

Managing the Quality of Geolocation Data

Background

Access to geolocation data, such as the exact location of a sewer line, with additional information, such as the date of its last inspection and an indication of its structural condition, is an asset for project planning and management at the Ville de Montréal (the City). In a survey conducted as part of our work, 89% of respondents said they use geolocation data recorded in the City's SIGS (Système d'information géographique et spatiale). However, this data must be of good quality, i.e., complete, accurate and up to date. The absence of such quality assurance for geolocation data could result in users turning to other data sources, leading to an inefficient approach. Worse yet, they could base their decisions on inaccurate information.

At the time of our audit, the SIGS contained 366 data layers, consisting of datasets on different themes, including water and sewer systems, the road network and bike paths, road signage, the municipal real estate inventory (e.g., buildings, lots, park benches, garbage cans), the inventory of parks and natural areas, as well as the network of electric vehicle charging stations.

Purpose of the Audit

The purpose of this audit was to ensure that the City's geolocation data is of good quality and that it is made available to all of the business units.

Results

Due to several deficiencies in the governance of the geolocation data, including with respect to the assignment of roles and responsibilities, the lack of defined minimum data quality criteria, and incomplete attributes for the geolocation data, we conclude that not all of the geolocation data made available to City employees as part of their operations and activities is of good quality. In addition, due to the lack of an inventory of all geolocation data, not all of the data is known and available to employees.

The *Directive sur la gouvernance des données de la Ville de Montréal*, in force since 2016, tends to be oriented more towards the dissemination of open data accessible on the City's website. This leaves the business units to decide how to treat the geolocation data throughout its processing cycle for dissemination to the SIGS. There is no single, formal structure for ensuring and communicating the quality of geolocation data available to the employees in the SIGS. Adjustments are necessary, including the development of an administrative framework specific to the geolocation data that defines the roles and responsibilities of the various business units involved in processing geolocation data and the quality criteria that must be met for this type of data. Finally, the SIGS needs to be cleaned up, and only the necessary layers kept.

Main Findings

Governance

- The roles and responsibilities of the various business units involved in the geolocation data processing cycle are not perfectly aligned with the *Directive sur la gouvernance des données de la Ville de Montréal*, particularly with respect to responsibility for data quality.
- The person responsible for the geolocation data layer is unknown for 60% of the layers in the SIGS making it difficult to determine its utility.
- There is no comprehensive documentation describing the process for creating a geolocation data layer, from the collection of the data to its dissemination.
- Approximately 4% of geolocation data layers contain metadata, i.e., global information that applies to all data—thereby limiting the ability to know who is responsible for the layer and how often it is updated.

Generation, Update and Dissemination Ensuring the Quality of Information

- There are no minimum quality criteria that must be met for the geolocation data, which may limit users' confidence in the data and encourage the use of parallel databases.
- The absence of a requirement to document the geolocation data processing cycle is not conducive to knowledge of all the steps leading to its dissemination and thus limits quality controls during this process.
- Regarding the geolocation data layers examined, the correspondence between the data collected by the business units responsible for the activity or asset and the data available in the SIGS is not perfect. The SIGS is therefore not an accurate representation of reality, and any decision based on the data could be wrong.
- For all of the geolocation data layers examined, not all data attributes contain values, thus limiting the usefulness of viewing such geolocation data.
- Automated scheduling of a data display update in SIGS is not performed for all observed data layers. In the case of two of the five layers examined, the updates dated back to 2012 and to 2016. Consequently, the risk that the data is not up to date is significant.

Accessibility of the Geolocation Data

- There is no catalogue available to enable all of the employees to inquire about the existence of geolocation data available to the City.

In addition to these results, we have made various recommendations to the business units, which are presented in the following pages. These business units were given the opportunity to agree to the recommendations.