

Open Data Policy

Policy No. CIMS 002
Version No: 2.0
Approval Date: November 22, 2011

City Clerk's Office

An Information Management Policy

Subject: Open Data Policy

Policy No: CIMS 002
Version No: 2.0

Keywords: Information management, data, accessibility, open government, principles, accountability, transparency

Issued by: City Clerk's Office
Corporate Information Management Services

Issued on:

Contact Information:

Director, Corporate Information Policy Tel: (416) 397-0736

Corporate Information Management 13W City Hall
Services 100 Queen St. West,
Toronto, Ontario M5H 2N2

Table of Contents

1. INTRODUCTION	1
2. ALIGNMENT WITH INFORMATION MANAGEMENT FRAMEWORK.....	1
3. ALIGNMENT WITH ACCESS BY DESIGN	2
4. ALIGNMENT WITH PRIVACY BY DESIGN.....	2
5. PURPOSE	3
6. POLICY STATEMENT.....	3
7. GUIDING PRINCIPLES TO MANAGE CITY DATASETS	4
8. APPLICATION	5
9. DEFINITIONS.....	5
10. ROLES AND RESPONSIBILITIES.....	6
11. COMPLIANCE	7
12. AUTHORITIES.....	7
13. APPLICABLE POLICIES AND RESOURCES	8
14. APPROVED BY:	8
15. POLICY APPROVAL AND REVIEW	8
16. APPENDIX A	9

1. Introduction

The Open Data Policy outlines the principles, roles, and responsibilities related to the City of Toronto's efforts to make data routinely available in machine readable format for any public use. The Open Data Policy supports the City of Toronto's commitment to Open Government. Open Government assists the City of Toronto in enriching customer service and proactively addressing enquiries and complaints in a timely and accurate manner.

Open Government and the Open Data Program are changing the landscape of information management accountability and information accessibility. Open Government is about citizen engagement, customer service, transparency, accountability and the sharing of knowledge and information leading to greater collaboration and innovation. Open Data is one driving force of Open Government and its singular focus is making data publicly available in recognized and usable formats for anyone to re-use, re-purpose, and develop into digital applications for the benefit of the public. Data can be accessed and utilized; one person's use does not preclude someone else from also accessing it, utilizing it and potentially offering new or enriched data for the benefit of everyone. This new environment of open, accessible and reusable data establishes a foundation where stakeholders use such data to foster healthy debate and discussion on City issues.

The Open Data Program is an enterprise information management initiative; it demonstrates the City's commitment to better manage business information throughout the information lifecycle. Identifying and making data accessible helps to ensure that the public is informed and engaged in an open and accessible government.

The City of Toronto makes data available to the public, businesses, institutions, visitors, and other levels of government via www.toronto.ca/open. The City must comply with provincial and federal legislation. The City will not post datasets containing confidential, proprietary, and/or personal information. By offering datasets, the City supports unfiltered access to its information.

2. Alignment with Information Management Framework

The Information Management Framework (IMF) was approved by the City's Business Advisory Panel, June 2009 and endorsed by the City's Open Government Committee in 2011. The IMF is a corporately approved standards-based approach to managing information as a strategic corporate asset.

Its principles are:

Accountability: All employees are responsible for the proper management of information.

Openness: Information is open and accessible.

Lifecycle: Information is managed through all of its stages of usefulness in coordination with business planning.

Value: Information is current, accurate, relevant and easy to use.

For further reference:

(<https://www.toronto.ca/wp-content/uploads/2017/08/8fe1-IMFrameworkTorontoRevised.pdf>)

The Open Data Policy supports the principle that information is open and accessible.

3. Alignment with Access by Design

The Office of the Information and Privacy Commissioner of Ontario has developed a set of seven access principles that encourage public institutions to take a proactive approach to releasing information and making the disclosure of government-held information an automatic process wherever possible:

(http://www.ipc.on.ca/images/Resources/accessbydesign_7fundamentalprinciples.pdf).

Access by Design advances the view that government-held information should be made available to the public, and that any exceptions should be limited and specific. This concept aligns and supports the principles for open data set forth in this policy.

Further validation comes in the form of a collective statement made by the Information and Privacy Commissioners of Canada and the Provinces and Territories on September 1, 2010, which spoke to the need for Open Government. This included defining one of the tenets of open government as: “Open, accessible and reusable information” (http://www.priv.gc.ca/media/nr-c/2010/res_100901_e.cfm).

4. Alignment with Privacy by Design

The Office of the Information and Privacy Commissioner of Ontario has developed a working concept called Privacy by Design (PbD) that addresses the ever-growing and systemic privacy concerns of managing information within information technology, social media and communication technologies. PbD is a set of seven high-level principles for organizations to follow to establish and build privacy controls within their business processes. The principles are found here:

<http://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf>

The Open Data Policy recognizes these principles and requires that the release of City of Toronto datasets will not contain personal and/or private information.

5. Purpose

The purpose of the Open Data Policy is to remove barriers and set the rules by which City of Toronto data is made available to the public as valuable, machine readable datasets.

6. Policy Statement

The City of Toronto will:

1. share with everyone its open and accessible datasets while adhering to rights of privacy, security and confidentiality as identified in the *Municipal Freedom of Information and Protection of Privacy Act*, *Personal Health Information Protection Act, 2004* and other legislation.
2. publish datasets via www.toronto.ca/open allowing everyone to develop digital applications that may improve government transparency and public participation, enhance access to City services, and ultimately strengthen democracy and contribute to a more liveable city.
3. post on the Open Data website, an Open Data Licence, procedures, supported file formats, glossary, and other dataset context information to promote responsible use of City of Toronto information.

Executives will:

1. identify existing and potential datasets for release as part of the Open Data Program and work with the Open Data Team on the planning and development of new datasets, review of existing ones, publication of datasets, and archiving of superseded datasets if required;
2. plan, implement, update, and identify metrics and measures of current datasets when planning the creation of new datasets/databases, or scheduling technology enhancements;
3. release datasets to Open Data Team where a formal Freedom of Information request has already been made or is in the process of being disclosed; and,
4. release datasets to the Open Data Team where data/information in hard copy (printed reports, posters, brochures, etc) or soft copy (PDF, web content, etc) has already been released to the public.

Open Government Committee will:

1. ensure information at the City of Toronto is managed in ways that assist in creating a culture of Open Government and information sharing by way of providing open data governance and oversight;
2. provide support to the Open Data Team to develop awareness and training programs to help City staff incorporate open data initiatives into their business planning processes;
3. promote information transparency and accountability to build trust and confidence in government; and,
4. foster Open Government leadership in recognition of the evolving democratic process.

7. Guiding Principles to Manage City Datasets

In August 2010, the U.S. Sunlight Foundation, a non-profit organization that focuses on “transparency in Government”, further developed a set of guiding principles for open data. The City of Toronto accepted these principles in December 2010 as part of a consultant's report, "Open Data Framework – Final Report, Prepared for G4 (City of Toronto, Edmonton, Vancouver and Ottawa)", and acknowledges these principles as providing the necessary structure for public sector engagement with open government and to ensure that data is open, accessible and reusable.

A complete description of each principle is found in Appendix A at the end of this policy.

1. Completeness

Datasets will be as complete as possible while complying with legislative obligations regarding the release of personal information, proprietary, or other confidential information.

2. Primacy

Datasets will be primary source data with data collection methods documented.

3. Timeliness

Datasets will be available to the public in a timely fashion to maintain the business value of the data.

4. Accessibility

Datasets will be as accessible as possible, with accessibility defined as the ease with which information can be obtained.

5. Machine Readable

Datasets will be machine readable so that the public can create applications that can use the data for new services, research, or analysis.

6. Non-discrimination

Datasets are available to anyone, with no requirement for registration.

7. Non-proprietary

No entity has exclusive control over the datasets.

8. Licence Free

Datasets are not subject to any copyright, patent, trademark or trade secret regulation.

9. Long Term Preservation of Datasets

Datasets made available online should remain online, with appropriate version-tracking and archiving over time where applicable and available.

10. Usage Costs

Datasets are free-of-charge.

8. Application

This policy applies to all City of Toronto divisions and offices.

9. Definitions

Dataset: means a collection of raw, non-manipulated data usually presented in tabular form with associated metadata, and which is machine readable.

Examples:

What is a raw dataset: a structured file format (including geospatial formats) that can be read by a machine, such as spreadsheets, comma delimited, Extensible Markup Language (XML), or JavaScript Object Notation (JSON).

What is not a raw dataset: a report, a flyer, some web applications, a PDF document, anything that cannot be exported or used by a machine.

Executives: are the City Manager, Deputy City Managers, General Managers, Division heads, City Solicitor, and City Clerk.

Machine Readable Data: means data that, in order to be understood, must be translated by a computer or other type of equipment. Portable document format (PDF) is not machine readable.

Enterprise Information Management: means a set of business processes, disciplines and practices used to manage the information created from an organization's data.

Open Data: is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.

Open Government: is a means to promote transparency, accountability and accessibility of good governance and fosters a culture of collaboration and improved service to the public.

Primary Source Data: original information created or collected by the City, details on how the data was created or collected and the original source documents recording the creation or collection of the data.

10. Roles and Responsibilities

Executives

1. The Executives are accountable for ensuring compliance with this policy.
2. The mandate for information management policy for the organization is assigned to the City Clerk by the City Manager.
3. The City Clerk and Chief Information Officer, designated by Business Advisory Panel as corporate leads for Open Data or their delegates (Executive Director, CIMS and Director, Strategic Planning and Architecture, Information & Technology) are jointly responsible for Open Data awareness, training and issue resolution.
4. Executives are responsible to provide final approval to release datasets for publication and ensure the preservation and access to all datasets. Executives shall formally respond to the Open Data Team with their articulated, documented reasons for not providing data to the Open Data Team.

5. Executives will determine the frequency at which published datasets are reviewed and updated, and communicate these schedules to the Open Data Team.
6. City Clerk's Office and Information & Technology Division are responsible for maintaining the Open Data Licence.

Open Government Committee

7. The Open Government Committee will provide governance for open data and oversight for the Open Data Program.

Open Data Team

8. Open Data Team includes staff from the City Clerk's Office and Information & Technology Division whose mandate is to assess, prioritize, release and monitor datasets in accordance with this policy.
9. The Open Data Team will work with Executives and their staff to identify and assess datasets for publication, assist Division staff in the completion of the Open Data Approval to Publish form, release the datasets under www.toronto.ca/open and, review datasets against the City's privacy protection requirements.
10. If the Open Data Team cannot resolve Executive and Divisional staff non-compliance with the Open Data Policy, it shall forward its concerns to the Executive Director, CIMS and Director, Strategic Planning and Architecture, Information & Technology for issues resolution.
11. The Open Data Team shall monitor community feedback, community requests for datasets and where possible, the applications developed from City of Toronto datasets.

11. Compliance

Where Executives determine they cannot comply with their roles and responsibilities outlined in the Open Data Policy they shall bring their non-compliance issues to the Open Government Committee for review. The Chair of the Open Government Committee will recommend to the City Manager an agreed upon course of action.

12. Authorities

City of Toronto Act, 2006
Municipal Freedom of Information and Protection of Privacy Act
Personal Health Information Protection Act, 2004
City of Toronto Municipal Code, Chapter 169

13. Applicable Policies and Resources

- I&T Policies (<http://insideto.toronto.ca/itweb/policiesstandards/index.htm>)
- Access & Privacy
(http://wi.toronto.ca/intra/clerks/cco_policies.nsf/IntranetBrowse?OpenView)
- Records and Information Management
(http://insideto.toronto.ca/records_archives/pdf/records_management.pdf)
- Open Data Licence
(<https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-licence>)
- Sunlight Foundation Open Data Principles
(<http://sunlightfoundation.com/policy/documents/ten-open-dataprinciples>)

14. Approved by:

Joseph P. Pennachetti
City Manager
Date: November 22, 2011

15. Policy Approval and Review

This policy will be reviewed yearly or sooner if necessary. Approval follows the process in effect at the time of review.

16. Appendix A

The “Ten Principles for Opening Up Government Information” referenced from (<http://sunlightfoundation.com/policy/documents/ten-open-dataprinciples>) are as follows:

1. Completeness

Datasets released by the government should be as complete as possible, reflecting the entirety of what is recorded about a particular subject. All raw information from a dataset should be released to the public, except to the extent necessary to comply with federal law regarding the release of personally identifiable information. Metadata that defines and explains the raw data should be included as well, along with formulas and explanations for how derived data was calculated. Doing so will permit users to understand the scope of information available and examine each data item at the greatest possible level of detail.

2. Primacy

Datasets released by the government should be primary source data. This includes the original information collected by the government, details on how the data was collected and the original source documents recording the collection of the data. Public dissemination will allow users to verify that information was collected properly and recorded accurately.

3. Timeliness

Datasets released by the government should be available to the public in a timely fashion. Whenever feasible, information collected by the government should be released as quickly as it is gathered and collected. Priority should be given to data whose utility is time sensitive. Real-time information updates would maximize the utility the public can obtain from this information.

4. Ease of Physical and Electronic Access

Datasets released by the government should as accessible as possible, with accessibility defined as the ease with which information can be obtained, whether through physical or electronic means. Barriers to physical access include requirements to visit a particular office in person or requirements to comply with particular procedures (such as completing forms or submitting FOIA requests). Barriers to automated electronic access include making data accessible only via submitted forms or systems that require browser-oriented technologies (e.g., Flash, Javascript, cookies or Java applets). By contrast, providing an interface for users to download all of the information stored in a database at once (known as "bulk" access) and the means to make specific calls for data through an Application Programming Interface (API) make data much more readily accessible. (An aspect of this is "findability," which is the ability to easily locate and download content.)

[Note: The City of Toronto has renamed this principle as "Accessibility"]

5. Machine readability

Machines can handle certain kinds of inputs much better than others. For example, handwritten notes on paper are very difficult for machines to process. Scanning text via Optical Character Recognition (OCR) results in many matching and formatting errors. Information shared in the widely-used PDF format, for example, is very difficult for machines to parse. Thus, information should be stored in widely-used file formats that easily lend themselves to machine processing. (When other factors necessitate the use of difficult-to-parse formats, data should also be available in machine-friendly formats.) These files should be accompanied by documentation related to the format and how to use it in relation to the data.

[Note: The City of Toronto has renamed this principle as "Machine Readable"]

6. Non-discrimination

"Non-discrimination" refers to who can access data and how they must do so. Barriers to use of data can include registration or membership requirements. Another barrier is the uses of "walled garden," which is when only some applications are allowed access to data. At its broadest, non-discriminatory access to data means that any person can access the data at any time without having to identify him/herself or provide any justification for doing so.

7. Use of Commonly Owned Standards

Commonly owned (or "open") standards refers to who owns the format in which data is stored. For example, if only one company manufactures the program that can read a file where data is stored, access to that information is dependent upon use of the company's processing program. Sometimes that program is unavailable to the public at any cost, or is available, but for a fee. For example, Microsoft Excel is a fairly commonly-used spreadsheet program which costs money to use. Freely available alternative formats often exist by which stored data can be accessed without the need for a software licence. Removing this cost makes the data available to a wider pool of potential users.

[Note: The City of Toronto has renamed this principle as "Non-proprietary"]

8. Licensing

The imposition of "Terms of Service," attribution requirements, restrictions on dissemination and so on acts as barriers to public use of data. Maximal openness includes clearly labeling public information as a work of the government and available without restrictions on use as part of the public domain.

[Note: The City of Toronto has renamed this principle as "Licence Free"]

9. Permanence

The capability of finding information over time is referred to as permanence. Information released by the government online should be sticky: It should be available online in archives in perpetuity. Often times, information is updated, changed or removed without any indication that an alteration has been made. Or, it is made available as a stream of data, but not archived anywhere. For best use by the public, information made available online should remain online, with appropriate version-tracking and archiving over time.

[Note: The City of Toronto has renamed this principle as "Long Term Preservation of Datasets"]

10. Usage Costs

One of the greatest barriers to access to ostensibly publicly-available information is the cost imposed on the public for access--even when the cost is *de minimus*. Governments use a number of bases for charging the public for access to their own documents: the costs of creating the information; a cost-recovery basis (cost to produce the information divided by the expected number of purchasers); the cost to retrieve information; a per page or per inquiry cost; processing cost; the cost of duplication etc.

Most government information is collected for governmental purposes, and the existence of user fees has little to no effect on whether the government gathers the data in the first place. Imposing fees for access skews the pool of who is willing (or able) to access information. It also may preclude transformative uses of the data that in turn generates business growth and tax revenues.

[Note: The City of Toronto, at this time, has no intention of charging for datasets]