcare, and we can sell this know-how to others – we already have best-in-class intellectual capital at sites such as the Institute for Clinical Evaluative Sciences – and provide this insight to governments from Beijing to Bangalore. The world is changing toward an ecosystem of people from different industries that want to make healthcare more accessible and more elegant for the public to understand. For many years, Alan Katz reminds us, the potential power of administrative data sets for quality improvement has been well documented. Lisa M. Lix highlights the challenges of hospital discharge data alone, which have limited value on their own simply because hospitalization is an acute event. The chapter by Roger Shafe, Pamela Spencer, Melissa Hudson, Kamini Milnes and Terrence Sullivan makes this “linkage” case convincingly in the domain of cancer control, where database linkage provides profound understanding of improved surveillance, system planning and budgeting and performance and quality improvement.

Let us not fall behind the rest of the world. If so, we could, for example, make the same mistake in delaying our decision to harmonize data use standards across provinces for the very same reason that we have been on the wrong side of the evolution of maturing e-health decision tools at the point of care. Hub national organizations such as the Canadian Partnership against Cancer, as suggested by Chafe and colleagues, could be crucial to promoting best practices in data sharing.

Here I have a humble suggestion: I believe that the highest role of researchers is not to own data or set standards on who enjoys privileged access but, rather, to help set the critical questions that the crowds of 20- and 30-somethings can answer. For example, we should unleash the raw granular data sets from an EHR, encrypt and de-associate the data from all personal information, link a bunch of data sets and then support a “hack-a-thon” with 50 coders over a weekend (cost = beer and pizza). And we should ask them, “How do we use these data to crack the question, what open source data from social networks and other online spaces are available that can be mashed onto these unlocked data that can help us answer where people are who have weak or non-existent social ties that will hurt them when they fall through the cracks and lose their job or suffer ill health?”

Across all the excellent contributions to this important volume lies a common theme: we need a cultural change that keeps with the times and ensures that roadblocks to what we all know to be valuable – quality patient-focused care generated through rich, secure, linked data sets – melt away. We don't have all the answers yet; but we know the problems, and the

authors of this book have asked the right questions with the right equanimity at the right time. Thank you. **HQ**

References

Rizo, C., A. Deshpande, A. Ing and N. Seeman. 2011. "A Rapid, Web-Based Method for Obtaining Patient Views on Effects and Side-Effects of Antidepressants." *Journal of Affective Disorders* 130(1-2):290-93.

Seeman, N. 2008. "Web 2.0 and Chronic Illness: New Horizons, New Opportunities." *Electronic Healthcare* 6(3): 104-10. Retrieved April 23, 2012. <<http://www.longwoods.com/content/19506>>.

Seeman, N. 2010. *From Ehealth to Mhealth: Celebrating the Mobile Phone at 5 Billion* [Essay]. Toronto, ON: Longwoods Publishing Corporation. Retrieved April 23, 2012. <<http://www.longwoods.com/content/21873>>.

Seeman, N. 2012. *N of 1* [Essay]. Toronto, ON: Longwoods Publishing Corporation. Retrieved April 23, 2012. <<http://www.longwoods.com/content/22739>>.

Wikipedia. 2012a. *Cloud Computing*. Wikimedia Foundation. Retrieved April 23, 2012. <http://en.wikipedia.org/wiki/Cloud_computing>.

Wikipedia. 2012b. *HTTP Cookie*. Wikimedia Foundation. Retrieved April 23, 2012. <http://en.wikipedia.org/wiki/HTTP_cookie>.

Wikipedia. 2012c. *Web Scraping*. Wikimedia Foundation. Retrieved April 23, 2012. <http://en.wikipedia.org/wiki/Web_scraping>.

About the Reviewer

Neil Seeman is founder and chief executive officer (CEO) of the RIWI Corporation, a global Internet technology company that captures opinion and intent data 24/7 around the world, and engages and moves millions of people to action for corporate and non-governmental organization campaigns. He is CEO of the Health Strategy Innovation Cell at Massey College, in Toronto, Ontario; senior resident in health system innovation; and co-author of *XXL: Obesity and the Limits of Shame*, a finalist for the \$50,000 Donner Book Prize, awarded to the best book on public policy by a Canadian. He has taught health law and policy and advises numerous Web start-ups.

Better Healthcare
... everywhere

Longwoods.com