Privacy Protection System and Big Data

Major Research Paper

Presented by

Runqi Tu

Under the supervision of

Professor David C.G. Brown

April 2017

Table of Content

Abstract

With the development of data calculating and storage technology, big data become a buzzword and application of big data analytics have spread all over the world. Although the prosperous of big data bring a lot of benefit to us, many people worry about the privacy issue generated by it, like data breaches, identity fraud and the security of data infrastructure. Many people also think the current privacy protection system should be adjusted to adapt to the big data. This research paper aims to research whether and how we should change our privacy protection system with the coming of big data according to the experience of Canada, China mainland and Hong Kong. This paper argues that we should change our privacy protection system in a way that could protect people's privacy as well as would not hinder reasonable data analytics. The result shows that the current privacy protection system could protect privacy generally, but some changes also should be adapted to reduce risks. On the other hand, due to the new features of big data analytics, the system would impede some certain types of data analytics. This study concludes the feature of this type of data analytics, and call for more research on this topic.

Key words: Big Data, Data Analytics, Privacy, Privacy Protection

1 Introduction

1.1 Research purpose

Nowadays, the applications of big data in the fields of commerce, medicine and government bring us a great deal of benefit, it makes our society better and more comfortable. However, some applications of big data are designed to work based on personal data. As a result, many companies, institutions and agencies collect personal information in our daily life. For example, in some cities of China, governments install cameras along the road or in some business areas. On one hand, this collected information can provide solutions to solve traffic problems or insecurity events using data analytic. (Xu et al. 2014) However, this means that people's activities are recorded by government, and some of the information is personal and very sensitive. Besides the government, a huge amount of data is also collected by all kinds of organizations, and privacy incidents such as identity fraud and data breaches sometimes happen. The development of big data and increase in privacy incidents at the same time make more and more people start to be concerned about the privacy issue relating to big data and to worry that the current privacy protection system cannot protect their privacy.

Sometimes, the current privacy protection system will also prevent some kinds of data analytics. Some privacy laws protect people's privacy from abusive commercial use, but also prevent scientific research. For example, using data in unanticipated ways has been a hallmark of the big-data revolution. Current the regulations of in OECD countries require that all data user must inform data owners what they will collect and what is their

purpose in making this collection. (Government of HK SAR. 2013, Government of Canada. 1985) But for the big data era, data users cannot guarantee whether the data will be used for any other purpose without harming the owner's privacy in the future. If a data collector or user must inform and get approval from the data owner about their purpose every time, it will reduce the efficiency and hinder the development of big data.

Because of such concerns, some people suspect that the current privacy protection system may not protect people's privacy well, they think the system is so loose that some parts of the regulation are not enforced. Other people, who argue the current system hinders reasonable data analytics, think our system is too strict. So, this research paper will try to analyze whether it is necessary to change the current privacy protection system, by researching whether it can protect privacy or, on the other hand, hinder reasonable data analytics. Based on the opinions above, although there is a tension between "loose" and "strict", this paper would argue that we need to change our privacy protection system, both because the system cannot protect privacy very well and also because it would hinder some kinds of data analytics. I will find evidence to support my argument in the following sections of the paper.

1.2 Significance of the topic

Information Technologies have been applied and played important roles everywhere of the world, and no one can live without the technology. Data collecting is also related to everyone's daily life. When organizations own tons of data, the privacy problems will become the core concern of the whole public.

Besides, the development of data analytic technology benefits society a lot. With the development of big data technology, data analytics is being adopted in both the public and private sectors, providing support to decision-making and other activities. But sometimes, due to the regulation of privacy protection, some analysis cannot be implemented well, and the current privacy laws limit the ability of both the public and private sectors to expand and utilize their datasets for analytic purposes. (Rubinstein 2013) This will hinder the development of data analytics and the use of big data.

So, it is very important to make sure the privacy protection system can protect people's privacy and will not block reasonable data analytics.

1.3 Literature review

First, until now, scholars still do not have a very clear definition for Big Data. (Lenard and Rubin 2014, Perera et al. 2015) But whatever the definition is, many scholars agree Big Data will bring a lot of benefit to the world. Big Data technologies have been applied a lot in the private sectors. Besides, although governments do not pursue profits, they can use Big Data to promote public goods, including providing public services and assisting decision-making. (Kim et al. 2014)

Privacy concern about personal data collection is a continuing topic since the emergence of the Internet. In 1999, Culnan and Armstrong addressed the tensions between the collection and use of personal information and the privacy concerns of the data owner. As the development of Big Data occurred, people became aware it may cause some problems about privacy and our current privacy protection system might not work

well. For example, it could increase the risks associated with identity fraud and data breaches. Big data will lead to some discriminatory decisions in markets and it is harmful to consumers. For example, if the result of data analytic shows that a certain gene will lead to a severe disease, the insurance company may refuse to accept the application of people who have this gene. And it may reduce the amount of information that people can see. (Mehmood et al. 2016, Lenard and Rubin 2014)

For the privacy problems of big data, every part of the society expresses their concerns, both the public and the government; the White House released two reports about this concern in 2014. (Gaff et al. 2014) The Information and Privacy Commissioner of Ontario indicate they are facing several challenges brought by big data and are researching how to make change to the privacy protection system. (IPC 2017) But both White House reports and some journals also points out that some aspects of our current privacy protection system may block the development of big data, (Executive Office of the President 2014; Lenard and Rubin 2014), for example that the regulation requiring that data owners be informed about the specific purpose of the data collection and use reduces efficiency. Using data in unanticipated ways has been a hallmark of the big-data revolution, however the current privacy protection system limiting the reuse or sharing of data would be particularly harmful if applied to big data because they are inconsistent with the innovative ways in which data are being used. (Lenard and Rubin 2014) They claim that the current privacy system should be changed with the coming of big data. Rubinstein (2013), on the other hand, thinks the wave of big data will overwhelm the

core privacy principles of informed choice and data minimization of our aging privacy protection system. Because nowadays it is the organizations that control personal data, then, he argues a new business model shifting control over both the collection and use of data from firms to individuals and regulators should encourage businesses to adopt new models premised on consumer empowerment.

Most of the related literature expresses the opinion that the development of big data will harm privacy. However, most of the literature about big data privacy and development is based on speculative and hypothetical analysis. This research paper will analyse the real system of selected countries and territories and examine the research question in that context.

1.4 Methodology

To research the question, first, this article will summarize the key applications and privacy problems of big data. Then, I will choose 3 countries and territories, Canada, China mainland and Hong Kong Special Administrative Region (HKSAR) and research their privacy protection system. In the following part, this paper will try to analysis whether the privacy protection system in those countries and territories could prevent the privacy problems and would not hinder the applications summarized in the first part. Some data from the privacy protection agencies of Hong Kong and Canada is also adopted to support my argument. There is nearly no data from China mainland because there is less information available about China mainland.

I choose Canada, China mainland and Hong Kong as the research object because

they have different characteristics. Canada and Hong Kong are ex-colonies of the United Kingdom, so their political and administrative systems are a version of the Westminster System. At the same time, Canada is a well-developed country and Hong Kong is a place of Chinese culture, and the different culture may affect the approach to privacy protection. After Hong Kong was handed over to Chinese government in 1997, admittedly, the Chinese government have had some level of control over the Hong Kong government for almost 20 years. So, in a word, Hong Kong is a combination of both Western and Eastern political ideas and culture, and could be an interesting case. Besides, China now is the second-largest economy and owns Hong Kong, therefore it is very interesting to see how China could protect their people's privacy.

This research will be based on government documents of each of the countries and journal articles written by writers from the different countries. If I cannot find the English version of an article, I translate the title or the part of article that is used in this paper and attach them at the end of the paper.

1.5 Structure

This first section of the paper will make an introduction about big data, introducing the idea, the reason and the application of big data, both in the private and public sectors. Then, the following section will consider privacy problems in the application of big data, starting by an analysis of the definition of privacy and general description of the privacy protection system.

In the next section, we will analysis the current privacy protection systems of

Canada, China mainland and Hong Kong SAR. In the final two sections, this paper will examine the research question by analysing the current system of these countries and territories.

2 Data and Big Data

2.1 Data and data life cycle

According to Oxford dictionary and Merriam-Webster dictionary, data is “facts and statistics collected together for reference or analysis” as well as “factual information used as a basis for reasoning, discussion, or calculation”. In the digital era, as these two dictionary, data is the information output by a sensing device or organ, including quantities, characters, or symbols, which may be stored and transmitted in the form of electrical signals and recorded. Data includes “both useful and irrelevant or redundant information and must be processed to be meaningful”.

The lifecycle of data has about 6-7 phases (the number varies with the different versions of the data life cycle). According to the version of the UK Data Archive (2017), data life cycle is divided into creating, processing, analyzing, preserving, giving access to data and reusing data. We always design, plan and execute data collection in the data-creating phase. After collection, processing data means entering or digitizing, checking and validating data, sometimes anonymizing data in this phase, and in the end storing them. In the analyzing phase, data will be interpreted and derived, and we will produce research outputs and prepare data for preservation. Preserving data involves

migrating data to the best format and medium, then providing back-up, storing and archiving data. In the giving access phase, data will be distributed and shared and access control and copyright will be established. The last phase, re-using data, sometimes happens when research undertakes reviews or scrutinizes findings. Also, re-using data will happen in the follow-up and new research phase, which means a new life cycle is opening. The data life cycle will be a great help in resolving issues about data.

2.2 Big Data and Open Data

Although we can define data clearly, until now we do not have a very specific definition of big data. (Lenard and Rubin 2014, Perera et al. 2015) McKinsey defines big data as referring to "datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze." (McKinsey & Company. 2011, 1) In the book "*Big Data: A Revolution That Will Transform How We Live, Work, and Think*", Mayer-Schonberger and Cukier (2013, 6) define big data as "things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more."

Although there are different definitions, one thing is clear that big data always means a huge amount of data, and with the new technology which can analyse the amount of data, it can provide us greater insight and more information.

In the big data era, there is another term about data, open data. According to the definition of Government of Canada (2014), Open data is structured data that is

machine-readable, freely shared, used and built on without restrictions. Open data is not only about government, but also every organization would freely share their data with the public. It can support innovation and research for the whole society, and for government it can advance accountability and democratic reform. Open data about the whole society becomes an important source of data analytics and advances the development of big data.

2.3 The success of big data

In general, advanced technology and the need for mass quantity data processing makes big data analytics become possible. The quickly-updated calculation, storage and communication technologies provide the hardware basis for big data. On the other hand, with the development of the Internet and the popularity of smart phones and social network, a huge amount of data is created every year, and the evolution from "data" to "big data" has happened very quickly. For example, in 1946, the earliest electronic general-purpose computer in the world, "ENIAC" had just less than 10 KB memory. In contrast, currently, every day we create 2.5 quintillion bytes of data and so much so that 90% of the data in the world today has been created in the last two years alone. (IBM 2017) The global number of Internet users was more than 3 billion and Smart phone users were more than 2.5 billion in 2015. The huge number of users and their data provide a plentiful resource to big data analytics, leading to a strong need for companies or governments to process these data, to analysis them and to extract useful information.

2.4 Features of big data

Mayer-Schonberger and Cukier (2013), describe three features of big data in their

book. The first feature of big data is we can process the entire dataset, not just a part of it. One of the most important parts of contemporary statistics is sampling. Because statistical analysis does not need to deal with the whole dataset, the quality of sampling will determine the quality of the statistics. But in the big data era, with the advanced technology, it is possible to process all the dataset without sampling, so, at this time, the quality of analytics is determined by the amount of data collected by the researcher. (Mayer-Schonberger and Cukier 2013, 19)

Second, we care less about exactitude. (Mayer-Schbnberger and Cukier 2013) Big data is messy, with different varieties, such as even emails, content of social networks, video websites, and it is distributed among countless servers around the world. Because of the quantity and complexity, we cannot make sure that every item of data is accurate. For big data analytics, the increase of quantity will reduce the error, enable us to get more accurate results. However, "we don't give up on exactitude entirely; we only give up our devotion to it". (Mayer-Schbnberger and Cukier 2013, 32) Caring less about exactitude does not mean we do not want accurate data, we just do not spend more time thinking about how to make data collection more accurate. But for some data collection that needs to be accurate, exactitude is still a requirement, for example, sophisticated scientific research or medical inspection.

Third, in big data era, we focus more on patterns and correlations in the data, rather than the age-old search for causality. (Mayer-Schonberger and Cukier 2013, 50) We can discover patterns and correlations in the data that offer us novel and invaluable insights.

As “prediction” is the core of big data, some big data applications always predict the future by providing us with data correlations. The correlations may not tell us precisely why something is happening but they alert us that it is happening. The correlation-based predictions based on historical data are being adopted widely in the commercial field and in decision-making models.

2.5 Data analytics and reasonable data analytics

Nowadays, big data has been applied to many field in the form of data analytics, and it will play more important roles in the future. Before I start, I will make a definition of reasonable data analytics, which is helpful for this research.

Big data analytic can be divided into analytics based on non-personal data and on personal data. Most of applications of big data are lawful, and can fit the current privacy protection laws, as most lawful data analytics are based on non-personal data, and the rest are generally consistent with the current privacy laws.

However, there are some applications of big data that violate privacy laws while the benefits they provided are so huge to the whole society that we do not have a good reason to stop their use, although we may change our privacy laws and find a new way to protect privacy. All of these kinds of data analytics are based on personal data and the specific examples will be given in the following part.

So, in this paper, when I talk about reasonable data analytics, I mean the lawful applications and those applications violating the privacy laws but whose advantages far outweigh their privacy concerns.

3 The application of big data

Big data technology has been widely adopted by the private sector for several years and recently also by public sectors. As some writers point out, the core of big data applications is prediction. (Mayer-Schonberger and Cukier 2013, 11) The roles big data plays in these applications is usually to formulate the model and predict, based on the relevant historical data, then to provide support to decision-making and other issues.

3.1 Big data in commerce

The commercial field here is a complex field and consist of lots of sub-fields such as retail, finance or the Internet. Big data has been applied in these field a lot, and for nearly each application, prediction is at the core.

For the retail industry, big data can be used to predict the need for commodities, according to different factors, such as weather, location or fashion trend. It always helps managers decide their business strategy.

The Internet industry is an emerging industry, and several innovations have been made based on big data. Predicting price is one application, for example, Farecast; with this website, users can know when the flight ticket price is lowest for each route, so they can get to their destination with cheapest ticket. This prediction is made by Farecast based on the historical data of several years and billions of flight-price records. There are also other applications of big data that can predict other price, such as agricultural products, seafood, using the same mechanism.

In the Internet industry, some companies such as Google, Facebook or Baidu in

China, hold huge amounts of data, and they can do more work with these data. For example, Google can predict the trend of flu based on user search history and Facebook can predict one's location or habits with the browsing and click history. (Mayer-Schbnberger and Cukier 2013,1) The Internet industry was and will be the main stage of big data.

Big data analysis has also been adopted in the financial industry, such as credit card companies protecting the security of their credit card. A credit card company can set up models to find a fraud transaction, applying this model to analyse each transaction and update this model frequently. Based on the huge number of credit card transactions, this model can be very useful. (Lenard and Rubin 2014) On the other hand, big data-based credit score evaluation is also become more and more popular. ZestFinance and other company use more variables about borrowers than the traditional approach and compare the borrower's situation with the company's huge database about credit records to decide whether they should lend money to the borrower. (ZestFinance 2017) In China, after the credit evaluation market opened for private companies, many shopping websites and online payment companies started to build their own credit evaluation platform with their own records about their users, which have accumulated in the past 10-15 years and are about half a billion people.

On the other hand, some applications in legal grey areas of the commercial field have caused some dispute. One example is big data-based debt collection in China. Recently, because of the simple qualification review procedure of some emerging small

loan platforms, their charge-off rate is always high, the higher charge-off rate means more loan cannot be collected. The high rate leads to the emergence of Internet debt collection companies, using big data technology to collect debts; Cuitx (催天下) is one of these companies (DTCaijing 2017). Usually, this kind of company would contact a debtor by collecting and integrating personal information from all publically-available information from the Internet, such as from government documents, business registration or some legal documents. But sometimes, they will use leaked information, such as data hacked by a hacker (DTCaijing 2017). Even when they just bring public information together, the usage of leaked information and finding people with personal information is hard to justify.

3.2 Big data in medicine

In the early phase, most medical data was in paper form. But with the development of storage, computing and Internet technology, nowadays large amounts of medical data are collected from hospitals in digital form. The electronic health records are the main systems to record medical data and many countries are busy building their EHR system, like the United States, Canada, China mainland and Hong Kong. On the other hand, the emergence of wearable devices make it possible to record health information, such as blood pressure and heart rate. The amount of medical data is very large. In 2011, the total volume of health record information was about 150EB (1 EB=1024PB, 1 PB=1024 TB, 1 TB=1024 GB) (Quintero et al. 2016, 4), and in 2014 a health record company in California, Kaiser Permanente, held electronic health record information of 9.6 million

members. (Kaiser Permanente 2014) So, the boom in health records require the application of big data technology.

Prediction is also a core application of big data in medicine. The increasingly accurate data about disease will enable researchers to monitor disease more precisely, making treatment more useful and efficient. Based on more historical data, the application of big data could predict and estimate the trends of some diseases or patients, for example, predicting the type or the trend of flu spread, or estimating which patients are vulnerable to new diseases or complications of surgery. (V. Raghupathi and W. Raghupathi 2014)

The huge amount of data will also promote the development of research. More data will enable researchers to do more analysis, leading to the development of evidence-based medicine and genome analysis. In addition, it will also accelerate the development of personal health care, based on personal health records or even gene information, and the application of big data could make health monitoring more precise and provide more suggestions about how to reduce the risk of disease and cure disease quickly. (V. Raghupathi and W. Raghupathi 2014)

The application of the results of genetic testing can be also seem to be in the legal grey zone. After the Human Genome Project was finished in 2000, more and more human genes can be understood by researcher. (National Human Genome Research Institute 2003) With several years' development, especially with the help of computerized database comparison, gene testing has become a mature technology and popular in the

market. In 2008, the Retail DNA Test was chosen by Time magazine as the No.1 of the Best Inventions of that year. (Hamilton 2008) The results could let people know about congenital physical conditions, risk of certain diseases, or even know about personality. Then doctors can make precautionary measurements in advance for certain diseases.

3.3 Big data in government

The biggest difference between big data in government and in the private sector is government always uses data to pursue the public interest and the private sector always to pursue their own interest, to make a profit. That also leads to people feeling more confident to provide information to government than to companies, although sometimes people also will suspect the government will leak or abuse their privacy. On the other hand, the data collected by a company is based on the voluntary action of the user, however when government collects information, it always has the right to force the collection of information, such as video surveillance or the information collected when people fill out government documents. That's why government agencies have a very large amount of data and also why it will benefit greatly from big data technology. Governments can either outsource data analytics to the private sector or do data analytics by themselves, and they can do analytics based on either their own data or on the data provided by private sector.

What can big data do for government? The book *Smart Government* (智慧政府) (Xu et al. 2014) concluded there are five general aspects. First, big data increases transparency and promotes information sharing. Second, the application of big data will

encourage society to innovate. These two points mean big data can increase the efficiency of the whole society and can seem to be Open Data on the part of government. As governments hold an enormous amount of data, making non-confidential data open to the public can increase data re-use and encourage more data analysis. Third, big data can increase the efficiency of the public sector by evaluating its performance. Fourth, big data can make demographic segmentation possible, so that government can enact specific policy for specific groups and provide personalized public service. Compared with the traditional approach of having the same policy for an entire area, the new way can meet special needs, increasing the efficiency of the policy. Fifth, big-data-assisted decisions can reduce the occurrence of errors. Based on the core of prediction and strong calculating power, big data technology can provide useful suggestions when making decisions, even predicting the results of each policy.

For the concrete applications of big data in public sector, many scholars have started to try, and even some private companies are trying, to do some of the work of government. Some people use large-scale, real-time data to track and forecast economic activity, providing a new measure beside the official statistics. (Lenard and Rubin 2014) Other scholars can produce the real-time Consumer Price Indexes with transaction data from online retail websites and some scholars can provide indexes about unemployment and consumer confidence. (Einav and Levin 2014) All of these applications are based on the big data.

Not only can the application of big data be used against credit card fraud, but it can

also be used to protect national security, such as protecting the country from terrorist attack. The mechanism for doing this is the same as the mechanism against credit card fraud: comparing each event, communicating with the models that conclude the features of terrorist activity and taking action. (Mayer-Schbnberger and Cukier 2013,27, Lenard and Rubin 2014) The same mechanism can also be used in tax fraud and postage fraud.

Big data can help government make policy or take decisions. For example, big data can be a tool to provide proposals for urban planning based on data of traffic flows, population migration or even social media. Some real-time transportation information and location-based social network data can be also used to solve traffic problems, or public security problems, with the prediction of future demands and risks.

Beside governmental activities, big data technology has also been introduced into political activities. During the 2016 United States election campaign, Donald Trump hired Cambridge Analytica to market his election campaign. This company used a specific and sophisticated model to describe the personality of individual Americans with the data indicating the preferences on Facebook postings. Based on behavioral science, big data analysis, and ad targeting, they made personalized advertising, aligned as accurately as possible to the personality of an individual consumer. This company also played an important role during the Brexit Campaign of the United Kingdom. (VICE 2016) As the CEO of Cambridge Analytica stated, "Pretty much every message that Trump put out was data-driven," and the power of very accurate personalized advertising and big data analysis helped Trump succeed. The writer argues Cambridge Analytica may

have helped Trump keep Clinton voters away from the ballot box, and in any case they legally bought lots of American personal information. (Grassegger and Krogerus 2017)

From this passage, we can believe the writer and the researcher in the article are against the use of a psychological-profiling approach in election campaigns. It is hard to say the use of psychology is unfair, but it may interfere in people's political choice. The use of big data in political activities will be a topic in the future.

4 Privacy

4.1 The idea of privacy

Privacy is an old idea, but most of the legislation related to privacy has been enacted in the past 30 - 40 years. There are many versions of the definition of privacy. Some people think privacy is closely related to "confidentiality", just like preventing information from being disclosed. Some conceive of privacy as having four 'dimensions': privacy of the body; privacy of personal behavior; privacy of personal communications; and privacy of personal data. And a very old version is in the book Privacy and Freedom, written by Westin (1967). In this book, Westin proposes four categories of privacy: solitude, intimacy, anonymity and reserve.

The current privacy legislation in most countries is relatively newer, compared with other types of legislation. The privacy legislations in nearly all OECD country, except the United States, is derived from *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, which was developed in the late 1970s and adopted

in 1980. The eight privacy principles in the guideline include: Collection Limitation Principle, Data Quality Principle, Purpose Specification Principle, Use Limitation Principle, Security Safeguards Principle, Openness Principle, Individual Participation Principle, and Accountability Principle. These have become the core of privacy law in nearly all OECD countries, and even in some non-OECD countries or territories, such as Hong Kong. (OECD 1980)

As the guideline mentions, with the development of information technology and the spread of computerized processing of personal data, the public interest has tended to focus on the related risks and implications. (OECD 1980) We can conclude that it is the development of technology that makes people worry more about privacy. For big data, it is not hard to imagine that the development of big data will raise concerns about privacy again.

Besides the definition and legislation, we can believe that privacy is an idea generated by western culture, and most people in western countries have a strong awareness of privacy. (Stanaland and Lwin 2013, Lim et al. 2009) The timing of privacy legislation enactment could also reflect this point. As Japan is an old member of OECD, its privacy legislation was enacted in 1988. (See the Act on Protection of Personal Information of Japan) However, although South Korea joined OECD in 1996, its privacy legislation was actual finished in 2011, 15 years after it became a member, and strengthen in 2016. (See Personal Information Protection Act of South Korean) For Singapore, its first privacy legislation was passed in 2012, and became effective in 2014. (See the

Personal Data Protection Act of Singapore)

4.2 Privacy and the data life cycle

Only personal data will raise privacy issues. When it comes to the relationship between privacy and the data life cycle, different phases of the data life cycle will have different privacy risks. The data-creating phase is related to the personal data collection issue: personal data collection must be consistent with privacy laws, and data owners must be informed about the collection's purpose and provide their consent. For processing data, the anonymized procedure may be not enough, data storage may be insecure and processing data sometimes will ignore data integrity. For the analyzing, preserving and giving access phases, because of the presentation of data, personal data may be disclosed without the consent of the data owner, and access control may prevent data owners from accessing their data. If new research is conducted, the re-use of personal data usually is unauthorized if the users do not get the consent of the data owner again, because of the changing purpose.

5 Analysis of privacy problems of big data

As described above, big data is very useful for governments and they also collect a lot of data. Government can collect data, because they have a right to information about citizens. But, citizens also have privacy rights. Uncontrolled information collection or information controlled in a wrong way will lead to privacy problems.

This section will try to identify the possible privacy problems in the big data era.

The first part will consider the general problems with big data privacy, faced by both the private sector and government. The second part will identify specific privacy problems that government needs to solve.

5.1 The general privacy risks or problems of big data

(1) The risk of identity fraud and data breaches increased

This risk usually occurs in the processing phase of the data life cycle.

There is a concern that the development of big data will result in more identity fraud and data breaches. Because in the big data era more and more organizations, like companies and governments, want to collect data, the amount of collected data is increased. In general, for the same risk level, the more data collected, the more data will be exposed to risks. On the other hand, more data means that more work based on data can be conducted, which means that more people will have the right to know the data and more research reports will be published. If these activities cannot protect the privacy of data owner, the risks of identity fraud and data breaches will increase.

(2) Increasing the possibility of spam and advertisement

With big data technology, companies and retailers can load people's personal information, such as age, location, religion, health, financial status and shopping habits into the algorithms. Then they can divide consumers into several groups, providing more personalize advertisements or other marketing approaches for specific groups. But people are afraid that the more accurate marketing will lead to greater popularity of the use of the approach. Then more people will be disturbed and

possibly even without their prior consent.

(3) Individuals will be forced to reveal personal information

This risk usually happens in the data creating phase of life cycle.

There two types of data collection. The first one is active data generation, which means that the data owner is willing to provide the data to a third party, while passive data generation refers to the situations where the data are generated by the data owner's online activity (e.g., browsing) and the data owner may not even be aware of that the data is being collected by a third party. (Mehmood et al. 2016)

For active data generation, unless data owners input their information, no one can have the right to know them. But for passive data collection, due to the fact that collection is done silently, without reminder or alarm, the data user decides what information and when it should be collected. Although there are always some clauses in the terms of use, nearly no one reads it carefully, and if someone denies the terms, they cannot use the service. So, protecting data owners from passive data collecting is another challenge for protecting privacy. In addition, government surveillance can be seen as a form of passive data collection, although surveillance is different from passive collection from websites as it is always justified on security grounds.

(4) The security of big data infrastructure

The data security risk exists during the entire life cycle.

With billions of data items, how to ensure the security of data is always a challenge. More data means more storage devices and more servers, and the system

will be more vulnerable. Because if one bug is found in one system, it will also exist in all other similar systems. If the big data storage system is compromised, it can be very harmful, as individuals' personal information can be disclosed. In modern information systems, data centers play an important role of performing complex computations and retrieving large amounts of data. In a distributed environment, an application may need several datasets from different data centers and therefore faces the challenge of privacy protection. (Mehmood et al. 2016) In the same way as for storage systems, the security of process and transmission devices is also important. Big data analysis always requires the support of cloud computing, so the security of data centers and the related transmission devices should be ensured to protect privacy.

(5) The problem of data integrity

Data integrity is also a problem throughout the entire data life cycle. When collecting data, the data user should make sure data is inputted accurately. Preserving data, analyzing data and even re-using data should all require storage and use of complete data, with no change able to be made to the data. Data integrity also includes using up-to-data data. For personal data, these requirements are even more important. Data analytics with incomplete data or incorrect data will result in the wrong result, especially when the data analysis affects the data owner, as in the calculation of an insurance premium or credit score.

(6) Privacy problems brought about by open data

Privacy problem related to open data only relate to open personal data, as most

open data is anonymous data. Due to the different needs of the anonymized level, part of sensitive information may be deleted. But this kind of data anonymization sometimes is also harmful to the data owner. For example, the debt collection company referred to earlier just using data opened to the public to contact the debtor. There should be a general regulation of open data about how to process sensitive personal information, in order to maximize protection data owner's privacy. In 2016, the Office of Privacy Commissioner of Canada conducted research about Open Data and Privacy, especially talking about the problem that individuals can be "re-identified" from imperfectly anonymized data. (Fewer 2016)

5.2 Specific privacy problems of big data

This part considers specific privacy problem for the commercial, medical and government fields. For all these three fields, privacy risks or problems exist in the entire data life cycle. We can conclude here that privacy risks can be found in the entire data life cycle.

(1) Commerce

Other than the general concern about big data, the main concern for the commercial field is about that unintended information release will cause some privacy problems. In daily life, some organizations will always record information about individuals, such as credit card transaction and bill payment records. Or with the rise of social networks, people usually share their words, photos and even locations on the Internet. Companies or other organization always want to use this

information to conduct research, to improve marketing or to help them make better business decisions. Admittedly, some of this information seems anonymous, but if a huge amount of data is gathered about a person, the data user could identify the owner of the personal information. For example, de Montjoye et al. (2013) from MIT found that if they know four locations of one mobile phone, they could identify the user of the mobile phone. Adam Sadilek and John Krumm found that, based on a huge amount of "anonymous" data, they could predict where one individual will go in the next 80 weeks, with an accuracy rating as high as 80%. (Sadilek and Krumm 2012) Of course, the application of big data in commercial activity brings us a lot of benefits, but we also should recognize the privacy issues coming with the broadly application of big data.

(2) Medicine

As an important privacy concern about big data applications in medicine, patient privacy is an old topic. According to Ding (2014), patient health information privacy has two dimensions: one is confidentiality of health information and the other is sovereignty of health information. Protecting data confidentiality and respecting data ownership will be the core concerns in protecting privacy during medical data analysis.

As the main resource of data analytics in the field of medicine, health information of many individuals will be used in research. The first concern is that as the amount of personal health information is increased, more researchers have the

right to access the data and more research is conducted, the large scale will make regulation difficult. Second, anonymous data in research reports are also dangerous. If a research report relates to personal cases, the researcher will hide some personal information. But if personal data is revealed several times in several reports, and different parts of the data are hidden in each report, it is possible to "guess" who is the owner of this data. In this case, we cannot protect the person's privacy.

Protecting privacy in medical uses is also a process of weighing and balancing. Sometimes there is some conflict between protecting privacy and protecting the integrity of data. To protect their privacy, some patients do not want to add some medical test results, such as HIV test results, into their health record. But lack of results will harm the integrity of data and make some research inaccurate, and even cause some problems for the future treatment. In general, to ensure patient's health and the accuracy of future treatment, doctors need to make sure the data integrity of health record. However, the health record is patient's data and to some extent, they have the right to modify. So, it will be very complex to protect privacy for the medical field of data analytics.

The above risks are related to all phases of the data life cycle except the giving access phase. Access control over personal medical data is critical to protect patient privacy. The access control should not only decide who have the right to know, but also should decide which organizations have the right to collect, even including the government. Recently more and more people have taken genetic tests, as a result of

which they have found they have some gene that will lead to some disease such as cancer or Huntingtor's Disease, and once the insurance company or employer know the result, they have increased the premium or laid off the employee. (Bombard et al. 2016) Just few countries have legislation about protecting genetic information from use by insurance companies and employers. In March 2017, the Act to Prevent Genetic Discrimination, S-201, passed in Canada and became the law, but there are still many voices that argue against the law. (Davis 2017)

(3) Government activities

For government, most privacy issues arise in the data creating, processing and analyzing phase of the data life cycle.

Surveillance tools, such as surveillance cameras or ID cards, always are considered the biggest threats by citizens when government collects data. The use of cameras is becoming more and more pervasive. The Public Security Ministry of P. R. China launched the Skynet surveillance project(天网工程) around the year 2005. (Langfitt 2013) This project cost a very large amount of money to install surveillance cameras around traffic routes and nearly all public areas, such as metro stations, shopping centers and entrances of communities. Government officials said this project will help government to relieve traffic problems and help police solve criminal cases. However, many people are afraid that because the cameras are installed everywhere, the cameras will invade their privacy. The video and photo data collected by government is highly sensitive, because with these data, they can trace

the movement of any car and anyone.

The video and photo data collected by cameras is hard to analysis directly, but the ID card is a tool that can collect analyzable data. Many European and Asian countries already issue and use ID cards. (Lim et al. 2009) In China, if a resident wants to do a banking transaction, check-in to a hotel or take the coach, train or airplane, they are required to show their ID card and the service provider will save their information. In the end, this information will be handed in to the government. If not all the people can be identified in a video, with the ID card information, they can be identified accurately and quickly.

Besides surveillance cameras and ID cards, there are other ways government can interfere with privacy, such as Internet censorship. We can see the benefits of government surveillance in enhancing national security, especially in the post 9/11 environment. (Bailey and Caidi 2005) Several terrorists could be caught just because of the existence of government surveillance. However, we should also think about the privacy problems caused by surveillance.

When governments act as a data user, they are also subject to privacy regulations. Besides reasonable data analytics, sometimes government will force some organizations to provide data. A highly controversial case is the USA Patriot Act, which terminated at the beginning of June 2015. When this Act was in effect, if government thought some data was related to national security, they had the right to obtain the data, even without the consent of the data owner. Although this Act is no

longer in effect, we have to admit that government coercive powers sometimes will harm privacy.

For government, outsourcing data analysis to the private sector or doing data analytics by themselves are both good choices. Outsourcing data analytic can reduce government's burden to learn or buy big data technology. But in this case, organizations will have the right to access the data collected by government. Rather than having the projects within the government, governments only can control these projects weakly, but governments still need to ensure the security of personal data. On the other hand, when governments do analytics by themselves, sometimes they will need data provided by private sector, so how to ensure data integrity is also important.

Government data infrastructure is always a special topic. It isn't like the private sector with its flexible funds; building a new data center or updating aging data infrastructure always require having the budget approved by Parliament. Recently, Shared Services Canada indicated that if they do not get enough funding this year, the target of modernizing the government's technology networks by 2020 will be delayed. (Bagnall 2017) On the other hand, sometimes government will use the cloud service provided by a private company. Shared Services Canada plan to use services provided by Microsoft and Amazon. (Bagnall 2017) How to maintain ownership and security when data is preserved in a data center owned by the private sector is also a challenge.

6 The current data privacy protection system in Canada and China

6.1 Canada

(1) Overview

Because Canada is a federal country, privacy legislation in Canada has two levels, the federal and the provincial.

At the federal level, there are two privacy laws. The Privacy Act was passed by the Parliament in 1983 and covers the personal information-handling practices of federal government departments and agencies. The Act followed the enactment of the Canadian Charter of Rights and Freedoms, which ensures freedom of conscience, thought, expression and other fundamental freedoms and rights of Canadians and protects people from unreasonable search and seizure. In addition, the Privacy Act was passed 3 year after the implementation of the OECD privacy principles. The Federal Personal Information Protection and Electronic Documents Act (PIPEDA), passed in 2000, is the federal private-sector privacy law. (OPC 2014) We can believe that the Privacy Act and PIPEDA are inspired by *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.*

At the provincial level, each province and territory has its own public-sector legislation and the relevant provincial privacy act applies to provincial government agencies. For the private-sector, some provinces, such as Alberta, British Columbia and Québec, have privacy legislation that has been deemed “substantially similar” to PIPEDA, which means that it is applied instead of PIPEDA in some cases. PIPEDA is

also applied in health care field, but for health information, some provinces have legislation to protect health information, for example the Personal Health Information Protection Act in Ontario. In this case, if the health information protection law in one province is deemed "substantially similar" to PIPEDA, PIPEDA will not apply to that province's health sector. (OPC 2014)

(2) Features

The first feature in the Canadian privacy protection system is that the Privacy Act protects personal privacy from government activities. The Privacy Act permits governments to collect personal information if this information relates directly to a program or activity. When government agencies collect personal information, they must get approved from the data owners and must inform them about the purpose of the collection, and the personal information should only be used for that purpose, unless the individual consents. In addition, the government cannot disclose personal information, unless it has been given the consent of data owner. And every Canadian citizen or permanent resident has the right to assess their personal information under the control of a government agencies. The Privacy Act was also translated by Treasury Board into administrative policy that forces all government departments and agencies to follow the Act.

The second feature is that there are ten fair information principles in PIPEDA which clearly regulate how a company should collect information. (Government of Canada 2013) The fair information principles require it to be responsible for personal

information under its control. The collection, use, or disclosure of personal information must be known and consented to by the individual. The purpose of the collection must be identified at or before the time the information is collected, and only information related to that purpose can be collected. Data users only can retain data as long as necessary for the fulfillment of those purposes. Except with the consent of the data owner, their personal data cannot be used or disclosed for any other purposes. For organizations who collect personal data, they must ensure the security of the data, and make the data accurate, complete, and up-to-date. In the end, like under the Privacy Act, individuals have the right to know the existence, use, and disclosure of their personal information, and have the right to access them.

The third feature is that health information is specially protected in some province. For example, in Ontario The Personal Health Information Protection Act (PHIPA) was passed in 2004. (OPC 2014) Most of the content is similar to PIPEDA at the federal level, but this law provides guidelines for the use and disclosure of personal health information for research purposes. Research with health data must meet specified requirements and the research plan, whose format is also regulated specifically by the law, must be approved by a research ethics board, which consists of five members with different and specific background.

The next feature is that there are privacy commissioners at both the federal and provincial level to assist people to protect their privacy. The federal Office of the Privacy Commissioner of Canada (OPC) was established in 1983 following the

passage of the Privacy Act. In 2001, the duties of the OPC were extended to include private sector businesses subject to the Personal Information Protection and Electronic Documents Act (PIPEDA). The Privacy Commissioner of Canada is a non-partisan officer of the Parliament of Canada and reports directly to Parliament, which means he is independent from the government and an important part of the system of checks and balances to protect privacy rights. The Commissioner has the right to investigate complaints, conduct audits and pursue court action, and they also publicly report information about privacy practices. There are also privacy commissioners at the provincial level, who are responsible for protection of people's privacy under the provincial privacy act. (OPC 2016) Every year, the federal Privacy Commissioner needs to make a report to Parliament. And in Parliament, there is a parliamentary committee – the House of Commons Standing Committee on Access to Information, Privacy and Ethics – whose mandate includes to study matters related to reports of the Privacy Commissioner and to initiate discussion about those issues in Parliament. (Parliament of Canada 2017)

This last feature is that the way people can file a complaint about privacy issues is very clear. (OPC 2016)Individuals can do so when they have privacy concerns either about a government agency or a business. First they are encouraged to try to resolve the issue directly with the institution. For example, for federal institutions, individuals can contact the Access to Information and Privacy Coordinator in each agency, or a privacy officer in a private organization that is regulated by PIPEDA.

Then, if they are not satisfied with the response, they can file a formal complaint with the Privacy Commissioner at the federal or provincial level according to the organization they are complaining about. The commissioner will assist them to resolve their concerns.

6.2 China

6.2.1 China Mainland

(1) Overview

Some fragments in several Chinese laws relate to privacy rights and enable victims of some cases about privacy to fight for their rights. In article 33 of the Constitution (宪法) 1982, first, human rights are protected, which includes the right to privacy in most countries around the world. Then, articles 39 and 40 also protect citizens from illegal searching and entering of their residence and protect the freedom and privacy of correspondence of citizens. In some specific laws, like the General Principles of Civil Law (民法通则), part of privacy rights is protected under the categories of reputation right, portraiture right and name right. In the Criminal Law, illegally providing personal information of citizens, illegally searching the body, residence or invading the correspondence right of a citizen, if the circumstances are serious, will result in a sentence to fixed-term imprisonment and a fine.

For online privacy protection, the Standing Committee of the National People's Congress passed the Decision on Strengthening the Protection of Online Information (全国人大常委会关于加强网络信息保护的决定) on 28 December 2012. It was the

first time that the national legislature promulgated a law specially to protect online information that may identify or affect the privacy of the data subject, sending out a message that information privacy has become an important concern of the state. This decision requires all Internet service providers and other organizations to comply with some rules when they collect and use personal information. There are two main aspects. First, when these organization are collecting personal information, they must get the consent of users, and they must keep the security of this information with adequate technological measures. When they find a security incident about personal information, they must take remedial measures immediately. Second, when users notice their personal information is revealed online, they have the right to require the provider to delete all related information. (See Appendix A)

For medical information, although without a general law, some specific laws protect patient's personal information. The Regulations on Prevention and Treatment of HIV/AIDS (艾滋病防治条例) 2006 and the Mental Health Law (精神卫生法) 2012 state that the personal information of an HIV-infected or suffering individual and his family and of mentally ill patients are protected by the law. No organizations and individuals can disclose the name, address, employer, portrait, medical history and other personal information of patients unless the disclosure is required by law.

(2) Features

To conclude discussion of the features of the privacy protection system of Mainland China, there is no statutory scheme or comprehensive set of legal rules to

protect citizen's privacy in China. In fact, privacy rights are hardly recognized by the Chinese government and legislature. The Tort Law (侵权责任法) 2009 is the first time that the Chinese national legislature explicitly recognised the right of privacy. (Ding 2014) The law enforcement agency of the related law is just the court, and it is very hard for normal citizens to argue for their rights. Although there is the Online Information Protection Decision, it is just a general regulation and there is not very specific content about how to implement and protect the personal information of citizens or who will be responsible to address a privacy issue.

In the end, to form an effective privacy protection system, Chinese policy-makers and legislators have a lot of work to do.

6.2.2 Hong Kong S.A.R.

(1) Overview

Hong Kong is a Special Administrative Region (S.A.R.) of China. As an ex-colony of the United Kingdom, the political and legal system of Hong Kong could be seen as a version of the Westminster system, just like Canada. Privacy is not a very important part of traditional Asian culture, especially East Asian culture. But due to the influence of the British, more people in Hong Kong care about their privacy than on mainland China. So, both political and cultural factors result in Hong Kong having a more specific privacy protection system than mainland China.

In the Basic Law of the Hong Kong Special Administrative Region of the People's Republic of China, the "Small Constitution" of Hong Kong, several

fundamental rights of Hong Kong residents are also protected. The law protects the freedom of speech, of the press and of publication, and protects people in Hong Kong from unlawful search of the body and unlawful search of a resident's home or other premises. (Government of People's Republic of China 1990)

Unlike Canada, there is just one law relating to privacy, covering both the public and the private sectors. The Personal Data (Privacy) Ordinance (Cap. 486),protects personal data and first came into force on 20 December 1996, before Hong Kong was handed over to Chinese government. The Ordinance was amended in 2012, adding a regulation about direct marketing and increasing the level of penalty. (Government of HK SAR 2013)

(2) Features

There are six data protection principles in the Ordinance, regulating the behavior of all data users, including government agencies and private companies. (Government of HK SAR 2013) The principles require that personal data must be collected in a lawful and fair way, for a purpose directly related to an activity of the data users, and personal data only can be used for that purpose. The data user must notify the data owner the purpose of collection and to whom the data may be transferred. They need to ensure personal data is accurate and safe and not kept longer than is necessary to fulfill the purpose for which it is collected. In addition, the data user must make its data policy public and indicate the types of personal data it holds and how the data is used. For data owners, they have the right to access their personal data and can

correct inaccurate data.

The amendment in 2012 requires data users before using personal data in direct marketing to inform the data owner about the marketing and provide a response channel to data owners that will enable them to decide whether they would accept the marketing. (Government of HK SAR 2013) That is a response to the increase in spam messages and email.

There also is a privacy commissioner in Hong Kong. The Office of the Privacy Commissioner for Personal Data is an independent statutory body set up to oversee the enforcement of the Personal Data (Privacy) Ordinance. The Privacy Commissioner is not an employee of government, and their key roles are enforcement the of Ordinance, monitoring and supervising compliance, promotion of the idea of privacy, and meeting changing needs. The Privacy Commissioner also has the responsibility to promote and assist data users to prepare codes of practice in compliance with the Ordinance and to carry out inspections of any personal data systems used by government or corporations. The commissioner is required to report his work to the Panel on Constitutional Affairs of the Hong Kong Legislative Council every year. (Government of HK SAR 2013)

For medical information, the Commissioner clearly indicates that personal medical information is also protected by the Ordinance. As they have said, “Patients’ health records in the System amounts to personal data, which is protected under the Personal Data (Privacy) Ordinance.” (PCPD 2015)

People in Hong Kong also have a clear way to file complaints about privacy issues to and assisted by the Privacy Commissioner. They are also encouraged to contact the related data user first, before filing a complaint.

6.3 Summary

In general, the Constitution in all three countries protects fundamental freedom and rights, such as the freedom of speech, expression and communication of its residents; the Constitution also protects them from unreasonable search and seizure.

Canada and Hong Kong have a sophisticated privacy protection system, perhaps because people living there care more about privacy. The privacy legislation in Canada and Hong Kong regulates the activities of the public and private sectors collecting personal data, although Canada has several laws while there is just one law in Hong Kong. The regulations in both Canada and Hong Kong require data users to inform data owners about the purpose of data collection, and the collected personal data only can be used for that purpose, unless they get the consent of data owners. Data users in these two area cannot disclose or transfer personal data in any form without the consent of the data owner, and they must ensure the security of data. All people in these areas have the right to assess their personal data.

The privacy laws in Canada and Hong Kong establish their privacy commissioners to protect privacy. The two privacy commissioners are responsible for inspecting the internal regulations in government. But in Hong Kong, the commissioner needs to provide support to government departments and even companies to set up the internal

regulation of privacy. They need to report to Parliament every year. But in Hong Kong the commissioner will report to a general committee, while in Canada, its commissioner will report to a specific committee on privacy and ethics.

In the end, for the development of big data, both Hong Kong and Canada have conducted several studies recently to think about privacy in the big data era. The Privacy Commissioner of Hong Kong has held several seminars about big data and privacy, the next seminar will be held in April 2017. (PCPD 2017) On the Canadian side, on January 28th, 2017, (i.e., on the privacy day), the information and privacy commissioner of Ontario made a presentation about big data and privacy. They are aware big data may collect information indirectly and the purpose may be unclear at the time of collection. In 2016, the IPC issued a de-identification guidance to guide anonymization, and a full set of Big Data Guidelines will be released in Spring 2017. (IPC 2017) In the study of the federal level of Office of Privacy commissioner, *The Consent Dilemma*, indicating the current consent model is no longer practicable and needs to enhance the consent model in order to make it more meaningful. (Therrien 2016) But obviously, they still do not have final decision on whether need to and how to change the current privacy protection system. They need to play the important role to lead the change of the system.

7 Analysis

7.1 Can the system protect privacy well?

This part will mainly talk about whether the current privacy protection system can

protect privacy well in the big data era, based on the system in Canada and Hong Kong. I do not use the system in China because an effective system does not exist, but maybe the system of Hong Kong and Canada can provide instruction to China for its privacy protection system.

In the big data era, as I describe above, the biggest difference is that the amount of data has increased and more kinds of data analysis are conducted. First, I will argue that, in the current system, if every government agency, company or other organization complies with privacy regulations, people's privacy can be well protected in the big data era. Because the current system in Hong Kong and Canada prohibit data users from collecting, using or transferring personal information without the consent of data owners, data owners can decide who they need to provide their information to. That's protects privacy in the data creating phase.

In addition, without the owner's consent, no organization can provide or disclose personal information to a third party. As a result, identity fraud and data breaches cannot happen, and a data owner would not be forced to provide data, as they have the right to fully control their personal data. On the other hand, when collecting data, the data user must inform the data owner about the purpose of this collection and the collected data only can be used for that purpose, so spam and extra advertisements would not exist because the data owner has the right to decline the requests of unwelcome purpose. In the end, because regulations require data users to protect the security of data, as long as data users update big data infrastructure frequently, fix security bugs and regularly check

system status, data security can be protected. These measures protect people's privacy in the processing, analyzing and preserving phases of the data life cycle. A requirement to use data only for the purpose of collection could protect people's privacy in the re-use phase, but it is a controversial point. (OPC 2016)

As the regulations require giving data owners the right to access their data and prohibit disclosure to other users without the consent of the data owner, this protects privacy in the giving access phase of data life cycle. In the end, the current privacy protection system protects privacy in every phase of data life cycle. To conclude, big data just makes the amount of data increase, and creates new pressure on our privacy protection system, but the system can still work well if organizations comply with the regulations. Data life cycle is always a useful tool to analysis data-related problem, but for the privacy and data life cycle, there is less research focus on. Further study could focus more on the relationship between data life cycle and privacy, and see whether it can provide more insight about the problem.

However, identity fraud, data breaches and other privacy issues still happen. This is not because the development of big data defeats our privacy protection but instead probably because not every organization obeys the privacy regulations and not everyone cares about their privacy. New data analytics in the legal grey zone also cause some challenges to the privacy protection system as a whole. Besides, the volume of data has increased exponentially in the big data era, which has created an enormous workload for privacy protection agencies, so that they cannot be sure that each organization complies

with the regulations. Indeed, privacy issues not only happen in the big data era but also exist in earlier phases, so the occurrence of privacy incidents cannot be blamed on big data. On the other hand, current research shows that there is no correlation between the development of big data and the increase of data breaches. In the United States, Privacy Rights Clearinghouse (PRC) and Identity Theft Resource Center (ITRC) have recorded the incidence of data breach for each year from 2005. The results for the year 2006-2015 are as Chart 1:

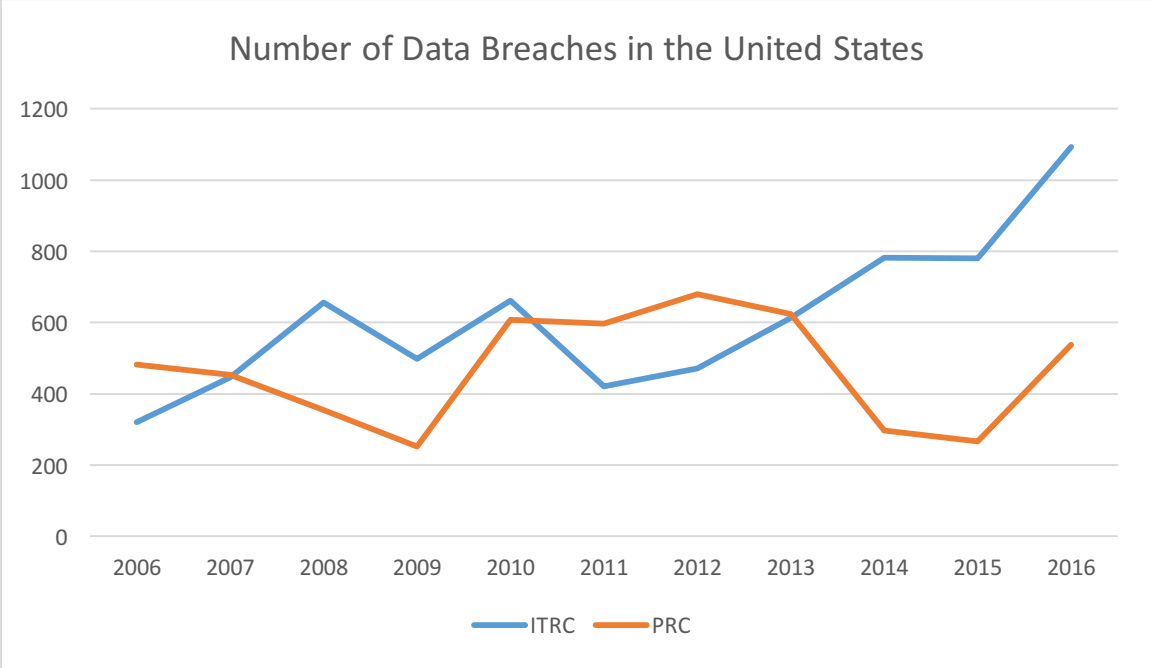

Chart 1

Source: Collecting from Privacy Rights Clearinghouse data breach database

Collecting from Identity Theft Resource Center data breach report 2007-2017

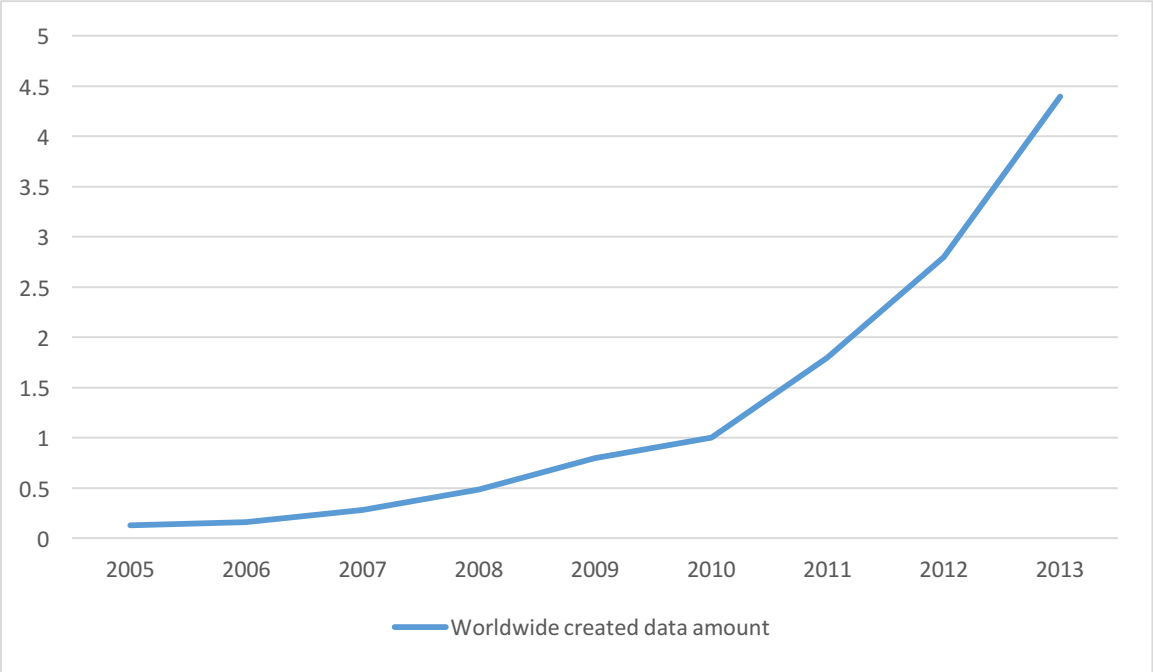

Chart 2

Source: EMC Digital Universe Research 2006-2014

As we can see from chart 1, there is no obvious upward trend in the data of Privacy Rights Clearinghouse, although the data from Identity Theft Resource Center, do indicate an upward trend and a tripling over the period studied. Chart 2 shows that the amount of data that was created worldwide in each year from 2005 to 2013; compared with the increase in the total amount of data, however, the rate of increase in data breaches is much lower than the increase in data volume and data analytics activity, so the development of big data is not the only factor in pushing the trend upward. It is very likely that with the amount of data increase, the pressure on privacy regulation has also increased. There is some need to enforce privacy regulation, but our current privacy protection system still works – if it didn't the number of data breaches would triple!

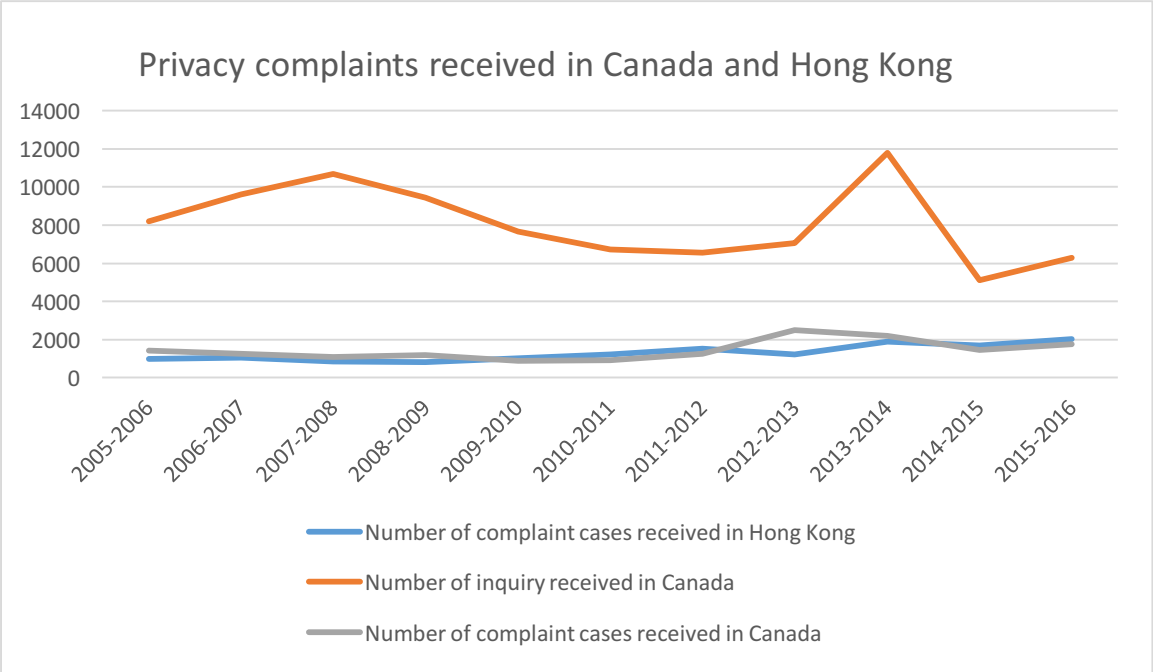

Chart 3

Through the data from the website of The Office of Privacy Commissioner of Canada and The Office of the Privacy Commissioner for Personal Data in Hong Kong, we can develop chart 3. From 2005-2016, there is no significant change in the inquiries and complaints received by the two privacy commissioners. If the current privacy protection system is unable to work well, the complaint cases will increase significantly. So, results from the United States, Canada and Hong Kong show, in general, that the current privacy protection system can protect privacy in the era of big data, although the pressure existed. The conclusion from experience so far is that the development of big data has not made our current privacy protection system fail, but it has increased the pressure on our system.

Although I think our privacy protection system still works, there are risks to privacy, as the system is not perfect. As I mention in the earlier discussion of the commercial,

medicine and even government field, sometimes with multiple "anonymous" data, people can identify the data owner. That is because the current privacy regulations do not require the anonymization of data, and the current anonymization approach instead focuses on eliminating part, but not all, of potentially identifiable personal information before releasing data. So, that's why sometimes "anonymous" data are not truly anonymous. If enforced data anonymization can be part of our privacy protection system, and data can be anonymous from the time the data is collected, our privacy protection system could be more effective. As a first step, the Information and Privacy Commissioner of Ontario issued a Guidance on anonymization or de-identification in June 2016, providing a step-by-step process for de-identifying data sets with common de-identification techniques. (IPC 2016) In the future, the requirement of de-identifying could become a part of Privacy legislation.

Recently, some new anonymization technologies have also become mature; Differential Privacy is one of them. Differential Privacy was first introduced by Cynthia Dwork (2006) of Microsoft Research, and since then, many organizations and researchers have started working on it. Finally, in 2016 Apple announced that they will adopt Differential Privacy for all data they collected. (Greenberg 2016) Differential Privacy is just like adding "noise" or "turbulence" to the original data, with the result that the data owner cannot be identified but in a way that will not harm the results of data analytics using this data. If these type of anonymization technologies can be successful and will be a required application for data users, the risks for the current privacy protection system

could be reduced or eliminated.

To conclude, although I observe that the current privacy protection system can protect privacy at all phases of the data life cycle and that privacy-related data do not show that there is a relationship between the development of big data and the occurrence of privacy issues, I believe there are still risks in our system. These include partial and imperfect anonymization. If an enforced anonymization requirement and a general standard of anonymization are introduced into the system, the system could be better. In the future, we can believe the worldwide created data amount will continue increase and even more rapidly, which will cause more pressure on the privacy protection system. But as more people start focus on the issue of big data and privacy, and more regulation agency start to think about the necessity to change the system, the occurrence of privacy issue will maintain a same level or gradually decline if better regulation can be enforced and reduce risks.

7.2 Will the privacy system hinder reasonable data analytics?

This part will talk about whether the current privacy protection system will hinder reasonable data analytics, which I define as lawful applications and those applications violating privacy laws but whose advantages far outweigh their disadvantages.

"Lawful" data analytics can be divided into two parts. For reasonable data analytics without using personal data, the privacy protection system will not hinder them at all, such as ticket price trends or flu trends. For reasonable data analytics using personal data

that are consistent with the laws, such as single purpose collection or research that will not disclose personal information, data users just need to get the approval from data owners, do research by themselves, remove identifiable features before publishing the results and delete data after the research end.

However, in the case of reasonable data analytics using personal information that will violate privacy laws, our privacy protection system will become the obstacle. It is difficult to find a very specific example, but in this paper, big data-based debt collecting and deciding insurance premium based on genetic testing obviously do not belong to reasonable data analytics. They did not bring much benefit to the society, maybe they just have some advantages for certain groups of people. This part will mainly talk about what feature of data analytic will be impeded by out current privacy protection system.

First, our current privacy protection system requires data users to obtain consent from data owners for the purpose of collection every time data is collected. This regulation is a double-edged sword. Admittedly, it can prevent the abuse of personal data and enable data owners to control their personal information. However, this regulation sometimes will increase costs and reduce efficiency. For a data user conducting a new reasonable data analysis with new purposes or changed purposes after the original use, the re-approval will cost a lot, as the "Consent dilemma" research conducted by the Office of the Privacy Commissioner of Canada. (OPC 2016) As described above, in the big data era, many personal data collection activities are indirect and cannot fully determine the purpose at the time of collection. In these cases, informing the owner of the

purpose at the time of collection becomes impossible. And because big data relies greatly on historical data, the old consent approach may reduce the analytic data volume, and hinder the development of big data.

Second, privacy regulations in both Canada and Hong Kong require data users to delete personal data after the informed purpose is finished. However, a great deal of big data analytics is based on historical data. Sometimes, after the original purpose is finished, the same data can also play an important role in the next data analysis. Again, the re-collection will cost a lot and reduce efficiency. If data users cannot keep data, the development of big data would be severely limited.

Third, under current privacy laws, data users cannot transfer or disclose data to a third party without the consent of data owners. But in the case of some projects, such as the genome project, this could produce major benefits, although it would also require cooperation among different research institutions. Under the current system, information sharing between these institutions is also prohibited if the data owner does not agree. This would also reduce efficiency and sometimes make some research impossible.

Besides, sometimes because these applications are in a legal grey zone, researchers and operators of data analytics may be afraid of the penalties of privacy law, as a result slowing the development of data analytics, or even stopping it.

In the end, I believe our current privacy system would hinder a type of reasonable data analytics, which is based on personal data but violates current privacy laws with the features above. Unreasonable data analytics can be considered to be a problem or risk in

the privacy protection system and should be eliminated. But this raises an issue: who should decide a data analytic is reasonable? And how to judge a data analytic project's advantages far outweigh its privacy concerns? This decision should be made independently and probably could be a new role for the privacy commissioner. Alternatively, the approach of Negative List could be an appropriate way for making this judgement. This issue could be a topic for research in the future.

8 Conclusion

This article mainly talks about with the coming of big data, whether our current privacy protection system could protect privacy well but also whether it would hinder data analytics. The analysis is mainly based on the regulations of Canada, China mainland and Hong Kong SAR of China. It also draws on experience in Hong Kong and Canada. With the research, we have found that Canada and Hong Kong have well-designed privacy protection systems and the main principles are roughly the same, probably influenced by OECD regulations. The privacy protection system in China needs considerable improvement, and probably people in China mainland are not aware of the importance of privacy protection.

In response to the research questions, I have found that although privacy breaches sometimes occur, and the system always faces a great deal of pressure with the development of big data, people's privacy can be protected in general. But, small adjustments, such as anonymization, are needed to reduce risk and enforce regulation. On

the other hand, to some extent, our privacy protection system is an obstacle for some type of data analytics. This is especially the case with reasonable analytics using personal data but in ways that violate our current privacy protection system; this paper concludes its feature, including continually re-using or changing the data purpose or conducting analytics that involve sharing personal data with a third party. These types of data analytics could bring benefit to our society, like in the medical field, but the judgement of what are reasonable data analytics is still a problem. Further researches could focus the standard of reasonable data analytics and who have the right to decide one data analytic is a reasonable data analytic.

In a word, changes are needed to our privacy protection system in Canada and Hong Kong to facilitate the development of big data while protecting privacy at the same time. For China mainland, it seems they have a long way to go to build their privacy protection system, and maybe they can draw conclusions on the merits of other countries and build a new system. Admittedly, no one system can be considered perfect, because our society is complex and the World is always changing. So, only if we keep adjusting our system will we can get a better system.

Appendix

Original Chinese text and translation of Chinese literature by Google Translate

Xu, Jihua, Qina Feng, and Zhenru Chen. 2014. Smart Government. Beijing: China CITIC Press

徐继华，冯启娜，和陈贞汝. 2014. 智慧政府. 北京: 中信出版社

一是实现信息透明和共享，使外部利益相关者和内部利益相关者都能提供自身的工作效率，产生积极的经济社会综合效益。

二是通过评估公共部门的绩效，增强内部竞争，激励工作表现，提高公共建设效率，提升行政服务质量，降低政府管理成本。

三是通过人口细分和定制政策，增强公共政策的针对性，提高工作效率和公众满意度，减少开支。

四是用政务智能替代或辅助人工决策，在纷繁复杂的数据中自动识别出不一致，错误和虚假的信息，减少出错成本和福利管理中的诈骗，缩小税收缺口。

五是引导公共部门内部和外部的创新，例如商业、非营利机构、第三方通过开发出大数据工具和分析，对公共服务进行反馈，为改善现有的方案提出建议，从而为公共部门创造新的价值。

First, the realization of information transparency and sharing, so that external stakeholders and internal stakeholders can provide their own work efficiency, resulting in a comprehensive economic and social benefits.

Second, by assessing the performance of the public sector, enhance internal competition, stimulate job performance, improve the efficiency of public construction, improve the quality of administrative services, reduce government management costs.

Third, through population segmentation and customization policies, enhance the relevance of public policy, improve work efficiency and public satisfaction, reduce expenses.

Fourth, the use of government intelligence to replace or assist in manual decision-making, in the complex data automatically identify inconsistent, false and false information to reduce the cost of fraud and welfare management fraud, narrow the tax gap.

Five is to guide internal and external innovation within the public sector, such as commercial, nonprofit, third parties, through the development of large data tools and analysis, feedback on public services, recommendations for improving existing programs to create new sectors for the public sector the value of.

The Standing Committee of the National People's Congress passed the Decision on Strengthening the Protection of Online Information (Extract)

全国人大常委会关于加强网络信息保护的决定（节选)

- 国家保护能够识别公民个人身份和涉及公民个人隐私的电子信息。
- 网络服务提供者和其他企业事业单位在业务活动中收集、使用公民个人电子信息，应当遵循合法、正当、必要的原则，明示收集、使用信息的目的、方式和

范围，并经被收集者同意，不得违反法律、法规的规定和双方的约定收集、使用信息。

- 公民发现泄露个人身份、散布个人隐私等侵害其合法权益的网络信息，或者受到商业性电子信息侵扰的，有权要求网络服务提供者删除有关信息或者采取其他必要措施予以制止。
- 任何组织和个人对窃取或者以其他非法方式获取、出售或者非法向他人提供公民个人电子信息的违法犯罪行为以及其他网络信息违法犯罪行为，有权向有关主管部门举报、控告；接到举报、控告的部门应当依法及时处理。被侵权人可以依法提起诉讼。

State protection can identify citizens 'personal identities and electronic information that involves citizens' personal privacy.

- The network service providers and other enterprises and institutions collect and use personal electronic information in their business activities. They should follow the principles of lawfulness, legitimacy and necessity, expressly collect and use the purpose, manner and scope of the information and agree with the collectors , Shall not violate the provisions of laws and regulations and the two sides agreed to collect and use information.
- Citizens are found to disclose personal information, disclose personal information such as privacy, or infringe on their legitimate rights and interests, or be subject to commercial electronic information, the right to require network service providers to delete the relevant information or take other necessary measures to be stopped.

- Any organization or individual who has acquired, sold or illegally illegally provided criminal electronic information to others, as well as other network information, has the right to report and accuse the relevant authorities; to report, The departments concerned shall handle them in a timely manner. The infringer may sue in accordance with the law.

DTCaijing. 2017. What is big data-based debt collection?. Baidu Baijia. http://dtcaijing.baijia.baidu.com/ article/759509 (Accessed April 15th 2017)

DT 财经：大数据“讨债”，到底是门什么生意？

零壹财经曾报道称：网贷行业的坏账率高达 12%。截至 2016 年 5 月，中国整个银行业不良率已经突破了 2%。金融行业坏账攀升，直接刺激了“催收”生意的火爆。

大数据在催收行业到底如何发挥作用？第一步是帮助找到人。百融金服副总裁王立新曾撰文指出，国内个人信贷市场中，借款人失联的情况非常普遍，低账龄逾期借款人中失联比例估计在 30-50% ,而进入到不良阶段 ,失联比例可达到 70%以上。那么，能帮助你找人的大数据哪里来？催收公司各家有各家的路数，比较常见的一类是通过整合政府公开发布的数据，建立自己的内部数据库，从中找到借款人的联系方式或资产，这类平台有催天下、搜赖网等。根据催天下给 DT 的反馈，其数据基础主要是其所属公司汇法集团积累多年的法律文书，他们通过技术手段把案件当事人的信息清洗出来，作为催天下失联修复的方式之一。

在数据的获取和使用中我们注意到了一些灰色现象，尤其是涉及个人隐私的部分。比如，有些催收公司的失联修复的功能是通过外部“大数据供应商”，把用户个人联系信息（如通讯方式，QQ 号，淘宝号等）补齐。黑客也会利用“社工库”和一些黑客手段，获取一些个人隐私信息。社工库，是指大量外泄的用户隐私数据集合地，也是黑客常用的“数据共享库”。

Zero Finance has reported that: net loan industry bad debt rate as high as 12%. As of May 2016, China's entire banking industry has broken through 2%. Financial sector bad debt rise, directly stimulated the "collection" business hot.

Big data in the collection industry in the end how to play a role? The first step is to help find people. In the domestic personal credit market, the borrower lost the situation is very common, low-age overdue borrower in the proportion of lost in 30-50%, and into the bad stage, lost The proportion can reach more than 70%. So, can you help you find someone's big data where? The collection of the various companies have the number of each way, the more common one is through the integration of government public data released, the establishment of their own internal database, from which to find the borrower's contact or assets, such platforms have reminders, Network and so on. According to the feedback to the world, the data base is mainly its own company, the accumulation of years of legal instruments, the means of the parties to the information of the parties to clean out the information, as a reminder of the world to repair one of the ways.

In the data acquisition and use, we have noticed some gray phenomena, especially those involved in personal privacy. For example, some of the company's lost partner repair function is through the external "large data providers", the user personal contact information (such as communication, QQ, Taobao, etc.) fill. Hackers will also use the "social workers library" and some hacking means to obtain some personal privacy information. Social workers library, refers to a large number of leaked user privacy data collection, but also hackers commonly used "data sharing library."

Reference

Bailey, Stuart G M, and Nadia Caidi. 2005. How Much Is Too Little? Privacy and Smart Cards in Hong Kong and Ontario. *Journal of Information Science* 31 (5): 354–64.

Bagnall, Jim. 2017. The Cloud looms on Shared Services' horizon. *Ottawasun*. March 19th,2017, http://www.ottawasun.com/2017/03/19/ bagnall-the-cloud-looms-on-shared-services-horizon

Bombard, Yvonne, Ronald Cohn, and Stephen Scherer. 2016. Why we need a law to prevent genetic discrimination. *The Globe and Mail*. Sep. 19, 2016

Culnan, M. J., and P. K. Armstrong. 1999. Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. *Organization Science* 10 (1): 104–15.

de Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. Unique in the Crowd: The Privacy Bounds of Human Mobility. *Scientific Reports* 3: 1376.

Davis, Mark Edward. 2017. Canada Passes Legislation Protecting Genetic Information. Data Protection Report. http://www.dataprotectionreport.com/ 2017/04/canada-passes-legislation-protecting-genetic-information/

Ding, Chunyan. 2013. PATIENT PRIVACY PROTECTION IN CHINA IN THE AGE OF ELECTRONIC HEALTH RECORDS. *Hong Kong Law Journal* 1: 245–178.

DTCaijing. 2017. What is big data-based debt collection? (大数据“讨债”，到底是门什么生意？). Baidu Baijia. http://dtcaijing.baijia.baidu.com/ article/759509 (Accessed April 15th 2017, see Appendix A)

Dwork, Cynthia. 2006. Differential Privacy. *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming*, 1–12.

Eckhoff, David, and Christoph Sommer. 2014. Driving for Big Data? Privacy Concerns in Vehicular Networking. *IEEE Security & Privacy* 12 (1): 77–79.

Einav, Liran, and Jonathan Levin. 2014. Economics in the age of big data. *Science*. Nov 07, 2014, http://science.sciencemag.org/content/346/6210/1243089

Executive Office of the President. President's Council of Advisors on Science and Technology. 2014. *Report to the President: Big Data and Privacy: A Technological Perspective.*

Fewer, David. 2016. Open Data, Open Citizens? Open Data and Privacy. Office of the Privacy Commissioner of Canada. https://www.priv.gc.ca/en/opc-actions-and-decisions/research/funding-for-privacy-research-and-knowledge-translation/completed-contributions-program-projects/2015-2016/p_201516_09/

Gaff, Brian M., Heather Egan Sussman, and Jennifer Geetter. 2014. Privacy and Big Data. *COMPUTER AND THE LAW* 47 (6): 7–9.

Government of Canada. 1985. *Privacy Act of Canada*. *ReVision*. http://laws-lois.justice.gc.ca/eng/acts/P-21/.

Government of Canada. 2013. *Personal Information Protection and Electronic Documents Act*, SC 2000, c 5. *ReVision*,

Government of Canada. 2013. Open Data 101. http://open.canada.ca/en/open-data-principles

Government of HK SAR. 2013. *Personal Data (Privacy) Ordinance*. Vol. 14. http://www.legislation.gov.hk/blis_pdf.nsf/4f0db701c6c25d4a4825755c00352e35/B4DF8B4125C4214D482575EF000EC5FF/$FILE/CAP_486_e_b5.pdf.

Government of People's Republic of China. 1990. The Basic Law of the Hong Kong Special Administrative Region of the People's Republic of China

Grassegger, Hannes, and Mikael Krogerus. The Data That Turned the World Upside Down. *VICE*. Jan 28 2017, https://motherboard.vice.com/en_us/article/how-our-likes-helped-trump-win

Greenberg, Andy. 2016. Apple's 'Differential Privacy' Is About Collecting Your Data—But Not Your Data. *Wired*. June 13th, 2016, https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/

Hamilton, Anita. 2008. Best Inventions of 2008. *Time*. Oct. 29, 2008, http://content.time.com/time/specials/packages/completelist/0,29569,1852747,00.html

Hardy, Keiran, and Alana Maurushat. 2016. Opening up Government Data for Big Data Analysis and Public Benefit. *Computer Law & Security Review*, no. July. Elsevier Ltd.

Hogan Lovalls. 2014. An Overview of Hong Kong's Personal Data (Privacy) Ordinance.

IBM. 2017. Bringing big data to the enterprise. IBM Corporation. https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

IPC. 2016. *De-identification Guidelines for Structured Data*. Information and Privacy Commissioner of Ontario. https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf (Accessed April 15th 2017)

IPC. 2017. *Government and Big Data: Privacy Risks and Solutions. Information and Privacy Commissioner of Ontario*. https://www.ipc.on.ca/wp-content/uploads/2017/01/2017-01-26-privacy-day-commissioner-presentation-final.pdf (Accessed April 15th 2017)

Kaiser Permanente. 2014. *Kaiser Permanente 2014 Annual Report*. Kaiser Permanente Corporation. https://share.kaiserpermanente.org/static/kp_annualreport_2014/ (accessed April 14, 2017)

Kennedy, Gabriela, and Heidi Gleeson. 2012. The Hong Kong Personal Data (Privacy) Ordinance Has Been Amended: Are Your Data Protection Practices and Policies Adequate? *Computer & Internet Lawyer* 29 (9): 17–21.

Kim, Gang-Hoon, Silvana Trimi, and Ji-Hyong Chung. 2014. Big-Data Applications in the Government Sector. *Association for Computing Machinery. Communications of the ACM* 57 (3): 78.

Kshetri, Nir. 2014. China's Data Privacy Regulations: A Tricky Tradeoff between ICT's Productive Utilization and Cybercontrol. *IEEE Security and Privacy* 12 (4): 38–45.

Langfitt, Frank. 2013. In China, Beware: A Camera May Be Watching You. *National Public Radio*. January 29, 2013, http://www.npr.org/2013/01/29/170469038/in-china-beware-a-camera-may-be-watching-you

Lenard, Thomas M, and Paul H. Rubin. 2014. BIG DATA, PRIVACY AND THE FAMILIAR SOLUTIONS. *JOURNAL OF LAW, ECONOMICS & POLICY* 1 (1).

Li, Hong. 2011. The Conflict and Balance between Government's Information Right and Citizen's Privacy Right. *Journal of Politics and Law* 4 (2): 104–8.

Lim, Sun Sun, Hichang Cho, and Milagros Rivera Sanchez. 2009. Online Privacy, Government Surveillance and National ID Cards. *Communications of the ACM* 52 (12): 116–20.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2014. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. British: John Murray

McKinsey & Company. 2011. Big Data: The next Frontier for Innovation, Competition, and Productivity. *McKinsey Global Institute*, no. June: 156.

Mehmood, Abid, Iynkaran Natgunanathan, Yong Xiang, Guang Hua, and Song Guo. 2016. Protection of Big Data Privacy. *IEEE Access* 4: 1821–34.

Mo, Jojo Y C. 2014. Are Data Protection Laws Sufficient for Privacy Intrusions? The Case in Hong Kong. *Computer Law and Security Review* 30 (4). Elsevier Ltd: 429–38.

Mo, Yun Ching Jojo, and A. K C Koo. 2014. A Bolder Step towards Privacy Protection in Hong Kong: A Statutory Cause of Action. *Asian Journal of Comparative Law* 9 (1): 345–80.

National Human Genome Research Institute. *The Human Genome Project Completion: Frequently Asked Questions*. National Human Genome Research Institute. https://www.genome.gov/11006943/ (last updated October 30, 2010)

Navetta, David. 2014. Legal Implications of Big Data. *The Computer & Internet Lawyer* 31 (1): 1–6.

OECD. 1980. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*.

Office of Privacy Commissioner of Canada. 2014. Overview of privacy legislation in Canada. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/02_05_d_15/ (Updated in May 2014)

Office of Privacy Commissioner of Canada. 2016. File a formal privacy complaint. https://www.priv.gc.ca/en/report-a-concern/file-a-formal-privacy-complaint/

Office of Privacy Commissioner of Canada. 2016. About OPC. https://www.priv.gc.ca/en/about-the-opc/

Parliament of Canada. 2017. About ETHI. http://www.parl.gc.ca/Committees/en/ETHI/About (Accessed April 15th 2017)

PCPD. 2015. Electronic Health Record Sharing System. The Privacy Commissioner for Personal Data Hong Kong. https://www.pcpd.org.hk/english/data_privacy_law/electronic_health_record_sharing_system/ehrss.html (Accessed April 15, 2017)

PCPD. 2017. 'Big Data, Artificial Intelligence and Privacy' Seminar. The Privacy Commissioner for Personal Data Hong Kong. https://www.pcpd.org.hk/spec_event/spec_event18_apply.php (Accessed April 15, 2017)

Perera, C, R Ranjan, L Wang, S U Khan, and A Y Zomaya. 2015. Big Data Privacy in the Internet of Things Era. *IT Professional* 17 (3): 32–39.

Quintero, Dino, Luis Bolinches, Aditya Gandakusuma Sutandyo, Nicolas Joly and Reinaldo Tetsuo Katahira. 2016. *IBM Data Engine for Hadoop and Spark*: International Business Machines Corporation. http://ibm.com/redbooks

Raghupathi, Wullianallur, and Viju Raghupathi. 2014. Big Data Analytics in Healthcare: Promise and Potential. *Health Information Science and Systems* 2: 3.

Rubinstein, Ira S. 2013. Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law* 3 (2): 74–87.

Sadilek, Adam, and John Krumm. 2012. Far Out: Predicting Long-Term Human Mobility. *26th AAAI Conference on Artificial Intelligence*, 814–20. doi:10.1.1.224.6709.

Shen, Anqi. 2012. Policing and Legal Protection of Privacy in China and the UK: A Comparative Study. *Journal of Law and Social Sciences* 2 (1): 1–8.

Sherer, James A, Jenny Le, and Amie Taal. 2015. Big Data Discovery, Privacy, and the Application of Diff Erential Privacy Mechanisms. *The Computer & Internet Lawyer* 32 (7): 10–17.

Stanaland, Andrea J S, and May O Lwin. 2013. Online Privacy Practices: Advances in China. *Journal of International Business Research* 12 (2): 33–47.

The White House. 2014. *Big Data: Seizing Opportunities, Preserving Values.*

Therrien, Daniel. 2016. The Consent Dilemma. Office of the Privacy Commissioner of Canada. https://www.priv.gc.ca/en/opc-news/speeches/2016/sp-d_20160705/

Westin, Alan. 1967. *Privacy And Freedom*. New York: Athenum

Williamson, Andy. 2014. "Big Data and the Implications for Government." *Legal Information Management* 14: 253–57.

Wu, Yanfang, Tuenyu Lau, David J. Atkin, and Carolyn A. Lin. 2011. A Comparative Study of Online Privacy Regulations in the U.S. and China. *Telecommunications Policy* 35 (7). Elsevier: 603–16.

Xu, Jihua, Qina Feng, and Zhenru Chen. 2014. *Smart Government(智慧政府)*. Beijing: China CITIC Press

Yao, Mike Z, and Jinguang Zhang. 2008. Predicting User Concerns about Online Privacy in Hong Kong. *CYBERPSYCHOLOGY & BEHAVIOR* 11 (6): 779–81.

ZestFinance. 2017. Our Story. https://www.zestfinance.com/our-story (Accessed April 15th, 2017)