
15 March 2019

Original: French

United Nations Group of Experts On

Geographical Names

First session

New York, 29 April – 3 May 2019

Item 15 of the provisional agenda

Toponymic data files and gazetteers (data processing and tools, database management, data dissemination: products and services) (Working Group on Toponymic Data Files and Gazetteers)

Rationalisation et normalisation des codes de statuts des toponymes

Submitted by Canada**

* GEGN.2/2019/1

** Prepared by Cindy Doyle, Canada Centre for Mapping and Earth Observation, Natural Resources Canada

Summary **

Le statut d'un toponyme doit indiquer s'il s'agit d'un nom officiel ou s'il s'agit d'un toponyme historique abrogé. Or, bien que la base de données des toponymes du Canada (BDTC) ait été entièrement reconstruite en 2015 dans un nouveau modèle spatial normalisé, il a été impossible, à ce moment, d'éliminer la redondance d'information implicite dissimulée dans les codes de statuts. On parle de redondance lorsqu'une information est enregistrée à deux endroits différents dans la base de données, elle peut mener à des incohérences et même des contradictions si l'information est mise à jour à un seul endroit ultérieurement. La multitude de codes de statuts des toponymes de la BDTC en constituait l'exemple le plus patent, complexifié dans un jargon que seul un initié pouvait décoder. Le secrétariat de la Commission de toponymie du Canada (CTC) s'est attaqué à ce problème en formant un comité chargé de redéfinir la liste des codes de statuts afin qu'ils soient clairs, concis, interopérables et exempts de toute redondance pouvant mener à des contradictions. Cet article décrit les efforts menés par le ministère des Ressources naturelles du Canada (RNCan) pour rationaliser l'ancienne liste de codes lourde et désuète afin de la rendre simple et intuitive tout en facilitant les échanges entre la BDTC et les partenaires provinciaux.

Contexte

La Commission de toponymie du Canada (CTC) est l'organisme national chargé de coordonner les normes et les politiques applicables aux dénominations géographiques au Canada. La Commission est établie par décret fédéral, et se compose de membres issus de ministères et d'organismes des gouvernements fédéral, provinciaux et territoriaux, qui, chacun, exercent des responsabilités particulières en fonction de leur compétence et de leur mandat respectifs. Regroupés au sein d'un organisme intergouvernemental de portée nationale, les membres de la CTC veillent à ce que les noms géographiques soient gérés de manière uniforme au Canada.

La CTC est appuyée par un secrétariat assuré par Ressources naturelles Canada (RNCan), un ministère du gouvernement du Canada. RNCan fournit l'infrastructure et le soutien à la Base de données toponymiques du Canada (BDTC), la base de données nationale des noms géographiques de référence et une composante essentielle de l'Infrastructure canadienne de données géospatiales (ICDG).

La liste des codes de statuts de la BDTC avant et après

Une image vaut mille mots, dit-on. Il suffit de juxtaposer les deux listes (Tableau 1), l'ancienne comprenant 52 codes indéchiffrables et la nouvelle ne comprenant que six codes, pour mesurer le degré de complexité que la classification des toponymes avait atteint.

Tableau 1 - Anciennes et nouvelles listes de codes de statuts

<p>Ancienne liste de codes de statuts</p> <p>A1, A10, A11, A12, A13, A14, A2, A3, A4, A5, A6, A7, A8, A9, A99, B1, B10, B11, B12, B13, B14, B15, B2, B3, B4, B5, B6, B9, C1, C10, C11, C2, C3, C4, C5, C7, C9, D48, D60, D62, E1, E10, E2, E3, E4, E5, E6, E7, G1, M1, M2, P1, Q1, Q2, S1, S2, S3, T1, U1, U2, U3, W1, X1, Y1, Y2</p>
<p>Nouvelle liste de codes de statuts</p> <ul style="list-style-type: none"> - official - previouslyOfficial - pending - alternate - unpublished - forInformation

À moins de disposer du dictionnaire de données permettant de déchiffrer les codes de statuts, il était impossible pour un profane de déterminer si le toponyme était officiel ou historique avec le code de statut de l'ancienne liste.

Comment en était-on arrivé là?

L'origine des codes de statuts de la BDTC

Pour comprendre d'où vient ce jargon, il faut remonter à l'époque où les octets de mémoire étaient une denrée rare et dispendieuse. Lorsque les bases de données étaient assemblées à cette époque, on abrégait tout ce qu'il était possible d'abrégier pour sauver de l'espace. L'exemple le plus courant consistait à ne mettre que les deux derniers chiffres de l'année pour indiquer une date, ce qui avait mené au fameux bogue de l'an 2000. Dans le même souci d'économie d'espace, la plupart des attributs d'un toponyme étaient codifiés pour abrégier l'information qu'ils contenaient et les rendre digestes pour la capacité limitée des ordinateurs de l'époque. Toutefois, l'abréviation des codes de statuts n'explique pas la présence de 52 variantes. La BDTC étant une mosaïque de données qui proviennent de dix provinces, de trois territoires et de quatre agences fédérales, chacun disposant de leur système indépendant, les codes de statuts n'étaient pas nécessairement harmonisés lors de leur importation, multipliant les variantes de codes ayant la même signification. De plus, les premières versions de l'application utilisée pour mettre à jour la BDTC n'étaient pas facilement adaptables ou programmables pour répondre à de nouveaux besoins de gestion. Le secrétariat de la BDTC s'est servi du code de statut en attribuant des codes bidons dans le seul but de les regrouper et de les retrouver facilement dans la base de données. Par exemple, afin de pouvoir retrouver un jeu de données correspondant à un lot de nom téléchargé à une date donnée, un attribut alors non-disponible pour la recherche, on donnait un code de statut fictif tel que D48, le D signifiant un contenu Digital et 48, révélant le code de la province de l'Alberta. Ce code permettait aux toponymistes de retrouver tous les noms téléchargés à une date donnée en utilisant le code du statut D48 dans les critères de recherches. Ces codes manipulés, bien que pertinents et utiles à l'époque, n'avaient aucun rapport avec le statut réel du toponyme et dénaturaient l'attribut censé révéler si le toponyme était officiel ou abrogé. Ces codes étaient généralement utilisés de façon temporaire, mais certains ont sombré dans l'oubli jusqu'à ce que la normalisation de 2018 fasse table rase de ce passé révolu.

Normalisation

Selon [Wikipédia](#), *le but essentiel de la normalisation est d'éviter les anomalies transactionnelles pouvant découler d'une mauvaise modélisation des données et ainsi éviter un certain nombre de problèmes potentiels tels que les anomalies de lecture, les anomalies d'écriture, la redondance des données et la contre-performance.*

Il existe plusieurs niveaux de normalisation et, sans entrer dans les détails, on dit des attributs se conformant à la première forme normale qu'ils sont atomique, c'est-à-dire que la subdivision d'une information n'en apporte pas de supplémentaire ou de complémentaire. Or l'ancienne liste de codes ne respectait pas ce critère. Outre révéler si le toponyme était officiel ou non, il donnait des indications complémentaires sur la provenance ou la localisation géographique du toponyme, sur la nature de l'entité nommée et même sur l'usage du toponyme et ce, en parfaite violation avec le premier niveau de normalisation.

Les deuxièmes et troisièmes formes normales interdisent les dépendances entre les attributs, c'est-à-dire que la valeur d'un attribut ne doit pas dupliquer la valeur contenue dans un autre attribut. Les codes de statuts de l'ancienne liste, en plus de donner trop d'information, dédoublaient des informations existantes dans d'autres attributs comme la province ou le type d'entité créant une dépendance entre eux. Ces attributs devenaient fréquemment incohérents lorsqu'un seul des deux était mis à jour subséquemment.

Par exemple, le code de statut d'un toponyme officiel de réserve indienne était A13. Il pouvait arriver que la réserve indienne perde son statut officiel de réserve mais que le toponyme soit encore en usage par la communauté. Un toponymiste bien intentionné changeait le code générique de la réserve indienne pour celui d'une communauté, mais omettait de changer le code de statut A13, censé être à l'usage exclusif des réserves indiennes. Deux informations contradictoires définissaient le toponyme : son code générique et son code de statut qui stipulaient qu'il s'agissait d'une communauté et d'une réserve indienne, respectivement. Dans un tel scénario, l'utilisateur se trouvait dans l'impossibilité de déterminer lequel du code générique ou du code de statut était valide à moins d'effectuer des recherches dans les décisions. Ces recherches nécessitaient parfois de devoir contacter l'autorité toponymique afin de valider l'information, occasionnant autant de perte de temps. Ce gaspillage de ressources était causé par une mauvaise modélisation des attributs qui permettait à ces incohérences d'exister. Bien que la refonte complète de la BDTC en 2015 ait nécessité de multiples consultations s'échelonnant sur un an, cette liste avait survécu, envers et contre toutes les règles de saine gestion d'une base de données relationnelles.

Ce que dissimulait les codes

On pourrait croire avec le tableau 2 ci-dessous que les codes de statuts étaient organisés et classifiés correctement. Or, ces catégories se recourent de sorte qu'un toponyme se qualifiait parfois dans quatre catégories différentes, rendant l'exercice de classification aléatoire en l'absence de directives claires.

Tableau 2 - Définition de la portion alphabétique du code de statut de l'ancienne liste

Code	Définition
A	Approuvé
B	Non approuvé
C	Jamais approuvé
D	Information Digitale
E	Toponyme du Québec
M	Toponyme du Manitoba
P	Toponyme pancanadien
Q	Nom de L'Ontario
S	Toponyme Statutaire
W	Eaux Internationales approuvées
U	Utilisés dans certains contextes
Y	Toponyme du Yukon

Dans quelle catégorie est classifié un toponyme approuvé du Québec? Un toponyme abrogé de l'Ontario? Un toponyme jumelé du Manitoba reçu en format digital? Impossible de le savoir à moins de détenir le dictionnaire de données qui révélait la définition de chacun des codes et leur mode d'emploi. Ce document tenait sur 15 pages de tableaux et de définitions, nécessitant une longue formation et une bonne dose d'expérience pour apprendre à les utiliser correctement.

Le tableau 3 est un échantillon des principaux codes de statuts de l'ancienne liste avec la définition telle qu'on la trouvait dans le dictionnaire, l'information implicite qu'ils contenaient de même que les codes normalisés de remplacement. Nous verrons ensuite les solutions apportées afin de pallier à la perte d'information résultant de la reclassification des anciens codes de statuts vers les six codes normalisés.

Tableau 3 - Principaux anciens codes de statuts

Code	Définition	Information cachée ou dédoublée contenu dans un autre attribut	Code normalisé (<i>en anglais</i>)
A1	Approuvé		official
A3	Double approbation	Toponyme approuvé par plus d'une autorité	official
A7	Toponyme jumelé	Entité pour laquelle on trouve deux noms officiels	official
A8	Changement de toponyme	Toponyme qui a remplacé un autre nom	official
A10	Toponyme officiel au Québec	Province du Québec	official
A13	Toponyme de réserve indienne	Type d'entité, réserve indienne	official
B2	Toponyme abrogé	Toponyme qui n'a pas été remplacé	previouslyOfficial
B3	Toponyme supprimé	Toponyme publié par inadvertance	forInformation
B4	Toponyme abrogé	Remplacé par un nouveau nom officiel	previouslyOfficial
B5	Toponyme de réserve indienne abrogé	Type d'entité, réserve indienne.	unpublished
B10	Toponyme abrogé au Québec	Province du Québec Toponyme non-remplacé	previouslyOfficial
B11	Ancienne variante diffusé	Province du Québec Toponyme remplacé	previouslyOfficial

Les solutions pour éviter la perte d'information

Dans la plupart des cas, l'information rattachée au code de statut était une information redondante déjà présente dans les autres attributs du toponyme, leur élimination ne causait donc aucune perte d'information. Les deux codes suivants sont des exemples où aucune mesure n'a été prise pour sauver l'information cachée puisque qu'elle était redondante et potentiellement nuisible (Tableau 4).

Tableau 4 - Information redondante non-conservée

Code	Définition	Information cachée ou dédoublée contenu dans un autre attribut	Code normalisé (<i>en anglais</i>)
A10	Toponyme officiel au Québec	Province du Québec	official
A13	Toponyme de réserve indienne	Type d'entité, réserve indienne	official

Le code A10 était attribué aux toponymes officiels du Québec. Or, la province est une information gérée dans l'attribut REGION_CL qui est un champ indexé sur lequel on peut faire des recherches. Aucune action n'a été prise pour préserver cette duplication d'information. On l'a tout simplement remplacé par le nouveau code official qui désigne un toponyme officiellement approuvé par l'autorité toponymique.

Il en va de même pour le code A13, attribué au toponymes de réserves indiennes officiels. Le type d'entité existe aussi dans une liste normalisée sous la forme *IR – Indian Reserve/Réserve Indienne*. De la même façon que les régions, le type d'entité est un attribut indexé sur lequel les recherches sont permises. Un usager qui veut retrouver les toponymes officiels de réserves indiennes n'a qu'à utiliser la combinaison de deux critères de recherches, le statut et le type d'entité.

Il en va autrement pour les cas présentés dans les tableaux 6 et 7 où l'information ajoutée a été jugé non-redondante et suffisamment importante pour la préserver dans des groupes toponymiques. Les groupes toponymiques sont des regroupements de toponymes en divers ensembles pour nous permettre de les retrouver. Ils ont l'immense avantage d'être entièrement automatisés, c'est-à-dire que leur mise à jour s'effectue de façon hebdomadaire par des processus qui analysent les données et reconstituent les groupes sans intervention manuelle.

Tableau 5 - Groupes toponymiques créés pour pallier à la perte d'information lors de la rationalisation des codes de statuts

ID	NAME	DESIGNATION_FR	DEFINITION_FR
1244	jointDecision	Décision Conjointe	Le groupe « <i>Décision Conjointe</i> » est utilisé pour identifier les noms dont l'approbation de plus d'une autorité est requise, i.e. l'autorité provinciale ou territoriale et l'autorité de l'agence fédérale compétente.
1245	multiple	Multiple	Le groupe « <i>Multiple</i> » est utilisé pour identifier les toponymes nommant des entités qui comportent plus d'un nom officiellement approuvé.
1247	rescinded	Abrogé	Le groupe « <i>Abrogé</i> » est utilisé pour identifier les toponymes dont la reconnaissance officielle a été révoquée et qui n'ont pas été remplacés par des nouveaux noms.

Leur second avantage est de permettre l'appartenance d'un toponyme à plusieurs groupes. Là où on devait choisir un seul code de statut pour un toponyme donné, on peut maintenant lui assigner plusieurs groupes. Ainsi, les toponymes français et anglais d'un parc national affectés par une décision conjointe de la province et de l'agence fédérale, peuvent faire partie des groupes *jointDecision* et *multiple*. Le groupe *jointDecision* indique que deux autorités ont émis une décision pour un toponyme alors que le groupe *multiple* indique que le nom officiel n'est pas unique, i.e. qu'il en existe plus d'un dans plusieurs langues. Avec l'ancienne liste de code, il était impossible de collecter toute cette information, le toponymiste était forcé de choisir entre le code A3 – *Double Approbation* ou le code A7 – *Toponyme jumelé* alors que les deux définitions s'appliquaient aux toponymes (Tableau 6). L'attribution du code de statut était inconstante et l'information incomplète.

Tableau 6 - Information préservée dans les groupes toponymiques *jointDecision* et *multiple*

Code	Définition	Information cachée préservée	Code normalisé (en anglais)	Groupe retenant l'information
A3	Double approbation	Toponyme approuvé par plus d'une autorité	official	<i>jointDecision</i>
A7	Toponyme jumelé	Entité pour laquelle on trouve deux noms officiels	official	<i>multiple</i>

Avec la normalisation, le toponymiste n'a plus à se soucier du nombre de noms approuvés ou du nombre de décisions émises par les différentes autorités, il n'a qu'à entrer les décisions et les noms avec leurs statuts officiels, les processus hebdomadaires peupleront les groupes toponymiques de façon automatisée.

Tableau 7 - Information préservée dans le groupe toponymique rescindé (Abrogé)

Code	Définition	Information cachée préservée	Code normalisé (en anglais)	Groupe retenant l'information
B2	Toponyme abrogé	Toponyme qui n'a pas été remplacé	previouslyOfficial	<i>rescinded</i>
B4	Toponyme abrogé	Remplacé par un nouveau nom officiel	previouslyOfficial	N/A
B10	Toponyme abrogé au Québec	Province du Québec Toponyme non-remplacé	previouslyOfficial	<i>rescinded</i>
B11	Ancienne variante diffusé	Province du Québec Toponyme remplacé	previouslyOfficial	N/A

Tous les codes du tableau 7 étaient attribués à des noms historiques. On leur attribuait les codes B2 ou B10 (au Québec) lorsque le nom abrogé n'était pas remplacé par un nouveau toponyme. Les codes B4 ou B11 (au Québec) étaient attribués lorsque les noms abrogés étaient remplacés. Or, le remplacement d'un toponyme pouvait se produire plusieurs années après son abrogation. Lors de l'ajout du nouveau toponyme, il était fréquent d'oublier de mettre à jour le statut du toponyme historique à B4 ou B11. Ce dernier conservait son statut indiquant qu'il n'y avait pas de toponyme l'ayant remplacé alors que c'était faux. La nouvelle liste de codes simplifie ces opérations, un seule code, *previouslyOfficial*, est attribué pour marquer que le toponyme est abrogé indépendamment du fait qu'il soit localisé au Québec et du fait qu'il soit remplacé ou non. Les processus automatisés de mise à jour des groupes toponymiques vont ajouter les noms non-remplacés au groupe *rescinded*. Si, des années plus tard, de nouveaux toponymes de remplacement sont approuvés, aucune action n'est nécessaire pour indiquer que les toponymes abrogés sont dorénavant remplacés. Les processus automatisés vont détecter les nouveaux toponymes partageant l'identifiant unique de l'entité et automatiquement retirer le toponyme historique du groupe *rescinded*.

Conclusion

Cette rationalisation des codes de statuts nous a permis d'éliminer de la redondance, des facteurs potentiels d'incohérence et surtout, d'immenses pertes de temps causées par la complexité du jargon utilisé, tant pour former des employés capables de les utiliser que pour maintenir de façon manuelle une information incomplète. De plus, les nouveaux codes de statuts sont explicites (Tableau 1), permettant au grand public de les comprendre aisément. Nous sommes fiers d'avoir établi ce consensus parmi les membres de la CTC afin de produire cette liste de codes toute simple et parfaitement compréhensible pour quiconque a à utiliser les toponymes du Canada.