Big Data & the Benefits of an Enterprise Data Solution for a Municipality

Jennifer Fraser-Lee

February 26, 2020

AMCTO – Executive Diploma in Municipal Management

Executive Summary

In today's world, data is no longer seen as a byproduct of technology but rather a valuable asset. It is generated from a variety of sources – from wearable technology to our refrigerators to our doorbells. When analyzed data can provide businesses powerful insights into what products are profitable, customer preferences and can assist in predicting future customer behaviour. Municipal governments are beginning to enter the big data and big data analytics world with a rapid increase in the amount of data they collect and store. Many municipal organizations are looking for ways to use this data to help improve and innovate service delivery and policy making while reducing costs. As governments begin to navigate this new data driven world it is important to highlight certain privacy and information management concerns that can be alleviated or mitigated with the implementation of an enterprise data solution. This paper provides a legislative review of applicable information management and privacy legislation, defines big data and data analytics, identifies information management and privacy concerns related to big data and the use of analytics and presents the benefits of an enterprise data solution in mitigating these concerns.

Table of Contents

Introduction

With the proliferation of technology, especially connected devices the data generation has significantly increased. With the growth of data has come the recognition that it can provide businesses with a competitive advantage and government organizations with opportunities to innovate service delivery while managing resources more efficiently (Reinsel, Gantz, Rynding, 2018). It is the collection of large data sets that is referred to as "big data" and its subsequent analysis as "data analytics" or "predictive analytics." To capitalize on the data in the custody and control of an organization, many organizations have begun to implement enterprise data solutions. An enterprise data solution aims to ensure an organization's data is ethically collected, used and stored while providing insight on its resources and clients.

There are numerous drivers behind the use of data analytics in municipal government. These drivers were highlighted Module #1 – Public Sector Trends of this Diploma program – doing more with less, responding to customers' changing expectations and technological changes and expectations (AMCTO, Module #1, 2018). As Canadian municipalities become more interested in using big data and data analytics as a way of being responsive in this changing landscape an enterprise data solution can assist in addressing certain privacy and information management concerns.

Scope and Methodology

This paper provides a legislative review of applicable information management and privacy legislation. It also defines big data, data analytics and enterprise data solution. Following the definitions of these key terms, this paper will outline the potential privacy and information management risks, identify the benefits of an enterprise data solution, and recommend approaches and solutions to address these concerns.

To understand the benefits of an enterprise data solution secondary source research was completed. This included the review of relevant and applicable legislation as well as industry literature related to big data and data analytics as it relates to the management of data and the protection of privacy.

Legislative Review

To understand privacy and information management concerns related to municipal use of big data and data analytics the following is a legislative review. Ontario municipalities are governed by three pieces of legislation that have components related to the management of information and protection of privacy.

These are the:

- *Municipal Act, 2001*
- *Municipal Freedom of Information and Protection of Privacy Act* (MFIPPA)
- *Personal Health Information Protection Act* (PHIPA)[1]

For the purposes of this paper only the *Municipal Act* and MFIPPA will be examined as both apply to all Ontario municipalities.

The Municipal Act

The *Municipal Act* sets forth the powers and duties of a municipality (Ontario, 2001). Although the *Municipal Act* outlines a variety of responsibilities, this paper will examine the municipality's duty to appropriately manage information. Section 253 of the *Municipal Act* prescribes how an institution must manage the records in its custody and control. It states that pursuant to the *Municipal Freedom of Information and Protection of Privacy Act* an individual has the right to inspect the records held by the Clerk (Ontario, 2001). Section 254 outlines an institution's obligation to retain records. It states:

[1] Note: Only municipalities with Health Information Custodians are governed by PHIPA.

> A municipality shall retain and preserve the records of the municipality and its local boards in a secure and accessible manner and, if a local board is a local board of more than one municipality, the affected municipalities are jointly responsible for complying with this subsection. 2001, c. 25, s. 254 (1). (Ontario, 2001).

As such, the *Municipal Act* requires all municipalities to retain records securely while making them accessible. This also includes data retained by the municipality.

Municipal Freedom of Information and Protection of Privacy Act

The *Municipal Freedom of Information and Protection of Privacy Act* (MFIPPA) sets out a municipality's obligation for providing individuals access to records held by the institution and rules around the protection of an individual's privacy. MFIPPA defines a record as:

> any record of information however recorded, whether in printed form, on film, by electronic means or otherwise, and includes,
>
> (a) correspondence, a memorandum, a book, a plan, a map, a drawing, a diagram, a pictorial or graphic work, a photograph, a film, a microfilm, a sound recording, a videotape, a machine readable record, any other documentary material, regardless of physical form or characteristics, and any copy thereof, and
>
> (b) subject to the regulations, any record that is capable of being produced from a machine readable record under the control of an institution by means of computer hardware and software or any other information storage equipment and technical expertise normally used by the institution; ("document") (Ontario, 2019).

Essentially this definition incorporates everything held by the institution that is recorded. Although "data" is not specifically listed in MFIPPA, it can be argued that data does fall within the scope of the definition of a record as it is a machine-readable record and is capable of being produced by computer hardware and software. This piece of legislation is dated in its terminology and does not tackle the intricacies of big data but the broad definition does not exclude data.

Access to Information

MFIPPA outlines an institution's obligation to provide individuals access to government held information. In particular, Section 4(1) states:

> Every person has a right of access to a record or a part of a record in the custody or under the control of an institution unless,
>
> (a) the record or the part of the record falls within one of the exemptions under sections 6 to 15; or
>
> (b) the head is of the opinion on reasonable grounds that the request for access is frivolous or vexatious. (Ontario, 2019).

In order to access government held information, an individual would submit a Freedom of Information request. The exemptions referred to in Section 4(1)(a) are broken down into mandatory and discretionary exemptions. Where mandatory exemptions apply (see appendix 1) the head of an institution must sever the contents of those records. For discretionary exemptions (see appendix 1) a head may use the criteria outlined in the legislation to determine if all or parts of the record should be severed. When exemptions are applied to sever records, they must be specific and limited. This ensures transparency and accountability in the activities of the municipality.

An individual can request specific data related to a subject (e.g. Emergency Medical Services response times, expenditures, etc.) or submit a broad request encompassing anything and everything on a specific matter, which would include data. As such, it is imperative for a municipality to manage its data in a manner that allows timely retrieval of the information.

Protection of Privacy

Part two of MFIPPA outlines a municipality's obligations to protect an individual's privacy and specific rules around the collection, use and disclosure of personal information. MFIPPA defines personal information as any "recorded information about an identifiable individual" (see Appendix 2 for full definition). As per MFIPPA a municipality can only collect personal information under certain and specific circumstances and must provide notice to the individual when collecting their personal information. Similarly, there are specific rules around the institution's use and disclosure of personal information. Generally, any personal information an institution has can only be used or disclosed as outlined in the notice of collection or under other specific and limited criteria outlined in the legislation (see Appendix 3 for complete criteria). The legislation also states that personal information should only be retained for one year after its use (Ontario, 2019). Limiting the collection, use, disclosure and retention of personal information is paramount in protecting an individual's privacy and maintaining public trust.

If an individual disagrees with an institution's decision regarding their freedom of information request or how the institution has handled their personal information they have the right to file an appeal or privacy compliant with the Information and Privacy Commissioner of Ontario (IPC). In 2018 the IPC received 1442 appeals related to freedom of information requests, 306 privacy complaints and 870 complaints related to the handling of personal health information (IPC, 2019). With the rapid increase in data generated and collected by institutions and the new uses of that data it is expected that these appeal and complaint numbers will increase.

In 2016, MFIPPA was amended to include requirements for an institution to take reasonable measures to preserve records in its custody or control. It applies to all records (as defined by MFIPPA) in all stages of the record's lifecycle. Additionally, the

new amendments made it an offence to alter, conceal or destroy a record with the intention of denying an individual the right of access to that record under MFIPPA (IPC, 2015). The guidance document released by the IPC in relation to the amendments reinforces that the definition of a "record" within MFIPPA is very broad and includes as examples "unique data collections" and "geospatial data" (IPC, 2015). These amendments strengthen the obligations of an institution when it comes to appropriately protecting the records, including data, in its custody or control.

Benefits of an Enterprise Data Solution

In addition to fostering accountability and transparency, there are numerous benefits for both the institution and individual when an institution has a comprehensive enterprise data solution.

Benefits for the Institution	**Benefits for the Individual**
Consistency in practices	Quicker access
Accuracy of the data	Improved privacy protection
Improved service delivery	Efficient government spending leads to reduced costs to residents
Reduced duplication	
Cost savings	
Secure data	
Quicker access	

Definitions

What is Big Data?

According to Press (2014) the term "big data" was first referenced in a 1997 NASA paper. The paper used the term to describe the problem of data sets being so large that they did not fit into computer or disk memory. Since this initial reference to the term big data has been used widely. Regardless of any nuances in the definitions, almost all include reference to the role of technology and the large size of the datasets. Some definitions include reference to the ability to predict behaviours and trends from

the data while others focus on the ability of computer system to compute or use the data. For the purposes of this paper the Oxford Dictionary definition of big data will be used. The Oxford Dictionary defines big data as "extremely large data sets that may be analyzed computationally to reveal patterns, trends and associations, especially relating to human behaviour and interactions" (Lexico, 2020).

Although there are multiple definitions that vary slightly from one another, there is consistency in identifying the "three V's" of big data. These characteristics are referred to as variety, velocity and volume (O'Reilly Media, 2012). The three V's of big data are described in the table below:

The three V's of Big Data:

Variety	**Velocity**	**Volume**
Data is collected with varying degrees of structure	Numbers come in fast and can be processed very quickly (e.g. real-time data collected from mobile devices)	Refers to the sheer amount of data available

**Chart from the Berkeley School of Information (2019).*

These three attributes characterize the uniqueness of big data and help define what it means. As the big data field matures and the technological landscape changes, the reference to "big" in terms of volume and the ability for computers to handle large datasets is becoming less relevant. The focus is moving from the volume of data to ensuring the data is accurate, accessible and useful for evidence based decision-making. In order for the data to be useable and beneficial, it must be accessible by those that need it, it must be organized and accurate, and any biases in the data must be identified (Berkeley, 2019). This shift to the reliability of the data is the drive behind organizations evaluating and implementing enterprise data solutions.

What is Data Analytics?

Data analytics is the process by which insights are derived from large data sets. Various tools and methods such as algorithms can be used to identify patterns, automate decision-making and predict behaviour (IPC, 2017). It is the identification of these patterns and trends that can assist organizations in tailoring products or services. For a government organization understanding patterns and trends could assist in a redistribution of resources to provide more streamlined and responsive services. This could also lead to reduced costs and the creation of new programs to meet previously unknown needs. In the municipal setting, the use of big data analytics is gaining momentum in police services, health care, transit, water and wastewater, and traffic management.

Deriving insights from data has benefits for both the municipality and the individual. That being said, it must be done in an ethical manner that upholds the trust of residents and meets any legislative requirements. In Module #1 of this course the instructor detailed the shift from the "Traditional Public Administration" in local government to the "New Public Management" model. With the shift there is a greater focus on good governance and how municipalities can increase transparency, accountability, sustainability and participation (AMCTO, Module #1, 2018). Providing access to data and using data in decision-making can assist municipalities in achieving these goals.

This chart highlights some of the benefits of using big data analytics:

Benefits for the Institution	**Benefits for the Individual**
Cost savings	Personalized, tailored offerings and interactions
Improved service delivery	Improved communications
Innovative services	New service offerings
Evidence based decision-making	Improved wait times
Assist with policy development – research (AMCTO, Module #11, 2019)	

Privacy and Information Management Concerns

As outlined above there are legal obligations municipalities must follow to protect privacy and manage information. In terms of implementing an enterprise data program there are some key privacy and information management concerns to address. Although there are a variety of privacy and information management concerns related to the use of big data and analytics in a municipality, this paper will focus on: accuracy of the data and identifying bias, using or disclosing the data beyond the original purpose, access to the data (both access rights and ease of access), sensitivity of the data and the potential to identify individuals, and the retention, storage and security of the data.

Regardless of the specific information management and privacy concerns identified in this paper it is essential that organizations always ensure the ethical collection, use and disclosure of any data that may contain personal identifiers or be sensitive in nature. Although we may be able to legally collect and use the data in a particular manner, with big data and data analytics we need to really ask ourselves the following questions:

- Is this analysis ethical? Is it the right thing to do?
- Are we using the data beyond the original purpose?
- Would the individual expect or accept their data being used or analyzed in this manner? Is it reasonable?
- Is there another way we can accomplish what we are trying to do that decreases the risk to privacy?
- Do we need to seek consent?
- How do we notify individuals?

Accuracy of the Data & Identifying Bias

One of the most cited arguments for using big data is that the data is objective because of its sheer volume. This can be true in certain instances but biases can still

exist in big data sets. The IPC argues that these biases can "exclude or single out certain groups or people" (IPC, 2017). If there are biases in the data that are not identified this could have significant harmful affects on individuals. It could result in certain groups of individuals not receiving services they may need. Biases should be minimized where possible to ensure purity of the data and where it is not possible to avoid bias then they should be clearly identified and stated.

Another prominent reason for using big data and analytics is its ability to allow for evidence-based decision-making. This argument relies heavily on the accuracy of the data available. The emphasis on accurate data has lead to the addition of a fourth characteristic to the definition of big data. This fourth "v" is veracity. Data veracity refers to the "accuracy or trustfulness of a data set" (Ved, 2019). In the Technology Vision 2018 Intelligent Enterprise presented by Yves Bernaert, Senior Managing Director at Accenture, data veracity is listed as the third trend. In particular, Bernaert points out that no one is talking of "big data" anymore and that the focus is on the quality of the data and not the volume (Biciuc, 2019). With accurate data, there is the ability to improve resource allocation in the municipal setting. For example, traffic signal timings can be improved or adjusted by analyzing traffic patterns or an analysis of historic weather patterns can assist in the administration of the winter roads maintenance work (Raynor, 2018).

When the accuracy of data is questionable there exists potential for inaccurate results or predictions. This can significantly impact individuals' lives by resulting in conclusions that marginalize individuals or groups of individuals. In addition to inaccurate data producing inaccurate results, the IPC highlights the importance of understanding the difference between correlation and causation in interpreting the analysis of the data (IPC, 2017). Correlation is the relationship between two statistical variables, which tend to increase or decrease together. Causation is that two variables are related by "necessity and that a change in one always brings a change in the other"

(IPC, 2017). So although there may be a correlation between two variables, this does not necessarily mean causation. Similarly, it is important to identify any biases related to the data when interpreting the results. Depending on how the data was collected the objectivity of the data can be skewed because of biases that were present during the collection. This is often times highlighted in predictive policing. Using historical crime data can increase police presence in certain neighbourhoods, which can lead to increased enforcement. This skews the data and becomes a cycle difficult to break. Without understanding biases in the data and how the algorithms work, it can negatively impact certain minority groups and strain relationships. It also does not provide reliable insights as to where policing resources are actually needed. It is imperative to verify the accuracy of the data, understand the difference between correlation and causation, identify bias and use critical thinking when examining the source data and the algorithms.

Use or Disclosure of the Data

As mentioned above, MFIPPA has specific rules about how personal information can be used or disclosed. In particular, the legislation only allows a municipality to use personal information if the individual has consented to its use, it is used for the purposes outlined in the notice of collection or if the institution uses it for a purpose outlined in Section 32 of MFIPPA (Ontario, 2019). Similarly, an institution may only disclose personal information in particular instances. See Appendix 3 for Section 32 of MFIPPA, which outlines the rules around the disclosure of personal information. Based on the services municipalities provide, it is unavoidable that they use data that contains some element of personal information.

With big data analytics, one of the biggest concerns and threats to privacy is the secondary use and disclosure of the data. If an institution has collected data containing personal information from an individual for a specific purpose (e.g. administering social assistance) and then another department uses that data (e.g. to determine who to

contact about upcoming road work in a specific area) this is not considered a consistent use or disclosure of the information – there would be other ways to identify who needs to be notified. Although this example highlights the use or disclosure of contact information, it is a common request within municipalities to share data sets containing personal information between departments. A large driver behind data analytics in municipalities is its use in policy development. Having data available when researching a problem or identifying a need is necessary in order to develop a quality policy that addresses the need (AMCTO, Module #11, 2019). Even though there is a business need, municipalities must ensure that the data used does not contain personal information and that the risk of re-identification is low if it was combined with other datasets. The IPC cautions that if "not properly managed the collection, use and disclosure of big data sets may be contrary to Ontario's privacy laws" (IPC, 2017).

Access

When discussing access to data there are two components that need to be considered that can impact privacy and information management practices – access rights and ease of access. The first deals with ensuring authorized access to the data. Access to data must be restricted based on the sensitivity of the data. The less people that have access to data containing personal information the more secure the data is and the better the institution will be at protecting privacy.

The other information management concern related to access to datasets is the ease of access to the data. As previously mentioned, volume and velocity are key characteristics of big data. These two unique attributes make it very important to properly organize the information in a manner that allows ease of access where permitted. If it is difficult to access data then an organization cannot realize the value of the data. Many organizations do not know where all of their data resides and its fragmentation does not allow for meaningful analysis.

It is important to recognize that individuals can seek access to government information through freedom of information requests. As previously mentioned, these requests have strict timelines and fee structures that can make requests for large datasets costly for the organization to respond to. As a result, many government institutions have launched open data programs as a way of providing access to government data for both internal and external customers. Open data programs are one of the trends shaping the public sector in the 21st century (AMCTO, Module #1, 2018). The City of Guelph's Open Data program is recognized as one of the most comprehensive and innovative programs that generates value for both the municipality, its residents and businesses (Younes, 2016). In order to provide these types of programs or to respond to freedom of information requests it is necessary to identify where the data exists and who can have access to it.

Sensitivity of Data

Municipalities are required to collect vast amounts of personal information from individuals in order to administer various programs and services. The sensitivity of the personal information varies depending on the program or service provided but can range from name and date of birth to detailed financial and medical information. Every piece of personal "data" needs to be protected in accordance with applicable privacy laws. There exists a struggle between protecting privacy by removing sensitive data and personal identifiers and having useable valuable data. A paper released by researchers at Harvard states that "less granular data protects privacy but is less valuable as an asset to promote transparency, enable innovation and aid research" (Harvard, 2017). The paper indicates that granular data "contains the most detailed information…[it] often includes personally sensitive information" (Harvard, 2017). As such a balance is needed between having usable valuable data to assist municipalities (which is often the granular data) in providing cost effective, innovative services for their residents while protecting their privacy.

One of the most common ways to protect an individual's privacy when using datasets is to de-identify the data. The IPC released guidelines in 2016 on de-identifying structured data. In these guidelines, de-identification is defined as "removing any information that (i) identifies an individual, or (ii) for which there is a reasonable expectation that the information could be used, either alone or with other information to identify an individual " (IPC, 2016). Regardless of the technique used to de-identify the data (e.g. masking or generalization), experts do agree that a risk-based approach should be used to determine an "acceptable level of re-identification risk" for a dataset (IPC, 2016). In order to protect an individual's privacy it is necessary to evaluate the sensitivity of the data before it is used and to determine how to remove any identifying or sensitive information from the dataset.

Storage, Retention and Security

As previously mentioned volume is one of the original defining characteristics of big data. With the amount of data generated continuing to drastically increase, organizations are faced with the problem of finding appropriate storage solutions for the data. According to the 2018 IDC report cloud data centres are becoming the primary enterprise data repository (IDC, 2018). The IDC predicts that "49% of the world's stored data will reside in public cloud environments" by 2025 (IDC, 2018). In terms of information management concerns, organizations will be faced with increased storage costs. Although cloud storage is cheaper than traditional centralized storage, it will still be costly for organizations to store their growing data. As a result and in order to ensure fiscal prudence, it will be important to keep the data only as long as necessary – no more, no less and to be thoughtful about what you actually need to collect.

There are legislative obligations that municipalities establish records retention bylaws and retain the records in their custody and control in accordance with the municipality's established bylaw. The recent amendment from Bill 8 reinforces an organization's responsibility to properly retain records. This includes retaining data in

accordance with established retention periods. This can become difficult if an organization does not know where their data is stored resulting in the destruction of data before its met its retention or keeping data longer than the specified retention period and increasing costs.

Proper storage and retention of data also impacts the security, accuracy and reliability of the data. Knowing where the data resides assists an organization in applying appropriate security rights and implementing the proper controls to protect that data from internal and external threats.

What is an Enterprise Data Solution?

An Enterprise Data Solution (EDS) does not refer to a single particular technology product but rather an organization's approach to effectively managing its data to harness its value. With the proliferation of technology, especially real-time devices, the amount of data created and stored has significantly increased in recent years. The International Data Corporation predicts that growth will continue to be exorbitant with the Global Datasphere growing from 33 zettabytes (ZB) to 175 ZB by 2025 (Reinsel, Gantz, Rynding, 2018). A significant portion of this data is generated from real-time devices and embedded sensors in Internet of Things devices. Although the amount of data created is rapidly growing, not all of the data created needs to be stored. An EDS assists an organization in determining which data is valuable, for what purposes and how to best keep it. It puts in place an appropriate governance structure that ensures consistency across the organization and that privacy and information management concerns are addressed.

Recommendations

Governance

In order to implement an enterprise data solution that allows a municipality to harness the value of the data it collects, solid governance is needed. The first step is to establish the need for an enterprise program and receive buy-in from various stakeholders (senior management to front-line staff). In order to launch an enterprise data program with the necessary governance in place, the creation of a business case would assist in articulating the driving needs and benefits. A business case would provide control, minimize risk and establish direction for the governance (AMCTO, Module #2, 2018). With proper governance there will be consistency across the organization in its data management practices. This will ensure privacy and information management concerns are identified and mitigated. In order to assist with the governance of an enterprise data program, it is necessary to have qualified staff in place to provide expertise on where the data comes from, how the algorithms work and how the results should be interpreted (Raynor, 2018).

There is recognition that legislative reform is needed to better protect individual's privacy when it comes to big data, big data analytics and artificial intelligence while still enabling organizations to use the data to generate value for the organization and assist organizations in tackling the management of the data. The *Simpler, Faster, Better Services Act* seeks to provide the public with more digital services from the Ontario government while assisting public sector organizations such as municipalities with guidance around data management practices. It mandates that any digital service provided by a government institution be designed with data collection and management as a priority from the start. It also furthers the transparency of government institutions by requiring the publication of lists of datasets and facilitating access to these datasets subject to sensitivity and privacy requirements (Ontario, 2019).

This legislation will also provide guidance on how data can be shared between ministries. Municipalities can incorporate these requirements and practices where applicable. This piece of legislation is the first step in the Ontario government's recognition of the importance of data to government organizations in order to provide responsive and innovative services to the public while protecting personal privacy. It also highlights the benefits to clients and the institution when responsibly sharing and integrating data between ministries or departments. With the direction moving towards sharing and integrating data across ministries the IPC has made it clear that best practices are to have a "government-wide solution for data integration [and that] such a solution would enable data linkages to support effective planning, analysis and evaluation while protecting personal privacy" (IPC, 2019). In the same report, the IPC cautions that "a fragmented approach to data integration could result in a proliferation of linked databases containing the same or similar information (IPC, 2019). Although this new piece of legislation and guidance from the IPC relates to the sharing and integration of data across ministries it is still relevant to municipalities. As municipalities begin to implement enterprise data programs and consider using big data and data analytics they can build in the guidance from the IPC by ensuring the program is corporate wide with consistent rules and application. This will help protect privacy, improve accuracy and reduce duplication and costs. Establishing corporate wide governance within an organization will become even more important to ensure legislative compliance.

Formalize sharing and access

Access to the data both from within and outside an organization is a key component in big data and the ability to analyze the data. As part of the governance structure, formalizing the process for sharing and accessing data by both internal and external customers will help ensure privacy is protected, the data is used in a valuable way and is managed appropriately. Ultimately, the sensitivity of the data should dictate

access rights and before any sharing or manipulation can occur a review of the collection notice should be completed to ensure the sharing is consistent with what the described use is.

Since a freedom of information request provides a strict timeline and cost structure that a municipality must follow it is recommended that the organization establish a routine disclosure and comprehensive open data program for data. This would enable access to data outside the MFIPPA process, which would provide the institution more flexibility in response times and cost recovery (if routine disclosure). Both approaches help strengthen transparency and accountability for municipalities.

Inventory

It is recommended that when implementing an enterprise data program an updated inventory be conducted on all data sources. As per requirements under MFIPPA, a municipality must keep an inventory of all personal information banks. This should be broadened to include all data (containing both personal and non-personal datasets). The inventory should document what data elements are collected in a dataset, where the data resides, who currently has access, how it is secured, how it was collected and when it was last updated. The inventory should be completed corporate-wide and be the starting point for the organization in determining any risks to the data, individual privacy and how to best initiate consistent corporate-wide data management practices. If the dataset contains personal information the collection notice should be reviewed to understand how the data can be used.

De-identify sensitive data

It is essential to de-identify all sensitive datasets. There are many approaches and techniques to de-identifying datasets. Both the IPC and the Harvard report on Open Data Privacy recommend implementing a risk-based approach to de-identifying datasets. Historically, the recommendation when de-identifying or anonymizing data

was to remove any data elements that contain personally identifiable information (PII). Recent research indicates that simply removing the traditional PII is not enough to protect an individual's privacy. According to the Harvard paper since there is so much data from a variety of sources and "because databases can be manipulated and combined in complex and unpredictable ways, information that might not be deemed PII can lead to the identification of a specific individual and enable inferences to be made about that individual" (Harvard, 2017).

The risk-based approach allows an organization to calculate an acceptable level of re-identification risk while weighing it against the benefit of the releasing or sharing the dataset. It explores the risk and benefit of the data attributes (vulnerability, asset), event (threat event, advantage event), source (threat source, advantage source), likelihood, impact and outcome (risk, benefit). From this a municipality can determine if the risk is greater than the benefit and vice versa. This will allow it to determine if a dataset should be released or shared and any mitigations that need to occur. From this analysis and based on the data elements that need to be de-identified there are a variety of methods to de-identify the data. These methods include but are not limited to: removing fields, aggregating data, k-anonymity, and creating anonymous identifiers (Harvard, 2017). Being a steward of the data, it is integral for municipalities to ensure the privacy of any individual represented in the data is protected.

Audit

To ensure the governance of an enterprise data solution is working and to identify areas of improvement it is necessary to conduct regular audits of the program. In particular, it is recommended the following areas be audited on an annual basis to ensure consistency and to assist in avoiding privacy risks:

- Access rights;
- Security threats;
- What data has met its retention and identify any data that has been disposed of;

- What datasets are being shared;
- How datasets are being integrated with other datasets – has this expanded from the original use?;
- Accuracy of the data used;
- If biases in the data have been identified and communicated;
- Collection notices - do our collection notices reflect our use or disclosure of the data?

Regular audits, will ensure appropriate privacy protections are incorporated into all aspects of managing data and will lay the foundation of a solid governance program.

Conclusion

Using data and data analytics will become a key priority for municipalities as we enter the new decade. It will provide opportunities to innovate services, programs and the way we interact with our residents. It will also provide costs savings, increase transparency and accountability, and modernize municipal institutions. That being said, in order for municipalities to use data and data analytics effectively they must do it responsibly. As trusted entities, any analysis, manipulation and sharing of data must be done ethically. We must not go beyond a person's reasonable expectation of acceptable use when using their data. Even if it is only shared internally or it is de-identified we must do it within the confines of the current legislative framework (even if it is outdated and does not account for the nuances of big data) and with an individual's privacy at the forefront. It is imperative that municipalities focus on creating enterprise data solutions rooted in sound governance that address information management and privacy concerns from the beginning and not as an after thought.

References

AMCTO. (2018). Executive Diploma in Municipal Management – Module #1 Public Sector Trends – Course Notes.

AMCTO. (2018). Executive Diploma in Municipal Management – Module #2 Building Business Cases – Course Notes.

AMCTO. (2018). Executive Diploma in Municipal Management – Module #11 Policy Formulation, Implementation & Evaluation – Course Notes.

Bicuic, Valentina. (2019). "10 Highlights of what should be from Web Summit 2018 – Day 2." *Wolfpack Digital.* https://medium.com/wolfpack-digital/10-highlights-from-the-web-summit-2018-day-2-643b844a6e37

Government of Ontario. (2001). *Municipal Act, 2001.* https://www.ontario.ca/laws/statute/01m25#BK2

Government of Ontario. (1990). *Municipal Freedom of Information and Protection of Privacy Act,* R.S.O. 1990, c. M. 56. https://www.ontario.ca/laws/statute/90m56?search=municipal+freedom+of+information#BK5

Government of Ontario (2019). *Simpler, Faster, Better Services Act, 2019,* S.O. 2019, c.7, Sched. 56. https://www.ontario.ca/laws/statute/19s07#BK8

Green, Ben, Gabe Cunningham, Ariel Ekblaw, Paul Kominers, Andrew Linzer, and Susan Crawford. 2017. Open Data Privacy (2017). Berkman Klein Center for Internet & Society Research Publication. https://dash.harvard.edu/bitstream/handle/1/30340010/OpenDataPrivacy.pdf

O'Reilly Media (2012). *Volume, Velocity, Variety: What You Need to Know About Big Data.* Forbes. https://www.forbes.com/sites/oreillymedia/2012/01/19/volume-velocity-variety-what-you-need-to-know-about-big-data/#3c5808611b6d

Office of the Information and Privacy Commissioner of Ontario. (2019). *2018 Annual Report – Privacy and Accountability for a Digital Ontario*. https://www.ipc.on.ca/wp-content/uploads/2019/06/ar-2018-e.pdf

Office of the Information and Privacy Commissioner of Ontario. (2017). *Big Data and Your Privacy Rights.* https://www.ipc.on.ca/wp-content/uploads/2017/01/fact-sheet-big-data-with-links.pdf

Office of the Information and Privacy Commissioner of Ontario. (2016). *De-Identification Guidelines for Structured Data.* https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf

Office of the Information and Privacy Commissioner of Ontario. (2015). *FIPPA and MFIPPA: Bill 8 – The Recordkeeping Amendments.* https://www.ipc.on.ca/wp-content/uploads/resources/bill8-new-recordkeeping-amendments.pdf

Press, Gil. (2014). *12 Big Data Definitions: What's Yours?* Forbes. https://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#7df2435a13ae

Raynor, Christopher. (2018). *Big Data Policy.* AMCTO Policy and Management Brief. https://staging.amcto.com/getattachment/3816747f-9af1-47e0-b1b4-db1701b83151/.aspx

Reinsel, David., Gantz, John and Rydning, John. (2018). "The Digitization of the World: From Edge to Core." *IDC*. https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

University of California: Berkeley School of Information (2019). *Big Data Isn't a Concept – It's a Problem to Solve.* https://datascience.berkeley.edu/blog/what-is-big-data/

Ved, Max. (2019). "Data Veracity: A New Key to Big Data." *Datafloq.* https://datafloq.com/read/data-veracity-new-key-big-data/6595

Younes, Joseph Bou. (2016). "Civic Innovation and Open Data at the City of Guelph." *Communitech News.* https://news.communitech.ca/civic-innovation-and-open-data-at-the-city-of-guelph/

Appendix 1 – Exemptions under MFIPPA

Mandatory Exemptions	**Discretionary Exemptions**
-Relations with governments	-Draft by-laws, etc.
-Third party information	-Advice or recommendations
-Personal Privacy	-Law enforcement
	-Economic and other interests
	-Solicitor-client privilege
	-Danger to health or safety
	-Information soon to be published

Appendix 2 – MFIPPA Definition of "Personal Information"

"personal information" means recorded information about an identifiable individual, including,

(a) information relating to the race, national or ethnic origin, colour, religion, age, sex, sexual orientation or marital or family status of the individual,

(b) information relating to the education or the medical, psychiatric, psychological, criminal or employment history of the individual or information relating to financial transactions in which the individual has been involved,

(c) any identifying number, symbol or other particular assigned to the individual,

(d) the address, telephone number, fingerprints or blood type of the individual,

(e) the personal opinions or views of the individual except if they relate to another individual,

(f) correspondence sent to an institution by the individual that is implicitly or explicitly of a private or confidential nature, and replies to that correspondence that would reveal the contents of the original correspondence,

(g) the views or opinions of another individual about the individual, and

(h) the individual's name if it appears with other personal information relating to the individual or where the disclosure of the name would reveal other personal information about the individual; ("renseignements personnels")

Appendix 3 – Use and Disclosure of Personal Information under MFIPPA

USE AND DISCLOSURE OF PERSONAL INFORMATION

Use of personal information

31 An institution shall not use personal information in its custody or under its control except,

(a) if the person to whom the information relates has identified that information in particular and consented to its use;

(b) for the purpose for which it was obtained or compiled or for a consistent purpose; or

(c) for a purpose for which the information may be disclosed to the institution under section 32 or under section 42 of the *Freedom of Information and Protection of Privacy Act*. R.S.O. 1990, c. M.56, s. 31.

Where disclosure permitted

32 An institution shall not disclose personal information in its custody or under its control except,

(a) in accordance with Part I;

(b) if the person to whom the information relates has identified that information in particular and consented to its disclosure;

(c) for the purpose for which it was obtained or compiled or for a consistent purpose;

(d) if the disclosure is made to an officer, employee, consultant or agent of the institution who needs the record in the performance of their duties and if the disclosure is necessary and proper in the discharge of the institution's functions;

(e) for the purpose of complying with an Act of the Legislature or an Act of Parliament, an agreement or arrangement under such an Act or a treaty;

Note: On a day to be named by proclamation of the Lieutenant Governor, clause 32 (e) of the Act is repealed and the following substituted: (See: 2019, c. 7, Sched. 41, s. 2 (1))

(e) where permitted or required by law or by a treaty, agreement or arrangement made under an Act or an Act of Canada;

(f) if disclosure is by a law enforcement institution,

(i) to a law enforcement agency in a foreign country under an arrangement, a written agreement or treaty or legislative authority, or

(ii) to another law enforcement agency in Canada;

(g) if disclosure is to an institution or a law enforcement agency in Canada to aid an investigation undertaken with a view to a law enforcement proceeding or from which a law enforcement proceeding is likely to result;

Note: On a day to be named by proclamation of the Lieutenant Governor, clause 32 (g) of the Act is repealed and the following substituted: (See: 2019, c. 7, Sched. 41, s. 2 (2))

(g) to an institution or a law enforcement agency in Canada if,

(i) the disclosure is to aid in an investigation undertaken by the institution or the agency with a view to a law enforcement proceeding, or

(ii) there is a reasonable basis to believe that an offence may have been committed and the disclosure is to enable the institution or the agency to determine whether to conduct such an investigation;

(h) in compelling circumstances affecting the health or safety of an individual if upon disclosure notification is mailed to the last known address of the individual to whom the information relates;

(i) in compassionate circumstances, to facilitate contact with the spouse, a close relative or a friend of an individual who is injured, ill or deceased;

(j) to the Minister;

(k) to the Information and Privacy Commissioner;

(l) to the Government of Canada or the Government of Ontario in order to facilitate the auditing of shared cost programs. R.S.O. 1990, c. M.56, s. 32; 2006, c. 19, Sched. N, s. 3 (5); 2006, c. 34, Sched. C, s. 15.

Section Amendments with date in force (d/m/y)

Consistent purpose

33 The purpose of a use or disclosure of personal information that has been collected directly from the individual to whom the information relates is a consistent purpose under clauses 31 (b) and 32 (c) only if the individual might reasonably have expected such a use or disclosure. R.S.O. 1990, c. M.56, s. 33.