

Gestion de la qualité des données géolocalisées

Mise en contexte

Accéder à des données géolocalisées, comme l'emplacement exact d'une conduite d'égout, et comportant des informations complémentaires comme la date de sa dernière inspection et un indice sur l'état de sa structure, est un atout pour la planification et la gestion de projets à la Ville de Montréal (la Ville), alors que 89 % des répondants à un sondage réalisé dans le cadre de nos travaux disent utiliser des données géolocalisées consignées au sein du Système d'information géographique et spatiale (SIGS) de la Ville. Toutefois, encore faut-il que ces données soient de qualité, c'est-à-dire complètes, précises et à jour. L'absence d'une telle assurance de qualité envers les données géolocalisées pourrait faire en sorte que les utilisateurs se tournent vers d'autres sources de données, entraînant une démarche inefficace, ou pire qu'ils appuient leurs décisions sur la base d'informations inexacts.

Au moment de nos travaux, le SIGS contenait 366 couches de données, soit des ensembles de données sur des thématiques différentes, dont les réseaux d'aqueduc et d'égout, le réseau routier incluant les pistes cyclables, la signalisation routière, l'inventaire immobilier municipal (p. ex. les bâtiments, les terrains, les bancs de parc, les poubelles), l'inventaire des parcs et des milieux naturels et le réseau de bornes de recharge pour les véhicules électriques.

Objectif de l'audit

Le présent audit avait pour objectif de s'assurer que la Ville possède des données géolocalisées de qualité et qu'elles sont mises à la disposition de l'ensemble des unités d'affaires.

Résultats

En raison de plusieurs lacunes dans la gouvernance des données géolocalisées, notamment en lien avec l'attribution des rôles et des responsabilités, avec l'absence de définition de critères minimums de qualité des données et avec des attributs incomplets pour ces données, nous concluons que les données géolocalisées accessibles aux employés de la Ville dans le cadre de leurs opérations et de leurs activités ne sont pas toutes de qualité. De plus, en raison de l'absence d'un inventaire de l'ensemble des données géolocalisées, ces dernières ne sont pas toutes connues et disponibles pour les employés.

La *Directive sur la gouvernance des données de la Ville de Montréal*, en vigueur depuis 2016, tend à être davantage orientée vers la diffusion de données ouvertes lesquelles sont accessibles sur le site internet de la Ville. Les unités d'affaires sont ainsi laissées à elles-mêmes quant à la manière d'encadrer les données géolocalisées tout au long de leur cycle de traitement, en vue de leur diffusion au SIGS. Il n'y a pas de structure formelle et unique permettant d'assurer et de faire connaître la qualité des données géolocalisées mises à la disposition des employés dans le SIGS. Des ajustements sont nécessaires, dont l'élaboration d'un encadrement administratif propre aux données géolocalisées qui définit notamment les rôles et les responsabilités des différentes unités d'affaires impliquées dans le traitement d'une donnée géolocalisée et les critères de qualité devant satisfaire ce type de données. Finalement, un nettoyage du SIGS doit être réalisé pour ne conserver que les couches nécessaires.

Principaux constats

Gouvernance

- Les rôles et les responsabilités des différentes unités d'affaires impliquées dans le cycle de traitement des données géolocalisées ne sont pas parfaitement arrimés avec la *Directive sur la gouvernance des données de la Ville de Montréal*, notamment en ce qui a trait à la responsabilité de la qualité des données.
- Le responsable d'une couche de données géolocalisées n'est pas connu pour 60 % des couches figurant au SIGS rendant ainsi difficiles les démarches pour déterminer son utilité.
- Il n'y a pas de documentation globale décrivant le processus à suivre pour la création d'une couche de données géolocalisées, de la cueillette des données à leur diffusion.
- Environ 4 % des couches de données géolocalisées contiennent des métadonnées, soit des informations globales s'appliquant à toutes les données, limitant ainsi la capacité de savoir qui est responsable de la couche et sa fréquence de mise à jour.

Génération, mise à jour et diffusion assurant la qualité de l'information

- Il n'existe aucun critère de qualité minimale devant être respecté pour les données géolocalisées, ceci peut limiter la confiance des utilisateurs envers les données et favoriser l'usage de bases de données parallèles.
- L'absence d'une exigence pour documenter le cycle de traitement des données géolocalisées ne permet pas de connaître toutes les étapes menant à leur diffusion et limitant ainsi les contrôles qualité durant ce processus.
- Pour les couches de données géolocalisées examinées, il n'y a pas une concordance parfaite entre les données recueillies par les unités d'affaires responsables de l'activité ou de l'actif et les données disponibles dans le SIGS. Ce dernier ne représente donc pas fidèlement la réalité et toute prise de décision basée sur ces données pourrait être erronée.
- Pour toutes les couches de données géolocalisées examinées, les attributs des données ne contiennent pas tous des valeurs, limitant ainsi l'utilité de consulter de telles données géolocalisées.
- La programmation automatisée d'une mise à jour de l'affichage de données dans le SIGS n'est pas effectuée pour toutes les couches de données observées. Pour 2 des 5 couches examinées, la mise à jour datait de 2012 et de 2016. Il y a donc un risque non négligeable que les données ne soient pas à jour.

Accessibilité des données géolocalisées

- Il n'y a pas un catalogue accessible à tous les employés afin que ceux-ci puissent s'enquérir de l'existence des données géolocalisées disponibles à la Ville.

En marge de ces résultats, nous avons formulé différentes recommandations aux unités d'affaires qui sont présentées dans les pages suivantes. Ces unités d'affaires ont eu l'opportunité de donner leur accord relativement aux recommandations.