



HAL
open science

Intégration et correction automatique de données OpenStreetMap

Matthieu Baley, Guillaume Touya

► **To cite this version:**

Matthieu Baley, Guillaume Touya. Intégration et correction automatique de données OpenStreetMap. Géomatique Expert, 2014, 96, pp.32-43. hal-02288775

HAL Id: hal-02288775

<https://hal.science/hal-02288775v1>

Submitted on 16 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Intégration et correction automatique de données OpenStreetMap

MATTHIEU BALEY, GUILLAUME TOUYA.
LABORATOIRE COGIT, IGN

Avec l'avènement du « web 2.0 », un nouveau phénomène est apparu : le *crowdsourcing* (Tapscott & Williams, 2007). Ce dernier consiste à utiliser les compétences d'un grand nombre d'utilisateurs pour réaliser des tâches à moindre coût. Le succès de la distribution gratuite des logiciels (*Open Source*) a permis d'étendre le mouvement aux bases de données, dont la licence détermine les conditions d'accès et de réutilisation ; d'autre part, les équipements GPS sont désormais abordables et se démocratisent. Ces deux facteurs ont suscité l'apparition de nouveaux producteurs de données, de simples individus, qui aiment partager le fruit de leur travail au sein d'une communauté. Aujourd'hui, le projet *OpenStreetMap* est le premier représentant de ce nouveau type de données, et s'enorgueillit de compter une communauté très active.

Le laboratoire *Cogit* de l'IGN explore l'automatisation des processus de cartographie à partir de bases de données

géographiques vectorielles. Il étudie notamment la généralisation cartographique, qui consiste à réduire automatiquement le niveau de détail des cartes à petite échelle pour en faciliter la compréhension. De plus, il s'intéresse aux problématiques de recherche liées au phénomène des données collaboratives.

Cet article se propose d'étudier la cartographie des données aux niveaux de détail hétérogènes ; il s'agit notamment de trouver un moyen pour harmoniser ces niveaux de détail à l'aide de traitements automatiques. Quand cette harmonisation consiste à omettre certaines données détaillées, on parle de généralisation cartographique, un sujet abondamment traité. Nous allons plutôt ici nous intéresser au cas inverse : l'enrichissement des données les moins détaillées, lorsque l'on souhaite réaliser des cartes à grande échelle, ce que l'on peut reformuler en : « comment réaliser une harmonisation du niveau de détail automatique en vue d'une cartographie à différentes échelles ? »

La cartographie issue de données géographiques collaboratives

Les données géographiques collaboratives

L'« *information géographique volontaire* », plus connue par son acronyme anglais VGI ou *Volunteered Geographic Information*, un terme inventé par Goodchild (2007), désigne une initiative communautaire de cartographie collaborative dans laquelle les communautés en ligne deviennent des producteurs de données géographiques. En français, on peut utiliser le terme « *données géographiques collaboratives* », même s'il n'est pas tout à fait équivalent.

Plusieurs contenus géographiques communautaires existent et disposent de portails *Internet* : *Google Map Maker* (GMM), *Wikimapia* et *OpenStreetMap* (OSM) sont trois exemples de projets exploitant le VGI, dont l'objectif consiste à cartog-

raphier le maximum d'entités géographiques du monde. On les qualifie de « *généralistes* » (Brando, 2013) dans leur contenu car ils ne visent aucune application spécifique *a priori*. Leur modèle de données a tendance à s'adapter aux nouveaux besoins exprimés au fur et à mesure de leur apparition.

SpotCrime et *Ushahidi* constituent d'autres exemples de VGI de types « *non-généralistes* » (Brando, 2013) car ils visent un usage particulier : la cartographie de la criminalité pour *SpotCrime* et la cartographie sociale pour *Ushahidi*. Leur modèle de données est prédéfini pour renseigner tous les attributs des objets pertinents à leur finalité.

L'IGN a, de son côté, mis en place son propre système de données partagées à l'aide d'une plateforme *web RiPart* (Remontée d'Information Partagée). Celle-ci est développée sur la base de l'API *Géoportail* de l'IGN afin de mettre à jour le RGE® (base de données centrale de l'IGN). Ce

projet s'appuie sur des partenariats administratifs et n'est pas ouvert à la communauté, contrairement aux projets précédents. Ce type de collaboration plus fermée se nomme *community sourcing*. Par ailleurs, l'IGN a mis en ligne sur le site *Géoportail* IGN un outil de signalement d'erreurs, d'évolutions ou d'omissions permettant au grand public de participer à la mise à jour du RGE™.

OpenStreetMap

Le projet communautaire *OpenStreetMap* (OSM) n'est plus à présenter. Il a été fondé par l'ingénieur anglais Steeve Coast, qui a été confronté à l'impossibilité d'utiliser les données géographiques existantes de l'*Ordnance Survey* (l'agence cartographique publique anglaise), en raison de restrictions légales ou techniques. La licence choisie pour les données OSM est l'*Open Database License* (ODbL), issue du projet *opendatacommons.org*. Elle permet à chacun d'exploiter publiquement, commercialement ou non, les données à condition

d'y maintenir la licence. Les modifications apportées doivent obligatoirement adopter la licence libre pour diffuser la nouvelle version. La force d'OSM repose sur le nombre important de contributeurs qui n'a cessé d'augmenter depuis sa création : on en recense aujourd'hui plus d'un million.

Cette communauté est connue pour sa réactivité lors de besoins spécifiques. Par exemple, à l'occasion du tremblement de terre en Haïti, la nouvelle carte de l'île a été produite en trois jours (à l'aide de photos aériennes récentes), ce qui a permis d'aider les secours sur place. Plus récemment, on peut retrouver sur OSM la carte d'un camp jordanien de réfugiés où vivent 130 000 Syriens ; sur *Google Maps*, le camp n'est même pas représenté.

Les données importées dans OSM par les contributeurs sont modelées et stockées avec des *tags* de primitives géométriques, de trois types : points ou nœuds des chemins ou polygones et des relations (figure 2). L'élément de base de la structure d'OSM est le nœud ; il est défini par sa latitude et sa longitude. Un point d'intérêt (POI), par exemple, est rattaché à un nœud (cabine téléphonique, arrêt de bus...) ; plusieurs nœuds reliés entre eux forment un chemin. Un chemin ouvert représente une ligne (route, chemin de fer...) alors qu'un chemin fermé correspond à une surface (étang, forêt...) ou à une ligne fermée (rond-point). Un élément de regroupement d'objets formant un autre objet définit une relation. Il existe différents types de relations ; par exemple, une piste cyclable est la relation de plusieurs chemins, une limite administrative est la relation d'une ligne de type frontière.

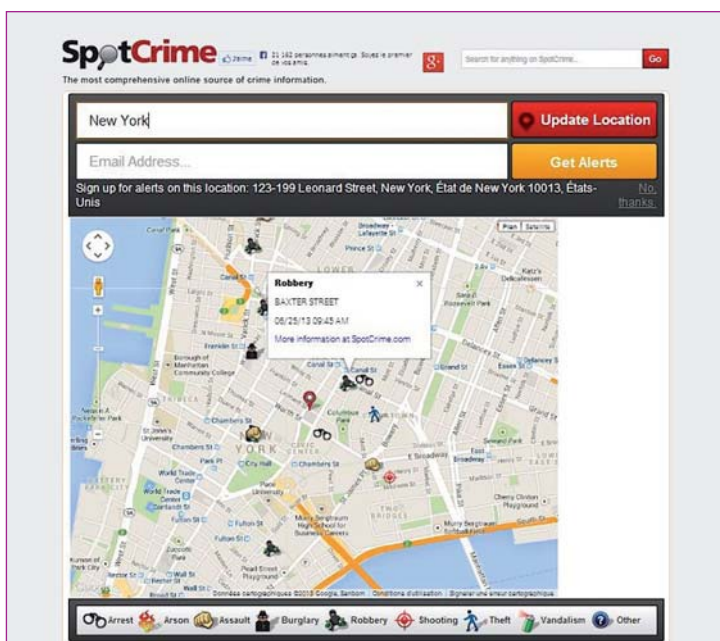
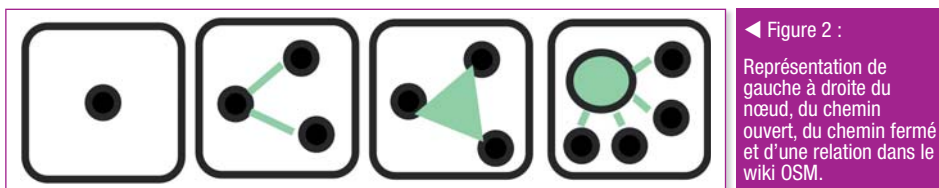


Figure 1 : Exemple de carte collaborative « *non-généraliste* » issu du site *web SpotCrime*.



◀ Figure 2 : Représentation de gauche à droite du nœud, du chemin ouvert, du chemin fermé et d'une relation dans le wiki OSM.

Tourisme (tourism)
 Utilisées pour cartographier les lieux et les choses d'intérêt spécifique pour les touristes: lieux à visiter, des lieux de séjour, des lieux offrant un soutien. Voir Key:tourism pour plus d'information. Voir Tourism pour une introduction générale.

Clé	Valeur	Élément	Commentaire	Représentation	Photo
tourism	alpine_hut		Refuge de montagne (Wikipedia).		
tourism	attraction		Attraction touristique.		
tourism	artwork		Œuvre d'art.		
tourism	camp_site		Camping.		
tourism	caravan_site		Aire pour caravanes.		

▲ Figure 3 : Exemples de descriptions de tags pour les entités touristiques. Un camping est représenté soit par un nœud ou une surface avec le tag : (clef) tourism = camp_site (valeur).

Les *tags* ne sont pas des éléments en tant que tels mais qualifient les trois éléments de base (nœud, chemin et relations) : un *tag*, ou une balise, comporte une clef et une valeur, et équivaut aux attributs dans les bases de données géographiques classiques.

OSM comporte dix-huit niveaux de zoom (maintenant dix-neuf depuis fin août). Une échelle de représentation et une distance par pixel sont associées à chaque niveau. Par ailleurs, le fond de carte *osmfr*, visible sur le serveur français, a été retravaillé pour les utilisateurs. Le contributeur *cquest* (Christian Quest), vice-président de la communauté OSM France, a adapté ce fond de carte en traduisant les toponymes, les symboles locaux (La poste, SNCF, baguette pour les boulangeries...), et en rajoutant deux niveaux de zoom supplémentaires (niveau de zoom 19 et 20) qui permettent de réaliser des cartes à très grande échelle.

La cartographie à différentes échelles

Le niveau de détail

Le niveau de détail dans une base de données géographiques est défini par la conjugaison du schéma conceptuel de la donnée, de la résolution sémantique, de la résolution géométrique, de la précision géométrique et de la « *taille du grain* » ou granularité (Ruas & Bianchin, 2002). Une base de données est organisée selon un schéma conceptuel composé de classes, d'attributs et de relations. Celui-ci reflète la manière de représenter la vérité du terrain dans la base de données géographiques ; par exemple, les représentations d'une forêt par une surface ou par des nœuds identifiant chaque arbre sont deux schémas conceptuels correspondant à des niveaux de détails différents.

La résolution sémantique désigne la quantité d'information attribuée aux données décrivant les entités

géographiques. Cet élément n'étant pas directement lié à la cartographie, nous ne reviendrons pas dessus dans la suite de ce mémoire. La résolution géométrique d'une base de données vectorielles se définit approximativement par la distance minimum entre deux éléments d'une même géométrie ; cette définition est ambiguë et à manier avec précaution. Il est plus pertinent de rajouter la précision géométrique et la granularité, comme le suggère Ruas & Bianchin (2002). La précision géométrique qualifie simplement l'écart entre la position réelle et la position représentée (valeur moyenne). La granularité caractérise la taille des plus petites formes géométriques décrites dans la base. Nous comprenons donc qu'un niveau de détail important cumule une grande précision, une haute résolution etc., et qu'un faible niveau de détail se compose lui d'une petite précision, résolution, etc.

La généralisation cartographique

Pour cartographier l'espace géographique à partir d'une base de données, il est important de savoir ce que l'on veut représenter et la manière de le faire. La difficulté repose sur les différentes représentations numériques possibles d'un même espace ; l'enjeu est d'avoir une cohérence entre ces représentations afin de rendre l'expérience de l'utilisateur parcourant la carte naturelle. On élabore pour cela des règles de sélection et des règles de représentation, qui constituent le schéma conceptuel.

Créer une base de données différente pour chaque représentation se révèle beaucoup trop coûteux en temps et en argent.

La solution classique consiste à constituer une unique base de données de précision importante, puis de dériver chaque représentation d'un processus de synthèse appelé généralisation. La réduction de l'échelle implique une simplification : il n'est pas possible de conserver tous les éléments présents sur la carte à grande échelle, car celle-ci deviendrait illisible.

La généralisation sélectionne les éléments pertinents à l'échelle choisie, puis applique éventuellement des opérateurs qui modélisent les transformations types utilisées pendant la généralisation : par exemple, une simplification élimine de l'information afin de créer une représentation plus agrégée ; les caractères sélectionnés, car jugés importants, sont caricaturés, c'est-à-dire exagérés pour être visibles. On distingue des algorithmes indépendants, qui ciblent un objet et modifient sa géométrie, et des algorithmes contextuels, qui s'appliquent sur un ensemble d'objets.

La généralisation inverse

La généralisation inverse est un sujet inédit. Proposé par Touya et Brando (2013), cette approche originale de la généralisation apporte de nouvelles problématiques dans le domaine de l'automatisation des processus cartographiques. Nous parlons de « *généralisation inverse* » car les objectifs visés sont différents de la généralisation : le but est tout d'abord de réaliser une carte à grande échelle et non une carte à petite échelle. Au lieu de simplifier les données, l'objectif est de les complexifier. Il ne s'agit pas d'inventer de l'information manquante mais plutôt de rétablir des relations spatiales entre objets détaillés et non détaillés.



Figure 4 : Données initiales d'OSM puis après traitement automatique.

La figure 4 représente les données initiales d'OSM (*initial data*) ; elle montre un groupe de bâtiments situé en dehors des limites d'une emprise urbaine (en gris) déduite d'une source (*Corine Land Cover*) à faible niveau de détail. En appliquant le principe de généralisation inverse, l'emprise corrigée inclut cette fois les bâtiments dans la zone (figure 4). La qualité de la donnée zone bâtie est améliorée à l'aide des données d'objets plus détaillées, ici les bâtiments.

Méthodologie de l'harmonisation du niveau de détail

Avant d'harmoniser le niveau de détail des données issues d'OSM, il faut dans un premier temps identifier des cas d'incohérence de niveau de détail à corriger.

Parcourir OpenStreetMap

Le moyen le plus simple est de parcourir la carte OSM et d'y déceler des problèmes, comme un utilisateur le ferait. La carte OSM est une représentation du monde extrêmement riche. On y trouve, par exemple, les zones urbaines, les zones rurales, les forêts, les bords de mer, les zones montagneuses... En outre, la densité de données varie en fonction du nombre de contributeurs actifs « *locaux* » (Touya & Girres, 2010). En France, par exemple, les grandes villes sont beaucoup plus complètes que les petites communes rurales. Il est intéressant de rechercher

les zones les plus denses, dans lesquelles se trouvent le plus d'entités différentes, et où on maximise la chance de déceler des problèmes d'hétérogénéités. Frederic Ramm met à disposition une liste des deux cents zones les plus denses d'OSM.

Outils de qualité

OSM n'impose pas de restrictions rigides ; les consignes sont dictées par la communauté à travers le *Wiki OSM* et les contributeurs sont seulement encouragés à les respecter. C'est pourquoi la communauté a mis en place un grand nombre d'outils externes de *quality assurance* (assurance qualité), qui aident les participants à améliorer la qualité des données. Ces outils ne corrigent pas les erreurs automatiquement, mais pointent les bugs et erreurs présents dans la base de données (dans une certaine mesure), ce qui permet d'orienter ceux qui ont décidé d'améliorer et de vérifier le travail réalisé par les autres contributeurs : 58,6 % des contributeurs sondés par Brando (2013) indiquent qu'ils sont impliqués dans la correction d'erreurs présentes dans les données.

Nous distinguons deux types d'outils de signalement d'erreurs. Certains détectent les erreurs automatiquement à l'aide d'outils *web* tel que *Keepright* et *Osmose* (figure 5) ; d'autres recensent les erreurs détectées par les contributeurs eux-mêmes (*OpenStreetBugs*). Il existe également des outils de suivi de zones, qui rapportent

les nouvelles éditions réalisées dans une zone précisée, pour le compte d'un contributeur voulant vérifier le travail effectué.

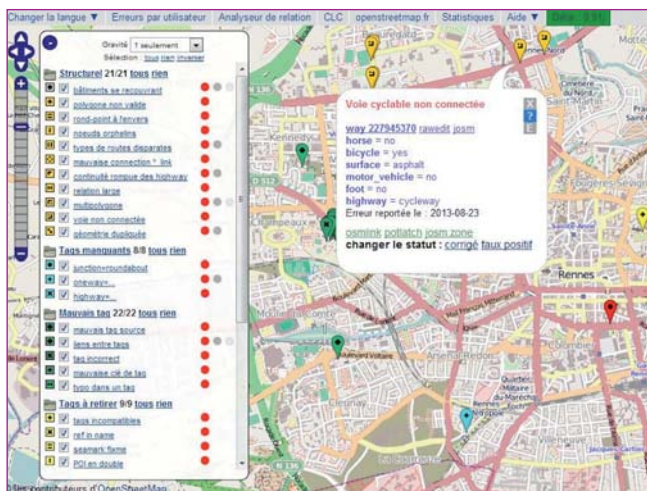


Figure 5 : L'outil *Osmose* signifie *OpenStreetMap Oversight Search Engine*. L'exemple est situé dans la ville de Rennes où *Osmose* indique une erreur lors de l'insertion de la piste cyclable qui se retrouve non connectée.

Ces outils peuvent servir de première approche dans la recherche de problèmes d'hétérogénéité. Ils permettent de comprendre une grande partie des bugs ou des erreurs présents sur OSM et de vérifier le travail de collaboration. En revanche, ces outils relèvent principalement des erreurs de topologie, non des problèmes liés aux niveaux de détail.

Sélection des cas d'hétérogénéités

Les problèmes rencontrés sous OSM ne sont pas tous issus d'hétérogénéité de niveau de détail. Après avoir trouvé un

problème cartographique (éventuellement à l'aide des outils déjà mentionnés), on utilise le *Géoportail* pour comparer la représentation d'OSM avec les images satellites et la carte IGN. Le portail géographique de l'IGN permet de s'assurer de l'authenticité du cas sélectionné. Si l'incohérence est avérée, il faut examiner les sources des données dans OSM pour en déduire leurs niveaux de détail. Par exemple, le chemin coupant le lac (figure 6) provient du cadastre alors que le lac est issu de *Corine Land Cover* (figure 7).

Pour terminer, si les niveaux de détail semblent hétérogènes, on extrait alors les données au format XML pour pouvoir harmoniser leurs niveaux de détail dans la plate-forme de travail utilisée à l'IGN, *CartAGen*.

La méthode systématique d'identification des cas

Cinq catégories de niveaux de détail, issues du *Géoportail*, ont été définies par Touya & Brando (2013) : *street*, *city*, *county*, *region* et *country*. Afin de détecter automatiquement les hétérogénéités de niveau de détail, le programme *Electre tri* (ET), décrit dans l'article (Touya & Brando, 2013) a été utilisé : ET

attribue, pour chaque donnée de la base, un niveau de détail individuel allant de 1 à 5, selon une analyse multicritère (sept critères) : chaque critère se voit attribuer une valeur normalisée allant de 0 à 1, la valeur 1 signifiant un faible niveau de détail. Figueira *et al.* (2005) précise que la méthode ET peut être employée lorsque les trois conditions suivantes sont réunies : la classification est ordonnée en catégories (ici du niveau de détail *street* au niveau de détail *country*), on utilise plus de trois critères (sept ici) et ceux-ci sont hétérogènes. La méthode produit également un indice de confiance. Par exemple, « *ce bâtiment a un niveau de détail de deux avec un indice de confiance de 75 %* ».

Connaître le niveau de détail de chaque élément constituant la carte est nécessaire pour détecter les erreurs, mais cela ne suffit pas. En effet, comme le soulignent Duchêne *et al.* (2012), les relations spatiales sont fondamentales dans l'automatisation cartographique et dans la généralisation. Brando (2012) propose l'idée de rajouter un programme de vérification des relations spatiales des données collaboratives dans les éditeurs (par exemple en vérifiant que le bâtiment importé ne croise pas une route ou un autre bâtiment pré-existant). Touya et Brando (2013) suivent cette idée et exposent la même approche pour identifier



Figure 6 : Détail de la donnée du chemin sur l'étang du Grand Gournava à Pleucadeuc dans OSM.



Figure 7 : Détail de la donnée de l'étang du Grand Gournava à Pleucadeuc dans OSM.

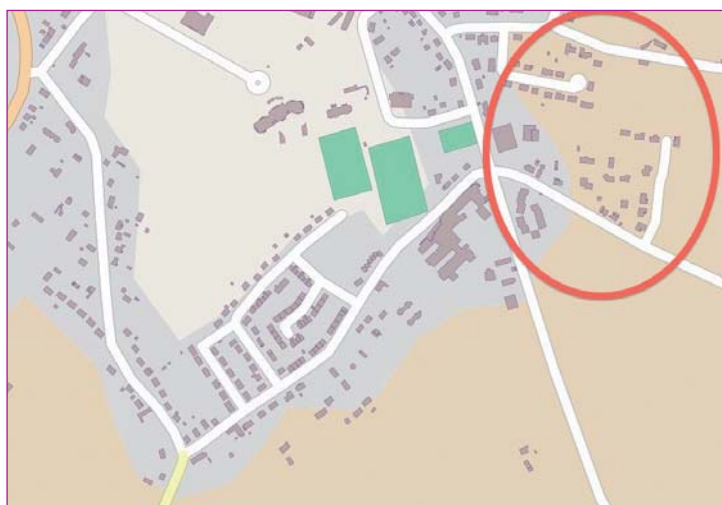


Figure 8 : Mauvaise délimitation de la zone bâtie à Loris, Loiret dans OSM.

les endroits où le niveau de détail hétérogène pose des problèmes de réalisation cartographique. Ainsi, lorsqu'une relation spatiale est presque vérifiée et que son niveau de détail est hétérogène, ce problème cartographique est une incohérence de niveau de détail. C'est le cas également si les deux objets vérifient une relation spatiale interdite. La figure 8 montre l'exemple d'une zone bâtie mal définie. On remarque certains bâtiments présents dans la zone beige, qui correspond à un terrain de ferme (cultures de céréales, élevages...) dans OSM (*landuse=farm*). De plus, la zone bâtie coupe certaines habitations.

Cette incohérence est présente dans la majorité des zones bâties. C'est l'un des problèmes les plus récurrents dans OSM. Pour le résoudre, on peut penser, dans un premier temps, détecter et sélectionner les bâtiments proches d'une zone bâtie mais n'y figurant pas, le but étant de les y intégrer. L'application d'un *buffer* regroupant les bâtiments peut être une des solutions. On vérifie par une contrainte spatiale que ce groupe est intégré complètement dans la zone bâtie. Si ce n'est pas le cas, on procède à un redimensionnement de la géométrie de la dite zone bâtie.

Implémentation et évaluation des résultats obtenus

La méthode de création de clairières

Les forêts, dans OSM, ont généralement un faible niveau de détail car la géométrie de celles-ci est soit issue de *Corine Land Cover*, soit dessinée grossièrement à l'aide de photos aériennes. Les bâtiments, en revanche, proviennent en grande partie des importations cadastrales, leur niveau de détail est donc plus important. Résultat, un groupe de bâtis présent dans une forêt est rarement représenté au sein d'une clairière. Or, à part certaines cabanes forestières ou bâtiments abandonnés, qui peuvent se retrouver au milieu de la végétation, nous avons supposé qu'un groupe de maisons dans une forêt est toujours situé dans une clairière.

La méthode de création de clairières ne s'applique pas sur les bâtis isolés mais sur des groupes de plus de quatre constructions car le niveau de zoom maximum d'OSM n'est pas suffisant ; de plus, dessiner une clairière pour un bâtiment unique n'est pas perti-

nent, voire gêne la compréhension de la carte. Chaque objet bâti se voit entourer d'un *buffer*, puis la fonction *BufferClustering* permet de regrouper ceux dont le *buffer* s'intercepte. Après avoir étudié différents paramètres, nous avons choisi de considérer que des bâtiments espacés de plus de quarante mètres ne sont pas dans la même clairière ; cela correspond à un *buffer* de vingt mètres. Une fois vérifié que ces groupes intègrent plus de quatre maisons, nous récupérons les forêts dans la base de données, puis testons que le groupe se trouve dans la forêt et que le niveau de détail est hétérogène entre les deux entités. Si les entités ainsi sélectionnées sont bien à l'origine d'un problème d'hétérogénéité de niveau de détail, les clairières vont être créées par traitement individuel de chaque groupe : la méthode rajoute la géométrie de l'enveloppe convexe du groupe de bâtiments dans la forêt pour créer la clairière.

La figure 9a montre un groupe de bâtiments dans la forêt de Rennes qui correspond aux critères pour créer une clairière. La clairière est créée à partir de l'enveloppe convexe du groupe de bâtiments mais cette méthode ne tient pas compte des chemins ou routes autour des bâtiments. La figure 9b illustre le « débordement » issu de l'enveloppe convexe. Le résultat final est visible figure 9c où les chemins sont pris en compte, afin de rendre la clairière plus réaliste.

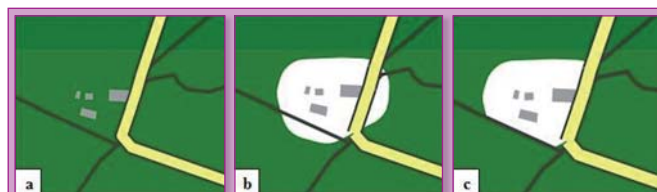


Figure 9 : Images issues de *CartAGEN* représentant un groupe de bâtiments dans la forêt de Rennes. Création d'une clairière.



Figure 10 : Image de la forêt de Rennes issue de *CartAGen* affichant deux clairières carrées conçues par la variante.

Les zones de clairières isolées par une route et sans bâtiment ont été supprimées. Une variante donne à la clairière une forme symbolique (ici un carré). La figure 10 illustre deux clairières carrées.

Après réflexion, si la surface délimitée paraît trop naturelle et réaliste, l'utilisateur ne s'aperçoit pas de la présence du traitement. En appliquant la variante, l'utilisateur réalise que la clairière a été générée automatiquement et repère que sa forme est bien trop géométrique pour être véritable.

Afin d'améliorer la méthode, il serait intéressant de prendre en compte au maximum les autres éléments présents en relation spatiale. Par exemple, quand le groupe de bâtiments n'est pas inclus complètement dans la forêt, la clairière n'est pas dessinée.

La méthode des chemins coupant les lacs

Tout comme les forêts, les lacs sont en grande partie importés de CLC ou dessinés depuis les images aériennes *Bing* (figure 11). La méthode *CrossingLakes* va redéfinir la géométrie du lac

lorsqu'un chemin le coupe, sous couvert que le chemin ait un niveau de détail plus important. Ainsi redimensionné, l'utilisateur n'a plus de problème d'interprétation lors de sa recherche d'itinéraire pour une balade en vélo par exemple.

La méthode récupère le ou les lacs (*WATER_AREAS_POP*) dans la fenêtre ainsi que les chemins (*ROADS_POP* de niveau 0, le niveau 1 correspondant à une petite route...). Ensuite, on sélectionne les lacs et les chemins en relation spatiale, c'est-à-dire les cas où un chemin coupe un lac. Comme pour chaque méthode implémentée, on vérifie que le

la route définissant sa largeur lors de sa représentation. Cette dernière est renseignée par son *tag* ; par exemple, une autoroute est codée *highway = motorway*. Pour cette raison, redéfinir la géométrie du lac directement à partir du chemin n'est pas très pertinent. La méthode découpe donc le lac suivant un *buffer* de trois mètres autour du chemin. Le choix des trois mètres résulte d'un compromis : il est rare qu'un chemin borde un plan d'eau à moins de cette distance.

Le chemin de la figure 12a coupe le lac (dans le carré rouge). Après application de la méthode *CrossingLakes* sur les données,



Figure 11 : Description du lac Altenweiher dans OSM dont la source est *Bing Aerial*™.

problème d'intersection est bien issu d'une hétérogénéité de niveau de détail entre les entités, puis la géométrie du lac est redéfinie en prenant en compte le chemin.

Comment le lac doit-il être redécoupé à partir du chemin ? Un chemin, une route ou une autoroute, dans OSM, est tracé à partir de son axe, la nature de

le résultat de la nouvelle géométrie du lac est illustré par la figure 12b, le chemin borde bien le contour du lac.

La méthode donne de bon résultats et augmente la qualité des lacs redessinés. Il serait intéressant de prendre en compte le cas de la figure 13 où un chemin traverse l'étang à deux endroits. Ceci est un cas peu fréquent où

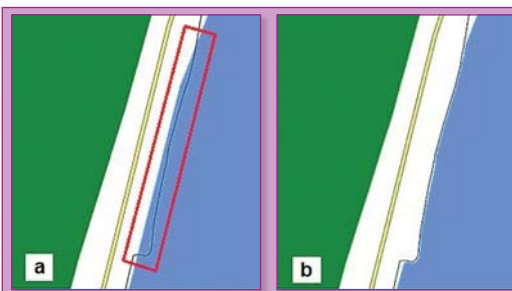


Figure 12 : Images issues de *CartAGen* avant et après que les données soient corrigées par la méthode. Données d'OSM du lac du Temple dans la réserve naturelle de la forêt d'Orient.

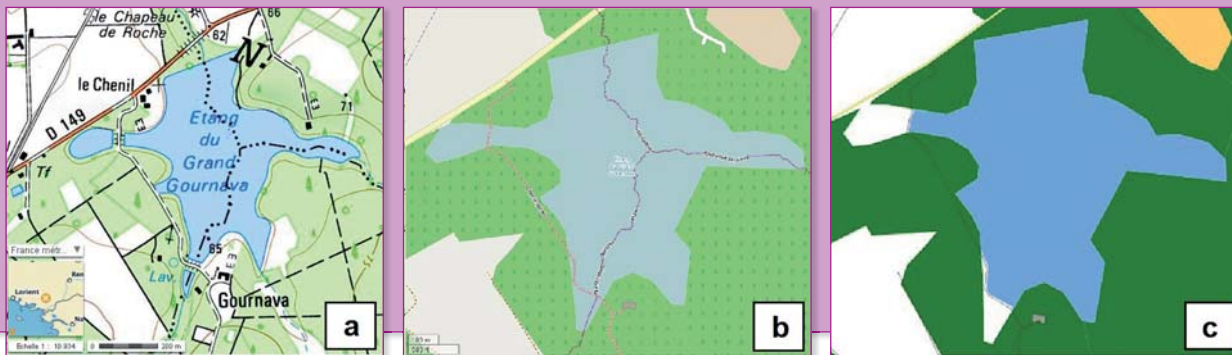


Figure 13 : La figure 13a (*Géoportail* IGN), la figure 13b (OSM) et la figure 13c (*CartAGen*) représentent l'étang du Grand Gournava à Pleucadeuc.

la représentation des données dans OSM (figure 13b) n'est pas totalement fausse car les traversées existent (figure 13a), mais le chemin devrait se trouver sur un accès non figuré. La méthode supprime automatiquement des parties de l'étang (figure 13c) et montre ainsi ses limites. Il serait plus judicieux de supprimer une bande autour du chemin, par exemple à l'aide d'un *buffer* pour signifier l'existence d'un passage, ou alors de ne pas détecter ces cas comme des incohérences en changeant la méthode *PathNotCrossingLake*.

La méthode du trait de côte

Dans OSM, le trait de côte, ou ligne de côte, est un élément particulier. Celui-ci délimite la transition entre la mer et la terre, dans un environnement riche et complexe : plages, ports, habitations, chemins, champs, exploitations ostréicoles... Tous ces éléments devraient avoir un niveau de détail homogène pour éviter les incohérences. Or, ces données ne sont pas rentrées par le même contributeur et encore moins avec le même niveau de détail. L'utilisateur éprouve donc des difficultés à comprendre la démarcation terre-mer.

Le premier écueil pour le contributeur qui souhaite intégrer un trait de côte repose sur sa définition même ; dans OSM, il est

défini comme la hauteur d'eau la plus importante atteinte lors des grandes marées de printemps. N'étant pas experts en hydrologie (ou océanographie), on constate des différences entre les contributeurs : certains importent leurs données GPS, d'autres retranscrivent des images aériennes. Le trait de côte est tracé, en grande majorité, au milieu des images satellites *Bing* ou *Landsat*.

Dans un premier temps, la côte a été intégrée à la base de données OSM depuis PGS (*Prototype Global Shoreline*), tracé extrait des images satellites du domaine public *Landsat* par un algorithme de reconnaissance d'images ; cette description « brute » de la côte sert de base aux améliorations suivantes. Les contributeurs

sont invités à la compléter pour améliorer le niveau de détail. En France, les sources varient, allant de PGS au cadastre. Christoph Hormann a réalisé une carte montrant la qualité du trait de côte dans le monde sur son blog. Le trait de côte global doit être fermé ; celui de la France se poursuit en Europe du Sud, Afrique puis Asie et revient par la Scandinavie. La polygone correspondante est orientée, la terre se trouvant du côté gauche.

La méthode *CoastLine* va traiter les problèmes de relations entre les occupations du sol et la ligne de côte. Les géométries des champs, des zones bâties, des forêts, empiètent sur la mer, gênant la lecture de la carte par un utilisateur. La figure 14 montre

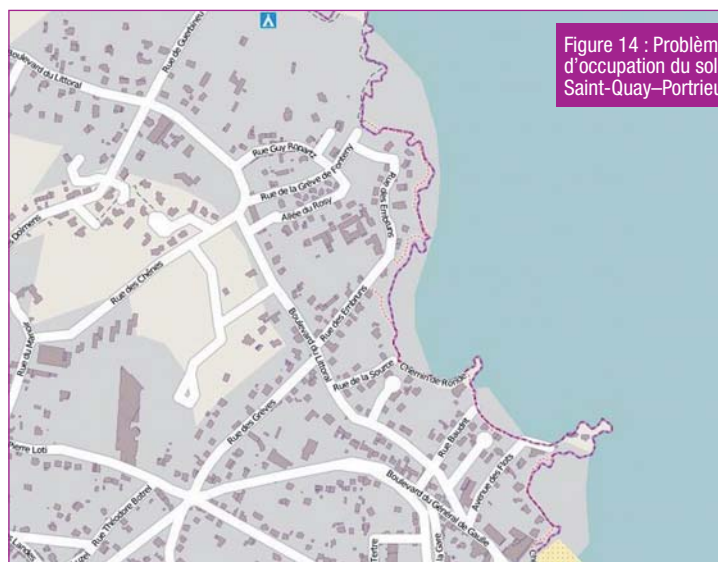


Figure 14 : Problème d'occupation du sol à Saint-Quay-Portrieux.

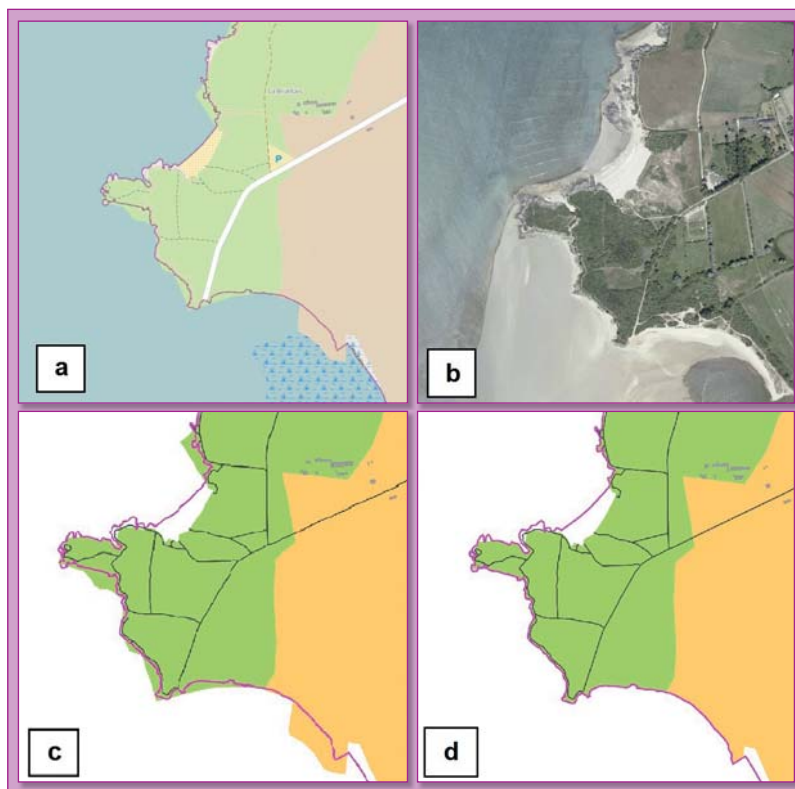


Figure 15 : Lancieux – 15a (OSM), 15b (Géoportail), 15c et 15d (CartAGen).

la zone bâtie de couleur grise de Saint-Quay-Portrieux, qui « déborde » dans la mer. Ces problèmes apparaissent assez souvent dans les zones ayant une côte découpée avec les zones d'occupations du sol importées de *Corine Land Cover*, comme en Bretagne par exemple.

Comme dans les autres méthodes, il faut dans un premier temps extraire les données dont le niveau de détail est à harmoniser. Les entités d'occupation

du sol (prairies, zones bâties, terrains agricoles...) proviennent de *LANDUSE_AREAS_POP*, et l'entité trait de côte à partir de *COASTLINE_POP*. Les éléments surfaciques intersectant le trait de côte sont sélectionnés, puis leur nouvelle géométrie est définie à partir de la ligne de côte (après vérification qu'il y ait une hétérogénéité de niveau de détail). Cette nouvelle géométrie est définie à l'aide d'une « zone tampon unilatérale » : la partie découpée se retrouvant en mer est supprimée.

La figure 15a est un aperçu de Lancieux sur OSM. On y repère un problème cartographique entre le trait de côte, la prairie (vert clair) et la zone agricole (marron clair). La figure 15c est une importation des données d'OSM dans *CartAGen* avant d'appliquer la correction à l'aide de la méthode. Le résultat de *CoastLine* sur ce jeu de données (figure 15d) améliore la représentation de la côte.

Cette méthode affine le rendu cartographique de la côte : l'utilisateur perçoit mieux la démarcation terre-mer. En outre, elle accroît la qualité des données des zones d'occupation du sol. Toutefois, certains cas résistent à cette méthode : la figure 16b montre toute une zone de ferme (en beige) ôtée après correction des données (fig. 16c) alors que la forêt (sur la gauche des images) est bien redéfinie. Ceci est dû à des bugs dans les calculs d'intersection entre polygones.

Afin d'améliorer la méthode, il serait intéressant d'ajouter d'autres éléments en relation avec le trait de côte, pour, par exemple, éliminer des « trous » d'eau (cf. figure 17) produits par des incohérences entre zones de plage et traits de côte.

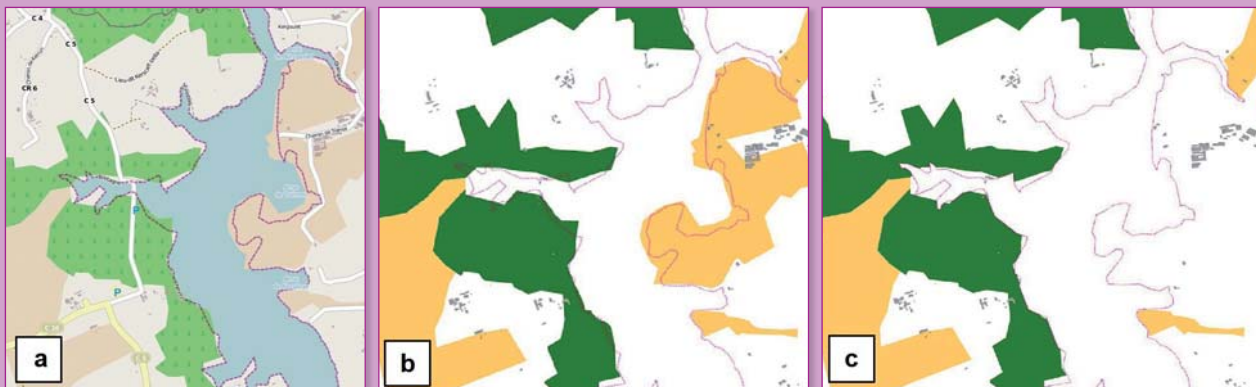


Figure 16 : l'anse de Tremor sur l'Aven. 16a (OSM), 16b et 16c (CartAGen).



Figure 17 : anse du Pisso à Pléneuf-Val-André.

La méthode des arbres sur les routes

Les contributeurs sont libres de rajouter les entités qu'ils veulent dans OSM afin d'enrichir au maximum la base de données. Ainsi, dans les zones denses à forte contribution, chaque arbre est localisé et figuré (surtout en ville). Dans OSM, un arbre est modélisé par un nœud, représenté par un carré vert et identifié par le *tag natural=tree*. Un trop grand nombre d'arbres saisis dans une zone surcharge la carte et nuit à la lisibilité ; cette surcharge visuelle s'appelle l'« effet de clutter » (figure 18).

La diminution de l'échelle entraîne en outre un empatement des routes ; les arbres présents sur les accotements ne changeant pas de place, ils se retrouvent donc au milieu de la chaussée : c'est un problème de généralisation. Le fond de carte *osmfr* sur *tile.openstreetmap.fr* a trouvé une solution ingénieuse pour éviter la saturation de la carte :

les arbres sont représentés par un disque vert plus ou moins transparent en fonction du niveau de zoom. L'emprise de l'arbre sur la carte ne gêne plus l'utilisateur (figure 19).

En dehors de ce problème dû à la généralisation, il arrive que certains arbres soient également saisis sur la route, lorsque le contributeur utilise des images satellites à petite échelle. La route ayant un niveau de détail plus élevé (car cette entité est considérée primordiale dans la cartographie), elle est mieux localisée.

Avant de traiter les données, il faut se placer à la même échelle qu'OSM. On travaille donc au niveau de zoom 18 correspondant à l'échelle 1:1 693. Il faut représenter sur *CartAGen* la même emprise de route que sur OSM pour un niveau de zoom donné ; chaque route possède une emprise différente, qui peut être déduite de son tag (exemple :

highway=tertiary pour les petites routes à l'intérieur des villes). Lorsqu'un arbre intercepte une route, il faut déterminer de quel côté il se trouve avant de le déplacer. Pour cela, la procédure superpose deux zones tampons unilatérales à la géométrie de la route ; ainsi, un arbre présent dans la zone tampon droite sera décalé à droite.

Nous supposons que les arbres présents sur les routes doivent être recalés sur les accotements. Des lignes parallèles à la géométrie de la route ont donc été tracées de chaque côté de celle-ci. Le décalage vaut la demi-largeur de la route plus la distance de représentation d'un arbre, c'est-à-dire quatre pixels dans notre cas.

Sur la figure 20a, les arbres se trouvent au milieu de la route au lieu de figurer, comme sur l'illustration 20b, au bord de celle-ci. En appliquant la méthode, le résultat obtenu avec les paramètres d'éloignement des arbres sur la figure 21b n'est pas satisfaisant. Les arbres sont trop loin du bord. Un deuxième essai illustré par la figure 21c a été effectué et se révèle meilleur.



Figure 18 : Yaoundé sur OSM avec une saturation d'arbres.

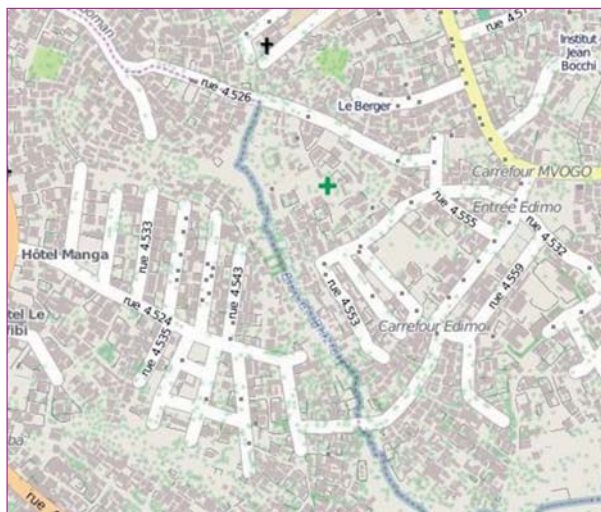


Figure 19 : Yaoundé avec le fond de carte *osmfr*.



Figure 20 : Les figures illustrent la cours Aristide Briand à Orange. 20a OSM, 20b Google Street View.

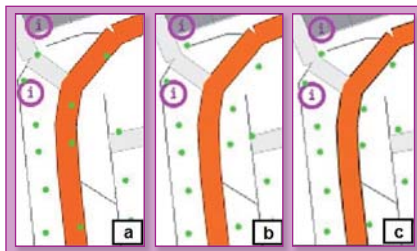


Figure 21 : Images de CartAGen. La figure 21a montre les données non traitées et les figures 21b et 21c sont les résultats de la méthode avec des paramètres de décalages différents.

La méthode appliquée donne de bons résultats dans l'ensemble mais il existe des configurations géographiques où le problème est multiple, par exemple lorsque les arbres ne sont pas alignés. En partant du principe que ceux-ci se trouvent tous à la même distance du bord, ils peuvent être regroupés dans un ensemble. On peut ainsi produire de nouvelles relations, comme égaliser l'écartement entre chaque arbre. Ceci demeure cependant difficile à implémenter à cause de nombreux cas particuliers : parc boisé, mobilier urbain, croisement routier peuvent altérer le résultat.

Autres cas non traités

De nombreux autres cas d'incohérences existent dans OSM. La base de données s'enrichissant de plus en plus en fonction des besoins et des propositions des contributeurs,

de nouveaux cas se font jour. Les paragraphes suivants décrivent d'autres cas identifiés qui n'ont pas encore été traités.

Les points d'intérêts

On peut intégrer dans OSM les points d'apport volontaire (verre, carton par exemple), les cabines téléphoniques, les signaux tricolores, les boîtes aux lettres, les parcs à vélo, les phares, les puits de pétrole, les pylônes électriques... Tous ces éléments constituent des points d'intérêts (POI) représentés par un symbole placé sur la carte à l'aide d'un nœud. L'utilisateur apprécie généralement ces éléments qui enrichissent la carte. Certains POI, de surcroît, entretiennent des relations spatiales qui forment un nouvel objet : les arrêts de bus et les pylônes électriques, par exemple, vont former respectivement des lignes de bus et des lignes électriques.

Lorsque ces POI sont mal localisés, le lecteur détecte une incohérence préjudiciable. La figure 22 représente un arrêt de bus positionné à distance de la rue ; impossible de savoir où il se situe exactement. Une méthode

recalant les arrêts de bus au bord de la route serait la bienvenue.

Les zones d'occupation du sol

Les zones d'occupation du sol dans OSM sont très diverses. Chaque type de zone est représenté par un polygone d'une couleur spécifique. Le niveau de détail de ces données est souvent faible, car elles proviennent soit d'importations massives (Corine Land Cover par exemple), soit des images aériennes (Bing par exemple). Les objets plus détaillés, tels que les bâtiments, entrent souvent en conflit, c'est-à-dire que les relations spatiales existantes ne sont pas représentées. L'utilisateur, qui est déjà sollicité pour comprendre la logique des différentes couleurs du fond de plan, doit également composer avec incohérences des bâtiments.

La figure 23 est une image d'OSM représentative des incohérences en bâtiments et fonds de carte. On constate des groupes de bâtiments entre deux zones ou coupés par l'une d'elles ou à moitié intégrés.

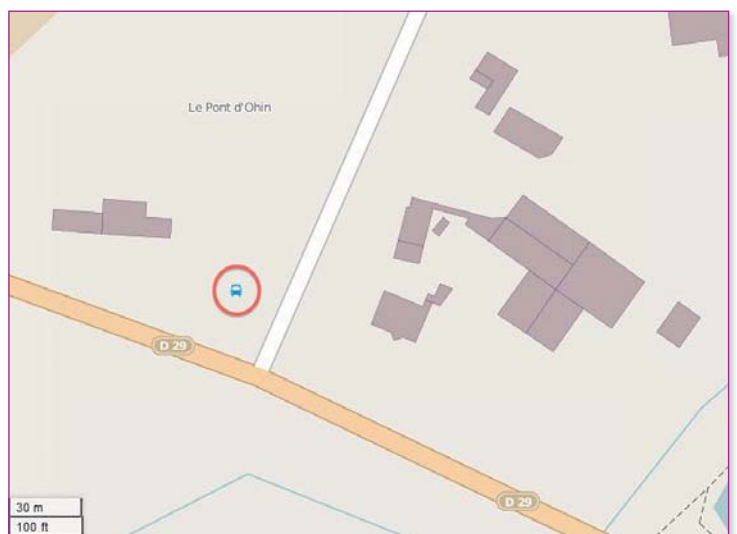


Figure 22 : Arrêt de bus se trouvant à près de trente mètres d'une rue dans OSM.



Figure 23 : Les différents types de zones d'occupation du sol à Saligny dans OSM.

Conclusion

La généralisation « *inverse* » est une nouvelle problématique dans l'automatisation de la cartographie qui n'a jamais été étudiée. On constate que les méthodes implémentées ne sont jamais parfaites et nécessitent des codes de plus en plus performants. Dédire les relations spatiales autour de l'objet dont le niveau de détail doit être augmenté est d'une importance capitale dans la généralisation « *inverse* ». On peut déjà penser que les difficultés rencontrées dans le processus automatique de généralisation vont surgir aussi ici.

Afin de trouver des méthodes pouvant fonctionner sur différents cas, une classification des différentes incohérences rencontrées en fonction des objets concernés peut être envisagée. Un grand nombre de cas non identifiés pourraient être traités automatiquement, ce qui augmenterait sensiblement la qualité de la base.

Certaines incohérences sont identifiables mais trop difficiles à

corriger avec certitude, lorsqu'il manque des informations impor-

tantes sur les données traitées. Par exemple, un arrêt de bus très loin d'une route ne peut pas être déplacé ; dans ce cas, on peut penser qu'un contributeur a mal positionné l'arrêt ou bien qu'une rue n'a pas été saisie. Afin d'exprimer cette incertitude, un objet diffus indique l'erreur mais montre aussi l'impossibilité de la corriger.

Les perspectives de ce travail sont multiples, elles ont pour but, à l'IGN, de comprendre et de gérer les données VGI de sorte qu'un système collaboratif puisse se mettre en place afin de compléter les bases de données. |

Bibliographie

- Brando C. : Édition collaborative d'un contenu géographique : une approche fondée sur des spécifications formelles et des relations spatiales explicites. Journées de la Recherche IGN, 2012
- Brando C. : Coalla : un modèle pour l'édition collaborative d'un contenu géographique et la gestion de sa cohérence. Thèse en Informatique. Université Paris-Est-Marne-la-Vallée, 2013.
- Bucher B., Brasebin M., Buard E., Grosso E., Sébastien M. : GeOxygene: Built on top of the expertise of the french NMA to host and share advanced GI science research results. In Proceedings of the First International OpenSource Geospatial Research Symposium (OGRS'09). Springer, Berlin. À paraître.
- Duchêne C., Ruas A., Cambier C. : The CartACom model: transforming coartographic features into communicating agents for cartographic generalisation. International Journal of Geographic Information Science 26(9):1533-1562
- Figueira J., Mousseau V., Roy B. : ELECTRE methods. In Multiple Criteria Decision Analysis : State of the Art Surveys, p. 133-162. Springer, Berlin
- Girres J.-F., Touya G. : Quality assessment of the French OpenStreetMap dataset. Transactions in GIS 14(4):435-459
- Goodchild M. : Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), p. 211-221. Téléchargeable : <http://dx.doi.org/10.1007/s10708-007-9111-y>.
- Ruas A., Bianchin A. : Échelle et niveau de détail. In Généralisation et représentation multiple, p.25-44 et p. 75-90. Hermès Lavoisier, Paris
- Tapscott D., Williams A. D. : Wikinomics : How Mass Collaboration Changes Everything. New York, Portofolio Hardcover
- Touya G., Bucher B., Falquet G., Jaara K., Steiniger S. : Modelling geographic relationships in automated environments. Chapitre 3 in Abstracting Geographic Information in a Data Rich World. Springer, Heidelberg 2014. ISBN 978-3-319-00202-6
- Touya G., Brando C. : Detecting Level-of-Detail Inconsistencies in Volunteered Geographic Information Data Sets. Cartographica: The International Journal for Geographic Information and Geovisualization, Volume 48, Number 2, Summer 2013. p. 134-143