Research in Brief

From Patients to Petabytes: Genomic Big Data, Privacy, and Informational Risk

Julie Frizzo-Barker & Peter A. Chow-White
Simon Fraser University

Abstract *Genomic big data is an emerging information technology, which presents new opportunities for medical innovation, as well as new challenges to our current ethical, social, and legal infrastructure. Rapid, affordable whole genomic sequencing translates patients' most sensitive personal information into petabytes of digital health data. While a biomedical approach traditionally focuses on risks and benefits to the human body, the fields of Communication and Science and Technology Studies (STS) can provide some of the critical and theoretical tools necessary to navigate the newly emerging terrain of the human body as digital code. Core areas of expertise from these fields, including the Internet, the network society and the social constructions of technology, ground our discussion of the social implications of open access genomic databases, privacy, and informational risk*

Keywords *Genomics; Big data; Personalized medicine; Privacy; Informational risk*

Résumé *Le « Big Data » en génomique est une technologie de l'information émergente, qui offre de nouvelles possibilités pour l'innovation médicale et présente de nouveaux défis pour nos structures éthique, sociale et juridique. Un séquençage génomique rapide et abordable, convertit les renseignements personnels les plus sensibles des patients en pétaoctets de données numériques de santé. Tandis que l'approche biomédicale traditionnellement se concentre sur les risques et les bénéfices pour la santé, les Études de la Communication, de la Science et de la Technologie (STS) peuvent fournir certains outils critiques et théoriques nécessaires afin d'explorer le terrain émergent de la représentation numérique du corps humain. Les domaines principaux de ces champs d'étude dont l'Internet, la société en réseau et les constructions sociales de la technologie, forment la base de notre discussion sur les implications sociales de l'accès ouvert aux bases de données génomiques, la confidentialité et les risques liés au stockage et la diffusion de l'information.*

Mots clés *Génomique; Données volumineuses; Médecine personnalisée; Vie privée; Risque informationnel*

Introduction

The 1997 sci-fi film *Gattaca*, set in "the not-too-distant future," portrays a eugenics-

Julie Frizzo-Barker, Doctoral student, School of Communication, Simon Fraser University. Email: jfrizzob@sfu.ca . **Peter A. Chow-White**, Associate Professor, School of Communication, Simon Fraser University. Email: petercw@sfu.ca .

Canadian Journal of Communication Vol 39 (2014) 615–625

based society in which genetically manipulated, "improved" children are classified as "valids," while those born of traditional means are the less desirable "in-valids" (DeVito, 1997). Although genetic discrimination is supposedly prohibited in this society, the "in-valid" protagonist Vincent Freeman struggles to overcome genetic discrimination to fulfill his dream of space travel. Today, the genetic technology in the film seems less science fiction and more science fact as DNA sequencing technologies have transformed the complex whole genome sequencing (WGS) process from a billion-dollar, decade-long race to a relatively affordable service that costs close to $1000 and takes about a week (Burn-Murdoch, 2012). As patients are translated into petabytes of digital data, our shifting sociotechnical landscape is characterized by new opportunities for medical breakthroughs, as well as new informational risks to privacy. Biotechnologies are "disruptive to some of our most fundamental social categories and boundaries" (Gerlach & Hamilton, 2005, p. 80), human and digital, in vitro and in silico, the bench and the bedside. The expansion of genomic big data presents new challenges for the public, practitioners, and policymakers in terms of managing a new type of personal information in the public healthcare system and privacy. In this article, we argue that key concepts from the scholarly fields of Communication and Science and Technology Studies (STS) offer a useful, complementary approach to understanding the innovation and deployment of new information technologies, such as genomics.

According to Castells (2010), genomics is itself an information technology, due to the manipulation, networking, and recoding of human DNA as digital data. Genomic research today has more to do with the Internet and computational suites than tissue samples in petri dishes. The shift from paper-based medical charts to digital databases has profound and far-reaching implications that are currently racing ahead of social, legal and medical infrastructure in Canada and beyond. Open access DNA databases pave the way for large-scale medical collaboration, as well as new ethical dilemmas to do with privacy and informational risk. For example, a human genome is a unique identifier of an individual, and also a network identifier of familial relations and hereditary diseases that may be widely accessed from a digital database (Allyse, Karkazis, Lee, Tobin, Greely, Cho & Magnus, 2012; Lunshof, Chadwick, Vorhaus & Church, 2008). Scholars of communication and STS are uniquely positioned to shed light on how the Internet, open access databases, and genomic information technologies affect personal privacy in our networked society.

Conceptualizing links and convergences between communication and STS is both an intellectual and practical endeavour. The authors of this article are currently involved in a collaborative clinical genomics research project for cancer. We spent the past three years working closely with genome scientists, bioinformaticians, clinicians, and health economists. The genomic test will potentially be the first for the clinical setting in Canada to move genomics from the research lab bench to the hospital bedside. We wanted to understand the possible social impact of technology diffusion from the context of a small group of innovators to a larger population and institutional setting. There is a critical need to understand the Internet, digital databases, and the social shaping of information technologies in relation to scientific knowledge development and its operationalization in the healthcare context. We engaged with our colleagues

and a network of stakeholders to develop practical guidelines for handling genomic information in the public healthcare system. We also wanted to disseminate our findings to policymakers and the public to help facilitate discussions.

While a traditional biomedical approach focuses on risks and benefits to the human body, the fields of communication and STS can provide some of the critical and theoretical tools necessary to navigate the newly emerging terrain of the human body as digital code. Scholars of communication have analyzed the socio-cultural impacts of mediated information since the 1960s, networked computing since the 1970s, and media and information technologies since the 1990s (McLuhan, 1964; Williams, 1975; Castells, 2000). Meanwhile scholars of STS have illuminated how design, practice, and politics are intertwined, with a focus on engineering, knowledge production, and transportation (Latour, 1996; Pinch & Bijker, 1984; Winner, 1980). More recently, since the introduction of the Internet, scholars of STS have also turned to investigate media and information technologies, building on the field's foundational strengths in complex technologies and knowledge systems (Boczkowski & Lievrouw, 2007; Bowker & Star, 1999). Scholars of both communication and STS have traced the co-evolution of biomedicine and information technologies for over thirty years. Over the past decade, Eugene Thacker (2004, 2005) has investigated key issues at the intersection of biotechnology, genomics, and politics. Celeste Condit (2004, 2007) has researched culture, science reporting, and the role of genomics in popular understandings of race. And Chow-White and Garcia-Sancho (2012) have conceptualized DNA databases as "spaces of convergence for computing and biology" that have evolved in both form and function over the past fifty years, setting the stage for today's genomic research in which "the biological and computational are currently indivisible" (p. 128). Likewise, the fields of communication and STS provide a fruitful theoretical "space of convergence" for studying the social impacts and technological shaping of genomics.

How can communication and STS inform a greater understanding of the new opportunities, challenges, and risks introduced by genomic big data? We address this question guided by several complementary concepts: the Internet and the network society from communication, as well as the social construction of technology from STS. First, we define big data and its relationship to genomics. Then, we illustrate the social significance of genomic big data by looking at the core communication interests of open access information, privacy, and informational risk. Our analysis demonstrates the need for communication and STS scholarship in studying the design and use of new information technologies, such as genomics. In addition, we expand the literature by showing how key concepts from these distinct academic arenas can complement one another in timely, useful ways.

Genomics as big data

The influential role of big data in business, marketing, science, and medicine has led experts to label it "the industrial revolution of data" (Buxton, Hayward, Pearson, Kärkkäinen, Greiner, Dyson, Ito, Chung, Kelly, & Schillace, 2008). By "big data" we mean vast data sets collected from and/or via the Internet, social media networks, sensors and surveillance tools, that are processed, shared, and combined beyond tradi-

tional databases through networked computing and software algorithms (Lohr, 2012). Big data is all about predictions, connections, and relationships amongst vast data sets (Mayer-Schönberger & Cukier, 2013). With 2.5 quintillion bytes of data generated per day, 90 percent of the data in the world today has been created in the last two years alone, characterized by a significant increase in the four V's: volume, velocity, variety, and veracity (IBM, 2012). Big data employs all sorts of smart machines, or artificial intelligence algorithms, to parse out useful data linkages in information across disparate data trails from databases, social media networks, and surveillance sensors. According to Rod Smith, vice president for emerging Internet technologies at IBM, these networked linkages are key to big data's significance: "Big data is really about new uses and new insights, not so much the data itself" (Lohr, 2012). Most accounts of big data point to the technological features as we describe above; however, big data is also a cultural shift in terms of collaborations. For example, Chow-White and Garcia-Sanchez (2012) show how early big data ventures, such as genomics, brought together different groups in the academy that did not normally work together. This collaboration became a new scientific field with new methodological approaches.

Genomics is big data. A single human genome is made up of 6 billion chemical letters of information (DNA Dilemma, 2012). The file size of a single genome varies depending on metadata and can be as large as 200GB. This particular type of data holds new social significance, not only because it represents an individual's one-of-a-kind DNA, but also because it takes on a tangible, material life of its own as it enters the digital database. In communication terms, data are representations. They are "cultural objects that stand in for stimuli and mediate relations" (Chow-White & Green, 2013, p. 6). This is especially true for genomic data, which may be accessed, replicated, or analyzed in unforeseen ways with unknown implications.

The world's largest genomics research institute, China-based BGI, sequences the equivalent of 2,000 human genomes per day, limited only by the fact that "the ability to determine DNA sequences is starting to outrun the ability of researchers to store, transmit and especially to analyze the data" (Pollack, 2011). Strasser (2012) contextualizes the so-called data deluge, noting that, since the Renaissance, societies have invented new technologies to deal with perceptions of information overload. Yet he concedes a few distinct traits in today's life science research:

> the analysis of the data is carried out by researchers with different disciplinary backgrounds than those who produce it, the analysis is heavily dependent on statistical tools, and the analyzed data come from the laboratory, not the field. (Strasser, 2012, p. 86)

To this list, we would add data mining, which refers to the process of attempting to discover patterns and meanings from large data sets, with the Internet being the most obvious example of such a database (Manovich, 2001). Data mining involves gathering

> disparate types of information from users and consumers—sometimes with the users' knowledge, sometimes without—and turn[ing] this information into analytical data points for measurement, sorting, and classification to achieve different organizational and institutional goals. (Chow-White & Green, 2013, p. 556)

Big data is designed to be aggregated, mined, and networked (boyd & Crawford, 2012); that is, actors can identify patterns and relationships about a person, group or data set. This points us toward a larger societal shift at play: the network society.

Castells' (2010) theory of the network society articulates the significance of the Internet in contemporary society. It underscores how the practices and processes of everyday life are made possible through a constellation of digital and mobile connectivity, with key social structures and activities organized around electronically processed information networks. Lee Rainie and Barry Wellman (2012) suggest that "the triple revolution" of social networks, the Internet, and mobile technologies work in tandem to form the distinctly new social operating system we live in today. What does this new socio-technical landscape mean for genomic identities, healthcare, and privacy? On one hand, medical researchers and clinicians are eager to harness significant opportunities associated with genomics and predictive analytics to aid in medical discovery, to streamline policies and programs, and to evaluate critical data (Gordon & Pai, 2012). For example, the genome scientists and clinicians we work with are developing a genomic test for cancer. The goal is to identify the genetic signature of an acute form of cancer, acute myeloid leukemia, to make better treatment decisions. In the process, new challenges arise as this highly sensitive personal information is uploaded to open access Internet databases, and even social media websites. We explore this tension further in three interrelated aspects of genomics: the open access research model, informational risk, and networked privacy.

Open access genomics

The field of genomics represents one of the first public, global big data projects. The mapping of the first human genome, spanning from 1990 to 2003, set the precedent for genomic research as a collaborative field of big data research involving a global group of scientists, organizations, and funding (Collins, Morgan & Patrinos, 2003). Ultimately, the open-access approach of the multidisciplinary Human Genome Project (HGP) trumped the simultaneous attempts of Celera, a private company, to map and patent the first human genome (Marris, 2005). The significance of this research model reflected the ethos of the HGP itself: sharing such important data via Internet databases made it widely accessible for the greater good of society.

These ground-breaking scientific efforts dovetailed with equally significant developments in research and technology: the emerging open access model for information practices, and the early adoption of Internet technologies. In 1996, a group of leading scientists convened in Bermuda to pioneer a new set of principles requiring all DNA sequence data to be copyright-free and released within twenty-four hours of generation, in stark contrast to traditional scientific practices of releasing experimental data only after publication (Contreras, 2011). The resulting "Bermuda Principles" policy initiative shaped contemporary open access scientific practices and the very concept of viewing information as a global knowledge resource.

But who, realistically, is able to access such genomic big data, and to what end? As Internet technologies began to penetrate mainstream society in the early 2000s, communication research swiftly addressed digital inequalities, redefining "access" in social as well as technological terms:

> the pressing question will be *not* "who can find a network connection … from which to log on?," but instead, "what are people doing, and what are they *able* to do, when they go on-line?" (DiMaggio & Hargittai, 2001, p. 4)

A decade later, Internet researchers are revisiting these core questions in the context of open access, big data practices. Manovich (2012) identifies the players in the big data arena as those who upload the data, those who have the means to retrieve it, and those with the expertise to analyze it. Those experts represent the smallest group of innovators with the greatest influence in the big data game.

Open access genomic collaborations tend to have socially and scientifically progressive goals: the International HapMap Project traces genes associated with human disease, the 1000 Genomes Project aims to identify common genetic variants, and the $20 million International Genome Consortium is developing an open access database to decode genomes from 25,000 cancer cells (Gulland, 2010). At the present, Canadian researchers are soon to launch their own nationwide initiative called the Personal Genome Project Canada (PGP-Canada), which aims to gather and sequence the DNA of as many Canadians as possible through an independent public project by the Hospital for Sick Children and the University of Toronto's McLaughlin Centre for Molecular Medicine (Personal Genomes, 2012). According to the participant consent form, the study aims to

> explore ways to connect human genetic information with human trait information in a public fashion so that such data may be used for hypothesis-generating research and other scientific, clinical and commercial development efforts worldwide. (Personal Genomes, 2012, p. 2)

The project has been green-lighted by the most rigorous research ethics board in Canada. Ontario's Privacy Commissioner, Ann Cavoukian, fears, however, that participants may not understand the ramifications at stake in consenting to allow their most sensitive personal information, and indirectly their families', to be uploaded, networked, replicated, and retrieved for a wide variety of purposes as yet unknown (Abraham, 2012). For example, one's genomic data may be used, without any further consent for research, in ways that may run counter to the participants' cultural or moral values (Allyse et al., 2012).

Digital databases of sensitive personal information are nothing novel in and of themselves. The Canadian RCMP's National DNA Databank includes approximately 160,000 samples from convicted offenders and 50,000 DNA samples from crime scenes (DNA Dilemma, 2012). Canada also maintains an extensive electronic health records (EHR) system, providing shared access to various authorized healthcare providers. EHR systems are characterized as: *complete*, integrating information across various health providers; *life-long*, storing information spanning the life of a citizen; *accessible*, available to healthcare providers across geographical borders, and; *secure*, protected against public access (Williams & Weber-Jahnke, 2010). What is novel about PGP-Canada and other genomic databases is their open access model, which will span the geographical and social borders outlined by EHR systems. The database, including details such as physical attributes and medical history, may be accessed by researchers, government, and apparently, average citizens seeking family connections or related health information.

Indeed, those who access and use the PGP-Canada database will shape the technology as it develops. Feenberg (1992) notes the users' pivotal role in terms of "technical code," which defines "the object in strictly technical terms in accordance with the social meaning it has acquired" (p. 178). In short, technologies take on various meanings divergent from their original design as they become socially embedded and influenced by their user. This characterizes open access genomic data, which remains open to various applications and exploitations. The translation of genomic data from bench to bedside represents something more significant than a new mode of scientific discovery and deployment. The shift involves enormous social and economic implications. As such, STS concepts, such as technical code, are useful in linking technology development with social meaning and values, and in turn for advising on ethical and legal best practices for genomic data. New definitions of personal privacy and new informational risk arise in the network society. We now turn to explore those further.

Privacy and informational risk in the network society

Applying big data methodologies to personal health information raises thorny issues. Biotechnologies challenge our traditional notions of privacy and introduce new informational risks. For one, open access genomic databases fall outside of traditional social, ethical, and legal infrastructure, beyond bricks and mortar institutions and geographical borders. In addition, the governance of genomic big data remains unclear and inconsistent even amongst the researchers and medical practitioners who daily navigate this domain. As is often the case with new technologies, information and privacy legislation is not yet abreast of the current social realities. Canada remains the only G8 country without legislation to prevent the use of genetic information for non-health and non-research purposes. The United States, for example, enacted the Genetic Information Non-Discrimination Act (GINA) in 2008, which prevents health insurers and employers from accessing genetic information.

In Canada and the United States, discourses of privacy have historically been defined by individuals' rights, characterized by secrecy, anonymity, and solitude in personal matters (Gavison, 1980; Solove, 2008). Iceland, in contrast, considers its citizens' DNA as a part of communitarian biotechnological innovation. In 1998, the Icelandic government partnered with deCode Genetics to map the genome of the nation as part of a broad computerized medical database (Palsson & Rabinow, 2001). But, as open access genomic databases continue to develop globally, we are forced to redefine traditional notions of privacy created in earlier techno-social contexts, and consequently, address new informational risks.

Early regulation on human research, based on the Tuskegee experiments of the 1970s, focused mainly on protecting participants against physical and psychological harm, and thus informational risk is only mentioned tangentially (Hudson, 2011). Informational risk refers to the economic and social consequences involved in making one's private data public, the impact of incidental findings that may also affect one's family members, and the long-term participatory risks for research participants (Allyse et al., 2012). The informational risks associated with genomic big data are not yet widely understood, perhaps because they still seem futuristic and hypothetical. Some risks are known, such as the limits of control over genomic information once it is uploaded

to a public database. At this point, the primary researcher cannot guarantee privacy or the ability for the participant to withdraw their data. Some risks are unknown, such as future discoveries and uses of genomic data. The very idea of "biotechnology" can be confusing to the general public as it "often encompasses existing, emerging and imaginary scientific techniques" (Gerlach & Hamilton, 2005, p. 80). We can see the lack of knowledge of informational risk in one study of patients with diabetes, in which participants were more concerned with the privacy of their physical tissue samples entering biobanks than with their digital health data entering an online registry, despite similarities in their purposes and long-term uses (Gibson, Brazil, Coughlin, Emerson, Fournier, Schwartz & Willison, 2008). The physical samples were seen as more tangible, bearing more serious privacy implications, where the digital data was seen as more anonymous.

As the Internet has become increasingly enmeshed in everyday life over the past decade, the general public has become more comfortable sharing highly personal information online. We see this in the rise of social media and online banking. Latour (2007) describes one of the effects of this shift, which further exposes our subjectivities to the realm of empirical inquiry: "it is as if the inner workings of private worlds have been pried open because their inputs and outputs have become thoroughly traceable" (p. 2). More and more of our personal data is accessible to be downloaded, shared, analyzed and sorted. Yet big data methodologies pose an epistemological dilemma: raw data does not ascribe social meaning. In other words, big data is extremely useful in terms of *what* to analyze, but not *why* or *how* we ought to go about it. Again, this is where we see concepts from communication and STS as powerful sociocultural tools. Bowker and Star (1999) and Lyon (2003) have exposed the informational risks involved when classification systems and algorithms are used as social tools for surveillance. According to boyd (2010), the future of data is characterized by "networked privacy," and we therefore ought to shift to "a model that focuses on usage and interpretation. Who has the ability—and the right—to interpret what data they think they see?" (p. 349). These new sensibilities will require a solid understanding of the Internet, the network society, and the social constructions of technology.

Informational risk has only recently emerged as a visible public concern, with numerous privacy loopholes identified across public databases containing genomic data (Hayden, 2013). True de-identification of genomic data cannot be guaranteed since DNA is inherently one-of-a-kind and thus a small number of genetic variants can uniquely identify a participant (Lunshof, Chadwick, Vorhaus & Church, 2008; Lin, Owen & Altman, 2004). It has already been proven that supposedly anonymous participants can be re-identified from public genetic data, if one already knows the person's genetic makeup (Schadt, Woo & Hao, 2012). In the most recent breach, the US National Institute of General Medical Sciences (NIGMS) reacted by simply removing some data from public view without addressing the larger issue at hand. In short, open access DNA databases are not well protected by law. But simple updates to existing laws will not be sufficient to govern informational risk introduced by big data. Rather,

> new institutions and professionals will need to emerge to interpret the complex algorithms that underlie big-data findings, and to advocate for

> people who might be harmed by big data. (Mayer-Schönberger & Cukier, 2013, p. 172)

This presents a vital opportunity for scholars of communication and STS to engage with practitioners, policymakers, and the public.

Conclusion

Genomic big data is rapidly emerging as an information technology with real promise for medical innovation, while presenting new challenges for privacy, including informational risk. Open access databases of highly sensitive personal information have profound, long-term implications that are not currently addressed by the social, legal, or medical infrastructure in Canada and beyond. As we have explored in this article, the scholarly fields of Communication and Science and Technology studies provide some of the critical and theoretical tools necessary to understand the human body as digital code in contemporary society. While communication has historically focused on technological developments, such as the Internet, as well as social shifts, such as the network society, the phenomenon of genomic big data motivates the field to engage in a more fruitful relationship with STS in order to address the social constructions of the technology of genomics as an information science. The outcome of this burgeoning scholarly relationship is more than theoretical; rather, it provides the groundwork for researchers in both of these academic fields to make a tangible social impact in terms of public policy around genomics and privacy.

References

Abraham, Carolyn. (2012). Why Your DNA is a goldmine for marketers. *The Globe and Mail.* URL: http://www.theglobeandmail.com/news/national/time-to-lead/why-your-dna-is-a-goldmine-for-marketers/article6293064/ [December 20, 2012].

Allyse, Megan, Karkazis, Katrina, Lee, Sandra S., Tobin, Sara L., Greely, Henry T., Cho, Mildred K., & Magnus, David. (2012). Informational Risk, Institutional Review, and Autonomy in the Proposed Changes to the Common Rule. *IRB: Ethics & Human Research, 34*(3), 17–19.

Boczkowski, Pablo, & Lievrouw, Leah A. (2007). Bridging STS and Studies: Scholarship on media and information technologies. In E. J. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), *The handbook of science and technology studies* (pp. 949–977). Cambridge, MA: MIT Press.

Bowker, Geoffrey, & Star, Susan L. (1999). *Sorting things out: Classification and its consequences.* Cambridge, MA: MIT Press.

boyd, danah. (2010). Networked Privacy. *Surveillance & Society, 10*(3/4), 348–350.

boyd, danah & Crawford, Kate. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679.

Burn-Murdoch, John. (2012). Big Data: What is it and how can it help? *The Guardian.* URL: http://www.guardian.co.uk/news/datablog/2012/oct/26/big-data-what-is-it-examples [November 6, 2012].

Buxton, Bill, Hayward, Vincent, Pearson, Ian, Kärkkäinen, Leo, Greiner, Helen, Dyson, Esther, Ito, Joi, Chung, Anshe, Kelly, Kevin, & Schillace, Sam. (2008). Big data: The next Google. Interview by Duncan Graham-Rowe. *Nature*, *455*(7209), 8–9.

Castells, Manuel. (2000). The information age: Economy society and culture. Oxford, U.K.: Blackwell. (Originally published in 1996)

Castells, Manuel. (2010). *The Rise of the Network Society* (2nd Edition). Oxford, U.K.: Blackwell.

Chow-White, Peter A., & Garcia-Sancho, M. (2012). Bidirectional Shaping and Spaces of Convergence: Interactions between biology and computing from the first DNA sequencers to global genome databases. *Science, Technology & Human Values, 37*(1), 124–164.

Chow-White, Peter A., & Green, Jr., Sandy E. (2013). Data Mining Difference in the Age of Big Data: Communication and the social shaping of genome technologies from 1998 to 2007. *International Journal of Communication, 7*, 556–583.

Collins, Francis S., Morgan, Michael, & Patrinos, Aristides. (2003). The Human Genome Project: Lessons from large-scale biology. *Science, 300*, 286–90.

Condit, Celeste. (2007). How Culture and Science Make Race "Genetic": Motives and strategies for discrete categorization of the continuous and heterogeneous. *Literature and Medicine, 26*(1), 240–268.

Condit, Celeste M., Parrott, Roxanne L., Harris, Tina M., Lynch, John, & Dubriwny, Tasha. (2004). The role of "genetics" in popular understandings of race in the United States. *Public Understanding of Science, 13*(3), 249–272.

Contreras, Jorge L. (2011). Bermuda's legacy: Policy, patents, and the design of the genome commons. *Minnesota Journal of Law, Science & Technology*, *12*, 61–97.

DeVito, Danny. (1997). *Gattaca*. [Film]. Andrew Niccol, Director. Columbia Pictures.

DiMaggio, Paul, & Hargittai, Eszter. (2001). From the 'digital divide' to 'digital inequality': Studying Internet use as penetration increases. *Princeton University Center for Arts and Cultural Policy Studies, Working Paper Series, Number 15*.

DNA Dilemma. (2012). The DNA Dilemma: Why science wants your genome. *The Globe and Mail.* URL: http://www.theglobeandmail.com/news/national/time-to-lead/the-dna-dilemma-why-science-wants-your-genome/article5921666/ [December 10, 2012].

Feenberg, Andrew. (1992). From Information to Communication: The French experience with video-text. In M. Lea (Ed.), *Contexts of Computer-Mediated Communication* (pp.168–187). London: Harvester Wheatsheaf.

Gavison, Ruth. (1980). Privacy and the limits of law. *The Yale Law Journal, 89*(3), 421–471.

Gerlach, Neil, & Hamilton, Sheryl N. (2005). From Mad Scientist to Bad Scientist: Richard Seed as biogovernmental event. *Communication Theory, 15*, 78–99.

Gibson, Elaine, Brazil, Kevin, Coughlin, Michael D., Emerson, Claudio, Fournier, Francois, Schwartz, Lisa & Willison, Donald J. (2008). Who's minding the shop? The role of Canadian research ethics boards in the creation and uses of registries and biobanks. *BMC medical ethics*, *9*(1), 17.

Gordon, Dan, & Pai, Aditya. (2012). Big Data, Big Opportunity. *Canadian Healthcare Manager*, 1(2), 25–27.

Gulland, Anne. (2010). *Project to decode genomes in cancer samples promises new treatments.* URL: http://www.bmj.com/content/340/bmj.c2149 [May 24, 2013].

Hayden, Erika C. (2013). Privacy loophole found in genetic databases. *Nature.* URL: http://www.nature.com.proxy.lib.sfu.ca/news/privacy-loophole-found-in-genetic-databases-1.12237 [February 13, 2013].

Hudson, Kathy L. (2011). Genomics, Health Care, and Society. *New England Journal of Medicine, 365*(11), 1033–1041.

IBM. (2012). What is Big Data? *IBM*. URL: http://www-01.ibm.com/software/data/bigdata/ [August 10, 2013].

Latour, Bruno. (1996). *Aramis, or the Love of Technology*. Cambridge, MA: Harvard University Press.

Latour, Bruno. (2007). Beware, your imagination leaves digital traces. *Times Higher Literary Supplement.* URL: http://www.bruno-latour.fr/node/245 [2013-10-07].

Lin, Zhen, Owen, Art B., & Altman, Russ B. (2004). Genomic research and human subject privacy. *Science, 305*(*183*), 183.

Lohr, Steve. (2012, 29 March). New U.S. research will aim at flood of digital data. *New York Times, 29*.

Lunshof, Jeantine E., Chadwick, Ruth, Vorhaus, Daniel B. & Church, George M. (2008). From genetic privacy to open consent. *Nature Reviews Genetics*, *9*(5), 406–411.

Lyon, Daniel. (2003). *Surveillance as social sorting: Privacy, risk and digital discrimination.* London, U.K.: Routledge.

Manovich, Lev. (2001). *The Language of New Media.* Cambridge, MA: MIT Press.

Manovich, Lev. (2012). Trending: The promises and the challenges of big social data. In M. Gold (Ed.), *Debates in the Digital Humanities* (pp. 460–473). Minneapolis, MN: University of Minnesota Press.

Marris, Emma. (2005). Free genome databases finally defeat Celera. *Nature*, *435*(7038), 6.

Mayer-Schönberger, Viktor, & Cukier, Kenneth. (2013). *Big data: A revolution that will transform how we live, work and think*. London, U.K.: John Murray.

McLuhan, Marshall. (1964). *Understanding Media: The extensions of man*. New York, NY: McGraw Hill.

Palsson, Gisli, & Rabinow, Paul. (2001). The Icelandic Genome Debate. *Trends in Biotechnology, 19*(5), 166–171.

Personal Genomes. (2012). URL: http://www.personalgenomes.ca/consent/PGPCanada_Full_Consent_Form_Jan2012.pdf [July 13, 2013].

Pinch, Trevor J., & Bijker, Wiebe E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science, 14,* 399–441.

Pollack, Andrew. (2011, November 30). DNA sequencing caught in deluge of data. *New York Times*, 1.

Rainie, Lee. & Wellman, Barry. (2012). *Networked: The new social operating system*. Cambridge, MA: MIT Press.

Schadt, Eric E., Woo, Sangsoon, & Hao, Ke. (2012). Bayesian method to predict individual SNP genotypes from gene expression data. *Nature Genetics, 44*(5), 603–608.

Solove, Daniel. (2008). *Understanding Privacy*. Cambridge, MA: Harvard University Press.

Strasser, Bruno. (2012). Data-driven sciences: From wonder cabinets to electronic databases. *Studies in History and Philosophy of Biological and Biomedical Sciences, 43,* 85–87.

Thacker, Eugene. (2004). *Biomedia*. Minneapolis, MN: University of Minnesota Press.

Thacker, Eugene. (2005). *The Global Genome: Biotechnology, politics, and culture*. Cambridge, MA: The MIT Press.

United States of America. (2008). *Genetic Information Nondiscrimination Act (GINA)*. Washington, DC: U.S. Equal Employment Opportunity Commission.

Williams, Raymond. (1975). *Television: Technology and cultural form*. New York, NY: Schocken Books.

Williams, James, & Weber-Jahnke, Jens. (2010). Regulation of Patient Management Software. *Health Law Journal*, *18*, 73.

Winner, Langdon. (1980). Do Artifacts Have Politics? *Daedalus, 109*(1), 121–136.