Component of Statistics Canada Catalogue no. 11-522-X
Statistics Canada's International Symposium Series: Proceedings

Article

Symposium 2008:
Data Collection: Challenges, Achievements and New Directions

Use It or Lose It: Mining Municipal Administrative Data

by Joshua Bates, Pat Steenberg, Jacob Ritchie,
and Sarah Cannon

2009

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2008
Data Collection: Challenges, Achievements and New Directions

Use It or Lose It: Mining Municipal Administrative Data

Joshua Bates[1], Pat Steenberg[2], Jacob Ritchie[3], and Sarah Cannon[4]

Abstract

The Federation of Canadian Municipalities' (FCM) Quality of Life Reporting System (QOLRS) is a means by which to measure, monitor, and report on the quality of life in Canadian municipalities. To address that challenge of administrative data collection across member municipalities the QOLRS technical team collaborated on the development of the Municipal Data Collection Tool (MDCT) which has become a key component of QOLRS' data acquisition methodology. Offered as a case study on administrative data collection, this paper argues that the recent launch of the MDCT has enabled the FCM to access reliable pan-Canadian municipal administrative data for the QOLRS.

Key Words: Municipal, Administrative, Data, Web tool, QOLRS.

1. Introduction

1.1 Description

Led by the Federation of Canadian Municipalities (FCM), the Quality of Life Reporting System (QOLRS) is a means by which to measure, monitor, and report on the quality of life in Canadian municipalities. QOLRS indicators are structured into ten thematic domains, each domain including between 4 and 10 indicators of change (see Table 1.1-1). These data and the processes used to capture them on a national scale are the focus of this paper.

Starting with 16 municipalities in 1999, the QOLRS has grown to include 23 municipalities, comprising some of Canada's largest urban centres and many of the suburban municipalities surrounding them. The QOLRS marks the first time that municipal governments have worked together to develop a national policy and planning system for quality of life issues. Using data that are both credible and comparable across cities, the QOLRS provides a snapshot of the critical social, economic and environmental issues and intervention outcomes that affect quality of life, and enables municipalities to track these over time. Within this context, the QOLRS provides the municipal sector with an important instrument to build its case for increased federal investment.

Collecting and consolidating data from a variety of different municipal data sources poses a significant challenge. To address that challenge, members of the QOLRS technical team collaborated on the development of the Municipal Data Collection Tool (MDCT) which has become a key component of QOLRS' data acquisition methodology. The MDCT is a collaborative, on-line database that allows QOLRS members to accumulate and share administrative data in areas of municipal responsibility.

Offered as a case study on administrative data collection, this paper argues that the recent launch of the MDCT has enabled the FCM to access, for the first time, reliable pan-Canadian municipal administrative data for its reporting system. The paper looks at the context and development of the MDCT, including past collection efforts, collection process design, application design and development, types of municipal administrative data collected and some of the benefits provided by the new QOLRS collection tool.

[1] Joshua Bates, Federation of Canadian Municipalities, 24 Clarence Street, Ottawa, Ontario, Canada, K1N5P3
[2] Pat Steenberg, Canadian Council on Social Development, 190 O'Connor Street, Ottawa, Ontario, Canada, K2P2R3
[3] Jacob Ritchie, Acacia Consulting & Research, 430 Parkdale Avenue, Ottawa, Ontario, Canada, K1Y1H1
[4] Sarah Cannon, Halton Region, 1151 Bronte Road, Oakville, Ontario, Canada, L6M3L1

Table 1.1-1
QOLRS domains and indicators

Demographic & Background Information (DBI)	Affordable, Appropriate Housing (AAH)	Civic Engagement (CE)	Community & Social Infrastructure (CSI)	Education (ED)	Employment & Local Economy (ELE)	Natural Environment (NE)	Personal & Community Health (PCH)	Personal Financial Security (PFS)	Personal Safety (PS)
DBI1 Population Growth	AAH1 Rental Housing Affordability	CE1 Voter Turnout	CSI1 Social Housing Waiting Lists	ED1 Education Levels	ELE1 Business Bankruptcies	NE1 Air Quality	PCH1 Low Birth Weight Babies	PFS1 Families Receiving Social Assistance	PS1 Young Offenders
DBI2 Household Size	AAH2 Homeowner Affordability	CE2 Women in Municipal Government	CSI2 Rent-Geared-to-Income Housing	ED2 Literacy Levels	ELE2 Consumer Bankruptcies	NE2 Commuting Distance	PCH2 Teen Births	PFS2 Families Receiving EI	PS2 Violent Crimes
DBI3 Family Composition	AAH3 Core Housing Need	CE3 Volunteering	CSI3 Subsidized Child Care	ED3 Classroom Size	ELE3 Hourly Wages	NE3 Mode of Transportation	PCH3 Premature Mortality	PFS3 Incidence of Low Income Families	PS3 Property Crimes
DBI4 Average Income	AAH4 Substandard Units	CE4 Charitable Donations	CSI4 Community & Social Services Occupations	ED4 Composite Learning Index	ELE4 Change in Family Income	NE4 Density	PCH4 Infant Mortality	PSF4 Children Living in Poverty	PS4 Criminal Code Offences
DBI5 Renters & Owners	AAH5 Changing Face of Homelessness		CSI5 Recreation Facilities	ED5 Educational Occupations	ELE5 Building Permits	NE5 Water Consumption	PCH5 Body Mass Index	PFS5 Income Gap	PS5 Police Per Capita
DBI6 Population Mobility	AAH6 Vacancy Rates		CSI6 Cultural Facilities		ELE6 Unemployment	NE6 Wastewater Treatment	PCH6 Smoking Status	PFS6 Social Assistance Rates	PS6 Weapons Violations
DBI7 Immigration	AAH7 Rental Housing Starts		CSI7 Long Term Care Facilities		ELE7 Recent Immigrant Unemployment	NE7 Waste Diversion	PCH7 Life Expectancy	PFS7 Working Poor	PS7 Drug Violations
DBI8 Language Spoken at Home	AAH8 Monthly Rent		CSI8 Recreation Programs		ELE8 Quality of Employment	NE8 Recreational Water Quality	PCH8 Physical Activity	PFS8 Community Affordability	PS8 Drug Use
DBI9 Visible Minorities	AAH9 Cost of Housing		CSI9 Libraries		ELE9 Labour Force Replacement	NE9 Drinking Water Quality	PCH9 Prevalence of Asthma		PS9 Traffic Incidents
DBI10 Aboriginal Population	AAH10 Overcrowding		CSI10 Access to Health Care Professionals			NE10 *(Ecological Footprint)*	PCH10 Mental Health		PS10 Emergency Services Occupations

1.2 Context

QOLRS publications are unique in that they identify and explore issues and challenges facing Canadian cities with particular emphasis on ten key components — natural environment, affordable/appropriate housing, employment/local economy, civic engagement, community social infrastructure, crime/safety, transportation, education, personal/community health, and personal financial security. The QOLRS is notable for its national scale and scope and for enabling municipal level longitudinal analysis.

The QOLRS has been able to develop a suite of QOL indicators for Canada that otherwise would not exist. Moreover, the QOLRS network has collectively agreed on a common methodology for collecting, reporting and analyzing data in order to produce indicators that are comparable across urban areas. Criteria include:

- Data reliability: extent to which statistical sources are reliable, presence/absence of time-series data, size of survey sample (was large enough to cover all QOLRS member municipalities) and availability/lack of appropriate geography (municipal jurisdiction vs. Census Metropolitan Areas).
- Data consistency: comparability across 7 provinces and 23 jurisdictions, determining the highest quality common denominator to enable reporting across jurisdictions.
- Policy relevance: ability of indicators to "tell a story", bearing on policies that affect quality of life and have a degree of relevance to FCM member jurisdictions.
- Uniqueness: each QOLRS indicator should be substantively unique.
- Simplicity (Parsimony): least number of variables that tell the story and can be understood clearly by the reader.

2. Administrative data

2.1 Sources of municipal administrative data

National government agencies compile and disseminate a wide variety of locally-gathered administrative data. For example, Statistics Canada looks to municipal governments to provide data on building permit activity, waste management, and infrastructure inventories. Environment Canada surveys municipalities every three years to populate its Municipal Water & Wastewater Use Database (MUD). Human Resources & Social Development Canada collects data from individual shelters across the country for its Homeless Individuals & Families Information System (HIFIS). The QOLRS uses these datasets wherever possible, that is, where they have been disaggregated at the municipal level. Moreover, by collecting and making available supplementary data in the same areas, the QOLRS helps to compensate for some of their limitations, such as time-insensitivity.

The QOLRS uses a wide variety of municipally- and nationally-collected administrative data to support its indicators including income tax filer data, immigrant landing data, recreational usage, homeless shelter data, real estate sales data, building permit data, and police data. Other potentially rich sources of these types of data exist, but for a variety of reasons are either inaccessible or unusable. For example, while hospitals and primary health care centres collect a huge amount of public health data, access is generally restricted to users in the health care system. Procurement of local school board data requires the coordination of multiple school boards, across jurisdictions, making it difficult to ensure data consistency and to organize considering resource and time constraints.

Administrative data is a rich source of information for municipal planners and researchers. However, while massive amounts are collected, much sits unanalyzed and/or unnoticed in databases, either because the data were not collected 'for research purposes' or because the host organization simply has no capacity for analysis.

3. Quality of Life Reporting System

3.1 Past QOLRS data collection process

In 2003, the QOLRS surveyed its member municipalities to acquire a variety of municipal administrative data, including data on homelessness, social housing, child care, solid waste, recreational water quality and elections. The information was used to prepare the Quality of Life Report for the 19 participating municipalities.

The data collection instrument—a 40-page Microsoft Word document—was developed by the QOLRS Municipal Survey Working Group and consisted of 10 unique sub-surveys, each corresponding to a particular area of municipal responsibility and requiring input from multiple departments. Each participating QOLRS municipality had at least one representative on the technical team responsible for ensuring the data were collected and returned to FCM in a timely manner; they administered the survey in their own municipalities. To accomplish this, they were required to liaise with internal staff, other levels of municipal government and other organizations.

Each designated respondent received the entire survey, but were, for the most part, instructed to fill out only the sections relevant to their own department. These multiple versions of the completed survey were then returned electronically—or in some cases in hard copy—to the administrator, to be compiled and returned to FCM.

Administering the survey created a great deal of work for QOLRS technical team members. Locating the right data sources, answering questions about the system and compiling responses required an enormous amount of administrator time.

The sheer length of the survey was not in line with the QOLRS criterion of simplicity/parsimony. Many survey questions were too complex, making it next to impossible to disaggregate the responses into discrete pieces of information. All entries had to be cut and pasted into a central file and answers to the surveys that were completed by hand had to be manually entered into the system. Lack of attention to detail at any point throughout this painstaking process potentially led to errors in data entry. Moreover, everyone recognized that poor local collection would jeopardize the quality of the entire survey.

The collection and compilation process took place between April and September and was followed by four months of data validation by researchers at FCM who followed up with data providers by phone or email to verify their responses. Once validated, survey results were used to populate the QOLRS indicator tables. The review process raised several important questions about the survey results. Reviewers questioned (1) to what extent the data sources were reliable; (2) to what degree results were informative, given the absence of time-series data; (3) whether the survey sample was adequate; and (4) whether or not the data were provided at the most appropriate level(s) of geography (CD/CSD vs. CMA).

It was agreed that indicators with limited policy relevance for municipal government would be dropped from the QOLRS list. Only those indicators that could “tell a story” would be included.

Three survey sections — planning, education and community newspapers — were ultimately removed because of poor data quality. In the area of education, the collection team was unable to obtain responses from a critical mass of local school boards in the larger municipalities; in many cases, data were provided only by the smallest boards. Populating the community newspaper indicator was problematic, because there was no universally acceptable definition of “community newspaper” and only limited availability of reliable circulation numbers.

3.2 The 2008 Municipal Data Collection Tool (MDCT)

Many of the themes and questions included in the 2003 survey were incorporated into the 2008 Municipal Data Collection Tool (MDCT), including sections on homelessness, social housing, child care, solid waste, recreational water quality and elections. Given an increase to 23 municipalities, and the large number of target departments within those municipalities, QOLRS decided to administer the 2008 survey using a web-based tool. This significantly reduced the amount of time and effort required of the QOLRS team.

Early in the development stage, it became clear that "Municipal Survey" no longer accurately described the instrument being created and that the real goal of the MDCT was to create a living 'archive' of municipal administrative data on designated QOL reporting topics, for all QOLRS municipalities, that was immediately accessible and updatable as new data became available.

The new collection instrument would be required to:

- Identify which member within a given municipality entered the data online.
- Make municipal collection collaborative by enabling different individuals within any member municipality to view and modify the input of other individuals within their municipality.
- Accept input from organizations associated with an individual member municipality and integrate those with data entered by the municipality.
- Differentiate data entered by lower tier municipalities from those entered at the regional level—e.g. questions on municipal elections and lower level services cannot be answered by regional government staff. (Eight QOLRS members include lower-tier municipalities.)
- Both segregate and integrate data from different sources.
- Enable monitoring of data input and MCDT functionality by a third-party technical consultant.
- Make data from previous surveys accessible.

To meet these requirements, the MCDT:

- Assigns unique passwords to each of the QOLRS member municipalities. Outside organizations associated with that municipality (e.g. emergency shelters, social housing providers, provincial government departments) were provided with the same password.
- Directs lower-tier municipalities to the section containing questions uniquely directed at them.
- Identifies the input source using the email address and telephone phone number of the individual entering the data.
- Permits authorized third-parties to access (in his case the FCM consultants) to all databases and the capacity to cross-reference email addresses/phone numbers with data entry.
- Assures quality control by encouraging collaborative peer review among colleagues within a municipality and through consultant oversight.
- Permits authorized outside (i.e. non member municipality) users to incorporate data into the database.

Future development of the MDCT will enable all QOLRS municipalities to view (but not modify) the compiled data sets of all participating members. This will give them the information required to compare data across municipalities.

As no 'off-the-shelf' product could provide the required functionality, a custom product was developed under the direction of the FCM QOLRS technical team, incorporating server side scripting and database technology. A MySQL server database stores all the data for the MDCT on a web server, while PHP scripting performs the necessary logic for the users to manipulate and access data. Customization allowed the MDCT to look, feel, and operate in a manner that satisfied the largest number of municipal participants, which in turn increased the response rate. Feedback from beta testing helped to maximize functionality (minimize administrative workload) and ensure that questions were simple and commonly understood (ensure comparability across municipalities, while maintaining relevance to local municipality).

4. Data collected

The MCDT collects data on the following issue areas:

4.1 Homelessness

Specifically the number of permanent beds in emergency shelters by shelter type and the number of users, including youth, families, single men, single women, or couples, and combinations of these demographics. Throughout the member municipalities this data was either already being collected by the municipality or was available through the local HIFIS administrator. Where this is not the case, the MDCT relies on administrative reports produced locally by provincial governments or the emergency shelters.

4.2 Social housing

Specifically, we collect data on the total number of social housing units, subsidized (RGI) units and households on waiting lists; minimum and maximum wait times; total number of households placed in social housing during the year; and average age of social housing stock. This information is typically collected by local social housing registries; where this is not the case, it is usually provided by provincial housing ministries.

4.3 Child care

Municipal social services departments are able to provide data regarding the number of spaces available and the number of these spaces which are subsidized. This information will be used to accurately quantify the availability of child care in Canadian cities and, when combined with population statistics for children, can give a picture of the quality of coverage over time.

4.4 Recreation

Data regarding the quantity of recreation facilities are collected, focusing on the 6 most common facilities and programs. This section is intended to assess (and compare) the level of public accessibility to municipal recreational facilities and programs. The intent is not to produce a comprehensive inventory of recreational facilities but to use common facilities as a bellwether of recreation in the municipalities surveyed. These are: indoor pools, outdoor pools, indoor Arenas, outdoor natural or artificial ice rinks, indoor gymnasiums, and outdoor sports fields.

4.5 Culture

This section is similar to the recreation section in that 7 common types of cultural facilities are included and intended to act as a bellwether of cultural facilities throughout the municipalities in questions. The seven types of cultural facilities are: museums, art galleries, theatres, artist studios (live-work or work), indoor performance spaces, designated heritage buildings, and public art installations. Information compiled is used to gauge the level of access to municipal arts and culture facilities in QOLRS municipalities, as well as the level of municipal investment in arts and culture.

4.6 Long term care

Data collected in this section enable municipalities to measure the total number of long term care beds from all service providers (public and private, municipal and non-municipal) as a percentage of the population aged 75 years and over. Data collected includes only the total number of long term care beds available in each year. This data is available in most QOLRS municipalities through Health Services departments but in some cases this is a provincial concern.

4.7 Solid waste

Data collected in this section includes tonnes of residential solid waste, recyclable waste, organic waste, and hazardous waste collected. This data is used to calculate residential recycling rates and organic separation rates. A secondary question gauges the use of captured landfill gas across QOLRS municipalities. Municipal waste services departments are tapped to contribute this data to the MDCT.

4.8 Recreational water quality

This section asks members to contribute the number of days, per serviced water body, that municipally-serviced water bodies are closed. This indicator is intended to give insight into the effect human settlement has on water body health around Canadian Municipalities, although it is understood that this is a slightly simplistic view of the issues around water quality. As with Long Term Care, this data is available in most QOLRS municipalities through Health Services departments but in some cases this is a provincial concern.

4.9 Municipal elections

Data collected in this section documents participation rates in municipal elections; the QOLRS uses this data as part of the Civic Engagement theme. Additionally there is one indicator of female representation on municipal councils. The indicators include data for both upper-tier and lower tier municipal elections.

5. Collection process

The QOLRS grew out of the desire to build and strengthen a national network of municipal governments and the MDCT exemplifies its philosophy of partner engagement. The QOLRS network, in turn, is connected to a variety of other community networks of individuals and organizations working in the relevant issue-areas.

The MDCT success depends, in large part, on the relationships that exist between municipal staff and their counterparts in emergency shelters, social housing registries, provincial government departments and so forth. One of the principal strengths of the MDCT and the QOLRS is the existence of these positive and dynamic local networks that bring municipal staff together with a diversity of local organizations. QOLRS Technical Team Representatives serve as MDCT Coordinators within their respective communities. As such, they are responsible for the following tasks:

• Contacting the individuals within their municipal administration and in outside organizations who are required to respond to some or all MDCT questions and providing them with log-in information;
• Serving as the principal point of contact for those participating in the data collection exercise;
• Monitoring the MDCT to ensure accurate and complete responses are provided and contacting the MDCT administrator, as required;
• Notifying the MDCT administrator when individual sections of the MDCT are complete; and
• Following up directly with respondents to seek clarification on responses.

In its most recent iteration, the MCDT is no longer seen as a ‘one-time’ survey, with a specified timeframe and reporting deadlines. Instead, the MDCT functions as a database that can be added to, or amended, on an ongoing basis. However, everybody understands that there will be occasions when stricter deadlines must be imposed, for example, in anticipation of a QOLRS thematic report, such as the one on homelessness. At such times, additional resources will be invested to accelerate the entry of required data.

6. Outcomes

The 2008 MDCT was a vast improvement over the 2003 Municipal Survey. It reduced the amount of effort required from the QOLRS municipal team and vastly improved the data collection process. Improvements in the questionnaire’s structure and data collection process led to improved reliability of the data collected.

6.1 Administrative benefits

Within the MDCT centralized data entry platform, each QOLRS municipality had a unique username and password; this was used by both municipal staff and staff of participating outside organizations or agencies. Respondents were

directed to a secure site where answers could be entered (and seen) by everyone using the same username and password. The ability to view answers provided by one's colleagues enabled respondents to see whether a section had already been completed, thus avoiding multiple answers to the same question.

Moreover, the tool was designed to allow users to navigate quickly to the appropriate section, thus eliminating the need to read through the entire survey. However, while users could see and modify data entered by others using the same password and user name, they could not see or edit data entered by other municipalities. Database managers at FCM are the only entities with the right to see and alter all data. Ultimately, the MDCT will allow any municipality to view, but not alter, the tabulated data of all other participating municipalities.

Significant benefits have been realized at the compilation stage. As data are entered into the MDCT, they are automatically compiled and stored in a central database. This eliminates errors caused by entering data provided in hard-copy and by the need to consolidate answers from many electronic files into a single file. Not only did this save a tremendous amount of time, it enabled MDCT administrators to track which parts of the survey had been completed, without having to wade through multiple copies of a lengthy word document.

6.2 Data benefits

While Canadian municipalities face many of the same challenges, they all function in different ways. Therefore, the administrative data collected from any two municipalities will differ, however slightly, in its collection methodology and how it is defined. Differing terminology and methodology among municipalities can be addressed through survey design. Multiple terms can be used to refer to a particular item of solicited information and respondents can be encouraged to explain their terminology and methodologies. In the MCDCT, respondents use comment boxes to clarify their definitions. While considerable effort was made to provide very clear instructions about the type of information required, the persistence of definitional differences is inevitable. For example:

Child care: some measure number of children receiving subsidized childcare, others the number of subsidized childcare spaces.

Recreation: User fees for municipal recreational facilities vary from annual family membership for all facilities, to annual family membership for specific facilities only, no family membership, no annual membership, monthly only, etc.

Culture: The definition of 'cultural facility' varies widely.

Where definitional differences are deemed not to affect the meaning of the response, answers are retained with explanatory footnotes. In cases where differences in definition are considered too great, responses are invalidated.

Reaching a consensus on data concept definitions is a complicated process, requiring a great deal of debate and compromise on the part of QOLRS municipalities. However, each recognized that, in the end, the data must be useful to all of them.

In short, the MDCT is a tool that enables the QOLRS to accumulate and consolidate similar data from dissimilar databases. The tool's conception, design and implementation would not have been possible without intensive and ongoing collaboration among members of QOLRS team. The MDCT Technical Working Group, made up of representatives from the City of Calgary, Halifax Regional Municipality, Peel Region, Halton Region, City of Ottawa, Niagara Region, City of Saskatoon, City of Hamilton, and City of Toronto, oversaw every aspect of the MDCT development.

The team's familiarity with municipal planning processes and with their local communities, ensured they knew what indicators were significant, what information should be tracked, who collected that information and where and how that information was stored. Moreover, they were able to agree on how they would obtain the required information (question content) and how it would be characterized (what the variables would be called), as well as on tool functionality and graphic design. The leadership of the FCM was also critical to its success.

While the MDCT was designed to accommodate existing municipal administrative data collection practices, one hopes that, over the longer term, it will begin to reconcile administrative data definitions and collection methodologies across organizations, jurisdictions and levels of government.